Springer Series in Operations Research and Financial Engineering

Sidney I. Resnick

# Extreme Values, Regular Variation, and Point Processes

Springer

# Springer Series in Operations Research and Financial Engineering

*Series Editors:*
Thomas V. Mikosch
Sidney I. Resnick
Stephen M. Robinson

# Springer Series in Operations Research and Financial Engineering

*Altiok:* Performance Analysis of Manufacturing Systems
*Birge and Louveaux:* Introduction to Stochastic Programming
*Bonnans and Shapiro:* Perturbation Analysis of Optimization Problems
*Bramel, Chen, and Simchi-Levi:* The Logic of Logistics: Theory, Algorithms, and Applications for Logistics and Supply Chain Management (second edition)
*Dantzig and Thapa:* Linear Programming 1: Introduction
*Dantzig and Thapa:* Linear Programming 2: Theory and Extensions
*de Haan and Ferreira:* Extreme Value Theory: An Introduction
*Drezner (Editor):* Facility Location: A Survey of Applications and Methods
*Facchinei and Pang:* Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I
*Facchinei and Pang:* Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume II
*Fishman:* Discrete-Event Simulation: Modeling, Programming, and Analysis
*Fishman:* Monte Carlo: Concepts, Algorithms, and Applications
*Haas:* Stochastic Petri Nets: Modelling, Stability, Simulation
*Klamroth:* Single-Facility Location Problems with Barriers
*Muckstadt:* Analysis and Algorithms for Service Parts Supply Chains
*Nocedal and Wright:* Numerical Optimization
*Olson:* Decision Aids for Selection Problems
*Pinedo:* Planning and Scheduling in Manufacturing and Services
*Pochet and Wolsey:* Production Planning by Mixed Integer Programming
*Resnick:* Extreme Values, Regular Variation, and Point Processes
*Resnick:* Heavy Tail Phenomena: Probabilistic and Statistical Modeling
*Whitt:* Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues
*Yao (Editor):* Stochastic Modeling and Analysis of Manufacturing Systems
*Yao and Zheng:* Dynamic Control of Quality in Production-Inventory Systems: Coordination and Optimization
*Yeung and Petrosyan:* Cooperative Stochastic Differential Games

**Forthcoming**
*Muckstadt and Sapra:* Models and Solutions in Inventory Management

Sidney I. Resnick

# Extreme Values, Regular Variation and Point Processes

Springer

Sidney I. Resnick
Cornell University
School of Operations Research and
Industrial Engineering
Ithaca, NY 14853
U.S.A.
sirl@cornell.edu

*Series Editors:*
Thomas V. Mikosch
University of Copenhagen
Laboratory of Actuarial Mathematics
DK-1017 Copenhagen
Denmark
mikosh@act.ku.dk

Stephen M. Robinson
University of Wisconsin-Madison
Department of Industrial Engineering
Madison, WI 53706
U.S.A.
smrobins@facstaff.wise.edu

Sidney I. Resnick
Cornell University
School of Operations Research and
Industrial Engineering
Ithaca, NY 14853
U.S.A.
sirl@cornell.edu

ISBN 978-0-387-75952-4 ISBN 978-0-387-75953-1 (eBook)
DOI 10.1007/978-0-387-75953-1

Library of Congress Control Number: 2007936937

Mathematics Subject Classification (2000): 60G55, 60G70, 62G30, 62G32, 60F05, 60B10

Printed on acid-free paper.


**Originally published by Springer-Verlag New York, Inc. in 2008**

First Printing, Hardcover, *Applied Probability*, Vol. 4, ©1987.



9 8 7 6 5 4 3 2 1

springer.com (IBT)

# Preface to the Soft Cover Edition

I always thought books should be like kids: you prepare them as well as you can and then send them out into the world and observe the interactions with pride and affection but without much interfering. However, despite my reluctance to write, Springer insisted I put some perspective on this new soft cover edition.

I am not sure when the first printing of *Extremes Values, Regular Variation and Point Processes* became unavailable for purchase, but I have become increasingly conscious in the past few years of people telling me they could not obtain a copy. There is usually a vague undercurrent to the comment as if I have let someone down. So I inquired about the status at Springer and was told that in this era of new technologies books do not go out of print.

Of course, in October 2006 Springer published my book *Heavy-Tail Phenomena, Probabilistic and Statistical Modeling* and it is fair to wonder what is the difference between the two books. There is some overlap but I labored to keep the overlap minimal. This older text holds up rather well as an account of basic, foundational, mathematical material on extreme value theory. It describes the interplay between the analytical theory of regularly varying functions and extremes, and also the probabilistic interplay between point processes, extremes and weak convergence. It presents an approach to the study of extremes that is still quite current and useful. The interplay between regular variation, point processes and extremes is a clear theme leading to a coherent view of the subject. The book is a rather mathematical treatment and for the most part proofs are presented completely and in a self-contained manner. Applications are hinted at but not explicitly discussed. In particular, no explicit treatment of statistical topics is present.

In the heavy-tails book, the viewpoint is similar in that point processes and random measures are basic but some mathematical foundational details have been omitted and referred to either in the 1987 book or to other sources such as the 2006 Springer book by de Haan and Ferreira entitled *Extreme Value Theory: An Introduction*. The focus in *Heavy-Tail Phenomena* is on a subset of extreme value theory which I find particularly intriguing and which is critical to applied probability modeling in finance, networks, queues and insurance. Some of this heavy-tailed applied probability modeling is presented in my 2006 book. *Heavy-Tail Phenomena* also surveys semi-parametric statistical inference methodology, something which the 1987 book did not attempt. Both the 1987 and 2006 books advance the point of view that much of the subject is dimensionless if viewed in the correct framework.

So I hope people will continue to have access to *Extremes Values, Regular Variation and Point Processes*, and that they will continue to find it companionably useful.

Ithaca, NY

Sidney Resnick
August, 2007

# Preface to the Hard Cover Edition

Extreme value theory is an elegant and mathematically fascinating theory as well as a subject which pervades an enormous variety of applications. Consider the following circumstances:

Air pollution monitoring stations are located at various sites about a city. Government regulations mandate that pollution concentrations measured at each site be below certain specified levels.
A skyscraper is to be built near Lake Michigan and thus will be subject to wind stresses from several directions. Design strength must be sufficient to withstand these winds. Similarly, a mechanical component such as an airplane wing must be designed to withstand stresses from several sources.
Dams or dikes at locations along a body of water such as a river or sea must be built high enough to exceed the maximum water height.
A mining company drills core samples at points of a grid in a given region. Continued drilling will take place in the direction of maximum ore concentration.
Athletic records are frequently broken.

A common feature of these situations is that observational data has been or can be collected and the features of the observations of most interest depend on largest or smallest values; i.e., on the extremes. The data must be modeled and decisions made on the basis of how one believes the extreme values will behave.

This book is primarily concerned with the behavior of extreme values of independent, identically distributed (iid) observations. Within the iid framework there are surprising depth, beauty, and applicability. The treatment in this book is organized around two themes. The first is that the central analytic tool of extreme value theory is the theory of regularly varying functions, and the second is that the central probabilistic tool is point process theory and in particular the Poisson process. Accordingly we have presented a careful exposition of those aspects of regular variation and point processes which are essential for a proper understanding of extreme value theory.

Chapter 0 contains some mathematical preliminaries. Some authors might relegate these to appendices, but I believe these should be read first, before plunging into the following sections, in order that readers can get used to my way of doing things. Chapter 0 also contains a derivation of the three families of classical, Gnedenko limit distributions for extremes of iid variables and an account of regular variation and its extensions.

Chapter 1 discusses thoroughly questions of domains of attraction. If iid random variables have common distribution $F$, what criteria on $F$ or its density $F'$ guarantee that suitably scaled and centered extremes have limit distributions as the sample size gets large? What are suitable scaling and centering constants? These results provide a theoretical underpinning to statistical practice as discussed, for example, in Gumbel (1958): Suppose data are obtained such that each observation is an extreme. For example, our data may consist of maximal yearly flow rates at a particular site on the Colorado River over the last 50 years. Suppose by stretching the imagination, one believes the data to be modeled adequately by the iid assumption. The underlying distribution of the model is unknown, so the parametric assumption is made that the data comes from a limiting extreme value distribution. Usually it is the Gumbel, also called *double exponential*, distribution $\Lambda(x) = \exp\{-e^{-x}\}$ that is chosen, and the estimation problem reduces to choosing location and scale constants; this is sometimes done graphically using loglog paper. The underlying distribution may not be extreme value, but we robustly hope it is at least in a domain of attraction, so that the distribution of extremes is close to a limiting extreme value distribution.

By the end of Chapter 1 much analytic technique has been developed and this is exercised in the specialist Chapter 2. If normalized extremes of iid random variables have a limit distribution, when do moments and densities of normalized extremes have limits? We also discuss rates of convergence to the limit extreme value distributions and large deviation questions which emphasize sensitivity to the quality of the approximation of the right tail of the distribution of the maximum of $n$ iid random variables by the tail of the limit extreme value distribution.

Chapter 3 shifts the focus from the analytic to the probabilistic, and is a thorough discussion of those aspects of point processes (and in particular the Poisson process) which are essential, in my view, for a proper understanding of the structural behavior of extremes. The core of the probabilistic results is in Chapter 4, which views records and maxima of iid random variables as stochastic processes. In a sequence of iid observations from a continuous distribution, the records (i.e., those observations bigger than all previous ones) form a Poisson process, and the indices when records occur are approximately a Poisson process. Several extensions to these ideas are discussed.

Also in Chapter 4 is an account of extremal processes. If maxima of $n$ iid random variables are viewed as a stochastic process indexed by $n$, there is a continuous parameter process called an *extremal process*, which is a useful approximation. The structural properties of such processes are studied and the uses for weak convergence problems are detailed. We also give an account in Sections 3.5, 4.4, and 4.5 of a weak convergence technique called the *point process method*, which has proved invaluable for weak convergence problems involving heavy tailed phenomena. If it is necessary to prove some functional (say the maximum) of $n$ heavy tailed random variables converges weakly as $n \to \infty$, it is often simplest to first prove a point process based on the $n$ variables converges ($n \to \infty$) and then to get the desired result by continuous mapping theorems. The power of this technique is illustrated in Section 4.5, where it is applied to extremes of moving averages.

The last chapter examines some multivariate extreme value problems. In one dimension notions such as maximum and record have unambiguous meanings. In higher dimensions this is no longer the case. The maximum of $n$ multivariate observations could be the convex hull, or it could be the vector of componentwise maxima depending on the application. We concentrate on the latter definition, which seems most natural for the applications mentioned at the beginning of the introduction. We discuss characterizations of the limiting multivariate extreme value distributions and give domains of attraction criteria. A theory of multivariate regular variation is needed, and this is developed. Criteria for asymptotic independence are given, and it is proved that a concept of positive dependence called *association* applies to limiting extreme value distributions.

Notation will ideally seem clear and simple. One quirk that needs to be mentioned is that if a distribution $F$ has a density, it is denoted by $F'$ (even in the multivariate case) and never by $f$. The symbol $f$ is reserved for the auxiliary function of a class $\Gamma$ monotone function. Sometimes, in the completely separate context of point processes, $f$ denotes a bounded, continuous real function, but the important point to remember is that $f$ is not the density of $F$; rather $F'$ is the density of $F$.

Extreme value results are always phrased for maxima. One can convert results about maxima to apply to minima by using the rule

$$-\max - = \min.$$

For example, $2 = \min\{2, 3\} = -\max\{-2, -3\} = -(-2) = 2$. We denote $\max\{x_i\colon 1 \le i \le n\}$ by $\bigvee_{i=1}^{n} x_i$ and similarly min is denoted by $\bigwedge$. Also it is usually clear how to adapt weak convergence results for maxima so that they apply to the $k$th largest of a sample of size $n$ ($k$ fixed, $n \to \infty$). The point process method usually makes this adaptation transparent. See, for instance, Section 4.5.

The best plan for reading this book is to start from the beginning and read each page lovingly until the end. There is only one section that is tedious. The second best plan is to start from the beginning and go through, passing lightly over certain material depending on background, taste, and interests but slowing down for the important results. Chapter 2, parts of 3, 4.4.1, and part of 4.6 may be skimmed, but the motto to be kept in mind is "skim; don't skip." This includes the exercises, which contain complementary material and alternative approaches. The extent to which readers will actively attempt the exercises will determine the extent of their progress from observer to practitioner. If plans 1 and 2 seem too ambitious, readers could consider making a module of Chapters 0, 1, and 5 and another module of a skimmed Chapter 3 and a heavily studied Chapter 4.

There are a number of things this book is not. It is not an encyclopedia and it is not a history book. Using a literary analogy, think of this as a novel. There is a story to be told, and readers should pay attention to matters of style and exposition and to how cosmic themes and characters relate. This book provides excellent coverage of problems arising from iid observations and offers good grounding in the subject, but does not pretend to offer comprehensive coverage of the whole subject of extreme values. This is now so broad and vast that it is doubtful that one book would do it justice. Consequently, a reader needing a rounded view of the whole subject is encouraged to consult other books and sources, as well as this one. For instance, with the exception of Section 4.5 on extremes of moving averages, I do not give attention to the important case of extremes of dependent variables. Fortunately, there is already a superb book on this subject by M.R. Leadbetter, G. Lindgren, and H. Rootzen, entitled *Extremes and Related Properties of Random Sequences and Processes*. It is very well written and elegant and is highly recommended.

Chapters 0, 1, and 2 bear the intellectual influence of my colleague and friend Laurens de Haan, with whom I have had the privilege and pleasure of working and learning since 1972. Professor de Haan has had enormous influence on the subject, and his 1970 monograph remains, despite the huge quantity of research it stimulated, an excellent place to learn about the relationship of extreme value theory and regular variation.

Now the acknowledgments. It is customary at this point for authors to make a maudlin statement thanking their families for all the sacrifices which made the completion of the book possible. This may be rather out of tune in these pseudo-quasi-semiliberated eighties. I will merely thank Minna, Nathan, and Rachel Resnick for a cheery, happy family life. Minna and Rachel bought me the mechanical pencil that made this project possible, and Rachel generously shared her erasers with me as well as providing a back-up mechanical pencil from her stockpile when the original died after 400 manuscript pages. I appreciate the fact that Nathan was only moderately aggressive about attacking my Springer-Verlag correspondence with a hole puncher.

Robyn Kelley (with timely assistance from Waydene Casey and Pat Key) provided excellent typing worthy of a raise, and cheerfully coped with substandard penmanship, transatlantic mailings, the differences between $8\frac{1}{2} \times 11$ inch and A4 paper, European hole punchers that punch in the wrong places for U.S. binders, and a professor who though on leave of absence kept sending her work.

The U.S. National Science Foundation has been magnificent in its continuing support. As a young pup I was allowed to schnor from other people's grants, and then at a crucial stage in my career NSF made me co-principal investigator of a series of grants that continue to the present. Other institutions which have generously provided support which has generated time away from teaching for thinking and learning are Erasmus University, Rotterdam; University of Amsterdam; Mathematics Center, Holland; Lady Davis Fellowship Trust; Technion, Haifa, Israel; United Kindgom Science and Engineering Research Council (SERC). The first two-thirds of the book were written while I was on sabbatical support—gratefully acknowledged—from Colorado State University. Duane Boes has been an enlightened and supportive chairman at CSU. Sussex University, Brighton, United Kingdom, provided a hospitable and pleasant environment, where the last third was finished while I was being supported by the SERC Fellowship.

As they have been so often in the past, Joe Gani and Chris Heyde were very helpful and encouraging during the preparation of this book.

Fort Collins, Colorado
April 1987

Sidney I. Resnick

# Contents

## Chapter 0

# Preliminaries

Some of the topics discussed here are sometimes relegated to appendices. However, since these topics must be well understood before arriving at the core of the subject, it seems sensible to discuss the preliminary topics first. The reader is advised to skim 0.1, 0.2 according to taste and background, but slow down for 0.3, which discusses the possible limiting distributions for normalized maxima of independent, identically distributed (iid) random variables. Section 0.4 treats the basic facts in the theory of regular variation and some important extensions. Regular variation is the basic analytical theory underpinning extreme value theory, and its importance cannot be overemphasized.

## 0.1. Uniform Convergence

If $f_n$, $n \geq 0$ are real valued functions on $\mathbb{R}$ (or any metric space) then $f_n$ converges uniformly on $A \subset \mathbb{R}$ to $f_0$ if

$$\sup_{x \in A} |f_n(x) - f_0(x)| \to 0$$

as $n \to \infty$.

If $U_n$, $n \geq 0$ are nondecreasing real valued functions on $\mathbb{R}$ then it is a well known and useful fact that if $U_0$ is continuous and $U_n(x) \to U_0(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ then $U_n \to U_0$ locally uniformly; i.e., for any $a < b$

$$\sup_{x \in [a,b]} |U_n(x) - U_0(x)| \to 0. \tag{0.1}$$

One proof of this fact is outlined as follows: If $U_0$ is continuous on $[a,b]$, then it is uniformly continuous. For any $x$ there is an interval-neighborhood in $[a,b]$, $0_x$ containing $x$, on which $U$ oscillates by less than $\varepsilon$. This gives an open cover of $[a,b]$. Compactness allows us to prune $\{0_x, x \in [a,b]\}$ to obtain a finite subcover. Using this finite collection and the monotonicity of the functions leads straightaway to the desired uniform convergence.

Another proof of (0.1) is obtained by using the concept of continuous convergence (Kuratowski, 1966). Suppose $\mathscr{X}$, $\mathscr{Y}$ are two complete, separable metric spaces and $f_n: \mathscr{X} \to \mathscr{Y}$, $n \geq 0$. Then $f_n$ converges to $f$ continuously if whenever $x_n \in \mathscr{X}$, $n \geq 0$ and $x_n \to x_0$ we have $f_n(x_n) \to f_0(x_0)$.

The connection with uniform convergence is this: If $\mathscr{X}$ is compact and $f_0$ is continuous then $f_n \to f_0$ continuously iff $f_n \to f_0$ uniformly on $\mathscr{X}$. This equivalence sometimes provides a convenient way of proving uniform convergence since it allows us to prove convergence of a sequence of points in $\mathscr{Y}$ rather than having to deal with functions.

The equivalence of the two concepts is seen readily: Let $d$ be the metric on $\mathscr{Y}$. If $f_n \to f_0$ uniformly on $\mathscr{X}$ and $x_n \to x_0$ then we have

$$\begin{aligned} d(f_n(x_n), f_0(x_0)) &\leq d(f_n(x_n), f_0(x_n)) + d(f_0(x_n), f_0(x_0)) \\ &\leq \sup_{x \in \mathscr{X}} d(f_n(x), f_0(x)) + d(f_0(x_n), f_0(x_0)). \end{aligned}$$

The first term goes to zero as $n \to \infty$ by uniform convergence and the second term vanishes by continuity. Conversely suppose $f_n \to f_0$ continuously but not uniformly. Then there is a subsequence $\{n(k')\}$ and $\varepsilon > 0$ such that for all $n(k')$

$$\sup_{x \in \mathscr{X}} d(f_{n(k')}(x), f_0(x)) > 2\varepsilon.$$

Using the definition of sup we find points $\{x_{k'}\} \subset \mathscr{X}$ such that

$$d(f_{n(k')}(x_{k'}), f_0(x_{k'})) > \varepsilon. \tag{0.2}$$

Since $\mathscr{X}$ is assumed compact there is a limit point $x_0$ and a subsequence $\{x_k\} \subset \{x_{k'}\}$ with $x_k \to x_0$. Continuous convergence and continuity of $f_0$ require

$$d(f_{n(k)}(x_k), f_0(x_k)) \leq d(f_{n(k)}(x_k), f_0(x_0)) + d(f_0(x_0), f_0(x_k)) \to 0$$

in violation of (0.2). The contradiction occurs because we supposed $f_n$ did not converge to $f_0$ uniformly.

We now check (0.1) by using continuous convergence: Suppose $\{x_n, n \geq 0\} \subset [a, b]$ and $x_n \to x_0$. We check $U_n(x_n) \to U_0(x_0)$. It suffices to consider two cases: (a) $x_n > x_0$. (b) $x_n < x_0$. (If necessary, partition $\{x_n\}$ into two subsequences.) We consider only (a). The following are evident: There exists $\eta > 0$ such that

$$|U_0(x_0 + \eta) - U_0(x_0)| < \varepsilon \tag{0.3}$$

because $U_0$ is continuous. Furthermore there is $n_0$ such that if $n \geq n_0$

$$|x_n - x_0| < \eta \tag{0.4}$$

and

$$|U_n(x_0 + \eta) - U_0(x_0 + \eta)| \vee |U_n(x_0) - U_0(x_0)| < \varepsilon \tag{0.5}$$

since $U_n \to U_0$ pointwise. We then have for $n \geq n_0$ on the one hand

$$\begin{aligned} U_n(x_n) &\leq U_n(x_0 + \eta) && \text{(from (0.4))} \\ &\leq U_0(x_0 + \eta) + \varepsilon && \text{(from (0.5))} \\ &\leq U_0(x_0) + 2\varepsilon && \text{(from (0.3))} \end{aligned}$$

and on the other hand

$$\begin{aligned} U_n(x_n) &\geq U_n(x_0) && \text{(monotonicity)} \\ &\geq U_0(x_0) - \varepsilon && \text{(from (0.5))}. \end{aligned}$$

Continuous convergence follows.

If $F_n$, $n \geq 0$ are distribution functions on $\mathbb{R}$ (always understood to be nondefective) then $F_n \to F_0$ pointwise and $F_0$ continuous imply uniform convergence on $\mathbb{R}$. Local uniform convergence comes from (0.1), and off a large interval $[a, b]$ there is not much possibility of oscillation. Given $\varepsilon$ pick $b$ such that $F_0(b) > 1 - \varepsilon$ and there exists $n_0$ such that for $n \geq n_0$

$$|F_n(b) - F_0(b)| < \varepsilon$$

whence for $x \geq b$

$$|F_n(x) - F_n(b)| \leq 1 - F_n(b) \leq 1 - F_0(b) + |F_0(b) - F_n(b)| \leq 2\varepsilon$$

and therefore for $n \geq n_0$

$$\begin{aligned} \sup_{x>b} |F_n(x) - F_0(x)| &\leq \sup_{x>b} |F_n(x) - F_n(b)| \\ &\quad + |F_n(b) - F_0(b)| + |F_0(b) - F_0(x)| \\ &\leq 2\varepsilon + \varepsilon + \varepsilon. \end{aligned}$$

Similarly for $x < a$. Combined with uniform convergence on $[a, b]$ this gives convergence uniformly on $\mathbb{R}$.

Alternatively since $F_n(\infty) = 1$, $F_n(-\infty) = 0$ for all $n \geq 0$ we may compactify $\mathbb{R}$ and work on $[-\infty, \infty]$. If $F_n \to F_0$ pointwise on $[-\infty, \infty]$ and $F_0$ is continuous, local uniform convergence coincides with uniform convergence.

EXERCISES

0.1.1. Suppose $U_n^{(i)}$, $n \geq 0$ are real valued functions on $\mathbb{R}$ and as $n \to \infty$

$$U_n^{(i)} \to U_n^{(i)}$$

locally uniformly on $\mathbb{R}$ for $i = 1, 2$. Prove

(a) $U_n^{(1)} + U_n^{(2)} \to U_0^{(1)} + U_0^{(2)}$
locally uniformly.

(b) $U_n^{(1)} \cdot U_n^{(2)} \to U_0^{(1)} \cdot U_0^{(2)}$
locally uniformly.

(c) If $g\colon R \to R$ is bounded and continuous then $g(U_n^{(1)}) \to g(U_n^{(1)})$ locally uniformly.

Use continuous convergence.

## 0.2. Inverses of Monotone Functions

Suppose $H$ is a nondecreasing function on $\mathbb{R}$. With the convention that the infimum of an empty set is $+\infty$ we define the (left continuous) inverse of $H$ as

$$H^{\leftarrow}(y) = \inf\{s\colon H(s) \geq y\}.$$

To check $H^{\leftarrow}$ is left continuous at $x \in \mathbb{R}$, suppose $x_n \uparrow x$ but $H^{\leftarrow}(x_n) \uparrow H^{\leftarrow}(x-) < H^{\leftarrow}(x)$. Then there exist $\delta > 0$ and $y$ such that for all $n$

$$H^{\leftarrow}(x_n) < y < H^{\leftarrow}(x) - \delta.$$

The left inequality and the definition of $H^{\leftarrow}$ yield $H(y) \geq x_n$ for all $n$, and hence letting $n \to \infty$ we get $H(y) \geq x$ whence again by the definition of $H^{\leftarrow}$ we get $y \geq H^{\leftarrow}(x)$, which coupled with $y < H^{\leftarrow}(x) - \delta$ leads to the desired contradiction.

In case the function $H$ is right continuous we have the following interesting properties:

$$A(y) := \{s: H(s) \geq y\} \qquad \text{is closed} \tag{0.6a}$$

$$H(H^{\leftarrow}(y)) \geq y \tag{0.6b}$$

$$\begin{cases} H^{\leftarrow}(y) \leq t \text{ iff } y \leq H(t) \\ t < H^{\leftarrow}(y) \text{ iff } y > H(t). \end{cases} \tag{0.6c}$$

For (0.6a) observe if $s_n \in A(y)$ and $s_n \downarrow s$ then $y \leq H(s_n) \downarrow H(s)$ so $H(s) \geq y$ and $s \in A(y)$. If $s_n \uparrow s$ and $s_n \in A(y)$ then $y \leq H(s_n) \uparrow H(s-) \leq H(s)$ and $H(s) \geq y$ so $s \in A(y)$ again and $A(y)$ is closed. Since $A(y)$ is closed, $\inf A(y) \in A(y)$, i.e., $H^{\leftarrow}(y) \in A(y)$, which means $H(H^{\leftarrow}(y)) \geq y$. Last, (0.6c) follows from the definition of $H^{\leftarrow}$.

The probability integral transform follows: Let $([0,1], \mathscr{B}[0,1], m)$ be the Lebesgue probability space; $m$ is Lebesgue measure. Suppose $U$ is the identity function on $[0,1]$: i.e., $U$ is a uniformly distributed random variable. If $F$ is a distribution function (df) then $F^{\leftarrow}(U)$ is a random variable on $[0,1]$ with df $F$. This is readily checked: For $t \in \mathbb{R}$

$$\begin{aligned} m[F^{\leftarrow}(U) \leq t] &= m[U \leq F(t)] \qquad \text{(from (0.6c))} \\ &= F(t). \end{aligned}$$

A slight variant of this involves an exponential distribution rather than the uniform: Let $X$ be a real random variable with distribution $F$. Set $R = -\log(1 - F)$. If $P[E > x] = e^{-x}$, $x > 0$ then $R^{\leftarrow}(E)$ and $X$ have the same distribution which we write as

$$R^{\leftarrow}(E) \stackrel{\mathrm{d}}{=} X.$$

To check this is simple: For $x \in \mathbb{R}$

$$\begin{aligned} P[R^{\leftarrow}(E) > x] &= P[E > R(x)] \\ &= \exp\{-R(x)\} = 1 - F(x). \end{aligned}$$

We now discuss convergence of monotone functions. For any function $H$ denote

$$\mathscr{C}(H) = \{x \in \mathbb{R}: H \text{ is finite and continuous at } x\}.$$

A sequence $\{H_n, n \geq 0\}$ of nondecreasing functions on $\mathbb{R}$ converges weakly to

$H_0$ if as $n \to \infty$

$$H_n(x) \to H_0(x)$$

for all $x \in \mathscr{C}(H_0)$. We will denote this by $H_n \to H_0$, and no other form of convergence for monotone functions will be relevant. If $F_n$, $n \geq 0$ are (non-defective) df's then a myriad of names give equivalent concepts: complete convergence, vague convergence, weak* convergence, narrow convergence. If $\{X_n, n \geq 0\}$ are random variables and $X_n$ has df $F_n$, $n \geq 0$ then

$$X_n \Rightarrow X_0$$

means

$$F_n \to F_0.$$

**Proposition 0.1** (cf. Billingsley, 1979, page 287). *If $H_n$, $n \geq 0$ are nondecreasing functions and $H_n \to H_0$, then $H_n^{\leftarrow} \to H_0^{\leftarrow}$.*

Proof. Fix $\varepsilon > 0$ and $t \in \mathscr{C}(H_0^{\leftarrow})$. Since the discontinuities of the monotone function $H_0$ are at most countable, there exists $x \in (H_0^{\leftarrow}(t) - \varepsilon, H_0^{\leftarrow}(t))$ and $x \in \mathscr{C}(H_0)$. Since $x < H_0^{\leftarrow}(t)$ we have by definition of $H_0^{\leftarrow}$ that $H_0(x) < t$. Since $x \in \mathscr{C}(H_0)$ entails $H_n(x) \to H_0(x)$ we have for large $n$ $H_n(x) < t$, and again using the definition of inverse we get $x \leq H_n^{\leftarrow}(t)$ for large $n$. Thus

$$H_0^{\leftarrow}(t) - \varepsilon < x \leq H_n^{\leftarrow}(t)$$

for large $n$ implying, since $\varepsilon > 0$ is arbitrary, that

$$\liminf_{n \to \infty} H_n^{\leftarrow}(t) \geq H_0^{\leftarrow}(t).$$

(Note this half did not use $t \in \mathscr{C}(H_0^{\leftarrow})$.)

For a reverse inequality note that whenever $t' > t$ we may find $y \in \mathscr{C}(H_0)$ and

$$H_0^{\leftarrow}(t') < y < H_0^{\leftarrow}(t') + \varepsilon. \tag{0.7}$$

The left-hand inequality in (0.7) and the definition of the inverse give

$$H_0(y) \geq t' > t.$$

Since $y \in \mathscr{C}(H_0)$ we have $H_n(y) \to H_0(y)$, and so for large $n$, $H_n(y) \geq t$ and therefore $y \geq H_n^{\leftarrow}(t)$. From (0.7)

$$H_0^{\leftarrow}(t') + \varepsilon > y \geq H_n^{\leftarrow}(t)$$

for large $n$ and hence

$$\limsup_{n \to \infty} H_n^{\leftarrow}(t) \leq H_0^{\leftarrow}(t')$$

since $\varepsilon$ is arbitrary. Let $t' \downarrow t$ and use the continuity of $H_0^{\leftarrow}$ at $t$ to get

$$\limsup_{n \to \infty} H_n^{\leftarrow}(t) \leq H_0^{\leftarrow}(t).$$

This completes the sandwich and gives the desired result. $\square$

Proposition 0.1 allows us to prove easily the one-dimensional version of Skorohod's (1956) theorem relating convergence in distribution to almost sure convergence.

**Skorohod's Theorem.** *For $n \geq 0$ suppose $X_n$ is a real random variable on $(\Omega_n, \mathscr{B}_n, P_n)$ such that $X_n \Rightarrow X_0$. Then there exist random variables $\{\tilde{X}_n, n \geq 0\}$ defined on the Lebesgue probability space* $([0,1], \mathscr{B}[0,1], m)$ *such that*

(i) $\tilde{X}_n \stackrel{d}{=} X_n$ *for each* $n \geq 0$ *and*
(ii) $\tilde{X}_n \to \tilde{X}_0$ *almost surely with respect to m.*

By changing spaces and ignoring dependencies between $X$'s we get almost sure convergence. Note it is not true that $\{X_n, n \geq 0\} \stackrel{d}{=} \{\tilde{X}_n, n \geq 0\}$ as random elements of $\mathbb{R}^\infty$.

PROOF. Let $U$ be the identity function on $[0,1]$, so that $U$ is uniformly distributed with respect to $m$. Suppose the distribution of $X_n$ is $F_n$, $n \geq 0$ and define

$$\tilde{X}_n = F_n^{\leftarrow}(U).$$

The probability integral transform discussed previously shows $\tilde{X}_n \stackrel{d}{=} X_n$, $n \geq 0$ giving (i).

For (ii) note that $F_n \to F_0$ entails $F_n^{\leftarrow} \to F_0^{\leftarrow}$ by Proposition 0.1. Therefore

$$\begin{aligned} 1 &\geq m\{0 \leq u \leq 1 \colon \tilde{X}_n(u) \to \tilde{X}_0(u)\} \\ &= m\{u \colon F_n^{\leftarrow}(u) \to F_0^{\leftarrow}(u)\} \\ &\geq m(u \colon u \in \mathscr{C}(F_0^{\leftarrow})\} = 1 \end{aligned}$$

since the discontinuities of $F_0^{\leftarrow}$ are at most countable. □

EXERCISES

0.2.1. For a monotone function $U$, check

$$U_r^{\leftarrow}(y) := \inf\{s \colon U(s) > y\}$$

is right continuous. If $U$ is uniformly distributed on $[0,1]$, check $F_r^{\leftarrow}(U) \stackrel{d}{=} X$ where $X$ is a random variable with distribution $F(x)$.

0.2.2. If $U$ is monotone define

$$U^+(x) = \lim_{y \downarrow x} U(y)$$

$$U^-(x) = \lim_{y \uparrow x} U(y).$$

Verify:
(a) $(U^-)^{\leftarrow} = U^{\leftarrow}$
(b) $(U^{\leftarrow})^{\leftarrow} = U^-$
(c) If $U_n$, $n \geq 0$ are monotone then $U_n \to U_0$ implies $U_n^{\pm} \to U_0^{\pm}$.

0.2.3. Extend Proposition 0.1 to show $U_n \to U_0$ iff $U_n^{\leftarrow} \to U_0^{\leftarrow}$.

0.2.4. When is it true that
(a) $U(U^{\leftarrow}(t)) = t$
(b) $U^{\leftarrow}(U(t)) = t$?

## 0.3. Convergence to Types Theorem and Limit Distributions of Maxima

Two distribution functions $U(x)$ and $V(x)$ are of the same type if for some $A > 0$, $B \in \mathbb{R}$

$$V(x) = U(Ax + B)$$

for all $x$. For instance, $N(0, 1, x)$ (normal df with mean 0 and variance 1) is normal type as is $N(\mu, \sigma^2, x) = N(0, 1, \sigma^{-1}x - \sigma^{-1}\mu)$ for $\sigma > 0$ $\mu \in \mathbb{R}$. Affine transformations, weak convergence, and types are related as follows.

**Proposition 0.2.** *Suppose $U(x)$ and $V(x)$ are two distributions neither of which concentrates at a point.*
(a) *Suppose for $n \geq 1$ $F_n$ is a distribution, $a_n \geq 0$, $b_n \in \mathbb{R}$, $\alpha_n > 0$, $\beta_n \in \mathbb{R}$ and*

$$F_n(a_n x + b_n) \to U(x), \qquad F_n(\alpha_n x + \beta_n) \to V(x) \tag{0.8}$$

*weakly. Then*

$$\alpha_n/a_n \to A > 0, \qquad (\beta_n - b_n)/a_n \to B \in \mathbb{R} \tag{0.9}$$

*and*

$$V(x) = U(Ax + B). \tag{0.10}$$

*An equivalent formulation in terms of random variables*:
(a′) *Let $X_n$, $U$, $V$, $n \geq 1$ be random variables such that neither $U$ nor $V$ is almost surely constant. If*

$$(X_n - b_n)/a_n \Rightarrow U, \qquad (X_n - \beta_n)/\alpha_n \Rightarrow V \tag{0.8′}$$

*then* (0.9) *holds and*

$$V \stackrel{d}{=} (U - B)/A. \tag{0.10′}$$

(b) *If* (0.9) *holds then either of the two relations in* (0.8) (*or* (0.8′)) *implies the other and* (0.10) (*or* (0.10′)) *is true.*

Proof of (b). Suppose (0.9) holds and $Y_n := (X_n - b_n)/a_n \Rightarrow U$. We must show $(X_n - \beta_n)/\alpha_n \Rightarrow (U - B)/A$. By Skorohod's theorem there exist $\tilde{Y}_n$, $\tilde{U}$, $n \geq 1$, defined on $([0, 1], \mathscr{B}[0, 1], m)$ such that $\tilde{Y}_n \stackrel{d}{=} Y_n$, $n \geq 1$, $\tilde{U} \stackrel{d}{=} U$ and $\tilde{Y}_n \to \tilde{U}$ a.s. Define $\tilde{X}_n := a_n \tilde{Y}_n + b_n$ so $\tilde{X}_n \stackrel{d}{=} X_n$. Then

$$\begin{aligned}(X_n - \beta_n)/\alpha_n &\stackrel{d}{=} (\tilde{X}_n - \beta_n)/\alpha_n = (a_n/\alpha_n)\tilde{Y}_n + (b_n - \beta_n)/\alpha_n \\ &\to A^{-1}\tilde{U} - BA^{-1} \stackrel{d}{=} (U - B)/A\end{aligned}$$

so that $(X_n - \beta_n)/\alpha_n \Rightarrow (U - B)/A$.

PROOF OF (a). Using Proposition 0.1 the relations in (0.8) can be inverted to give

$$(F_n^{\leftarrow}(y) - b_n)/a_n \to U^{\leftarrow}(y), \qquad (F_n^{\leftarrow}(y) - \beta_n)/\alpha_n \to V^{\leftarrow}(y) \tag{0.11}$$

weakly. Since neither $U(x)$ nor $V(x)$ concentrates at one point we may find points $y_1, y_2$ satisfying

$$\begin{gathered} y_1, y_2 \in \mathscr{C}(U^{\leftarrow}) \cap \mathscr{C}(V^{\leftarrow}), \qquad y_1 < y_2, \\ U^{\leftarrow}(y_1) < U^{\leftarrow}(y_2) \quad \text{and} \quad V^{\leftarrow}(y_1) < V^{\leftarrow}(y_2). \end{gathered} \tag{0.12}$$

Therefore from (0.11) we have for $i = 1, 2$

$$(F_n^{\leftarrow}(y_i) - b_n)/a_n \to U^{\leftarrow}(y_i), \qquad (F_n^{\leftarrow}(y_i) - \beta_n)/\alpha_n \to V^{\leftarrow}(y_i) \tag{0.13}$$

and by subtraction

$$\begin{gathered} (F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1))/a_n \to U^{\leftarrow}(y_2) - U^{\leftarrow}(y_1) > 0, \\ (F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1))/\alpha_n \to V^{\leftarrow}(y_2) - V^{\leftarrow}(y_1) > 0. \end{gathered} \tag{0.14}$$

Divide the first relation in (0.14) by the second to obtain

$$\alpha_n/a_n \to (U^{\leftarrow}(y_2) - U^{\leftarrow}(y_1))/(V^{\leftarrow}(y_2) - V^{\leftarrow}(y_1)) =: A > 0.$$

Using this and (0.13) we get

$$(F_n^{\leftarrow}(y_1) - b_n)/a_n \to U^{\leftarrow}(y_1), \qquad (F_n^{\leftarrow}(y_1) - \beta_n)/a_n \to V^{\leftarrow}(y_1)A^{-1}$$

and so subtracting we obtain

$$(\beta_n - b_n)/a_n \to U^{\leftarrow}(y_1) - V^{\leftarrow}(y_1)A^{-1} =: B.$$

This gives (0.9) and (0.10) follows from (b). □

A nice by-product of this proof is that from (0.13) and (0.14) a suitable choice of the normalizing constants is

$$\begin{aligned} a_n &= F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1) \\ b_n &= F_n^{\leftarrow}(y_1). \end{aligned}$$

One of the nicest applications of the convergence to types result is to the derivation of the class of possible limit distributions for normalized maxima of iid random variables, and we now focus our attention on this result, which is one of the most basic in classical extreme value theory. Suppose $\{X_n, n \geq 1\}$ is an iid sequence of random variables with common distribution $F(x)$. Set $M_n = \bigvee_{i=1}^n X_i$. The distribution function of $M_n$ is $F^n(x)$ since

$$P[M_n \leq x] = P\left\{\bigcap_{i=1}^n [X_i \leq x]\right\} = \prod_{i=1}^n P[X_i \leq x] = F^n(x).$$

If we set

$$x_0 = \sup\{x\colon F(x) < 1\} \leq \infty, \tag{0.15}$$

then

$$\lim_{n\to\infty} \uparrow M_n = x_0 \qquad \text{a.s.} \tag{0.16}$$

To check this, observe that for $x < x_0$, $F(x) < 1$ and so

$$P[M_n \le x] = F^n(x) \to 0.$$

Therefore $M_n \xrightarrow{P} x_0$. Since $\{M_n\}$ is a nondecreasing sequence, convergence in probability implies convergence almost surely.

Analytic expressions for $F^n$ can be cumbersome even if $F$ is completely known; in statistical contexts that is often not the case. Just as the normal distribution is a useful approximation to the distribution of $\sum_{i=1}^n X_i$, we seek a limit distribution to act as an approximation to $F^n$. The relation (0.16) makes it clear that a nondegenerate limit distribution will not exist unless we normalize $M_n$. It is customary to use affine normalizations, which are also the most practical in statistical estimation problems.

**Proposition 0.3** (cf. Gnedenko, 1943; de Haan, 1970a, 1976; Weissman, 1975b). *Suppose there exist $a_n > 0$, $b_n \in \mathbb{R}$, $n \ge 1$ such that*

$$P[(M_n - b_n)/a_n \le x] = F^n(a_n x + b_n) \to G(x), \tag{0.17}$$

*weakly as $n \to \infty$ where $G$ is assumed nondegenerate. Then $G$ is of the type of one of the following three classes:*

(i) $\Phi_\alpha(x) = \begin{cases} 0 & x < 0 \\ \exp\{-x^{-\alpha}\} & x \ge 0 \end{cases}$
*for some $\alpha > 0$;*

(ii) $\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & x < 0 \\ 1 & x \ge 0 \end{cases}$
*for some $\alpha > 0$;*

(iii) $\Lambda(x) = \exp\{-e^{-x}\} \qquad x \in \mathbb{R}$.

*We refer to $\Phi_\alpha$, $\Psi_\alpha$, and $\Lambda$ as the extreme value distributions.*

Proof. For $t \in \mathbb{R}$ let as usual

$$[t] = \text{greatest integer less than or equal to } t.$$

From (0.17) we get for any $t > 0$.

$$F^{[nt]}(a_{[nt]}x + b_{[nt]}) \to G(x)$$

and also

$$F^{[nt]}(a_n x + b_n) = (F^n(a_n x + b_n))^{[nt]/n} \to G^t(x).$$

The convergence to types theorem applies and we are assured of the existence of $\alpha(t) > 0$, $\beta(t) \in \mathbb{R}$, $t > 0$ such that

$$\lim_{n\to\infty} a_n/a_{[nt]} = \alpha(t), \qquad \lim_{n\to\infty} (b_n - b_{[nt]})/a_{[nt]} = \beta(t) \tag{0.18}$$

and

$$G^t(x) = G(\alpha(t)x + \beta(t)). \tag{0.19}$$

From (0.18) it is immediately apparent that the functions $\alpha$, $\beta$ are measurable. For instance in the case of $\alpha$ we observe that limits of measurable functions are measurable so it suffices to check for a fixed $n$ that $a_n/a_{[n\cdot]}$ is measurable. But since the range of $a_{[n\cdot]}$ is the discrete set $\{a_j\}$ we have (assuming the $a_j$ are distinct)

$$\{t: a_{[nt]} = a_j\} = [jn^{-1}, (j+1)n^{-1})$$

and this amply demonstrates measurability.

Return to (0.19) and for $t > 0$, $s > 0$ we have on the one hand

$$G^{ts}(x) = G(\alpha(ts)x + \beta(ts))$$

and on the other

$$\begin{aligned} G^{ts}(x) &= (G^s(x))^t = G(\alpha(s)x + \beta(s))^t \\ &= G(\alpha(t)\{\alpha(s)x + \beta(s)\} + \beta(t)) \\ &= G(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)). \end{aligned}$$

Since $G$ is assumed nondegenerate we therefore conclude for $t > 0$, $s > 0$ (cf. Exercise 0.3.2):

$$\alpha(ts) = \alpha(t)\alpha(s) \tag{0.20}$$

$$\beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s), \tag{0.21}$$

the last step following by symmetry. We recognize (0.20) as the famous Hamel functional equation. The only finite measurable, nonnegative solution is of the following form (for example, see Seneta, 1976):

$$\alpha(t) = t^{-\theta}, \qquad \theta \in \mathbb{R}. \qquad \square$$

We now consider three cases: (a) $\theta = 0$, (b) $\theta > 0$, (c) $\theta < 0$.

*Case (a)*, $\theta = 0$. In this case $\alpha(t) \equiv 1$ and (0.21) becomes

$$\beta(ts) = \beta(t) + \beta(s),$$

a simple variant of the Hamel equation. The solution is of the form

$$\beta(t) = -c \log t, \qquad t > 0, \qquad c \in \mathbb{R}$$

and (0.19) is

$$G^t(x) = G(x - c \log t). \tag{0.22}$$

If $c$ were zero, it could not be the case that $G$ was nondegenerate. For any fixed $x$, $G^t(x)$ is nonincreasing in $t$, and we conclude therefore that $c > 0$.

If for some $x_0$, $G(x_0) = 1$ then from (0.22) we get

$$1 = G(x_0 - c \log t)$$

for all $t$, and changing variables gives $G(u) = 1$ for all $u$, a contradiction. Therefore $G(x) < 1$ for all $x$. Similarly it cannot be the case that $G(y) = 0$ for any $y$. Substitute $x = 0$ in (0.22) obtaining for $t > 0$

$$G^t(0) = G(-c \log t). \tag{0.23}$$

Set $\exp\{-e^{-p}\} = G(0) \in (0, 1)$ and $u = -c \log t$. Since the range of $t$ is $(0, \infty)$, the range of $u$ is $(-\infty, \infty)$ and changing variables in (0.23) gives

$$\begin{aligned} G(u) &= \exp\{-e^{-p}t\} = \exp\{-e^{-(c^{-1}u+p)}\} \\ &= \Lambda(c^{-1}u + p). \end{aligned}$$

*Case (b)*, $\theta > 0$. From (0.21)

$$\alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s)$$

so that $(t \neq 1, s \neq 1)$

$$\frac{\beta(s)}{1 - \alpha(s)} = \frac{\beta(t)}{1 - \alpha(t)},$$

i.e., the function $\beta(\cdot)(1 - \alpha(\cdot))^{-1}$ is constant equal to $c$ say. Therefore for $t \neq 1$

$$\begin{aligned} \beta(t) &= \beta(s)(1 - \alpha(s))^{-1}(1 - \alpha(t)) \\ &= c(1 - t^{-\theta}) \end{aligned}$$

and (0.19) becomes

$$\begin{aligned} G^t(x) &= G(t^{-\theta}x + c(1 - t^{-\theta})) \\ &= G(t^{-\theta}(x - c) + c); \end{aligned}$$

i.e., changing variables

$$G^t(x + c) = G(t^{-\theta}x + c).$$

Set $H(x) = G(x + c)$. Then $G$ and $H$ are of the same type so it suffices to solve for $H$. The function $H$ satisfies

$$H^t(x) = H(t^{-\theta}x) \tag{0.24}$$

and $H$ is nondegenerate. Set $x = 0$ and we get from (0.24), $t \log H(0) = \log H(0)$ for $t > 0$ so either $\log H(0) = 0$ or $-\infty$; i.e., either $H(0) = 0$ or 1. However, $H(0) = 1$ is impossible since it would imply the existence of $x < 0$ such that the left side of (0.24) is decreasing in $t$ while the right side of (0.24) is increasing in $t$. Therefore we conclude $H(0) = 0$.

Again from (0.24) we obtain $H^t(1) = H(t^{-\theta})$. If $H(1) = 0$ then $H \equiv 0$ and if $H(1) = 1$ then $H \equiv 1$, both statements contradicting $H$ nondegenerate. Therefore, $H(1) \in (0, 1)$. Set $\theta^{-1} = \alpha$, $H(1) = \exp\{-p^{-\alpha}\}$, $u = t^{-\theta}$ so that $u^{-\alpha} = t$. From (0.24) with $x = 1$ we get for $u > 0$

$$\begin{aligned} H(u) &= \exp\{-p^{-\alpha}t\} = \exp\{-(pu)^{-\alpha}\} \\ &= \Phi_\alpha(pu). \end{aligned}$$

*Case (c)*, $\theta < 0$. That this case leads to the type of $\Psi_\alpha$ is left as an exercise.

EXERCISES

0.3.1. Verify in the proof of Theorem 0.3 that the case $\theta < 0$ leads to type $\Psi_\alpha$.

0.3.2. The derivation in Proposition 0.3 uses the following fact: If $F$ is a nondegenerate distribution and $a > 0$, $c > 0$, $b \in \mathbb{R}$, $d \in \mathbb{R}$, and

$$F(ax + b) = F(cx + d)$$

then $a = c$ and $b = d$. Prove this two ways by
(a) Considering inverse functions;
(b) Showing it is enough to prove $F(Ax + B) = F(x)$ implies $A = 1$, $B = 0$ by iterating $F(T(x)) = F(x)$ (i.e., replacing $x$ by $(T(x)$ again and again) where $T(x) = Ax + B$.

0.3.3. Suppose $\{X_n, n \geq 0\}$ are iid and there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that for some $G$ nondegenerate

$$P\left[\sum_{i=1}^{n} X_i - b_n)/a_n \leq x\right] \to G(x)$$

as $n \to \infty$. Derive a functional equation for the characteristic function of $G$.

0.3.4. Suppose $Y_n$, $n \geq 1$ are random variables such that there exist $a_n > 0$, $b_n \in \mathbb{R}$ and

$$P[Y_n \leq a_n x + b_n] \to G(x),$$

nondegenerate and for each $t > 0$

$$P[Y_{[nt]} \leq a_n x + b_n] \to G_t(x)$$

nondegenerate. Then there exists $\beta(t) > 0$, $\alpha(t) \in \mathbb{R}$ such that

$$G(x) = G_t(\beta(t)x + \alpha(t))$$

and $\beta(t) = t^\theta$ and if $\theta = 0$ then $\alpha(t) = c \log t$ and if $\theta \neq 0$ then $\alpha(t) = c(1 - t^\theta)$ (Weissman, 1975b).

## 0.4. Regularly Varying Functions of a Real Variable

Having established what the possible limit laws for normalized maxima are we must next give criteria for convergence to each type. If $G$ is an extreme value distribution we say a distribution $F$ is in the domain of attraction of $G$ (written $F \in D(G)$) if there exists $a_n > 0$, $b_n$, $n \geq 1$ such that

$$F^n(a_n x + b_n) \to G(x)$$

weakly. It is also important to characterize $\{a_n\}$ and $\{b_n\}$.

Such domains of attraction questions are best understood within the framework of the theory of regularly varying functions, so before continuing with extreme value theory we pause for a brief account of regular variation. Mastering this subject is important for proper understanding of extreme value theory. The goal of this treatment is to provide a reader with a functional understanding of basics. It is not intended to be exhaustively complete, and in some cases proofs deal only with special cases. Pay particular attention to the material on de Haan's classes $\Pi$ and $\Gamma$, which are not nearly as well known as they deserve to be. Further information can be found in the following excellent references: Seneta (1976), de Haan (1970), Feller (1971), and Bingham, Goldie, and Teugels (1987).

## 0.4.1. Basics

Roughly speaking, *regularly varying functions* are those functions which behave asymptotically like power functions.

**Definition.** A measurable function $U: \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at $\infty$ with index $\rho$ (written $U \in RV_\rho$) if for $x > 0$

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho.$$

We call $\rho$ the *exponent of variation*. With obvious changes we may speak about regular variation at 0. The theories are equivalent: $U(x)$ is regularly varying at $\infty$ iff $U(x^{-1})$ is regularly varying at 0.

If $\rho = 0$ we call $U$ *slowly varying*. Slowly varying functions are generically denoted by $L(x)$. If $U \in RV_\rho$ then $U(x)/x^\rho \in RV_0$ and setting $L(x) = U(x)/x^\rho$ we see it is always possible to represent a $\rho$-varying function as $x^\rho L(x)$.

EXAMPLES. The canonical $\rho$-varying function is $x^\rho$. The functions $\log(1 + x)$, $\log\log(e + x)$ are slowly varying, as is $\exp\{(\log x)^\alpha\}$, $0 < \alpha < 1$. Any function $U$ such that $\lim_{x\to\infty} U(x) =: U(\infty)$ exists finite is slowly varying. The following functions are not regularly varying: $e^x$, $\sin(x + 2)$. Note $[\log x]$ is slowly varying, but $\exp\{[\log x]\}$ is not regularly varying.

In probability applications we are concerned with distributions whose tails are regularly varying. Examples are

$$1 - F(x) = x^{-\alpha}, \qquad x \geq 1, \qquad \alpha > 0,$$

and

$$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \qquad x \geq 0.$$

$\Phi_\alpha(x)$ has the property

$$1 - \Phi_\alpha(x) \sim x^{-\alpha} \quad \text{as } x \to \infty.$$

(Note, we write $g(x) \sim h(x)$ as $x \to \infty$ to mean $\lim_{x\to\infty} g(x)/h(x) = 1$.) A stable law with index $\alpha$, $0 < \alpha < 2$ has the property

$$1 - G(x) \sim cx^{-\alpha}, \qquad x \to \infty, \qquad c > 0$$

and more particularly the Cauchy density $f(x) = (\pi(1 + x^2))^{-1}$ has a df $F$ with the property

$$1 - F(x) \sim (\pi x)^{-1}.$$

If $N(x)$ is the standard normal df then $1 - N(x)$ is not regularly varying nor is $1 - \Lambda(x)$.

The definition of regular variation can be weakened slightly.

**Proposition 0.4** (de Haan, 1970; Feller, 1971).
(i) *A measurable function $U\colon R_+ \to R_+$ varies regularly if there exists a function $h$ such that for all $x > 0$*

$$\lim_{t\to\infty} U(tx)/U(t) = h(x).$$

*In this case $h(x) = x^\rho$ for some $\rho \in \mathbb{R}$ and $U \in RV_\rho$.*
(ii) *A monotone function $U\colon R_+ \to R_+$ varies regularly provided there are two sequences $\{\lambda_n\}$, $\{a_n\}$ of positive numbers satisfying*

$$\lambda_n \sim \lambda_{n+1} \quad as\ n \to \infty \tag{0.25}$$

$$a_n \to \infty \tag{0.26}$$

*and for all $x > 0$*

$$\lim_{n\to\infty} \lambda_n U(a_n x) =: \chi(x) \text{ exists positive and finite.} \tag{0.27}$$

*In this case $\chi(x)/\chi(1) = x^\rho$ and $U \in RV_\rho$ for some $\rho \in \mathbb{R}$.*

Proof. (i) The function $h$ is measurable since it is a limit of measurable functions. Then for $x > 0$, $y > 0$

$$\frac{U(txy)}{U(t)} = \frac{U(txy)}{U(tx)} \cdot \frac{U(tx)}{U(t)}$$

and letting $t \to \infty$ gives

$$h(xy) = h(y)h(x).$$

So $h$ satisfies the Hamel equation and is of the form $h(x) = x^\rho$ for some $\rho \in \mathbb{R}$.
(ii) For concreteness assume $U$ is nondecreasing. Since $a_n \to \infty$, for each $t$ there is a finite $n(t)$ defined by

$$n(t) = \inf\{m\colon a_{m+1} > t\}$$

so that

$$a_{n(t)} \le t < a_{n(t)+1}.$$

Therefore by monotonicity for $x > 0$

$$\left(\frac{\lambda_{n(t)+1}}{\lambda_{n(t)}}\right)\left(\frac{\lambda_{n(t)} U(a_{n(t)}x)}{\lambda_{n(t)+1} U(a_{n(t)+1})}\right) \le \frac{U(tx)}{U(t)} \le \left(\frac{\lambda_{n(t)}}{\lambda_{n(t)+1}}\right)\left(\frac{\lambda_{n(t)+1} U(a_{n(t)+1}x)}{\lambda_{n(t)} U(a_{n(t)})}\right).$$

Now let $t \to \infty$ and use (0.25) and (0.27) to get $\lim_{t\to\infty} \dfrac{U(tx)}{U(t)} = 1\dfrac{\chi(x)}{\chi(1)}$. The rest follows from part (i). □

EXERCISE 0.4.1.1. Proposition 0.4 (ii) remains true if we only assume (0.27) holds on a dense set. This is relevant to the case where $U$ is nondecreasing and $\lambda_n U(a_n x)$ converges weakly.

EXAMPLE. Suppose $X_n$, $n \geq 1$ are iid with common df $F(x)$. Find conditions on $F$ so that there exists $a_n > 0$ such that

$$P[a_n^{-1} M_n \leq x] = F^n(a_n x) \to \Phi_\alpha(x) \tag{0.28}$$

weakly and characterize $\{a_n\}$.

Set $x_0 = \sup\{x\colon F(x) < 1\}$ and we first check $x_0 = \infty$. Otherwise if $x_0 < \infty$ we get from (0.28) that for $x > 0$, $a_n x \to x_0$; i.e., $a_n \to x_0 x^{-1}$. Since $x > 0$ is arbitrary we get $a_n \to 0$ whence $x_0 = 0$. But then for $x > 0$, $F^n(a_n x) = 1$, which violates (0.28). Hence $x_0 = \infty$. Furthermore $a_n \to \infty$ since otherwise on a subsequence $n'$, $a_{n'} \leq K$ for some $K < \infty$ and

$$0 < \Phi_\alpha(1) = \lim_{n' \to \infty} F^{n'}(a_{n'}) \leq \lim_{n' \to \infty} F^{n'}(K) = 0$$

(since $F(K) < 1$) which is a contradiction.

In (0.28) take logarithms to get for $x > 0$, $\lim_{n \to \infty} n(-\log F(a_n x)) = x^{-\alpha}$. Now use the relation $-\log(1 - z) \sim z$ as $z \to 0$ and (0.28) is equivalent to

$$\lim_{n \to \infty} n(1 - F(a_n x)) = x^{-\alpha}, \qquad x > 0. \tag{0.29}$$

From (0.29) and Proposition 0.4 (ii) we get

$$1 - F(x) \sim x^{-\alpha} L(x), \qquad x \to \infty, \quad \text{for some } \alpha > 0. \tag{0.30}$$

To characterize $\{a_n\}$ set $U(x) = 1/(1 - F(x))$ and (0.29) is the same as

$$U(a_n x)/n \to x^\alpha, \qquad x > 0$$

and inverting we find via Proposition 0.1 that

$$\frac{U^{\leftarrow}(ny)}{a_n} \to y^{1/\alpha}, \qquad y > 0.$$

So $U^{\leftarrow}(n) = (1/(1 - F))^{\leftarrow}(n) \sim a_n$ and this determines $a_n$.

Conversely if (0.30) holds, define $a_n = U^{\leftarrow}(n)$ as previously. Then

$$\lim_{n \to \infty} \frac{1 - F(a_n x)}{1 - F(a_n)} = x^{-\alpha}$$

and we recover (0.29) provided $1 - F(a_n) \sim n^{-1}$ or what is the same, provided $U(a_n) \sim n$; i.e., $U(U^{\leftarrow}(n)) \sim n$. Recall from (0.6c) that $z < U^{\leftarrow}(n)$ iff $U(z) < n$ and setting $z = U^{\leftarrow}(n)(1 - \varepsilon)$ and then $z = U^{\leftarrow}(n)(1 + \varepsilon)$ we get

$$\frac{U(U^{\leftarrow}(n))}{U(U^{\leftarrow}(n)(1 + \varepsilon))} \leq \frac{U(U^{\leftarrow}(n))}{n} \leq \frac{U(U^{\leftarrow}(n))}{U(U^{\leftarrow}(n)(1 - \varepsilon))}.$$

Let $n \to \infty$ remembering $U = 1/(1 - F) \in RV_\alpha$. Then

$$(1 + \varepsilon)^{-\alpha} \le \liminf_{n\to\infty} n^{-1} U(U^{\leftarrow}(n)) \le \limsup_{n\to\infty} \le (1 - \varepsilon)^{-\alpha}$$

and since $\varepsilon > 0$ is arbitrary the desired result follows.

EXERCISES

0.4.1.2. Say that $1 - F$ is rapidly varying (de Haan, 1970), written $1 - F(x) \sim x^{-\infty}L(x)$ as $x \to \infty$, if

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\infty} := \begin{cases} 0 & \text{if} \quad x > 1 \\ \infty & \text{if} \quad 0 < x < 1. \end{cases}$$

Let $\{X_n, n \ge 1\}$ be iid random variables with common distribution $F(x)$. Prove the following result about relative stability (Gnedenko, 1943; de Haan, 1970): There exist constants $0 < b_n \uparrow \infty$ such that

$$M_n/b_n \xrightarrow{P} 1$$

iff $1 - F(x) \sim x^{-\infty}L(x)$. Also, give a characterization of $b_n$.

0.4.1.3. Let $\{X_n, n \ge 1\}$ be iid nonnegative random variables with $Ee^{-\lambda X_1} = \hat{F}(\lambda), \lambda > 0$ as the common Laplace transform. Give necessary and sufficient conditions on $\hat{F}$ for there to exist $0 < a_n \uparrow \infty$ such that

$$\sum_{i=1}^{n} X_i/a_n \xrightarrow{P} 1$$

and give a characterization of $a_n$. Also give necessary and sufficient conditions on $\hat{F}$ for there to exist $0 < a_n \uparrow \infty$ and a nondegenerate distribution $G(x)$ such that

$$P\left[\sum_{i=1}^{n} X_i/a_n \le x\right] \to G(x).$$

Characterize $a_n$ and the Laplace transform $\hat{G}(\lambda)$ (Feller, 1971).

0.4.1.4. $L: R_+ \to R_+$ is slowly varying iff

$$\lim_{t\to\infty} L(tx)/L(t) = 1$$

for all $x \ge 1$. If $L: R_+ \to R_+$ is monotone, then $L$ is slowly varying iff there exists one $x > 0$, $x \ne 1$, for which

$$\lim_{t\to\infty} L(tx)/L(t) = 1.$$

Show the last result is false without the assumption of monotonicity.

### 0.4.2. Deeper Results; Karamata's Theorem

The first result which is very useful is the uniform convergence theorem. (Cf. Exercise 0.4.2.1.)

**Proposition 0.5.** *If $U \in RV_\rho$ then*

$$\lim_{t\to\infty} U(tx)/U(t) = x^\rho$$

*locally uniformly on* $(0, \infty)$. *If $\rho < 0$ then uniform convergence holds on intervals of the form* $(b, \infty)$, $b > 0$. *If $\rho > 0$ uniform convergence holds on intervals* $(0, b]$ *provided $U$ is bounded on* $(0, b]$ *for all* $b > 0$.

If $U$ is monotone the result already follows from (0.1) since monotone functions are converging to a continuous limit. The extensions when $\rho < 0$ or $\rho > 0$ follow as in the end of Section 0.1. For the nonmonotone case see Seneta (1976).

The next set of results examines the integral properties of regularly varying functions. For purposes of integration, a $\rho$-varying function behaves roughly like $x^\rho$. We assume all functions are locally integrable and since we are interested in behavior at $\infty$ we assume integrability on intervals including 0 as well.

**Karamata's Theorem 0.6.** (a) *If $\rho \geq -1$ then $U \in RV_\rho$ implies $\int_0^x U(t)dt \in RV_{\rho+1}$ and*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \rho + 1. \tag{0.31}$$

*If $\rho < -1$ (or if $\rho = -1$ and $\int_x^\infty U(s)ds < \infty$) then $U \in RV_\rho$ implies $\int_x^\infty U(t)dt$ is finite, $\int_x^\infty U(t)dt \in RV_{\rho+1}$ and*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = -\rho - 1. \tag{0.32}$$

(b) *If $U$ satisfies*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \lambda \in (0, \infty) \tag{0.33}$$

*then $U \in RV_{\lambda-1}$. If $\int_x^\infty U(t)dt < \infty$ and*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = \lambda \in (0, \infty) \tag{0.34}$$

*then $U \in RV_{-\lambda-1}$.*

**Corollary** (The Karamata Representation). *$L$ is slowly varying* iff *$L$ can be represented as*

$$L(x) = c(x)\exp\left\{\int_1^x t^{-1}\varepsilon(t)dt\right\} \tag{0.35}$$

*for $x > 0$ where $c$: $R_+ \to R_+$, $\varepsilon$: $R_+ \to R_+$ and*

$$\lim_{x\to\infty} c(x) = c \in (0, \infty) \tag{0.36}$$

$$\lim_{t\to\infty} \varepsilon(t) = 0. \tag{0.37}$$

PROOF OF COROLLARY. If $L$ has a representation (0.35) then it must be slowly varying since for $x > 1$

$$\lim_{t\to\infty} L(tx)/L(t) = \lim_{t\to\infty} (c(tx)/c(t))\exp\left\{\int_t^{tx} s^{-1}\varepsilon(s)ds\right\}.$$

Given $\varepsilon$, there exists $t_0$ by (0.37) such that

$$-\varepsilon < \varepsilon(t) < \varepsilon, \qquad t \geq t_0,$$

so that

$$-\varepsilon \log x = -\varepsilon \int_t^{tx} s^{-1}\,ds \leq \int_t^{tx} s^{-1}\varepsilon(s)ds \leq \varepsilon \int_t^{tx} s^{-1}\,ds = \varepsilon \log x.$$

Therefore $\lim_{t\to\infty} \int_t^{tx} s^{-1}\varepsilon(s)ds = 0$ and $\lim_{t\to\infty} L(tx)/L(t) = 1$.

Conversely suppose $L \in RV_0$. By Karamata's theorem

$$b(x) := xL(x)\Big/\int_0^x L(s)ds \to 1$$

and $x \to \infty$. Note

$$L(x) = x^{-1}b(x)\int_0^x L(s)ds.$$

Set $\varepsilon(x) = b(x) - 1$ so $\varepsilon(x) \to 0$ and

$$\begin{aligned}\int_1^x t^{-1}\varepsilon(t)dt &= \int_1^x \left(L(t)\Big/\int_0^t L(s)ds\right)dt - \log x\\ &= \int_1^x d\left(\log \int_0^t L(s)ds\right) - \log x\\ &= \log\left(x^{-1}\int_0^x L(s)ds\Big/\int_0^1 L(s)ds\right)\end{aligned}$$

whence

$$\begin{aligned}\exp\left\{\int_1^x t^{-1}\varepsilon(t)dt\right\} &= x^{-1}\int_0^x L(s)ds\Big/\int_0^1 L(s)ds\\ &= L(x)\Big/\left(b(x)\int_0^1 L(s)ds\right)\end{aligned} \tag{0.38}$$

and the representation follows with

$$c(x) = b(x)\int_0^1 L(s)ds.$$

*Remark.* If $U \in RV_\rho$ then $U$ has the representation

$$U(x) = c(x)\exp\left\{\int_1^x t^{-1}\rho(t)dt\right\}$$

where $c(\cdot)$ satisfies (0.36) and $\lim_{t\to\infty}\rho(t) = \rho$. This is obtained from the corollary by writing $U(x) = x^\rho L(x)$ and using the representation for $L$.

Proof of Theorem 0.6(a). For certain values of $\rho$, uniform convergence suffices. If we wish to proceed, using elementary concepts, consider the following approach, which follows de Haan (1970).

If $\rho > -1$ we show $\int_0^\infty U(t)dt = \infty$. From $U \in RV_\rho$ we have $\lim_{s\to\infty} U(2s)/U(s) = 2^\rho > 2^{-1}$ since $\rho > -1$. Therefore there exists $s_0$ such that $s > s_0$ necessitates $U(2s) > 2^{-1}U(s)$. For $n$ with $2^n > s_0$ we have

$$\int_{2^{n+1}}^{2^{n+2}} U(s)ds = 2\int_{2^n}^{2^{n+1}} U(2s)ds > \int_{2^n}^{2^{n+1}} U(s)ds$$

and so setting $n_0 = \inf\{n: 2^n > s_0\}$ gives

$$\int_{s_0}^\infty U(s)ds \geq \sum_{n:2^n>s_0}\int_{2^{n+1}}^{2^{n+2}} U(s)ds > \sum_{n\geq n_0}\int_{2^{n_0+1}}^{2^{n_0+2}} U(s)ds = \infty.$$

Thus for $\rho > -1$, $x > 0$, and any $N < \infty$ we have $\int_0^t U(sx)ds \sim \int_N^t U(sx)ds$, $t \to \infty$, since $U(sx)$ is a $\rho$-varying function of $s$. For fixed $x$ and given $\varepsilon$, there exists $N$ such that for $s > N$

$$(1-\varepsilon)x^\rho U(s) \leq U(sx) \leq (1+\varepsilon)x^\rho U(s)$$

and thus

$$\begin{aligned}
\limsup_{t\to\infty}\frac{\int_0^{tx} U(s)ds}{\int_0^t U(s)ds} &= \limsup_{t\to\infty}\frac{x\int_0^t U(sx)ds}{\int_0^t U(s)ds}\\
&= \limsup_{t\to\infty}\frac{x\int_N^t U(sx)ds}{\int_N^t U(s)ds} \leq \limsup_{t\to\infty} x^{\rho+1}(1+\varepsilon)\frac{\int_N^t U(s)ds}{\int_N^t U(s)ds}\\
&= (1+\varepsilon)x^{\rho+1}.
\end{aligned}$$

An analogous argument applies for liminf and thus we have proved $\int_0^x U(s)ds \in RV_{\rho+1}$ when $\rho > -1$.

In case $\rho = -1$ then either $\int_0^\infty U(s)ds < \infty$ in which case $\int_0^x U(s)ds \in RV_{-1+1} = RV_0$ or else $\int_0^\infty U(s)ds = \infty$ and the previous argument is applicable. So we have checked that for $\rho \geq -1$, $\int_0^x U(s)ds \in RV_{\rho+1}$.

We now focus on proving (0.31) when $U \in RV_\rho$, $\rho \geq -1$. As in the development leading to (0.38), set

$$b(x) = xU(x)\bigg/\int_0^x U(t)dt$$

so that integrating $b(x)/x$ leads to the representations

$$\int_0^x U(s)ds = c\exp\left\{\int_1^x t^{-1}b(t)dt\right\}$$
$$U(x) = cx^{-1}b(x)\exp\left\{\int_1^x t^{-1}b(t)dt\right\}. \tag{0.39}$$

We must show $b(x) \to \rho + 1$. Observe first that

$$\begin{aligned}\liminf_{x\to\infty} 1/b(x) &= \liminf_{x\to\infty} \frac{\int_0^x U(t)dt}{xU(x)} \\ &= \liminf_{x\to\infty} \int_0^1 \frac{U(sx)}{U(x)}ds \qquad \text{(change variable } s = x^{-1}t)\end{aligned}$$

and by Fatou's lemma this is

$$\geq \int_0^1 \liminf_{x\to\infty} (U(sx)/U(x))ds = \int_0^1 s^\rho\, ds = \frac{1}{\rho+1}$$

and we conclude

$$\limsup_{x\to\infty} b(x) \leq \rho + 1. \tag{0.40}$$

If $\rho = -1$ then $b(x) \to 0$ as desired, so now suppose $\rho > -1$.

We observe the following properties of $b(x)$

(i) $b(x)$ is bounded on a semi-infinite neighborhood of $\infty$ (by 0.40),
(ii) $b$ is slowly varying since $xU(x) \in RV_{\rho+1}$ and $\int_0^x U(s)ds \in RV_{\rho+1}$,
(iii) $b(xt) - b(x) \to 0$ boundedly as $x \to \infty$.

The last statement follows since by slow variation

$$\lim_{x\to\infty} (b(xt) - b(x))/b(x) = 0$$

and the denominator is ultimately bounded.

From (iii) and dominated convergence

$$\lim_{x\to\infty} \int_1^s t^{-1}(b(xt) - b(x))dt = 0$$

and the left side may be rewritten to obtain

$$\lim_{x\to\infty} \left\{\int_1^s t^{-1}b(xt)dt - b(x)\log s\right\} = 0. \tag{0.41}$$

From (0.39)

$$c\exp\left\{\int_1^x t^{-1}b(t)dt\right\} = \int_0^x U(s)ds \in RV_{\rho+1}$$

and from the regular variation property

$$(\rho + 1)\log s = \lim_{x\to\infty} \log\left\{\frac{\int_0^{xs} U(t)dt}{\int_0^x U(t)dt}\right\}$$

$$= \lim_{x\to\infty} \int_x^{xs} t^{-1}b(t)dt = \lim_{x\to\infty} \int_1^s t^{-1}b(xt)dt$$

and combining this with (0.41) leads to the desired conclusion that $b(x) \to \rho + 1$.

Proof of (b). We suppose (0.33) holds and check $U \in RV_{\lambda-1}$. Set

$$b(x) = xU(x)\Big/\int_0^x U(t)dt$$

so that $b(x) \to \lambda$. From (0.39)

$$U(x) = cx^{-1}b(x)\exp\int_1^x t^{-1}b(t)dt\Big\}$$

$$= cb(x)\exp\left\{\int_1^x t^{-1}(b(t) - 1)dt\right\}$$

and since $b(t) - 1 \to \lambda - 1$ we see that $U$ satisfies the representation of a $(\lambda - 1)$-varying function. □

The previous results described the effect of integration on a regularly varying function. We now describe what happens when a $\rho$-varying function is differentiated.

**Proposition 0.7.** *Suppose $U: R_+ \to R_+$ is absolutely continuous with density $u$ so that*

$$U(x) = \int_0^x u(t)dt.$$

(a) *Von Mises: If*

$$\lim_{x\to\infty} xu(x)/U(x) = \rho, \tag{0.42}$$

*then $U \in RV_\rho$.*
(b) *Landau,* 1916; *de Haan,* 1970, *p.* 23, 109; *Seneta,* 1973, *p.* 1057: *If $U \in RV_\rho$, $\rho \in \mathbb{R}$, and $u$ is monotone then* (0.42) *holds and if $\rho \neq 0$ then* $(\operatorname{sgn}\rho)u(x) \in RV_{\rho-1}$.

Proof. (a) Set

$$b(x) = xu(x)/U(x)$$

and as before we find

$$U(x) = U(1)\exp\left\{\int_1^x t^{-1}b(t)dt\right\}$$

so that $U$ satisfies the representation theorem for a $\rho$-varying function.
(b) Suppose $u$ is nondecreasing. An analogous proof works in the case $u$ is nonincreasing. Let $0 < a < b$ and observe

$$(U(xb) - U(xa))/U(x) = \int_{xa}^{xb} u(y)dy/U(x).$$

By monotonicity we get

$$u(xb)x(b-a)/U(x) \geq (U(xb) - U(xa))/U(x) \geq u(xa)x(b-a)/U(x). \tag{0.43}$$

From (0.43) and the fact that $U \in RV_\rho$ we conclude

$$\limsup_{x\to\infty} xu(xa)/U(x) \leq (b^\rho - a^\rho)/(b-a) \tag{0.44}$$

for any $b > a > 0$. So let $b \downarrow a$, which is tantamount to taking a derivative. Then (0.44) becomes

$$\limsup_{x\to\infty} xu(xa)/U(x) \leq \rho a^{\rho-1} \tag{0.45}$$

for any $a > 0$. Similarly from the left-hand equality in (0.43) after letting $a \uparrow b$ we get

$$\liminf_{x\to\infty} xu(xb)/U(x) \geq \rho b^{\rho-1} \tag{0.46}$$

for any $b > 0$. Then (0.42) results by setting $a = 1$ in (0.45) and $b = 1$ in (0.46). □

Say $U\colon R_+ \to R_+$ is regularly varying with index $\infty$ ($U \in RV_\infty$) if for every $x > 0$

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\infty := \begin{cases} 0 & x < 1 \\ 1 & x = 1 \\ \infty & x > 1. \end{cases}$$

(Cf. Exercise 0.4.1.2.) Similarly $U \in RV_{-\infty}$ if

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{-\infty} := \begin{cases} \infty & x < 1 \\ 1 & x = 1 \\ 0 & x > 1. \end{cases}$$

The following proposition collects useful properties of regularly varying functions and is modeled after the list in de Haan (1970).

**Proposition 0.8.** (i) *If* $U \in RV_\rho$, $-\infty \leq \rho \leq \infty$, *then* $\lim_{x\to\infty} \log U(x)/\log x = \rho$ *so that*

$$\lim_{x\to\infty} U(x) = \begin{cases} 0 & \text{if } \rho < 0 \\ \infty & \text{if } \rho > 0. \end{cases}$$

(ii) *Suppose* $U \in RV_\rho$, $\rho \in \mathbb{R}$. *Take* $\varepsilon > 0$. *Then there exists* $t_0$ *such that for* $x \geq 1$

*and* $t \geq t_0$

$$(1-\varepsilon)x^{\rho-\varepsilon} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\rho+\varepsilon}.$$

(iii) *If* $U \in RV_\rho$, $\rho \in \mathbb{R}$, *and* $\{a_n\}$, $\{a_n'\}$ *satisfy*, $0 < a_n \to \infty$, $0 < a_n' \to \infty$, *and* $a_n \sim a_n' c$, $0 < c < \infty$, *then* $U(a_n) \sim c^\rho U(a_n')$. *If* $\rho \neq 0$ *the result also holds for* $c = 0$ *or* $\infty$. *Analogous results hold with sequences replaced by functions.*
(iv) *If* $U_1 \in RV_{\rho_1}$ *and* $U_2 \in RV_{\rho_2}$ *and* $\lim_{x\to\infty} U_2(x) = \infty$ *then*

$$U_1 \circ U_2 \in RV_{\rho_1\rho_2}.$$

(v) *Suppose* $U$ *is nondecreasing*, $U(\infty) = \infty$, *and* $U \in RV_\rho$, $0 \leq \rho \leq \infty$. *Then*

$$U^{\leftarrow} \in RV_{\rho^{-1}}.$$

(vi) *Suppose* $U_1$, $U_2$ *are nondecreasing and* $\rho$*-varying*, $0 < \rho < \infty$. *Then for* $0 \leq c \leq \infty$

$$\begin{aligned} &U_1(x) \sim cU_2(x), && x \to \infty \\ \text{iff}\quad & \\ &U_1^{\leftarrow}(x) \sim c^{-\rho^{-1}} U_2^{\leftarrow}(x), && x \to \infty. \end{aligned}$$

(vii) *If* $U \in RV_\rho$, $\rho \neq 0$, *then there exists a function* $U^*$ *which is absolutely continuous, strictly monotone, and*

$$U(x) \sim U(x)^*, \qquad x \to \infty.$$

PROOF. (i) We give the proof for the case $0 < \rho < \infty$. Suppose $U$ has Karamata representation

$$U(x) = c(x)\exp\left\{\int_1^x t^{-1}\rho(t)dt\right\}$$

where $c(x) \to c > 0$ and $\rho(t) \to \rho$. Then

$$\log U(x)/\log x = o(1) + \int_1^x t^{-1}\rho(t)dt \Big/ \int_1^x t^{-1}\,dt \to \rho.$$

(ii) Using the Karamata representation

$$U(tx)/U(t) = (c(tx)/c(t))\exp\left\{\int_1^x s^{-1}\rho(ts)ds\right\}$$

and the result is apparent since we may pick $t_0$ so that $t > t_0$ implies $\rho - \varepsilon < \rho(ts) < \rho + \varepsilon$ for $s > 1$.
(iii) If $c > 0$ then from the uniform convergence property in Proposition 0.5

$$\lim_{n\to\infty} \frac{U(a_n)}{U(a_n')} = \lim_{n\to\infty} \frac{U(a_n'(a_n/a_n'))}{U(a_n')} = \lim_{t\to\infty} \frac{U(tc)}{U(t)} = c^\rho.$$

(iv) Again by uniform convergence, for $x > 0$

$$\lim_{t\to\infty}\frac{U_1(U_2(tx))}{U_1(U_2(t))}=\lim_{t\to\infty}\frac{U_1(U_2(t)(U_2(tx)/U_2(t)))}{U_1(U_2(t))}$$
$$=\lim_{y\to\infty}\frac{U_1(yx^{\rho_2})}{U_1(y)}=x^{\rho_2\rho_1}.$$

(v) Let $U_t(x)=U(tx)/U(t)$ so that if $U\in RV_\rho$ and $U$ is nondecreasing then $(0<\rho<\infty)$

$$U_t(x)\to x^\rho,\qquad t\to\infty,$$

which implies by Proposition 0.1

$$U_t^{\leftarrow}(x)\to x^{\rho^{-1}}\qquad \text{as } t\to\infty;$$

i.e.,

$$\lim_{t\to\infty}U^{\leftarrow}(xU(t))/t=x^{\rho^{-1}}.$$

Therefore

$$\lim_{t\to\infty}U^{\leftarrow}(xU(U^{\leftarrow}(t)))/U^{\leftarrow}(t)=x^{\rho^{-1}}.$$

This limit holds locally uniformly since monotone functions are converging to a continuous limit. Now $U\circ U^{\leftarrow}(t)\sim t$ as $t\to\infty$ (cf. Section 0.4.1), and if we replace $x$ by $xt/U\circ U^{\leftarrow}(t)$ and use uniform convergence we get

$$\lim_{t\to\infty}\frac{U^{\leftarrow}(tx)}{U^{\leftarrow}(t)}=\lim_{t\to\infty}\frac{U^{\leftarrow}((xt/U\circ U^{\leftarrow}(t))U\circ U^{\leftarrow}(t))}{U^{\leftarrow}(t)}$$
$$=\lim_{t\to\infty}\frac{U^{\leftarrow}(xU\circ U^{\leftarrow}(t))}{U^{\leftarrow}(t)}=x^{\rho^{-1}}$$

which makes $U^{\leftarrow}\in RV_{\rho^{-1}}$.

(vi) If $c>0$, $0<\rho<\infty$ we have for $x>0$

$$\lim_{t\to\infty}\frac{U_1(tx)}{U_2(t)}=\lim_{t\to\infty}\frac{U_1(tx)U_2(tx)}{U_2(tx)U_2(t)}=cx^\rho.$$

Inverting we find for $y>0$

$$\lim_{t\to\infty}U_1^{\leftarrow}(yU_2(t))/t=(c^{-1}y)^{\rho^{-1}}$$

and so

$$\lim_{t\to\infty}U_1^{\leftarrow}(yU_2\circ U_2^{\leftarrow}(t))/U_2^{\leftarrow}(t)=(c^{-1}y)^{\rho^{-1}}$$

and since $U_2\circ U_2^{\leftarrow}(t)\sim t$

$$\lim_{t\to\infty}U_1^{\leftarrow}(yt)/U_2^{\leftarrow}(t)=(c^{-1}y)^{\rho^{-1}}.$$

Set $y=1$ to obtain the result.

(viii) For instance if $U \in RV_\rho$, $\rho > 0$ define

$$U^*(t) = \int_1^t s^{-1} U(s) ds.$$

Then $s^{-1}U(s) \in RV_{\rho-1}$ and by Karamata's theorem

$$U(x)/U^*(x) \to \rho.$$

$U^*$ is absolutely continuous and since $U(x) \to \infty$ when $\rho > 0$, $U^*$ is ultimately strictly increasing. □

EXERCISES

0.4.2.1. Use Karamata's representation of a slowly varying function to prove the uniform convergence in Proposition 0.5. Hint: Use continuous convergence.

0.4.2.2. If $\rho > 0$ verify Theorem 0.6a by using uniform convergence (Proposition 0.5).

0.4.2.3. Supply the proofs of omitted cases in Proposition 0.8.

0.4.2.4. If $U \in RV_\infty$ and $U$ is monotone, if $a_n \to \infty$, $a_n' \to \infty$, and $a_n \sim a_n' c$ $(c \neq 1$, $0 \le c \le \infty)$ then

$$\lim_{n\to\infty} U(a_n)/U(a_n') = c^\infty.$$

0.4.2.5. Give the Karamata representation of the slowly varying functions

$$(1 + x^{-1}) \log x$$

and

$$\exp\{(\log x)^\alpha\}, \qquad 0 < \alpha < 1.$$

0.4.2.6. Give an example of a slowly varying function $L(x)$ such that

$$\lim_{x\to\infty} L(x)$$

does not exist. (Hint: Use the Karamata representation.)

0.4.2.7. Suppose $U$ is integrable on $[0, N]$ for every $N$ and

$$\lim_{x\to\infty} x^{-1} \int_0^x U(t) dt = \rho$$

exists finite. Then show

$$\exp\left\{\int_0^x \frac{U(t)}{t} dt\right\} \in RV_\rho$$

(Seneta, 1976, page 88; Aljancic and Karamata, 1956).

0.4.2.8. Suppose $F(x)$ is a distribution on $R_+$ and

$$1 - F(x) \sim x^{-\alpha} L(x).$$

For $\eta \ge \alpha$ show by integrating by parts or using Fubini's theorem that

$$\lim_{x\to\infty} \frac{\int_0^x u^\eta F(du)}{x^\eta(1 - F(x))} = \frac{\alpha}{\eta - \alpha}.$$

Formulate and prove a converse. (Cf. Feller, 1971, page 283.)

0.4.2.9. If $L_i$: $R_+ \to R_+, i = 1, 2$, are slowly varying; so is $L_1 + L_2$ (Tucker, 1968, page 1382).

0.4.2.10. Suppose $a(x) \in RV_\rho$, $\rho \neq 0$. If $N_n$, $n \geq 0$ are nonnegative random variables such that

$$N_n/n \xrightarrow{P} N$$

then $a(N_n)/a(n) \xrightarrow{P} N^\rho$.

0.4.2.11. Suppose $L(x)$ is slowly varying and $\alpha > 0$. Then as $x \to \infty$

$$x^\alpha L(x) \sim \sup_{0<t\leq x} t^\alpha L(t).$$

So a regularly varying function with a positive exponent is asymptotic to a monotone nondecreasing (regularly varying) function (Karamata, 1962).

### 0.4.3. Extensions of Regular Variation: Π-Variation, Γ-Variation (de Haan, 1970, 1974a, 1976a)

In extreme value theory domain of attraction criteria for $\Phi_\alpha$ and $\Psi_\alpha$ can be satisfactorily handled with a knowledge of regularly varying functions. However characterizations for the domain of attraction of $\Lambda(x)$ require extensions which we now discuss. Restriction to monotone functions will be adequate. The revelance of the two following definitions is made precise in Proposition 0.10 later.

**Definition.** A nondecreasing function $U$ is Γ-varying (written $U \in \Gamma$) if $U$ is defined on an interval $(x_l, x_0)$, $\lim_{x\uparrow x_0} U(x) = \infty$ and there exists a positive function $f$ defined on $(x_l, x_0)$ such that for all $x$

$$\lim_{t\to x_0} \frac{U(t + xf(t))}{U(t)} = e^x. \tag{0.47}$$

The function $f$ is called an *auxiliary function* and is unique up to asymptotic equivalence. There are several ways to check this, but perhaps the most straightforward is to define for $t > 0$, $x > 0$

$$F_t(x) = 1 - U(t)/U(t + x)$$

so that $F_t(x)$ is a family of distributions. If (0.47) is satisfied for both $f_1$ and $f_2$ then

$$F_t(f_i(t)x) \to 1 - e^{-x}$$

for $i = 1, 2$, and hence by the convergence to types Proposition 0.2 we have

$$f_1(t) \sim f_2(t).$$

Conversely, if (0.47) is satisfied with $f$, and $f_1(t) \sim f(t)$, then (0.47) holds with $f_1$.

In probability applications we set $U = 1/(1 - F)$ where $F$ is a probability distribution with right-end $x_0 = \sup\{x: F(x) < 1\}$.

**Definition.** A nonnegative, nondecreasing function $V(x)$ defined on a semi-infinite interval $(z, \infty)$ is $\Pi$-varying (written $V \in \Pi$) if there exist functions $a(t) > 0$, $b(t) \in \mathbb{R}$ such that for $x > 0$

$$\lim_{t\to\infty} \frac{V(tx) - b(t)}{a(t)} = \log x. \tag{0.48}$$

Note in (0.48) we may take $b(t) = V(t)$ since

$$\frac{V(tx) - V(t)}{a(t)} = \frac{V(tx) - b(t)}{a(t)} - \frac{V(t) - b(t)}{a(t)} \to \log x - \log 1 = \log x. \tag{0.49}$$

Furthermore putting $x = e$ in (0.49) shows we may take

$$a(t) = V(te) - V(t).$$

The function $a(\cdot)$ is unique up to asymptotic equivalence: If $a(\cdot)$ satisfies (0.48) and $a(t) \sim a_1(t)$, then $a_1(t)$ satisfies (0.48). Any function $a(t)$ satisfying (0.48) is called an *auxiliary function.* Similarly if

$$(b(t) - b_1(t))/a(t) \to 0$$

then $b_1$ satisfies (0.48).

There is a convenient relationship between $\Pi$ and $\Gamma$.

**Proposition 0.9.** (a) *If $U \in \Gamma$ with auxiliary function $f$ then $U^{\leftarrow} \in \Pi$ with auxiliary function $a(t) = f \circ U^{\leftarrow}(t)$.*
(b) *If $V \in \Pi$ with auxiliary function $a(\cdot)$ then $V^{\leftarrow} \in \Gamma$ with auxiliary function $f(t) = a \circ V^{\leftarrow}(t)$.*

Proof (a). If $U \in \Gamma$ then (0.47) holds. Inverting (0.47) using Proposition 0.1 gives for $y > 1$

$$\lim_{t\uparrow x_0} \frac{U^{\leftarrow}(yU(t)) - t}{f(t)} = \log y$$

and so replacing $t$ by $U^{\leftarrow}(t)$ we get

$$\lim_{t\to\infty} \frac{U^{\leftarrow}(yU(U^{\leftarrow}(t)) - U^{\leftarrow}(t)}{f(U^{\leftarrow}(t))} = \log y. \tag{0.50}$$

The convergence in (0.50) is locally uniform so if $U(U^{\leftarrow}(t)) \sim t$ we will get

$$\lim_{t\to\infty} \frac{U^{\leftarrow}(ty) - U^{\leftarrow}(t)}{f(U^{\leftarrow}(t))} = \log y;$$

i.e. $U^{\leftarrow} \in \Pi$. To check $U \circ U^{\leftarrow}(t) \sim t$ recall that $y < U^{\leftarrow}(x)$ implies $U(y) < x$ and $y > U^{\leftarrow}(x)$ implies $U(y) \geq x$. Set $y = U^{\leftarrow}(t) \pm \varepsilon f(U^{\leftarrow}(t))$ where $\varepsilon > 0$ and we get

$$\frac{U(U^{\leftarrow}(t) - \varepsilon f(U^{\leftarrow}(t)))}{U(U^{\leftarrow}(t))} \leq \frac{t}{U(U^{\leftarrow}(t))} \leq \frac{U(U^{\leftarrow}(t) + \varepsilon f(U^{\leftarrow}(t)))}{U(U^{\leftarrow}(t))}.$$

Since $U \in \Gamma$, if we let $t \to \infty$ we get

$$e^{-\varepsilon} \leq \liminf_{t\to\infty} t/U \circ U^{\leftarrow}(t) \leq \limsup_{t\to\infty} \leq e^{\varepsilon}$$

and since $\varepsilon > 0$ is arbitrary the result follows.

The proof of (b) is similar. Analogous to the preceding step where one proves $U \circ U^{\leftarrow}(t) \sim t$, in (b) we need to show

$$\lim_{s\to V(\infty)} (s - V(V^{\leftarrow}(s)))/a(V^{\leftarrow}(s)) = 0.$$

Cf. Exercise 0.4.3.6. □

The relevance of $\Pi$ and $\Gamma$ to the study of domains of attraction of $D(\Lambda)$ is given in the next proposition. Recall $F \in D(\Lambda)$ means there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) \to \Lambda(x). \tag{0.51}$$

The following formulation is in terms of $U = 1/(1 - F)$. This is largely a matter of tradition as $1/(-\log F)$ is equally suitable from the theoretical point of view.

**Proposition 0.10** (Mejzler, 1949; de Haan, 1970). *For a distribution function $F$ set*

$$U := 1/(1 - F)$$

*so that $U^{\leftarrow}$ is defined on* $(1, \infty)$. *The following are equivalent:*

(i) $F \in D(\Lambda)$
(ii) $U \in \Gamma$
(iii) $U^{\leftarrow} \in \Pi$

Proof. We first show (i) implies (iii). If (i) holds, so does (0.51). In (0.51), take logarithms and use $-\log z \sim 1 - z$, $z \uparrow 1$, to get

$$n(1 - F(a_n x + b_n) \to e^{-x}, \qquad x \in \mathbb{R}$$

which can be reexpressed as

$$n^{-1} U(a_n x + b_n) \to e^{x}, \qquad x \in \mathbb{R}.$$

This implies by inversion (Proposition 0.1) that

$$(U^{\leftarrow}(ny) - b_n)/a_n \to \log y, \qquad y > 0;$$

i.e.,

$$(U^{\leftarrow}(ny) - U^{\leftarrow}(n))/a_n \to \log y. \tag{0.52}$$

If we set $a(t) = a_{[t]}$ it is easy to see (0.52) is the same as saying $U \in \Pi$ with auxiliary function $a(\cdot)$. For if $\varepsilon > 0$ and $t$ is sufficiently large

$$\begin{aligned}\frac{U^{\leftarrow}([t]y) - U^{\leftarrow}([t])}{a(t)} &- \left(\frac{U^{\leftarrow}([t](1+\varepsilon)) - U^{\leftarrow}([t])}{a(t)}\right)\\ &\le (U^{\leftarrow}([t]y) - U^{\leftarrow}([t]+1))/a(t)\\ &\le (U^{\leftarrow}(ty) - U^{\leftarrow}(t))/a(t)\\ &\le (U^{\leftarrow}(([t]+1)y) - U^{\leftarrow}([t]))/a(t)\\ &\le (U^{\leftarrow}([t]y(1+\varepsilon)) - U^{\leftarrow}([t]))/a(t)\end{aligned}$$

and letting $t \to \infty$ and using (0.52) we get

$$\begin{aligned}\log y - \log(1+\varepsilon) \le \liminf_{t\to\infty} (U(ty) - U(t))/a(t) &\le \limsup\\ &\le \log y + \log(1+\varepsilon)\end{aligned}$$

and since $\varepsilon > 0$ is arbitrary we see $U^{\leftarrow} \in \Pi$.

We next check that (iii) implies (ii). From Proposition 0.9(b), if $U^{\leftarrow} \in \Pi$ then $(U^{\leftarrow})^{\leftarrow} \in \Gamma$. Since Exercise 0.2.2 informs us that $(U^{\leftarrow})^{\leftarrow}(x) = U^{-}(x) = \lim_{t\uparrow x} U(t)$ it remains to see that $U^{-} \in \Gamma$ implies $U \in \Gamma$. Suppose for $x \in \mathbb{R}$

$$\lim_{t\uparrow x_0} U^{-}(t + xf(t))/U^{-}(t) = e^{x}.$$

For $\varepsilon > 0$ we have

$$\frac{U^{-}(t + xf(t))}{U^{-}(t)} \le \frac{U(t + xf(t))}{U^{-}(t)} \le \frac{U^{-}(t + (x+\varepsilon)f(t))}{U^{-}(t)}.$$

Let $t \uparrow x_0$. We see that

$$e^{x} \le \liminf_{t\uparrow x_0} \frac{U(t + xf(t))}{U^{-}(t)} \le \limsup_{t\uparrow x_0} \le e^{x+\varepsilon}$$

and since $\varepsilon > 0$ is arbitrary we conclude

$$\lim_{t\uparrow x_0} \frac{U(t + xf(t))}{U^{-}(t)} = e^{x} \tag{0.53}$$

for $x \in \mathbb{R}$. Set $x = 0$ in (0.53) and we see $U(t) \sim U^{-}(t)$ and hence $U \in \Gamma$.

Last we check (ii) implies (i). Given

$$\frac{U(t + xf(t))}{U(t)} \to e^{x}.$$

Recalling from the proof of Proposition 0.9a that $U(U^{\leftarrow}(t) \sim t$ we see that

$$\frac{U(U^{\leftarrow}(n) + xf(U^{\leftarrow}(n))}{n} \to e^{x},$$

which is the same as (set $a_n = f(U^{\leftarrow}(n))$, $b_n = U^{\leftarrow}(n)$)

$$n(1 - F(a_n x + b_n)) \to e^{-x}$$

and this is equivalent to (0.51). □

This proposition shows that a proper understanding of $\Pi$ and $\Gamma$ is essential for the study of $D(\Lambda)$. We now analyze the structure of a $\Pi$-varying function. The first fact is the analogue of Proposition 0.7 and gives a connection between $\Pi$ and $RV_{-1}$.

**Proposition 0.11** (de Haan, 1976b). *Suppose $V\colon \mathbb{R}_+ \to \mathbb{R}_+$ is absolutely continuous with density $v$.*
(a) *If $v(x) \in RV_{-1}$ then $V \in \Pi$ with auxiliary function $xv(x)$.*
(b) *If $v(x)$ is nonincreasing and $V \in \Pi$ then $v \in RV_{-1}$.*

Proof. (a) We have for $x > 1$ (a similar argument applies if $0 < x < 1$)

$$(V(tx) - V(t))/tv(t) = \int_1^x (v(ts)/v(t))ds.$$

Since the integrand tends to $s^{-1}$ uniformly on $[1, x]$ we find $V \in \Pi$.
Before proving (b) we note the following result:

**Proposition 0.12.** *If $V \in \Pi$ with auxiliary function $a(t)$ then $a(\cdot) \in RV_0$.*

Proof. For $x > 0$

$$\begin{aligned}\lim_{t\to\infty} a(tx)/a(t) &= \lim_{t\to\infty} \left(\frac{V(tx) - V(t)}{a(t)}\right) \Big/ \left(-\left(\frac{V(tx\cdot x^{-1}) - V(tx)}{a(tx)}\right)\right) \\ &= \log x/(-\log x^{-1}) = 1.\end{aligned}$$

□

Proof of Proposition 0.11(b). Suppose $V \in \Pi$ with auxiliary function $a(t)$. Then for $x > 1$

$$\frac{V(tx) - V(t)}{a(t)} = \frac{V(tx) - V(t)}{tv(t)} \cdot \frac{tv(t)}{a(t)} = \frac{tv(t)}{a(t)} \int_1^x \frac{v(ts)}{v(t)} ds$$

and therefore

$$\frac{tv(t)}{a(t)} = \frac{V(tx) - V(t)}{a(t)} \Big/ \int_1^x \frac{v(ts)}{v(t)} ds$$

and using the fact that $v$ is nondecreasing we obtain the bounds

$$(V(tx) - V(t))/(x - 1)a(t) \le tv(t)/a(t) \le \{(V(tx) - V(t))/a(t)\} \Big/ \frac{v(tx)}{v(t)}(x - 1).$$

In the left-hand inequality, let $t \to \infty$ to get $(\log x)/(x - 1) \le \liminf_{t\to\infty} tv(t)/a(t)$ and letting $x \downarrow 1$ gives $1 \le \liminf_{t\to\infty} tv(t)/a(t)$. The preceding right-hand inequality leads to

$$tv(tx)/a(t) \leq \{(V(tx) - V(t))/a(t)\}/(x-1) \tag{0.54}$$

so that

$$\limsup_{t\to\infty} tv(tx)/a(t) = \limsup_{t\to\infty} tx^{-1}v(t)/a(tx^{-1}) = \limsup_{t\to\infty} tx^{-1}v(t)/a(t)$$

(since $a(t) \in RV_0$)

$$\leq \log x/(x-1) \qquad \text{(from 0.54)}$$

i.e.,

$$\limsup_{t\to\infty} tv(t)/a(t) \leq x \log x/(x-1)$$

and letting $x \downarrow 1$ gives $\limsup_{t\to\infty} tv(t)/a(t) \leq 1$. We conclude that

$$v(t) \sim t^{-1}a(t) \in RV_{-1}$$

as required. □

*Remark.* Virtually the same proof shows that Proposition 0.11(b) is true if one merely assumes that $x^t v(x)$ is monotone in $x$ for some $t \in \mathbb{R}$.

Now a technical lemma necessary to derive a representation of $V \in \Pi$.

**Lemma 0.13.** *Given $V \in \Pi$ with auxiliary function $a(\cdot)$, for any $0 \leq \eta < 1$ there is $t_0 = t_0(\eta)$, $c > 0$ such that for $y \geq e$ (say), $t \geq t_0$*

$$(V(ty) - V(t))/a(t) \leq cy^{\eta}.$$

Proof. Given $\varepsilon > 0$ there is $t_0$ such that $t \geq t_0$ implies

$$(V(te) - V(t))/a(t) \leq 1 + \varepsilon \tag{0.55}$$

and

$$a(te)/a(t) \leq 1 + \varepsilon. \tag{0.56}$$

Therefore for any integer $n \geq 1$

$$\begin{aligned}
&(V(te^n) - V(t))/a(t) \\
&= \sum_{j=1}^{n} \left(\frac{V(te^j) - V(te^{j-1})}{a(te^{j-1})}\right) \frac{a(te^{j-1})}{a(t)} \\
&\leq \sum_{j=1}^{n} (1+\varepsilon) a(te^{j-1})/a(t) \\
&= \sum_{j=1}^{n} (1+\varepsilon) \prod_{i=1}^{j-1} (a(te^i)/a(te^{i-1})) \\
&\leq (1+\varepsilon) \sum_{j=1}^{n} \prod_{i=1}^{j-1} (1+\varepsilon) \leq (1+\varepsilon) \sum_{j=1}^{n} (1+\varepsilon)^{j-1} \\
&\leq \bar{c}(1+\varepsilon)^n \qquad \text{for } \bar{c} > 0.
\end{aligned}$$

For any $y \geq e$, $y = \exp\{\log y\} \leq \exp\{[\log y] + 1\}$ so

$$\begin{aligned}
(V(ty) - V(t))/a(t) &\leq (V(te^{([\log y]+1)}) - V(t))/a(t) \\
&\leq \bar{c}(1+\varepsilon)^{([\log y]+1)} \leq c(1+\varepsilon)^{\log y} \qquad \text{(for some } c > 0) \\
&= ce^{(\log(1+\varepsilon))\log y} = cy^{\log(1+\varepsilon)} = cy^{\eta}.
\end{aligned}$$

□

**Proposition 0.14.** *If $V \in \Pi$ with auxiliary function $a(t)$ then*

$$\begin{aligned}
a(t) &\sim t\left[\int_t^\infty V(s)s^{-2}\,ds - t^{-1}V(t)\right] \\
&= t\int_t^\infty u^{-1}V(du), \qquad \textit{as } t \to \infty.
\end{aligned}$$

Proof. Since

$$\lim_{t\to\infty} (V(tx) - V(t))/a(t) = \log x$$

we have from Lemma 0.13 and dominated convergence that

$$\lim_{t\to\infty} \int_1^\infty \frac{V(tx) - V(t)}{x^2 a(t)}\,dx = \int_1^\infty x^{-2}\log x\,dx. \tag{0.57}$$

The right side of (0.57) is

$$\begin{aligned}
\int_{x=1}^\infty \left(\int_{u=1}^x u^{-1}\,du\right)x^{-2}\,dx &= \int_{u=1}^\infty \left(\int_{x=u}^\infty x^{-2}\,dx\right)u^{-1}\,du \\
&= \int_1^\infty u^{-1}u^{-1}\,du = \int_1^\infty u^{-2}\,du = 1.
\end{aligned}$$

The left side of (0.57) is

$$\frac{t\int_t^\infty V(u)u^{-2}\,du - V(t)}{a(t)} = \frac{t\int_t^\infty u^{-1}V(du)}{a(t)},$$

the last step following by Fubini's theorem. □

**Proposition 0.15.** $V \in \Pi$ iff

$$K(x) := \int_x^\infty u^{-1}V(du) = \int_x^\infty u^{-2}V(u)du - x^{-1}V(x) \tag{0.58}$$

*is finite and $-1$ varying. In this case the auxiliary function $a(t)$ satisfies*

$$a(t) \sim tK(t)$$

*and we have representation*

$$V(x) - V(1) = \int_1^x K(u)du - xK(x) + K(1). \tag{0.59}$$

*(This emphasizes the relation between* $\Pi$*-variation and* $-1$ *variation introduced in Proposition* 0.11.)

PROOF. If $V \in \Pi$ then from Proposition 0.14

$$K(t) \sim t^{-1}a(t) \in RV_{-1}$$

since $a(t) \in RV_0$.

For the converse we first show we can express $V$ in terms of $K$ as in (0.59). We have

$$\begin{aligned}
\int_1^x K(u)du &= \int_{t=1}^x \int_{u=t}^\infty u^{-1}V(du)dt \\
&= \int_{u=1}^\infty \left(\int_{t=1}^{u\wedge x} dt\right)u^{-1}V(du) \\
&= \int_{u=1}^x (u-1)u^{-1}V(du) + \int_{u=x}^\infty (x-1)u^{-1}V(du) \\
&= V(x) - V(1) - (K(1) - K(x)) + (x-1)K(x)
\end{aligned}$$

and this gives (0.59). If now $K \in RV_{-1}$ then

$$\begin{aligned}
\frac{V(tx) - V(t)}{tK(t)} &= \int_1^x \frac{K(ut)du}{K(t)} - \frac{xK(tx)}{K(t)} + 1 \\
&\to \int_1^x u^{-1}\,du - xx^{-1} + 1 = \log x
\end{aligned}$$

so that $V \in \Pi$. $\square$

*Equivalence Classes*

For regular variation asymptotic equivalence is the appropriate equivalence relation. If $V_1 \in \Pi$ with auxiliary function $a(t)$, say $V_1$ and $V_2$ are $\Pi$-equivalent (written $V_1 \overset{\Pi}{\sim} V_2$) if

$$(V_1(t) - V_2(t))/a(t) \to c \in \mathbb{R}$$

as $t \to \infty$. In this case $V_2 \in \Pi$ with auxiliary function $a(t)$.

If $V \in \Pi$ we may construct smoother versions which are $\Pi$-equivalent to $V$.

**Proposition 0.16.** *If* $V \in \Pi$ *there exists a continuous, strictly increasing* $V_1 \overset{\Pi}{\sim} V$ *such that*

$$V_1(t) > V(t)$$

*and*

$$(V_1(t) - V(t))/a(t) \to 1.$$

*In fact, there exists a twice differentiable $V_2 \overset{\Pi}{\sim} V$ with $V_2(t) > V(t)$ and*

$$-\frac{1}{xV_2''(x)} \in RV_1, \qquad -xV_2''(x) \sim V_2'(x) \qquad \textit{as } x \to \infty.$$

PROOF. Set $V_1(t) = t\int_t^\infty V(u)u^{-2}\,du$ so that by monotonicity $V_1 > V$ and (0.57) translates to

$$(V_1(t) - V(t))/a(t) \to 1.$$

Note that almost everywhere

$$\begin{aligned} V_1'(t) &= -t^{-1}V(t) + \int_t^\infty u^{-2}V(u)du \\ &= K(t) \qquad \text{from (0.58).} \end{aligned} \tag{0.60}$$

Next set $V_0 = V$ and

$$V_2(t) := t\int_t^\infty V_1(u)u^{-2}\,du.$$

Then $V_2 \overset{\Pi}{\sim} V_1$ and $V_1 \overset{\Pi}{\sim} V$ so $V_2 \overset{\Pi}{\sim} V$. Finally to prove the last assertion of the proposition differentiate to get

$$\begin{aligned} V_2'(t) &= -t^{-1}V_1(t) + \int_t^\infty V_1(u)u^{-2}\,du \\ &= t^{-1}[-V_1(t) + t\int_t^\infty V_1(u)u^{-2}\,du] \\ &= t^{-1}[V_2(t) - V_1(t)] \end{aligned} \tag{0.61}$$

so that

$$V_2''(t) = \{t[V_2'(t) - V_1'(t)] - (V_2(t) - V_1(t))\}/t^2$$

and substituting the expression for $V_2 - V_1$ from (0.61) gives

$$\begin{aligned} V_2''(t) &= \{t[V_2'(t) - V_1'(t)] - tV_2'(t)\}/t^2 \\ &= -V_1'(t)/t \end{aligned}$$

and from (0.60)

$$V_2''(t) = -K(t)/t \tag{0.62}$$

and hence

$$-1/(xV_2''(x)) = 1/K(x) \in RV_1$$

since $K \in RV_{-1}$.

For the last claim of the proposition note by (0.61) $V_2' \in RV_{-1}$ so that $V_2'(t) \to 0$ as $t \to \infty$. Hence

$$V_2'(t) = \int_t^\infty -V_2''(s)ds = \int_t^\infty s^{-1}K(s)ds$$

by (0.62). Therefore,

$$\frac{V_2'(t)}{-tV_2''(t)} = \int_t^\infty s^{-1}K(s)ds/K(t)$$

and since $K \in RV_{-1}$ an appeal to Karamata's theorem 0.6 completes the proof. $\square$

## Exercises

0.4.3.1. For a monotone function $V \in \Pi$ prove $V \in RV_0$, and if the auxiliary function is $a(t)$ show $V(t)/a(t) \to \infty$ (de Haan, 1970).

0.4.3.2. Let $\pi$: $\mathbb{R}_+ \to \mathbb{R}_+$. We do not necessarily suppose $\pi$ is monotone. Say $\pi$ is $\Pi^+$ varying (written $\pi \in \Pi^+$) if there exists $a(t)$: $\mathbb{R}_+ \to \mathbb{R}_+$, $b(t)$: $\mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{t\to\infty} (\pi(tx) - b(t))/a(t) = \log x. \tag{0.63}$$

Similarly say $\pi \in \Pi^-$ if

$$\lim_{t\to\infty} (\pi(tx) - b(t))/a(t) = -\log x. \tag{0.64}$$

Take as fact, the statement that (0.63) and (0.64) hold locally uniformly (Balkema, 1973). Prove

(a) $\pi \in \Pi^+ \cup \Pi^-$ iff for every $r \in RV_1$ we have $\pi \circ r \in \Pi^+ \cup \Pi^-$. The auxiliary function of $\pi$ or is $a \circ r$ if the auxiliary function of $\pi$ is $a(\cdot)$. Moreover

$$\pi \circ r \overset{\Pi}{\sim} \pi \text{ iff } \lim_{x\to\infty} x^{-1}r(x) = c > 0.$$

(b) $\pi \in \Pi^+$ iff $1/\pi \in \Pi^-$. The auxiliary function of $1/\pi$ is $a(\cdot)/\pi^2$.

(c) Suppose $L \in RV_0$ and $\pi \in \Pi^\pm$. Then $L(t)\pi(t) \in \Pi^\pm$ with auxiliary function $L(t)a(t)$ iff

$$\lim_{t\to\infty} \left(\frac{L(tx)}{L(t)} - 1\right)\frac{\pi(t)}{a(t)} = 0 \qquad \text{for all } x > 0$$

(de Haan and Resnick, 1979a).

0.4.3.3. Suppose $V \in \Pi$ and for $x > 0$

$$\lim_{t\to\infty} (V(tx) - V(t)) = \rho \log x, \qquad \rho > 0.$$

Show $V(x) = c(x) + \int_1^x t^{-1}\rho(t)dt$ where

$$\lim_{x\to\infty} c(x) = c$$

$$\lim_{x\to\infty} \rho(x) = \rho.$$

0.4.3.4. If $V_1, V_2 \in \Pi$ show $V + V_2 \in \Pi$. Hint: Use (0.58). What is an auxiliary function of $V_1 + V_2$?

0.4.3.5. Let $U(x) = 2\log x + \sin(\log x)$. Then $U \in RV_0$ but $U$ is not in $\Pi$.

0.4.3.6. Prove Proposition 0.9(b) and check that

$$\lim_{s\to\infty} (s - V(V^{\leftarrow}(s)))/a(V^{\leftarrow}(s)) = 0.$$

0.4.3.7. Suppose $f: \mathbb{R}_+ \to \mathbb{R}_+$ is differentiable, $\lim_{t\to\infty} f'(t) = 0$, $1/f$ is integrable and define

$$H(x) = \exp\left\{\int_0^x (1/f(u)du\right\}.$$

Show $H \in \Gamma$ with auxiliary function $f$.

0.4.3.8. Suppose $f$ is as described in the previous exercise and $U: \mathbb{R}_+ \to \mathbb{R}_+$ is monotone. Then as $t \to \infty$ we have for $x > 0$

$$\frac{U(t + xf(t)) - U(t)}{U(t + f(t)) - U(t)} \to x \tag{0.65}$$

iff

$$U = V_1 \circ H$$

where $H$ is described in 0.4.3.7. and $V_1 \in \Pi$. Hint: If $U$ satisfies (0.65) show $U \circ H^{\leftarrow} \in \Pi$.

0.4.3.9. Give examples of nondecreasing $V \in \Pi$ such that $V(\infty) = \infty$ or $V(\infty) < \infty$.

0.4.3.10. If $U \in \Gamma$ with auxiliary function $f$ so that

$$\lim_{t\to\infty} U(t + xf(t))/U(t) = e^x$$

prove $\lim_{t\to\infty} f(t)/t = 0$. Hint: Use 0.4.3.1. and Proposition 0.9.

0.4.3.11. Suppose $U_1, U_2 \in \Gamma$ with the same $f$ and $x_0$. Show

$$U_1 = U \circ U_2$$

where $U \in RV_1$. Extend this result to the case where $U_i \in \Gamma$ with auxiliary function $f_i$, $i = 1, 2$ and $f_1(x) \sim f_2(x)$ as $x \uparrow x_0$ (de Haan, 1974a).

0.4.3.12. If $U \in \Gamma$ then $U(x) \sim (U^{\leftarrow})^{\leftarrow}(x)$ as $x \uparrow x_0$.

0.4.3.13. Suppose $U_1$ is monotone and $U_1 \in RV_\rho$, $\rho > 0$. If $U_2 \in \Gamma$ show $U_1 \circ U_2 \in \Gamma$. Express the auxiliary function of $U_1 \circ U_2$ in terms of the auxiliary function of $U_2$ (de Haan, 1970).

0.4.3.14. Suppose $U_1 \in \Gamma$ and $U_2$ is absolutely continuous with density $U \in RV_\rho$, $\rho > -1$. Show $U_1 \circ U_2 \in \Gamma$ (de Haan, 1970).

0.4.3.15. If $U_1 \in \Gamma$ and $U_2$ is absolutely continuous with density belonging to $\Gamma$ then $U := U_1 \circ U_2 \in \Gamma$ (de Haan, 1970).

0.4.3.16. Prove the remark which follows the proof of Proposition 0.11b.

0.4.3.17. Suppose $V$ is real valued with positive derivative $V'$ and $\lim_{x\to\infty} \uparrow V(x) = \infty$. If

$$\lim_{x\to\infty} \log V(x)/(xV'(x)) = 1$$

then $V \in \Pi$ with auxiliary function $\log V$ (de Haan and Hordijk, 1972).

0.4.3.18. Suppose $U$ is a twice differentiable real valued function with positive derivative $U'$ and $\lim_{x\uparrow\infty} \uparrow U(x) = \infty$. Define

$$q(t) = \log U(t)/U'(t)$$

and suppose $q'(t) \to 0$. Then

$$\lim_{t\to\infty} (U(t + xq(t)) - U(t))/\log U(t) = x$$

(de Haan and Hordijk, 1972).

# CHAPTER 1

# Domains of Attraction and Norming Constants

As in Chapter 0 suppose $\{X_n, n \geq 1\}$ is an iid sequence of random variables with common distribution $F(x)$. If $G$ is an extreme value distribution then according to Proposition 0.3, $G$ is of the form

$$G(x) = \Phi_\alpha(x) = \begin{cases} 0 & x < 0 \\ \exp\{-x^{-\alpha}\}, & x \geq 0 \end{cases}$$

or

$$G(x) = \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x < 0 \\ 1 & x \geq 0 \end{cases}$$

or

$$G(x) = \Lambda(x) = \exp\{-e^{-x}\}, \qquad x \in \mathbb{R},$$

where in the first two cases, $\alpha$ is a positive parameter.

We say $F \in D(G)$ if there exist normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) = P[M_n \leq a_n x + b_n] \to G(x) \tag{1.1}$$

where as usual, $M_n = \bigvee_{1 \leq i \leq n} X_i = \max\{X_1, \ldots, X_n\}$. The goals of this chapter are to give necessary and sufficient conditions for $F \in D(G)$ when $G$ is one of the three extreme value distributions and also to characterize $a_n$ and $b_n$. Recall that by taking logarithms and expanding (1.1) is equivalent to

$$n(1 - F(a_n x + b_n)) \to -\log G(x) \tag{1.1'}$$

for $x$ such that $G(x) > 0$. (Cf. the derivation of (0.29).)

## 1.1. Domain of Attraction of $\Lambda(x) = \exp\{-e^{-x}\}$

We begin our study with the double exponential distribution since we know from our discussion in Proposition 0.10 that the material about the function classes $\Pi$ and $\Gamma$ will be essential. We seek conditions for (1.1) to hold with $G = \Lambda$ and characterizations of $a_n$ and $b_n$.

Define the right end $x_0$ of the distribution $F$ to be

$$x_0 = \sup\{y: F(y) < 1\}.$$

If $F \in D(\Lambda)$, it is possible for $x_0$ to be either finite or infinite. An example where $x_0 = \infty$ is the exponential distribution $F(x) = 1 - e^{-x}, x > 0$, since in this case with $a_n = 1$ and $b_n = \log n$ we get for $x \in \mathbb{R}$ and $n$ sufficiently large

$$\begin{aligned} F^n(x + \log n) &= (1 - e^{-(x+\log n)})^n \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n \to \exp\{-e^{-x}\} = \Lambda(x), \end{aligned}$$

or in other notation, if $\{E_n, n \geq 1\}$ is iid from $1 - e^{-x}$, $x > 0$ and $Y$ has distribution $\Lambda(x)$,

$$\bigvee_{i=1}^{n} E_i - \log n \Rightarrow Y. \tag{1.2}$$

For an example where $x_0 < \infty$ consider (Gnedenko, 1943)

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp\{(-x/(1-x))\} & 0 \leq x < 1 \\ 1 & x > 1. \end{cases}$$

If $a_n = (1 + \log n)^{-2}$, $b_n = (\log n)/(1 + \log n)$ then it is checked readily from first principles that maxima $M_n$ from this distribution satisfy

$$P(M_n \leq a_n x + b_n) = F^n(a_n x + b_n) \to \Lambda(x).$$

A more illuminating approach is to derive this result from the previous one: If $g(x) = x/(1 + x)$: $[0, \infty) \to [0, 1)$ then $g'(x) = (1 + x)^{-2}$ and

$$M_n - b_n \stackrel{\mathrm{d}}{=} g\left(\bigvee_{i=1}^{n} E_i\right) - g(\log n)$$

(cf. Section 0.2), and by the mean value theorem the right side is

$$\left(\bigvee_{i=1}^{n} E_i - \log n\right) g'(\zeta_n)$$

where $\zeta_n$ is between $\log n$ and $\bigvee_{i=1}^{n} E_i$. Note $g' \in RV_{-2}$ and from (1.2) it is evident that

$$\bigvee_{i=1}^{n} E_i/\log n \xrightarrow{\mathrm{P}} 1.$$

Thus it follows that $\zeta_n/\log n \xrightarrow{\mathrm{P}} 1$ and hence from Proposition 0.8(iii) (cf. Exercise 0.4.2.8) $g'(\zeta_n)/g'(\log n) \xrightarrow{\mathrm{P}} 1$. Thus

$$\frac{M_n - b_n}{a_n} \stackrel{\mathrm{d}}{=} \frac{g(\bigvee_{i=1}^{n} E_i) - g(\log n)}{g'(\log n)} \Rightarrow Y.$$

We begin our study of $D(\Lambda)$ by considering a special case. A distribution $F_\#$ with right end $x_0$ is called a *Von Mises function* if it has the following representation: There must exist $z_0 < x_0$ such that for $z_0 < x < x_0$ and $c > 0$

$$1 - F_\#(x) = c\exp\left\{-\int_{z_0}^{x} (1/f(u))du\right\} \tag{1.3}$$

where $f(u) > 0$, $z_0 < u < x_0$, and $f$ is absolutely continuous on $(z_0, x_0)$ with density $f'(u)$ and $\lim_{u\uparrow x_0} f'(u) = 0$. Call $f$ an auxiliary function.

**Proposition 1.1.** (a) *If $F_\#$ is a Von Mises function with representation* (1.3) *then $F_\# \in D(\Lambda)$. The norming constants may be chosen as*

$$b_n = (1/(1 - F))^{\leftarrow}(n)$$

$$a_n = f(b_n)$$

*and $1/(1 - F_\#) \in \Gamma$ with auxiliary function $f$.*
(b) *Suppose $F$ is absolutely continuous with negative second derivative $F''$ for all $x$ in $(z_0, x_0)$. If*

$$\lim_{x\uparrow x_0} F''(x)(1 - F(x))/(F'(x))^2 = -1 \tag{1.4}$$

*then $F$ is a Von Mises function and $F \in D(\Lambda)$. We can set $f = (1 - F)/F'$. Conversely, a twice differentiable Von Mises function satisfies* (1.4).
*For the proof, we need two lemmas.*

**Lemma 1.2.** *Suppose as in the preceding definition that $f(u)$ is an absolutely continuous auxiliary function with $f'(u) \to 0$ as $u \uparrow x_0$.*
(a) *If $x_0 = \infty$ then $\lim_{t\to\infty} t^{-1}f(t) = 0$.*
(b) *If $x_0 < \infty$ then $f(x_0) = \lim_{t\uparrow x_0} f(t) = 0$ and $\lim_{t\uparrow x_0}(x_0 - t)^{-1}f(t) = 0$.*

*In either case*

$$\lim_{t\uparrow x_0} (t + xf(t)) = x_0$$

*for all $x \in \mathbb{R}$. (For* (a), *cf. Exercise* 0.4.3.10.)

PROOF. (a) We have as $t \to \infty$

$$t^{-1}f(t) \sim t^{-1}\int_{z_0}^{t} f'(u)du$$

and since the integrand goes to zero so does the Cesaro average. Therefore

$$t + xf(t) = t(1 + xf(t)/t) \sim t, \qquad \text{as } t \to \infty.$$

(b) If $x_0 < \infty$ then $1 - F(x_0) = 0$ and then from (1.3) for $z_0 < x < x_0$

$$\int_{x}^{x_0} (1/f(u))du = \infty$$

whence for all $x \in (z_0, x_0)$

$$\sup_{x \le u \le x_0} 1/f(u) = \infty$$

and thus

$$\inf_{x \le u \le x_0} f(u) = 0.$$

So by continuity there exists a sequence $u_n \uparrow x_0$ and $f(u_n) = 0$, whence $f(x_0) = 0$.

Next observe that since $f(x_0) = 0$

$$\lim_{t \uparrow x_0} f(t)/(x_0 - t) = \lim_{t \uparrow x_0} - \int_t^{x_0} (f'(u)/(x_0 - t))du.$$

Change variables $y = x_0 - u$ and $s = x_0 - t$ and the preceding becomes

$$\lim_{s \downarrow 0} - s^{-1} \int_0^s f'(x_0 - y)dy$$

which is clearly zero since $f'(x_0 - y) \to 0$ as $y \to 0$. Finally it is clear that since $f(t) \to 0$ as $t \to x_0$ we have $t + xf(t) \to x_0$ as $t \to x_0$. □

**Lemma 1.3.** *If $f$ satisfies the conditions in Lemma* 1.2

$$\lim_{t \uparrow x_0} \frac{f(t + xf(t))}{f(t)} = 1$$

*locally uniformly in* $x \in \mathbb{R}$.

Proof. We show continuous convergence. Let $x(t)$ be a function such that

$$\lim_{t \to x_0} x(t) = x \in \mathbb{R}.$$

Then

$$|f(t + x(t)f(t)) - f(t)| \le \left| \int_t^{t+x(t)f(t)} f'(u)du \right|.$$

Since from the previous lemma as $t \to \infty$ we have $t + x(t)f(t) \to x_0$ it follows from $f'(u) \to 0$ as $u \to x_0$ that given $\varepsilon$, for $t \ge t_0(\varepsilon)$

$$\left| \int_t^{t+x(t)f(t)} f'(u)du \right| \le \varepsilon |x(t)f(t)|$$

and therefore for $t \ge t_0(\varepsilon)$

$$\left| \frac{f(t + x(t)f(t))}{f(t)} - 1 \right| \le \varepsilon |x(t)|.$$

Since $\varepsilon$ is arbitrary and $|x(t)|$ is bounded the result follows. □

PROOF OF PROPOSITION 1.1(a). From the form of $F_{\#}$ given in (1.3) we have for $x \in \mathbb{R}$ and $t$ sufficiently large

$$\frac{1 - F_{\#}(t + xf(t))}{1 - F_{\#}(t)} = \exp\left\{-\int_t^{t+xf(t)} (1/f(u))du\right\}$$

$$= \exp\left\{-\int_0^x \{f(t)/f(t + sf(t))\}\right\}ds \qquad (s = (u - t)/f(t))$$

and since Lemma 1.3 ensures the integrand converges to 1 uniformly on $(0, x)$ we get

$$\lim_{t \to x_0} \frac{1 - F_{\#}(t + xf(t))}{1 - F_{\#}(t)} = e^{-x}$$

which says $1/(1 - F_{\#}) \in \Gamma$. So Proposition 0.10 asserts $F_{\#} \in D(\Lambda)$. In fact pick $b_n$ to satisfy

$$1 - F_{\#}(b_n) = n^{-1},$$

i.e.,

$$b_n = (1/(1 - F_{\#}))^{\leftarrow}(n),$$

and then since $1/(1 - F(b_n)) \sim n$ (cf. the proof of Proposition 0.9(a))

$$\lim_{n \to \infty} n(1 - F_{\#}(b_n + xf(b_n)) = e^{-x}$$

which is (1.1′) so that a suitable choice of $a_n$ is $f(b_n)$. □

PROOF OF (b). Set $1 - F = \exp\{-R\}$. Then the representation like (1.3) is possible with $f = 1/R'$ and $f' \to 0$ iff $(1/R')' \to 0$. But $R = -\log(1 - F)$ so $R' = F'/(1 - F)$ and $1/R' = (1 - F)/F'$ and

$$(1/R')' = (-(F')^2 - (1 - F)F'')/(F')^2 = -1 - ((1 - F)F'')/(F')^2$$

and the assertion follows. The converse is readily checked. □

EXAMPLES. (a) Let $F(x) = 1 - e^{-x}$, $x > 0$. Then $F'(x) = e^{-x}$, $x \geq 0$, and

$$f(x) = (1 - F(x)/F'(x) = e^{-x}/e^{-x} = 1.$$

Therefore $f'(x) \equiv 0$ and $F \in D(\Lambda)$.
(b) Let $F(x) = N(x)$, the standard normal distribution. We have

$$F'(x) = n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$F''(x) = \frac{-1}{\sqrt{2\pi}} x e^{-x^2/2} = -xn(x)$$

and using Mills' ratio (Feller, 1971) we have $1 - N(x) \sim x^{-1}n(x)$. Therefore

$$\lim_{x\to\infty} \frac{(1 - F(x))F''(x)}{(F'(x))^2} = \lim_{x\to\infty} \frac{-x^{-1}n(x)xn(x)}{(n(x))^2} = -1,$$

and

$$f(x) = \frac{1 - N(x)}{n(x)} \sim x^{-1}\frac{n(x)}{n(x)} = x^{-1}.$$

The next result gives a nice representation of $F \in D(\Lambda)$ due to Balkema and de Haan (1972).

**Proposition 1.4.** $F \in D(\Lambda)$ iff *there exists a Von Mises function* $F_\#$ *such that for* $x \in (z_0, x_0)$

$$1 - F(x) = c(x)(1 - F_\#(x)) = c(x)\exp\left\{-\int_{z_0}^{x} (1/f(u))du\right\} \tag{1.5}$$

*and*

$$\lim_{t\to x_0} c(x) = c > 0.$$

If (1.5) holds then from Proposition 1.1 there exists $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$n(1 - F_\#(a_n x + b_n)) \to e^{-x}$$

and thus

$$n(1 - F(a_n x + b_n)) \to ce^{-x}$$

so that

$$F^n(a_n x + b_n) \to \exp\{-ce^{-x}\}$$

and $F \in D(\Lambda)$. So it is only the converse which need concern us. We need the following lemma.

**Lemma 1.5.** *Suppose* $F \in D(\Lambda)$ *so that* $V := (1/(1 - F))^{\leftarrow} \in \Pi$. *Construct* $V_1$ *and* $V_2$ *as in Proposition* 0.16 *and define*

$$1/(1 - F_i) = V_i^{\leftarrow}, \qquad i = 1, 2.$$

*Then as* $x \to x_0$

$$1 - F(x) \sim c_i(1 - F_i(x)),$$

*where* $c_1 = e^{-1}$, $c_2 = e^{-2}$.

Proof. We have for $x > 0$ as $t \to \infty$

$$\frac{V(tx) - V_1(t)}{a(t)} = \frac{V(tx) - V(t)}{a(t)} + \frac{V(t) - V_1(t)}{a(t)} \to (\log x) - 1$$

and inverting we get for $y \in \mathbb{R}$

$$V^{\leftarrow}(ya(t) + V_1(t))/t \to \exp\{y + 1\}$$

and setting $y = 0$

$$\lim_{t\to\infty} V^{\leftarrow}(V_1(t))/t = e,$$

whence remembering $V_1$ is continuous and strictly increasing

$$\lim_{s\to\infty} V^{\leftarrow}(s)/V_1^{\leftarrow}(s) = e.$$

The result for $F_2$ is checked in an identical manner.

Now for the rest of the proof of Proposition 1.4. Assume $F \in D(\Lambda)$ and set $F_{\#} = F_2$ so $1 - F_{\#} = 1/V_2^{\leftarrow}$ and it suffices to show $F_{\#}$ is a Von Mises function. Write $R = -\log(1 - F_{\#})$ and we need to check

$$(1/R')' \to 0.$$

However

$$1/R' = \frac{1 - F_{\#}}{F'_{\#}} = \frac{1/V_2^{\leftarrow}}{(V_2^{\leftarrow})'/(V_2^{\leftarrow})^2} = V_2^{\leftarrow}/(V_2^{\leftarrow})' = V_2^{\leftarrow} \cdot V_2'(V_2^{\leftarrow})$$

so that

$$(1/R')' = V_2^{\leftarrow} \cdot \{V_2''(V_2^{\leftarrow})/V_2'(V_2^{\leftarrow})\} + V_2'(V_2^{\leftarrow})/V_2'(V_2^{\leftarrow}).$$

Therefore

$$\lim_{t\to x_0} (1/R'(x))' = \lim_{y\to\infty} (yV_2''(y)/V_2'(y)) + 1 = -1 + 1 = 0$$

by Proposition 0.16. The result is proved. □

*Remark.* A small point, glossed over in the proof of Lemma 1.5, is this: If $U \in \Gamma$ then $(U^{\leftarrow})^{\leftarrow} \sim U$. This was essentially proved within Proposition 0.10 and also assigned as Exercise 0.4.3.12. If you are hard to convince here are the details again: Suppose for convenience $U$ is right continuous, as will be the case if $U = 1/(1 - F)$. Then by definition

$$\begin{aligned}(U^{\leftarrow})^{\leftarrow}(x) &= \inf\{y\colon U^{\leftarrow}(y) \ge x\}\\ &\le \inf\{y\colon U^{\leftarrow}(y) > x\}\\ &= \inf\{y\colon y > U(x)\} \qquad \text{(by 0.6(c))}\\ &= U(x).\end{aligned}$$

On the other hand

$$(U^{\leftarrow})^{\leftarrow}(x) \ge \inf\{y\colon U^{\leftarrow}(y) > x - \varepsilon f(x)\}$$

where $\varepsilon > 0$ and $f$ is assumed the auxiliary function of $U \in \Gamma$. Therefore another application of (0.6(c)) yields

$$(U^{\leftarrow})^{\leftarrow}(x) \geq \inf\{y: y > U(x - \varepsilon f(x))\} = U(x - \varepsilon f(x))$$

so that

$$U(x - \varepsilon f(x))/U(x) \leq (U^{\leftarrow})^{\leftarrow}(x)/U(x) \leq 1.$$

Letting $x \to x_0$ gives

$$e^{-\varepsilon} \leq \liminf (U^{\leftarrow})^{\leftarrow}(x)/U(x) \leq \limsup \leq 1$$

and the result follows.

**Corollary 1.6.** *If $F \in D(\Lambda)$ then*

$$\lim_{x \uparrow x_0} (1 - F(x))/(1 - F(x-)) = 1.$$

Proof. Use (1.3). Since $F_{\#}$ is continuous

$$(1 - F(x))/(1 - F(x-)) = c(x)/c(x-)$$

and since $c(x) \to c$ the result is clear. □

Corollary 1.6 can sometimes be used to check that certain distributions are not attracted to $\Lambda(x)$ (or indeed, to any extreme value distribution).

Examples. (a) Although we have seen the exponential distribution $1 - e^{-x}$, $x > 0$ is a Von Mises function, the geometric distribution written as

$$1 - e^{-[x]}, \qquad x > 0$$

is not in $D(\Lambda)$ since

$$\lim_{x \to \infty} e^{-[x]}/e^{-[x-]} \neq 1$$

because

$$\lim_{n \to \infty} e^{-n}/e^{-(n-1)} = e^{-1}.$$

(b) The Poisson distribution is not attracted to $\Lambda(x)$ (Gnedenko, 1943). Set for $x > 0$, $\lambda > 0$

$$1 - F(x) = \sum_{k > x} e^{-\lambda}\lambda^k/(k!)$$

and let $N$ be a random variable with this distribution. Let $\Gamma_n = E_1 + \cdots + E_n$ where $\{E_i, i \geq 1\}$ are iid, $P[E_i > x] = e^{-\lambda x}$, $x > 0$. A well known relationship from renewal theory is

$$P[N \geq n] = P[\Gamma_n \leq 1]$$

so

$$\begin{aligned}
\lim_{n\to\infty} \frac{P[N \geq n]}{P[N \geq n+1]} &= \lim_{n\to\infty} \frac{P[\Gamma_n \leq 1]}{P[\Gamma_{n+1} \leq 1]} \\
&= \lim_{n\to\infty} \frac{\int_0^1 \lambda e^{-\lambda x}(\lambda x)^{n-1}\,dx/(n-1)!}{\int_0^1 \lambda e^{-\lambda x}(\lambda x)^n\,dx/n!} \\
&= \lim_{n\to\infty} \frac{n\int_0^\lambda e^{-y}y^{n-1}\,dy}{\int_0^\lambda e^{-y}y^n\,dy} = \lim_{n\to\infty}\left(1 + \frac{e^{-\lambda}\lambda^n}{\int_0^\lambda e^{-y}y^n\,dy}\right)
\end{aligned}$$

the last step following by integrating by parts. But observe

$$\frac{\int_0^\lambda e^{-y}y^n\,dy}{e^{-\lambda}\lambda^n} \leq \frac{\int_0^\lambda y^n\,dy}{e^{-\lambda}\lambda^n} = \frac{\lambda^{n+1}}{(n+1)e^{-\lambda}\lambda^n} = \frac{\lambda}{(n+1)e^{-\lambda}} \to 0$$

and therefore

$$\lim_{n\to\infty} P[N \geq n]/P[N \geq n+1] = \infty.$$

The representation in the following corollary sometimes offers more flexibility than the one given in Theorem 1.4.

**Corollary 1.7.** $F \in D(\Lambda)$ iff *there exists* $z_0 < x_0$ *and measurable functions* $c(x)$, $g(x)$, *and* $f(x)$ *such that*

$$\lim_{t\to x_0} c(x) = c_1 > 0, \qquad \lim_{t\to x_0} g(x) = 1$$

*and*

$$1 - F(x) = c(x)\exp\left\{-\int_{z_0}^{x} (g(t)/f(t))dt\right\}, \qquad z_0 < x < x_0 \tag{1.6}$$

*where* $f$ *is an auxiliary function with* $f > 0$ *on* $(z_0, x_0)$ *and* $f$ *is absolutely continuous with* $f'(x) \to 0$ *as* $x \to x_0$.

PROOF. If $F \in D(\Lambda)$ use Theorem 1.4 with $g \equiv 1$. Conversely if (1.6) holds then for any $x \in \mathbb{R}$

$$\begin{aligned}
\lim_{t\to x_0} \frac{1 - F(t + xf(t))}{1 - F(t)} &= \lim_{t\to x_0} \exp\left\{-\int_t^{t+xf(t)} (g(s)/f(s))ds\right\} \\
&= \lim_{t\to x_0} \exp\left\{-\int_0^x g(t + sf(t))\frac{f(t)}{f(t + sf(t))}ds.\right.
\end{aligned}$$

It is evident from Lemmas 1.2 and 1.3 that the integrand converges to one uniformly for $s \in (0, x)$ and thus

$$\lim_{t\to x_0} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x}$$

which is equivalent to $F \in D(\Lambda)$. $\square$

EXAMPLE. Let $Q$ be the rationals on $(0, \infty)$. Then

$$F(x) = 1 - \exp\left\{-\int_0^x ((1 + t^{-1} 1_Q(t))/\log t)dt\right\}$$

$\in D(\Lambda)$ by the representation of Corollary 1.7. It is not clear how to construct $F_{\#}$ of (1.3) and (1.5).

For most practical purposes the criteria given in Proposition 1.1 are enough. In the case $F$ is not differentiable, it still remains to give reasonable criteria for membership in $D(\Lambda)$ and useful characterizations of $a_n$, $b_n$.

We begin with an integrability lemma:

**Lemma 1.8.** *If $x_0 = \infty$ and $F \in D(\Lambda)$ then*

$$\int_1^\infty (1 - F(u))du < \infty, \qquad \int_1^\infty \int_s^\infty (1 - F(u))du\,ds < \infty.$$

*Of course the result is true for $x_0 < \infty$.*

PROOF. By Theorem 1.4 we may suppose without loss of generality that $1 - F$ has representation (1.3). Given any $\delta < 1/2$, there exists $u_0 > z_0$ such that for $u \geq u_0$

$$-\delta < f'(u) < \delta.$$

Therefore for $u > u_0$

$$-\delta(u - u_0) < f(u) - f(u_0) < \delta(u - u_0)$$

so that

$$\frac{1}{f(u_0) + \delta(u - u_0)} < \frac{1}{f(u)} < \frac{1}{f(u_0) - \delta(u - u_0)}$$

and hence for $x > u_0$ there is a constant $c'$ such that

$$\begin{aligned}
1 - F(x) &= c' \exp\left\{-\int_{u_0}^x (1/f(u))du\right\} \\
&\leq c' \exp\left\{-\int_{u_0}^x (1/(f(u_0) + \delta(u - u_0)))du\right\} \\
&= c' \exp\left\{-\delta^{-1} \int_{f(u_0)}^{f(u_0)+\delta(x-u_0)} s^{-1}\,ds\right\} \\
&= c' \exp\{-\log((f(u_0) + \delta(x - u_0))/f(u_0))^{\delta^{-1}}\} \\
&= c'(1 + \delta f(u_0)^{-1}(x - u_0))^{-\delta^{-1}} \sim c'' x^{-\delta^{-1}}
\end{aligned}$$

as $x \to \infty$ for $c'' > 0$. The result follows. □

**Proposition 1.9** (de Haan, 1970). $F \in D(\Lambda)$ iff

$$\lim_{x \uparrow x_0} (1 - F(x)) \int_x^{x_0} \int_y^{x_0} (1 - F(t))dt\,dy \Big/ \left(\int_x^{x_0} (1 - F(t))dt\right)^2 = 1 \quad (1.7)$$

and all the integrals in the preceding expression are finite. In this case $1/(1 - F) \in \Gamma$ and for the auxiliary function $f$ we may choose either

$$f(t) = \int_x^{x_0} \int_y^{x_0} (1 - F(t))dt\,dy \Big/ \int_x^{x_0} (1 - F(t))dt$$

or

$$f(t) = \int_x^{x_0} (1 - F(t))dt/(1 - F(x))$$

and

$$b_n = (1/(1 - F))^{\leftarrow}(n)$$

$$a_n = f(b_n)$$

are acceptable choices of normalizing constants.

PROOF: SUFFICIENCY. Suppose (1.7) holds. Set

$$1 - F_0(x) = \left(\int_x^{x_0} (1 - F(s))ds\right)^2 \Big/ \int_x^{x_0} \int_y^{x_0} (1 - F(t))dt\,dy.$$

From (1.7)

$$\lim_{t \uparrow x_0} \frac{1 - F_0(x)}{1 - F(x)} = 1$$

so that $1 - F_0(x) \to 0$ as $x \to x_0$. Furthermore

$$\begin{aligned}
&(1 - F_0(x))' \\
&= \left\{-\int_x^{x_0} \int_y^{x_0} (1 - F(t))dt\,dy\, 2\left(\int_x^{x_0} (1 - F(t))dt\right)(1 - F(x)) \right. \\
&\quad \left. + \left(\int_x^{x_0} (1 - F(t))dt\right)^2 \int_x^{x_0} (1 - F(t))dt\right\} \Big/ \left(\int_x^{x_0} \int_y^{x_0} (1 - F(t))dt\,dy\right)^2 \\
&= \frac{(\int_x^{x_0}(1 - F(t))dt)(1 - F(x))}{\int_x^{x_0} \int_y^{x_0}(1 - F(t))dt\,dy} \left\{-2 + \frac{(\int_x^{x_0}(1 - F(t))dt)^2}{(1 - F(x)) \int_x^{x_0} \int_y^{x_0}(1 - F(t))dt\,dy}\right\}
\end{aligned}$$

and because of (1.7) this is negative for sufficiently large $x$, say $x \geq z_0$. Thus $1 - F_0$ is a distribution tail. Now set

$$h(x) = (1 - F(x)) \int_x^{x_0} \int_y^{x_0} (1 - F(t))dt\,dy \Big/ \left(\int_x^{x_0} (1 - F(t))dt\right)^2$$

so that $h(x) \to 1$ as $x \uparrow x_0$. Check that

$$\begin{aligned}
&\frac{d}{dx}(-\log(1 - F_0(x))) \\
&\quad = F_0'(x)/(1 - F_0(x)) \\
&\quad = (2h(x) - 1)\Big/\left(\int_x^{x_0}\int_y^{x_0}(1 - F(s))ds\,dy\Big/\int_x^{x_0}(1 - F(s))ds\right)
\end{aligned}$$

and if we set $f(x) = \int_x^{x_0}\int_y^{x_0}(1 - F(s))ds\,dy/\int_x^{x_0}(1 - F(s))ds$ then

$$\frac{d}{dx}(-\log(1 - F_0(x))) = (2h(x) - 1)/f(x).$$

Note

$$\begin{aligned}
f'(x) &= \frac{-(\int_x^{x_0}(1 - F(s))ds)^2 + (\int_x^{x_0}\int_y^{x_0}(1 - F(s))ds\,dy)(1 - F(x))}{(\int_x^{x_0}(1 - F(s))ds)^2} \\
&= -1 + h(x) \to 0
\end{aligned}$$

as $x \to x_0$ from (1.7). Write $g = 2h - 1$ so $g(x) \to 1$ and we get for $x \geq z_0$

$$1 - F_0(x) = (1 - F_0(z_0))\exp\left\{-\int_{z_0}^x (g(t)/f(t))dt\right\}$$

whence

$$1 - F(x) = \left\{\frac{1 - F(x)}{1 - F_0(x)}(1 - F_0(z_0))\right\}\exp\left\{-\int_{z_0}^x (g(t)/f(t))dt\right\}$$

and the representation of (1.6) holds, proving $F \in D(\Lambda)$.

PROOF: NECESSITY. Suppose $F \in D(\Lambda)$. Suppose initially that $F$ is a Von Mises function with representation (1.3) so that $1/(1 - F)$ is continuous and strictly increasing in some neighborhood of $x_0$. We know from Proposition 0.10 that

$$(1/(1 - F))^{\leftarrow} =: V \in \Pi,$$

and from Proposition 0.14 the auxiliary function $a(t)$ for $V$ satisfies as $t \to \infty$

$$a(t) \sim t\int_t^{\infty} u^{-1}V(du) = t\int_{V(t)}^{x_0}(1 - F(s))ds,$$

where the last equality follows by the transformation theorem for Lebesgue integrals. The auxiliary function for $1/(1-F) \in \Gamma$ can be taken to be $a \circ 1/(1-F)$, and since auxiliary functions for $\Gamma$-varying functions are asymptotically unique we conclude the $f$ appearing in representation (1.3) must satisfy

$$f(t) \sim \int_t^{x_0}(1 - F(s))ds/(1 - F(t)) =: f_1(t). \tag{1.8}$$

Recalling Lemmas 1.2 and 1.3 we have

$$t + xf_1(t) = t + x(f_1(t)/f(t))f(t) \to x_0$$

for any $x$ and

$$\frac{f_1(t + xf_1(t))}{f_1(t)} \sim \frac{f(t + xf_1(t))}{f(t)} = \frac{f(t + x(f_1(t)/f(t))f(t))}{f(t)} \to 1 \tag{1.9}$$

locally uniformly as $t \uparrow x_0$. Define a distribution tail by

$$1 - F_3(t) = \int_t^{x_0} (1 - F(s))ds$$

so that $f_1(t) = (1 - F_3(t))/(1 - F(t))$ and (1.9) can be rewritten as

$$\frac{(1 - F_3(t + xf_1(t)))(1 - F(t))}{(1 - F(t + xf_1(t)))(1 - F_3(t))} \to 1;$$

$$\text{i.e.,} \quad \frac{1 - F_3(t + xf_1(t))}{1 - F_3(t)} \sim \frac{1 - F(t + xf_1(t))}{1 - F(t)} \to e^{-x}$$

as $t \to x_0$ for $x \in \mathbb{R}$. Therefore $1/(1 - F_3) \in \Gamma$ and mimicking the argument which led to (1.8) we obtain

$$f_1(t) \sim \int_t^{x_0} (1 - F_3(s))ds/(1 - F_3(t));$$

i.e., $\int_t^{x_0}(1 - F(s))ds/(1 - F(t)) \sim \int_t^{x_0}\int_s^{x_0}(1 - F(u))du\, ds/\int_t^{x_0}(1 - F(u))du$ which is equivalent to (1.7).

If $F \in D(\Lambda)$ but $F$ is not a Von Mises function then there exists by Theorem 1.4 a constant $c > 0$ and a Von Mises function $F_\#$ such that

$$1 - F(x) \sim c(1 - F_\#(x))$$

as $x \to x_0$. $F_\#$ satisfies (1.7) and it is readily seen that the tail equivalence of $F$ and $F_\#$ entails that $F$ satisfies (1.7) as well. □

Here is another criterion for $F \in D(\Lambda)$.

**Proposition 1.10** (de Haan, 1970). *$F \in D(\Lambda)$ iff*

$$r(x) := \frac{\int_x^{x_0}(1 - F(t))^\alpha dt}{(1 - F(x))\int_x^{x_0}(1 - F(t))^{\alpha-1}dt} \to \alpha^{-1}(\alpha - 1) \tag{1.10}$$

*as $x \to x_0$ for some $\alpha > 1$. In this case* (1.10) *is true for all $\alpha > 1$.*

PROOF. Suppose (1.10) holds for some $\alpha > 1$ and define

$$1 - F_4(x) := \int_x^{x_0} (1 - F(t))^\alpha dt \Big/ \int_x^{x_0} (1 - F(t))^{\alpha-1} dt.$$

We will show that for all sufficiently large $x$, $1 - F_4$ is a distribution tail. First of all we have from (1.10) that

$$\lim_{t \to x_0} (1 - F_4(x))/(1 - F(x)) = \alpha^{-1}(\alpha - 1)$$

so that $\lim_{t \to x_0} 1 - F_4(x) = 0$. Also differentiating we get

$$(1 - F_4(x))' = \frac{(1 - F(x))^\alpha}{\int_x^{x_0} (1 - F(t))^{\alpha-1} dt} \left\{ \frac{\int_x^{x_0} (1 - F(t))^\alpha dt}{(1 - F(x)) \int_x^{x_0} (1 - F(t))^{\alpha-1} dt} - 1 \right\}$$

$$= \frac{(1 - F(x))^\alpha}{\int_x^{x_0} (1 - F(t))^{\alpha-1} dt} \{r(x) - 1\} \tag{1.11}$$

and since $r(x) \to 1 - \alpha^{-1} < 1$ we have $(1 - F_4(x))'$ ultimately negative and hence $1 - F_4$ is ultimately decreasing. Next observe that

$$\frac{d}{dx}(-\log(1 - F_4(x)) = F_4'(x)/(1 - F_4(x))$$

and using (1.11) this is

$$\frac{((1 - F(x))^\alpha / \int_x^{x_0} (1 - F(t))^{\alpha-1} dt)(1 - r(x))}{\int_x^{x_0} (1 - F(x))^\alpha dt / \int_x^{x_0} (1 - F(t))^{\alpha-1} dt}$$

$$= \frac{(1 - F(x))^\alpha (1 - r(x))}{\int_x^{x_0} (1 - F(t))^\alpha dt}$$

$$= \frac{(1 - r(x))(1 - F(x))^\alpha / (1 - F_4(x))^\alpha}{\int_x^{x_0} (1 - F(t))^\alpha dt / (1 - F_4(x))^\alpha} =: \frac{h(x)}{f(x)}.$$

Now $h(x) \to \alpha^{-1}(\alpha/(\alpha - 1))^\alpha =: c_1$. Also

$$f'(x) = \frac{-(1 - F_4(x))^\alpha (1 - F(x))^\alpha - \int_x^{x_0} (1 - F(t))^\alpha dt\, \alpha (1 - F_4(x))^{\alpha-1} (1 - F_4(x))'}{(1 - F_4(x))^{2\alpha}}$$

$$= -((1 - F(x))/(1 - F_4(x)))^\alpha + \frac{\alpha \int_x^{x_0} (1 - F(t))^\alpha dt (1 - r(x))(1 - F(x))^\alpha}{(1 - F_4(x))^{\alpha+1} \int_x^{x_0} (1 - F(t))^{\alpha-1} dt}$$

$$= -\left(\frac{\alpha}{\alpha - 1}\right)^\alpha + o(1) + \frac{\alpha \int_x^{x_0} (1 - F(t))^\alpha dt (1 - r(x))(1 - F(x))^{\alpha+1}}{(1 - F(x)) \int_x^{x_0} (1 - F(t))^{\alpha-1} dt (1 - F_4(x))^{\alpha+1}}$$

$$= -\left(\frac{\alpha}{\alpha - 1}\right)^\alpha + o(1) + \alpha r(x)(1 - r(x)) \left(\frac{\alpha}{\alpha - 1}\right)^{\alpha+1} + o(1)$$

$$= o(1) - (\alpha/(\alpha - 1))^\alpha + \alpha(\alpha - 1)\alpha^{-1}\alpha^{-1}(\alpha/(\alpha - 1))^{\alpha+1} = o(1).$$

Set $h_1 = h/c_1$, $f_1 = f/c_1$, and for $z_0$ sufficiently large and $x \geq z_0$

$$1 - F_4(x) = c \exp\left\{ -\int_{z_0}^x (h_1(t)/f_1(t)) dt \right\}$$

where $c > 0$. Therefore

$$1 - F(x) = \{(1 - F(x))/(1 - F_4(x))c\}\exp\left\{-\int_{z_0}^{x} (h_1(t)/f_1(t))dt\right\}$$

satisfies the representation in Corollary 1.7 and thus $F \in D(\Lambda)$.

For the converse suppose $F_\# \in D(\Lambda)$ and $F_\#$ is a Von Mises function with representation

$$1 - F_\#(x) = c\exp\left\{-\int_{z_0}^{x} (1/f_\#(u))du\right\}$$

for $z_0 < x < x_0$. For any $\alpha > 1$

$$(1 - F_\#(x))^\alpha = c\exp\left\{-\int_{z_0}^{x} (1/(\alpha^{-1}f_\#(u)))du\right\}$$

so that $(1 - F_\#(x))^\alpha$ is the distribution tail of a Von Mises function with auxiliary function $\alpha^{-1}f_\#(u)$; similarly for $(1 - F_\#(x))^{\alpha-1}$. From Proposition 1.9

$$\alpha^{-1}f_\#(x) \sim \int_x^{x_0} (1 - F_\#(t))^\alpha dt/(1 - F_\#(x))^\alpha$$

and

$$(\alpha - 1)^{-1}f_\#(x) \sim \int_x^{x_0} (1 - F_\#(t))^{\alpha-1} dt/(1 - F_\#(x))^{\alpha-1}$$

whence dividing

$$\frac{\int_x^{x_0}(1 - F_\#(t))^\alpha dt}{(1 - F_\#(x))\int_x^{x_0}(1 - F_\#(t))^{\alpha-1}dt} \to \alpha^{-1}(\alpha - 1)$$

which is (1.10).

For the general case if $F \in D(\Lambda)$ then by Proposition 1.4 there exist a Von Mises function $F_\#(x)$ and a function $c(x) \to c > 0$ such that

$$1 - F(x) = c(x)(1 - F_\#(x))$$

and since $1 - F_\#(x)$ satisfies (1.10) it is readily seen that $1 - F(x)$ does also. □

## Exercises

1.1.1. (a) Prove if $F \in D(\Lambda)$ then $\int_0^{x_0} x^k F(dx) < \infty$ for every $k > 0$. Construct an example of $F \in D(\Lambda)$ where $\int_{-\infty}^0 |x| F(dx) = \infty$.

(b) If $x_0 < \infty$ and $F \in D(\Lambda)$ then for any $n$

$$\lim_{t \uparrow x_0} (x_0 - x)^{-n}(1 - F(x)) = 0.$$

So $F$ is differentiable at $x_0$ with $F'(x_0) = 0$. If $x_0 = \infty$ and $F \in D(\Lambda)$ then for any $n$

$$\lim_{x \to \infty} x^n(1 - F(x)) = 0.$$

1.1.2. Suppose $F \in D(\Lambda)$ with auxiliary function $f$. Given examples where

$$f(x) \to 0, \qquad x \to x_0$$

$$f(x) \to \infty, \qquad x \to x_0$$

and $\lim_{t \to x_0} f(x)$ does not exist.

1.1.3. Let $F(x)$ be the lognormal.

(i) Is $F \in D(\Lambda)$?

(ii) Do all moments exist? If so what are they? (Remember what the moment generating function of the normal is.)

(iii) Does the moment generating function exist?

(iv) Is $F$ determined by its moments?

(Cf. Feller, 1971, page 227.)

1.1.4. Suppose for $u \in \mathbb{R}$, $v > 0$ that as $x \to \infty$

$$1 - F(x) \sim cx^{-u}e^{-x^v}.$$

Check $F \in D(\Lambda)$. Find $a_n$ and $b_n$.

1.1.5. Let $F(x) = 1 - (\log x)^{-1}$, $x \geq e$. Show $F \notin D(\Lambda)$. Moment considerations should suffice.

1.1.6. Let $F(x) = 1 - e^{-x^\alpha}$, $x > 0$, $\alpha > 0$. Show $F \in D(\Lambda)$. Find $a_n$ and $b_n$.

1.1.7. Show if (1.6) holds that there exists a monotone $U \in RV_1$ such that

$$1/(1 - F) = U \circ (1/(1 - F_\#))$$

where $F_\#$ is given in (1.3). Use this to check $F \in D(\Lambda)$ (de Haan, 1974a).

1.1.8. Derive Lemma 1.8 from Lemma 0.13 and inversion.

1.1.9. If $x_0 = \infty$ show $F \in D(\Lambda)$ implies rapid variation: i.e.,

$$\lim_{t \to x_0} \frac{1 - F(tx)}{1 - F(t)} = x^{-\infty} = \begin{cases} 0 & x > 1 \\ \infty & 0 < x < 1 \end{cases}$$

so that $F \in D(\Lambda)$ implies weak stability. Cf. Exercise 0.4.1.2. Show the rapid variation two ways:

(a) Use the representation (1.5).

(b) If $F \in D(\Lambda)$

$$(M_n - b_n)/a_n \Rightarrow Y$$

("$\Rightarrow$" denotes convergence in distribution) where $Y$ has distribution $\Lambda$. Divide through by $b_n/a_n$ and use Exercise 0.4.3.1.

1.1.10. Suppose for $i = 1, \ldots, k$ that $X_i$ is a random variable with distribution $F_i \in D(\Lambda)$. Show $\bigwedge_{i=1}^k X_i$ has a distribution in $D(\Lambda)$ (Balkema, unpublished letter).

1.1.11. (a) Suppose $X \geq 0$ has distribution $F$ and $x_0 = \infty$. If $F \in D(\Lambda)$ then the distribution of $-X^{-1}$ is in $D(\Lambda)$.

(b) Suppose $T$ maps the open interval $I$ onto the interval $J$ and that $T$ is twice differentiable and $T'$ is strictly positive on $I$. If $X$ is a random variable with values in $I$ and whose distribution $F \in D(\Lambda)$, when does $Y := T(X)$ have a distribution in $D(\Lambda)$ (Balkema)?

1.1.12. If $F \in D(\Lambda)$ and $F^n(a_n x + b_n) \to \Lambda(x)$, prove $(1 - F(a_n))/(1 - F(b_n)) \to \infty$.

1.1.13. Suppose $X_n$, $n \geq 1$ are iid with common distribution $F$ with right end $x_0 > 0$. If there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$P\left[\left(\bigvee_{i=1}^{n} X_i - b_n\right)\Big/ a_n \leq x\right] \to \Lambda(x)$$

show

$$P\left[\left(\bigvee_{i=1}^{n} X_i^2 - b_n^2\right)\Big/ 2a_n b_n \leq x\right] \to \Lambda(x).$$

Do this two ways:

(a) If $F \in D(\Lambda)$ show $F(x^{1/2})$ for $x > 0$ is in $D(\Lambda)$. Use the representation (1.5) and compute norming constants for $F(x^{1/2})$ in terms of those for $F$.

(b) Write

$$\bigvee_{i=1}^{n} X_i^2 - b_n^2 = \left(\bigvee_{i=1}^{n} X_i\right)^2 - b_n^2 = \left(\bigvee_{i=1}^{n} X_i - b_n\right)\left(\bigvee_{i=1}^{n} X_i + b_n\right)$$

and use Exercise 1.1.9.

1.1.14. Suppose $X$ is a random variable with distribution $F \in D(\Lambda)$ and suppose the auxiliary function is $f(t)$. Prove if $f(t) \to \infty$ the moment generating function of $X^+$ does not exist; it does however if $f(t) \to c \in [0, \infty)$. (Cf. Exercise 1.1.3.)

## 1.2. Domain of Attraction of $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x > 0$

The domain of Attraction of $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x > 0$ is related to regular variation.

**Proposition 1.11** (Gnedenko, 1943). *$F \in D(\Phi_\alpha)$ iff $1 - F \in RV_{-\alpha}$. In this case*

$$F^n(a_n x) \to \Phi_\alpha(x)$$

*with*

$$a_n = (1/(1 - F))^{\leftarrow}(n). \tag{1.12}$$

So only distributions with infinite right end may qualify for membership in $D(\Phi_\alpha)$.

Proof. If $1 - F \in RV_{-\alpha}$ and $a_n = (1/(1 - F))^{\leftarrow}(n)$ then because $1 - F(a_n) \sim n^{-1}$ we have for $x > 0$

$$n(1 - F(a_n x)) \sim (1 - F(a_n x))/(1 - F(a_n)) \to x^{-\alpha}$$

as $n \to \infty$ since $a_n \to \infty$. Therefore for $x > 0$

$$n(-\log F(a_n x)) \to x^{-\alpha}$$

and $F^n(a_n x) \to x^{-\alpha}$. If $x < 0$ then $F^n(a_n x) \le F^n(0) \to 0 = \Phi_\alpha(x)$ since regular variation requires $F(0) < 1$.

Conversely suppose $F \in D(\Phi_\alpha)$. This means there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) \to \Phi_\alpha(x).$$

Taking logarithms, this leads to

$$n(1 - F(a_n x + b_n) \to x^{-\alpha}, \qquad x > 0$$

as $n \to \infty$. Set $U = 1/(1 - F)$ and $V = U^{\leftarrow}$. As in Proposition 0.10 we may invert the relation

$$U(a_n x + b_n)/n \to x^\alpha, \qquad x > 0$$

and switch to functions $a(t)$, $b(t)$ of a continuous variable to obtain

$$(V(ty) - b(t))/a(t) \to y^{1/\alpha}, \qquad y > 0$$

or in a more convenient form

$$(V(ty) - V(t))/a(t) \to y^{1/\alpha} - 1, \qquad y > 0. \tag{1.13}$$

First of all (1.13) implies $a(t) \in RV_{1/\alpha}$. To see this we mimic Proposition 0.12. For $x > 0$

$$\begin{aligned}\lim_{t\to\infty} a(tx)/a(t) &= \lim_{t\to\infty} \left(\frac{V(tx) - V(t)}{a(t)}\right) \Big/ \left(-\left(\frac{V(tx\cdot x^{-1}) - V(tx)}{a(tx)}\right)\right) \\ &= (x^{1/\alpha} - 1)/(-(x^{-1/\alpha} - 1) = x^{1/\alpha}.\end{aligned}$$

This means that for any fixed $y > 0$ the function $V(ty) - V(t)$, considered as a function of $t$, is also in $RV_{1/\alpha}$ and therefore by Karamata's theorem 0.6

$$\lim_{t\to\infty} \frac{\int_2^t (V(sy) - V(s))ds}{t(V(ty) - V(t))} = \frac{\alpha}{\alpha + 1}$$

and taking into account the fact that

$$\lim_{t\to\infty} \int_2^{2y} V(s)ds/(t(V(ty) - V(t))) = 0$$

(as a consequence of $t(V(ty) - V(t)) \to \infty$) we may rewrite the limit relation as

$$\lim_{t\to\infty} \frac{(ty)^{-1}\int_2^{ty} V(s)ds - t^{-1}\int_2^t V(s)ds}{a(t)} = \left(\frac{\alpha}{\alpha + 1}\right)(y^{1/\alpha} - 1).$$

Next observe that 1.13 is the convergence of monotone functions to a continuous limit and so the convergence is locally uniform. Therefore integrating over $y$ in (1.13) gives for $0 < \delta < 1$

$$\int_\delta^1 (V(ty) - V(t))dy/a(t) \to \int_\delta^1 y^{1/\alpha}dy - (1 - \delta);$$

i.e.,

$$\left((1-\delta)V(t) - t^{-1}\int_{\delta t}^{t} V(s)ds\right)\Big/ a(t) \to (1+\alpha)^{-1} + O(\delta)$$

where $O(\delta) = -\delta + \alpha\delta^{\alpha^{-1}+1}/(\alpha+1)$. This means

$$\begin{aligned}
&\left\{V(t) - t^{-1}\int_2^t V(s)ds - \delta\left[V(t) - (\delta t)^{-1}\int_2^{\delta t} V(s)ds\right]\right\}\Big/ a(t)\\
&= \left\{V(t) - t^{-1}\int_2^t V(s)ds - \delta\left[V(t) - t^{-1}\int_2^t V(s)ds + t^{-1}\int_2^t V(s)ds\right.\right.\\
&\quad \left.\left. - (\delta t)^{-1}\int_2^{\delta t} V(s)ds\right]\right\}\Big/ a(t)\\
&= \left\{(1-\delta)\left(V(t) - t^{-1}\int_2^t V(s)ds\right)\right.\\
&\quad \left. - \delta\left[t^{-1}\int_2^t V(s)ds - (\delta t)^{-1}\int_2^{\delta t} V(s)ds\right]\right\}\Big/ a(t)\\
&\to (1+\alpha)^{-1} + O(\delta)
\end{aligned}$$

and since

$$\begin{aligned}
&\left[t^{-1}\int_2^t V(s)ds - (\delta t)^{-1}\int_2^{\delta t} V(s)ds\right]\Big/ a(t)\\
&\quad \to \alpha(\alpha+1)^{-1}(1-\delta^{1/\alpha})
\end{aligned}$$

we get

$$\begin{aligned}
&\left(V(t) - t^{-1}\int_2^t V(s)ds\right)\Big/ a(t)\\
&\quad \to \frac{(1+\alpha)^{-1} + O(\delta) + \delta\alpha(\alpha+1)^{-1}(1-\delta^{1/\alpha})}{1-\delta}\\
&\quad = \frac{(1+\alpha)^{-1} - \delta + \alpha\delta^{\alpha^{-1}+1}(\alpha+1)^{-1} + \delta\alpha(\alpha+1)^{-1}(1-\delta^{1/\alpha})}{1-\delta}\\
&\quad = (1+\alpha)^{-1}
\end{aligned}$$

as $t \to \infty$. A final rephrasing is

$$V(t) - t^{-1}\int_2^t V(s)ds = \alpha(t) \tag{1.14}$$

where $\alpha(t) \sim (1+\alpha)^{-1}a(t) \in RV_{\alpha^{-1}}$. We now invert (1.14) and express $V$ in terms of $\alpha(t)$: Divide (1.14) by $t$ and integrate from 2 to $y$:

$$\int_2^y t^{-1}V(t)dt - \int_2^y t^{-2}\int_2^t V(s)ds\,dt = \int_2^y t^{-1}\alpha(t)dt. \tag{1.15}$$

The second term on the left side becomes after reversing the order of integration

$$\int_2^y \left(\int_s^y t^{-2}dt\right) V(s)ds = \int_2^y (s^{-1} - y^{-1})V(s)ds$$
$$= \int_2^y s^{-1}V(s)ds - y^{-1}\int_2^y V(s)ds$$
$$= \int_2^y s^{-1}V(s)ds + \alpha(y) - V(y) \qquad \text{(from (1.14))}.$$

So substituting in (1.15) we get

$$V(y) = \alpha(y) + \int_2^y t^{-1}\alpha(t)dt.$$

Therefore

$$\lim_{y\to\infty} V(y)/\alpha(y) = 1 + \lim_{y\to\infty} \int_2^y t^{-1}\alpha(t)dt/\alpha(y),$$

and applying Karamata's theorem this is

$$1 + \int_0^1 t^{\alpha^{-1}-1}dt = 1 + \frac{1}{1/\alpha} = 1 + \alpha.$$

Finally as $t \to \infty$

$$V(t) \sim \alpha(t)(1 + \alpha) \sim (1 + \alpha)^{-1}(1 + \alpha)a(t) = a(t).$$

Therefore $V \in RV_{1/\alpha}$ and from Proposition 1.8(v)

$$V^{\leftarrow}(t) \sim 1/(1 - F(t)) \in RV_\alpha$$

so that

$$1 - F \in RV_{-\alpha}$$

as required. □

*Remark.* If we start from the assumption that $F^n(a_n x) \to \Phi_\alpha(x)$ (instead of $F^n(a_n x + b_n) \to \Phi_\alpha(x)$) then it is more elementary to check $1 - F \in RV_{-\alpha}$. See the example after Proposition 0.4.

**Corollary 1.12.** $F \in D(\Phi_\alpha)$ iff *there exist measurable functions* $c(x)$ *and* $\alpha(x)$ *defined on* $(1, \infty)$ *such that*

$$\lim_{x\to\infty} c(x) = c > 0$$

$$\lim_{x\to\infty} \alpha(x) = \alpha > 0$$

*and*

$$1 - F(x) = c(x)\exp\left\{-\int_1^x t^{-1}\alpha(t)dt\right\}$$

*for* $x \geq 1$.

**Proof.** This is just the corollary to Karamata's Theorem 0.6 in disguise. □

## Exercises

1.2.1. Check $F \in D(\Phi_\alpha)$ where $F$ is Cauchy with df $F(x) = 1/2 + \pi^{-1} \arctan x$. Find $a_n$. Do the same for the Pareto distribution:

$$1 - F(x) = x^{-\beta}, \qquad x \geq 1, \qquad \beta > 0.$$

1.2.2. If $F \in D(\Phi_\alpha)$ and $X$ is a random variable with distribution $F$ then

$$E(X^+)^\gamma < \infty, \qquad 0 < \gamma < \alpha.$$

Use this or any other method you like to check that if

$$P[X = k] = \frac{C}{k(\log k)^2}, \qquad k \geq 2$$

where $C > 0$ is chosen appropriately, then $F \notin D(\Phi_\alpha) \cup D(\Lambda)$.

1.2.3. If $F \in D(\Phi_\alpha)$ why is it impossible for maxima of iid random variables distributed according to $F$ to be relatively stable?

1.2.4. If $F \in D(\Phi_\alpha)$ and $\lim_{x\to\infty} (1 - F(x))/(1 - G(x)) = c > 0$ for some distribution $G$, then $G \in D(\Phi_\alpha)$. What are suitable normalizing constants for maxima of iid random variables from $G$ to converge in distribution?

1.2.5. Let $X_1, \ldots, X_n$ be a sample of size $n$ from a continuous distribution $F(x)$ with $x_0 = \infty$ and let $X_n^{(1)}$ be the term of maximum modulus, i.e., the $X_i$ among $X_1, \ldots, X_n$ for which $|X_i|$ is the largest. (Ties among the $X$'s occur only with probability zero and can thus be neglected.)

(a) What is the distribution function of $X_n^{(1)}$?

(b) Prove the following relative stability result: There exist $b_n \to \infty$ such that $X_n^{(1)}/b_n \xrightarrow{P} 1$ iff $1 - F \in RV_{-\infty}$ and

$$1 - F(x) \sim P[|X_1| > x] \text{ as } x \to \infty.$$

(c) There exist $a_n \to \infty$ such that $X_n^{(1)}/a_n$ has a nondegenerate weak limit $X^{(1)}$ iff for some $\alpha \in (0, \infty]$ we have $1 - F \in RV_{-\alpha}$ and

$$\lim_{x\to\infty} (1 - F(x))/P[|X_1| > x] = C_+, \qquad \lim_{x\to\infty} \frac{F(-x)}{P[|X_1| > x]} = C_-$$

and if $\alpha = \infty$ both $C_+ > 0$, $C_- > 0$. In this case the limit satisfies for $x > 0$ and $\alpha < \infty$

$$P[X^{(1)} > x] = C_+(1 - e^{-x^{-\alpha}})$$

$$P[X^{(1)} < -x] = C_- e^{-|x|^{-\alpha}}$$

and if $\alpha = \infty$ we have

$$P[X^{(1)} = 1] = C_+, \qquad P[X^{(1)} = -1] = C_- \text{ (Cline, unpublished).}$$

Hint: Prove, using a change of variable and a Tauberian theorem, that for two distributions $F_1$, $F_2$

$$\lim_{x\to\infty} (1 - F_2(x))/(1 - F_1(x)) = l \in [0, \infty)$$

iff

$$\lim_{x\to\infty} n \int_{-\infty}^{\infty} F_1^{n-1}(x) F_2(dx) = l$$

(Maller and Resnick, 1984).

1.2.6. (a) Suppose $X > 0$ is a random variable with distribution $F \in D(\Phi_\alpha) \cup D(\Lambda)$. Then $\log X$ has distribution $F(e^x) \in D(\Lambda)$. Relate the norming constants of $\log X$ to those of $X$.

(b) Suppose $X$ is a random variable with distribution $F \in D(\Phi_\alpha)$. Suppose $g$ is defined on the range of $X$ and $g$ is continuous strictly increasing so that $g \circ X$ has distribution $F \circ g^{\leftarrow}$. Find conditions on $g$ which insure $F \circ g^{\leftarrow} \in D(\Lambda)$.

## 1.3. Domain of Attraction of $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$, $x < 0$

The last case is also related to regular variation. Suppose $\alpha > 0$.

**Proposition 1.13** (Gnedenko, 1943). *$F \in D(\Psi_\alpha)$ iff $x_0 < \infty$ and $1 - F(x_0 - x^{-1}) \in RV_{-\alpha}$, $x \to \infty$. In this case we may set*

$$\gamma_n = (1/(1 - F))^{\leftarrow}(n)$$

*and then*

$$F^n(x_0 + (x_0 - \gamma_n)x) \to \Psi_\alpha(x), \qquad x < 0.$$

Proof. Suppose $x_0 < \infty$ and $1 - F(x_0 - x^{-1}) \in RV_{-\alpha}$. Define

$$F_*(x) = \begin{cases} 0 & x < 0 \\ F(x_0 - x^{-1}), & x \geq 0. \end{cases}$$

Then $1 - F_*(x) \in RV_{-\alpha}$ and from Proposition 1.11 we may set $a_n = (1/(1 - F_*))^{\leftarrow}(n)$ and

$$F_*^n(a_n x) \to \Phi_\alpha(x), \qquad x > 0$$

i.e.,

$$F^n(x_0 - (a_n x)^{-1}) \to \exp\{-x^{-\alpha}\}, \qquad x > 0$$

whence

$$F^n(x_0 + a_n^{-1} y) \to \exp\{-(-y)^\alpha\}, \qquad y < 0.$$

Observe

$$\begin{aligned}
a_n &= \inf\{u\colon 1/(1 - F_*(u)) \geq n\} \\
&= \inf\{u\colon 1/(1 - F(x_0 - u^{-1})) \geq n\} \\
&= \inf\left\{\frac{1}{x_0 - s}\colon 1/(1 - F(s)) \geq n\right\} \\
&= 1/(x_0 - \inf\{s\colon 1/(1 - F(s)) \geq n\}) = 1/(x_0 - \gamma_n)
\end{aligned}$$

and therefore

$$F^n(x_0 + (x_0 - \gamma_n)y) \to \Psi_\alpha(y), \qquad y < 0$$

as required.

Conversely: Suppose there exist $a_n > 0$, $b_n$, $n \geq 1$ such that

$$F^n(a_n x + b_n) \to \Psi_\alpha(x).$$

Letting $U = 1/(1 - F)$ we find

$$U(a_n x + b_n)/n \to (-x)^{-\alpha}, \qquad x < 0$$

and inverting we have

$$(U^{\leftarrow}(ny) - b_n)/a_n \to -(y^{-1/\alpha}), \qquad y > 0.$$

Set $V = U^{\leftarrow}$ and switch to a continuous variable. We obtain

$$(V(ty) - V(t))/a(t) \to 1 - y^{-1/\alpha}, \qquad y > 0. \tag{1.16}$$

This relation implies $a(t) \in RV_{-1/\alpha}$ since for $x > 0$

$$\begin{aligned}
&\lim_{t\to\infty} (a(tx)/a(t)) \\
&= \lim_{t\to\infty} \left(\frac{V(tx) - V(t)}{a(t)}\right) \Bigg/ \left(-\left(\frac{V(tx\cdot x^{-1}) - V(tx)}{a(tx)}\right)\right) \\
&= (1 - x^{-1/\alpha})/(-(1 - x^{1/\alpha})) \\
&= x^{-1/\alpha}.
\end{aligned}$$

We next show $x_0 := V(\infty) < \infty$. It is enough to show $\lim_{n\to\infty} V(2^n) < \infty$. Pick $\delta < \alpha^{-1}$. From (1.16) and the fact that $a(t) \in RV_{-1/\alpha}$ it is clear there exists $n_0$ such that $n \geq n_0$ implies

$$(V(2^{n+1}) - V(2^n))/a(2^n) \leq 2(1 - 2^{-\alpha^{-1}})$$

and

$$a(2^{n+1})/a(2^n) < 2^\delta 2^{-1/\alpha}.$$

Then for any $k \geq 1$ (a product over an empty index set is 1)

$$\begin{aligned}
(V(2^{n_0+k}) - V(2^{n_0}))/a(2^{n_0}) &= \sum_{j=1}^{k} \frac{V(2^{n_0+j}) - V(2^{n_0+j-1})}{a(2^{n_0+j-1})} \prod_{i=n_0}^{n_0-j-2} \frac{a(2^{i+1})}{a(2^i)} \\
&\leq 2(1 - 2^{-1/\alpha}) \sum_{j=1}^{k} (2^{-(\alpha^{-1}-\delta)})^{j-1}
\end{aligned}$$

so that

$$\lim_{k\to\infty} (V(2^{n_0+k}) - V(2^{n_0}))/a(2^{n_0}) < \infty$$

and we readily conclude $V(\infty) < \infty$.

An elaboration of this argument, which mimics the argument in Lemma 0.13, produces a bound ($\varepsilon < \alpha^{-1}$)

$$\frac{V(ty) - V(t)}{a(t)} \le c(1 - 2^{-(\alpha^{-1}-\varepsilon)}y^{-(\alpha^{-1}-\varepsilon)})$$

valid for $y > 1$ and for all sufficiently large $t$. Dividing by $y^2$ produces a uniform bound which allows an application of the dominated convergence theorem. In (1.16) divide by $y^2$ and integrate over $(1, \infty)$ to get after setting $\bar{V}(t) = V(\infty) - V(t)$

$$\int_1^\infty y^{-2}(\bar{V}(ty) - \bar{V}(t))/a(t)dy \to \int_1^\infty (y^{-1/\alpha-2} - y^{-2})dy = -1/(1+\alpha);$$

i.e.,

$$\bar{V}(t) - t\int_t^\infty s^{-2}\bar{V}(s)ds = \alpha(t) \tag{1.17}$$

where

$$\alpha(t) \sim a(t)/(1+\alpha) \in RV_{-1/\alpha}, \qquad t \to \infty.$$

We now invert (1.17) and express $\bar{V}$ in terms of $a(t)$. Divide (1.17) through by $t$ and integrate:

$$\int_y^\infty t^{-1}\bar{V}(t)dt - \int_y^\infty \int_y^\infty s^{-2}\bar{V}(s)ds\,dt = \int_y^\infty t^{-1}\alpha(t)dt.$$

The second term, after a Fubini inversion, is

$$-\left\{\int_y^\infty s^{-1}\bar{V}(s)ds - y\int_y^\infty s^{-2}\bar{V}(s)ds\right\}$$

and hence from (1.17)

$$\bar{V}(y) - \alpha(y) = y\int_y^\infty s^{-2}\bar{V}(s)ds = \int_y^\infty s^{-1}\alpha(s)ds$$

so that applying Proposition 0.5

$$\bar{V}(y)/\alpha(y) = 1 + \int_1^\infty \frac{\alpha(ys)ds}{\alpha(y)s} \to 1 + \alpha$$

and

$$\bar{V}(y) \sim (1+\alpha)\alpha(y) \sim (1+\alpha)(1+\alpha)^{-1}a(y) = a(y)$$

as $y \to \infty$. Therefore

$$H(t) := 1/(V(\infty) - V(t)) \in RV_{1/\alpha}$$

so that

$$V(t) = V(\infty) - 1/H(t).$$

Inverting gives

$$1/(1 - F(y)) \sim H^{\leftarrow}((V(\infty) - y)^{-1})$$

and setting $s = (V(\infty) - y)^{-1}$ we get finally

$$1/(1 - F(V(\infty) - s^{-1})) \sim H^{\leftarrow}(s) \in RV_\alpha$$

(cf. Proposition 0.8(v)). □

We have the following representation.

**Corollary 1.14.** $F \in D(\Psi_\alpha)$ iff $x_0 < \infty$ *and there exist functions* $\delta$: $\mathbb{R}_+ \to \mathbb{R}_+$ *and a function* $c$: $\mathbb{R}_+ \to \mathbb{R}_+$ *and a constant* $c_0 > 0$ *such that*

$$\lim_{t \uparrow x_0} \delta(t) = \alpha, \qquad \lim_{t \uparrow x_0} c(t) = c_0 \tag{1.18}$$

*and for* $x < x_0$

$$1 - F(x) = c(x)\exp\left\{-\int_{x_0-1}^{x} \delta(t)/(x_0 - t)dt\right\}.$$

Proof. If $F \in D(\Psi_\alpha)$ then

$$1 - F(x_0 - x^{-1}) \in RV_{-\alpha} \quad \text{so}$$

$$1 - F(x_0 - x^{-1}) = \bar{c}(x)\exp\left\{-\int_1^x (\bar{\delta}(t)/t)dt\right\}$$

where $\bar{\delta}$ and $\bar{c}$ have the properties given in (1.18). Letting $y = x_0 - x^{-1}$ (so $x = (x_0 - y)^{-1}$) we get for $y < x_0$

$$1 - F(y) = \bar{c}((x_0 - y)^{-1})\exp\left\{-\int_1^{(x_0-y)^{-1}} (\bar{\delta}(t)/t)dt\right\}.$$

In the integral, change variables ($s = x_0 - t^{-1}$) to get

$$1 - F(y) = \bar{c}((x_0 - y)^{-1})\exp\left\{-\int_{x_0-1}^{y} (\bar{\delta}((x_0 - s)^{-1})/(x_0 - s))ds\right\}$$

$$= c(y)\exp\left\{-\int_{x_0-1}^{y} (\delta(s)/(x_0 - s))ds\right\}.$$ □

## 1.4. Von Mises Conditions

Some sufficient conditions for a distribution to belong to a domain of attraction were originally given by Von Mises (1936) and are often more convenient to verify than some of the conditions so far presented.

For $\Phi_\alpha$, inspecting the tails often suffices to check whether or not $F \in D(\Phi_\alpha)$. For example, $F(x) = 1 - x^{-\alpha}$, $x \geq 1$ is obviously in $D(\Phi_\alpha)$ and $F(x) = 1 - e^{-x}$, $x > 0$ is obviously not. Sometimes when a density $F'$ exists it is simple to check that $F'$ is regularly varying near $\infty$ and then a mental application of Karamata's theorem does the trick. This works, for example, with the Cauchy distribution whose density is in $RV_{-2}$. The following result is also useful. It is basically a minor rephrasing of the corollary on page 17.

**Proposition 1.15.** *Suppose $F$ is absolutely continuous with positive density $F'$ in some neighborhood of $\infty$.*
(a) *If for some $\alpha > 0$*

$$\lim_{x \to \infty} xF'(x)/(1 - F(x)) = \alpha \tag{1.19}$$

*then $F \in D(\Phi_\alpha)$. We may choose $a_n$ to satisfy $a_n F'(a_n) \sim \alpha/n$.*
(b) *If $F'$ is nonincreasing and $F \in D(\Phi_\alpha)$ then* (1.19) *holds.*
(c) *Equation* (1.19) *holds* iff *in the representation for* $1 - F$ *given in Corollary* 1.12, *$c(x)$ is ultimately constant, i.e.,* iff *for some $z_0$ and all $x > z_0$, we have*

$$1 - F(x) = c \exp\left\{-\int_{z_0}^{x} t^{-1}\alpha(t)dt\right\}$$

*where* $\lim_{t\to\infty} \alpha(t) = \alpha$.

PROOF. See the corollary on page 17, and its proof. □

For $\Psi_\alpha$ there is an analogous result.

**Proposition 1.16.** *Suppose $F$ has finite right endpoint $x_0$ and is absolutely continuous in a left neighborhood of $x_0$ with positive density $F'$.*
(a) *If for some $\alpha > 0$*

$$\lim_{x \uparrow x_0} (x_0 - x)F'(x)/(1 - F(x)) = \alpha \tag{1.20}$$

*then $F \in D(\Psi_\alpha)$.*
(b) *If $F'$ is nonincreasing and $F \in D(\Psi_\alpha)$ then* (1.20) *holds.*
(c) *Equation* (1.20) *holds* iff *$c(x)$ in the representation of Corollary* 1.14 *can be taken to be constant in some left neighborhood of $x_0$.*

PROOF. Recall $F \in D(\Psi_\alpha)$ iff $F_*(x) = F(x_0 - x^{-1}) \in D(\Phi_\alpha)$. But

$$\alpha = \lim_{x\to\infty} xF_*'(x)/(1 - F_*(x)) = \lim_{x\to\infty} xF'(x_0 - x^{-1})x^{-2}/(1 - F(x_0 - x^{-1})$$

iff

$$\alpha = \lim_{s \to x_0} F'(s)(x_0 - s)/(1 - F(s)).$$ □

For $\Lambda$ there are two results depending on how many derivatives $F$ is assumed to possess (cf. de Haan, 1970).

**Proposition 1.17.** *Let $F$ be absolutely continuous in a left neighborhood of $x_0$ with density $F'$.*
(a) *If*

$$\lim_{t\uparrow x_0} F'(x)\int_x^{x_0}(1-F(t))dt/(1-F(x))^2 = 1 \tag{1.21}$$

*then $F\in D(\Lambda)$. In this case we may take*

$$f(t) = \int_t^{x_0}(1-F(s))ds/(1-F(x)) \tag{1.22}$$

$$b_n = (1/(1-F))^{\leftarrow}(n), \qquad a_n = f(b_n).$$

(b) *If $F'$ is nonincreasing and $F\in D(\Lambda)$ then* (1.21) *holds.*
(c) *Equation* (1.21) *holds* iff *in representation* (1.6), *$c(x)$ may be taken constant, i.e.*, iff

$$1-F(x) = c\exp\left\{-\int_{z_0}^x (g(t)/f(t))dt\right\}, \qquad z_0 < x < x_0, \tag{1.23}$$

*where* $\lim_{t\uparrow x_0} g(x) = 1$ *and $f$ is absolutely continuous with density $f'(x)\to 0$ as $x\uparrow x_0$.*
(d) *Equation* (1.21) *or* (1.23) *are equivalent to* $tF'((1/(1-F))^{\leftarrow}(t))\in RV_0$ (*Sweeting*, 1985).

Proof. (a) With the choice of $f$ we have

$$f'(x) = \frac{-(1-F(x))^2 + \int_x^{x_0}(1-F(s))ds\,F'(x)}{(1-F(x))^2} \to -1+1 = 0.$$

Furthermore if $R = -\log(1-F)$ then $R' = F'/(1-F) = g/f$ where

$$g(x) = F'(x)\int_x^{x_0}(1-F(t))dt/(1-F(x))^2$$

and hence

$$\int_1^x R'(s)ds = R(x) - R(1) = \int_1^x (g(s)/f(s))ds$$

and

$$1-F(x) = e^{-R(1)}e^{-\int_1^x (g(s)/f(s))ds}$$

and so $F\in D(\Lambda)$ by Corollary 1.7.
(b) Since $F$ is nondecreasing, $F'\geq 0$. In fact in a left neighborhood of $x_0$, $F'(x)>0$ since otherwise if for $x_1<x_0$ $F'(x_1)=0$, then because $F'$ is nonincreasing we would have $F'(x)=0$, $x_1\leq x\leq x_0$ and hence $F$ would be

constant on $(x_1, x_0)$, which would contradict the definition of $x_0$. But $F'(x) > 0$ means $U := 1/(1-F)$ is continuous strictly increasing so that $U \circ U^{\leftarrow}(x) = U^{\leftarrow} \circ U(x) = x$.

Recall $F \in D(\Lambda)$ means $U \in \Gamma$ and a suitable auxiliary function is

$$f(t) = \int_t^{x_0} (1 - F(s))ds/(1 - F(t)).$$

(Proposition 1.9). Inverting via Proposition 0.9(a) we have $U^{\leftarrow} \in \Pi$ with auxiliary function

$$a(t) = f \circ U^{\leftarrow}(t).$$

On the other hand

$$\begin{aligned}(U^{\leftarrow})'(t) = 1/U'(U^{\leftarrow}(t)) &= (1 - F(U^{\leftarrow}(t)))^2/F'(U^{\leftarrow}(t)) \\ &= 1/(F'(U^{\leftarrow}(t))t^2)\end{aligned}$$

so that $t^2(U^{\leftarrow})'(t)$ is nondecreasing, and so by Proposition 0.11(b) and the remark following that proposition we have another choice for $a(t)$:

$$a(t) \sim t(U^{\leftarrow})'(t) = 1/(tF'(U^{\leftarrow}(t)).$$

Therefore since the two choices of $a(t)$ are asymptotically equivalent we have

$$\begin{aligned}1 &= \lim_{t\to\infty} f(U^{\leftarrow}(t))tF'(U^{\leftarrow}(t)) \\ &= \lim_{x\uparrow x_0} \frac{f(x)F'(x)}{1 - F(x)} = \lim_{x\uparrow x_0} \frac{\int_x^{x_0}(1 - F(t))dt\, F'(x)}{(1 - F(x))^2}.\end{aligned}$$

(c) The proof that (1.21) implies (1.23) was given in (a). If representation (1.23) holds, then from Corollary 1.7 and its proof a suitable auxiliary function is the $f$ from the representation. From Theorem 1.9 and the asymptotic uniqueness of auxiliary functions

$$\int_x^{x_0} (1 - F(s))ds/(1 - F(x)) \sim f(x) \sim f(x)/g(x) = (1 - F(x))/F'(x)$$

which is equivalent to (1.21).

(d) This is practically the same as the previous steps. If (1.21) holds, $U = 1/(1 - F) \in \Gamma$ and since $\int_x^{x_0}(1 - F(s))ds/(1 - F(x)) \sim (1 - F(x))/F'(x)$ we get from Proposition 1.9 and the asymptotic uniqueness of auxiliary functions that $f_0(x) = (1 - F(x))/F'(x)$ is a suitable auxiliary function. This means $U^{\leftarrow} \in \Pi$ with auxiliary function $f_0 \circ U^{\leftarrow}(x) = 1/(xF'((1/(1 - F))^{\leftarrow}(x)))$ and since auxiliary functions of $\Pi$-varying functions are slowly varying (Proposition 0.12) the result follows. Conversely suppose $tF'((1/(1 - F))^{\leftarrow}(t)) \in RV_0$. Then as in (b) we have

$$(U^{\leftarrow})'(t) = 1/(t^2F'(U^{\leftarrow}(t))) \in RV_{-1}$$

so that $U^{\leftarrow} \in \Pi$ with auxiliary $a$-function $(tF'(U^{\leftarrow}(t)))^{-1}$ (Proposition 0.11) and thus by inversion $(U^{\leftarrow})^{\leftarrow} = U \in \Gamma$ with auxiliary function $f_0(t) = (U(t)F'(U^{\leftarrow} \circ U(t)))^{-1} = (1 - F(t))/F'(t)$. Again by Proposition 1.9 a suitable auxiliary function is always $\int_t^{x_0} (1 - F(s))ds/(1 - F(t))$ and this must be asymptotic to $f_0$ giving (1.21). □

The last result of this section requires two derivatives.

**Proposition 1.18.** *Suppose $F$ has a negative second derivative $F''$ for all $x$ in some left neighborhood of $x_0$.*
(a) *If*

$$\lim_{x \uparrow x_0} F''(x)(1 - F(x))/(F'(x))^2 = -1 \tag{1.24}$$

*then $F \in D(\Lambda)$. We may take $f = (1 - F)/F'$.*
(b) *If $F''$ is nondecreasing, $F'(x) = \int_t^{x_0} (-F''(u))du$ and $F \in D(\Lambda)$, then* (1.24) *holds.*
(c) *Equation* (1.24) *holds* iff *$F$ is a twice differentiable Von Mises function so that* (1.3) *holds.*

Proof. (a) and (c) See Proposition 1.1(b).
(b) Observe $F'(x)$ is decreasing since $F''(x) < 0$. So by Proposition 1.17, since $F \in D(\Lambda)$, we get (1.21) holding. Define for $x$ sufficiently close to $x_0$

$$F_0(x) := 1 - F'(x)$$

so that $F_0$ is a distribution and $1 - F_0 = F'$. Rewriting (1.21) in terms of $F_0$ gives

$$\lim_{x \uparrow x_0} (1 - F_0(x)) \int_x^{x_0} \int_y^{x_0} (1 - F_0(s))ds\, dy/(1 - F_0(x))^2 = 1$$

so by Proposition 1.9, $F_0 \in D(\Lambda)$. But $F_0' = -F''$ is nonincreasing so applying again Proposition 1.17 gives

$$\lim_{x \uparrow x_0} F_0'(x) \int_x^{x_0} (1 - F_0(t))dt/(1 - F_0(x))^2 = 1$$

which translates into

$$\lim_{x \uparrow x_0} -F''(x)(1 - F(x)/(F'(x))^2 = 1$$

as required. □

It is useful to note how the various Von Mises conditions simplify the general representations of distributions in a domain of attraction. The Von Mises conditions will also be seen to play a role in local limit theory discussed in the next chapter.

EXERCISES

1.4.1. Prove Proposition 1.17(b) without resorting to inverse functions by imitating the methods of Proposition 0.12(b) and Proposition 0.7(b) (de Haan, 1970).

1.4.2. Generalize Proposition 1.17(b) to the case that $F$ has increasing or decreasing failure rate, i.e, to the case that

$$F'(x)/(1 - F(x))$$

is monotone.

1.4.3. If $F$ is absolutely continuous and

$$\lim_{x\to\infty} xF'(x)/(1 - F(x)) = \infty$$

then $1 - F$ is rapidly varying.

1.4.4. Check that the $t$ and $F$ densities satisfy a Von Mises condition.

## 1.5. Equivalence Classes and Computation of Normalizing Constants

Computing normalizing constants can be a brutal business, and any techniques which aid in this are welcome indeed. This is the focus of our discussion on equivalence classes.

We say two distributions $F$ and $G$ are *tail equivalent* if they have the same right endpoint $x_0$ and for some $A > 0$

$$\lim_{x\uparrow x_0} (1 - F(x))/(1 - G(x)) = A. \tag{1.25}$$

**Proposition 1.19.** *Let $F$ and $G$ be distribution functions and suppose $H_i$ is an extreme value distribution, $i = 1, 2$. Suppose that $F \in D(H_1)$ and that*

$$F^n(a_n x + b_n) \to H_1(x) \tag{1.26}$$

*for normalizing constants $a_n > 0$, $b_n \geq 1$. Then*

$$G^n(a_n x + b_n) \to H_2(x) \tag{1.27}$$

iff *for some $a > 0$, $b \in \mathbb{R}$*

$$H_2(x) = H_1(ax + b),$$

*$F$ and $G$ are tail equivalent with right endpoint $x_0$ and if*

(i) $H_1 = \Phi_\alpha$, *then* $b = 0$ *and* $\lim_{x\to\infty}(1 - F(x))/(1 - G(x)) = a^\alpha$;
(ii) $H_1 = \Psi_\alpha$, *then* $b = 0$ *and* $\lim_{x\to x_0}(1 - F(x))/(1 - G(x)) = a^{-\alpha}$;
(iii) $H_1 = \Lambda$, *then* $a = 1$ *and* $\lim_{x\to x_0}(1 - F(x))/(1 - G(x)) = e^b$.

*Remark.* Regarding the problem of calculating normalizing constants, this result suggests we switch to an easy tail equivalent distribution and compute constants for that one.

PROOF. Suppose first that $F$ and $G$ are tail equivalent and that (1.26) holds. An equivalent formulation of (1.26) is

$$n(1 - F(a_n x + b_n)) \to -\log H_1(x)$$

for $x$ such that $H_1(x) > 0$. For such $x$, $a_n x + b_n \to x_0$ and hence from tail equivalence

$$n(1 - G(a_n x + b_n)) \sim nA^{-1}(1 - F(a_n x + b_n)) \to -A^{-1} \log H_1(x);$$

i.e.,

$$G^n(a_n x + b_n) \to H_1^{A^{-1}}(x).$$

For the converse suppose we are given (1.26) and (1.27) and we wish to show $F$ and $G$ are tail equivalent. Set

$$V_F(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t), \qquad V_G(t) = \left(\frac{1}{1-G}\right)^{\leftarrow}(t)$$

and equivalent to (1.26) and (1.27) are the following two statements in terms of inverses:

$$\lim_{t\to\infty} (V_F(ty) - b(t))/a(t) = \left(\frac{1}{-\log H_1}\right)^{\leftarrow}(y) \tag{1.28}$$

$$\lim_{t\to\infty} (V_G(ty) - b(t))/a(t) = \left(\frac{1}{-\log H_2}\right)^{\leftarrow}(y) \tag{1.29}$$

for $y > 0$. Recall there are three mutually exclusive possibilities for $a(t)$:

$$\begin{array}{lll} a(t) \in RV_{1/\alpha} & \text{if} \quad H_1 = \Phi_\alpha, & \alpha > 0 \\ a(t) \in RV_0 & \text{if} \quad H_1 = \Lambda & \\ a(t) \in RV_{-1/\alpha} & \text{if} \quad H_1 = \Psi_\alpha, & \alpha > 0 \end{array}$$

(Propositions 0.12, 1.11, and 1.12).

Suppose first that $H_1(x) = \Phi_\alpha(x)$. Then $a(t) \in RV_{1/\alpha}$ and from (1.29) we must have $H_2(x) = \Phi_\alpha(ax + b)$. We check easily that $b = 0$ as follows. Recall from Proposition 1.11 that $V_F(t) \sim a(t)$. If we set $y = 1$ in (1.28) we get

$$\lim_{t\to\infty} (V_F(t) - b(t))/a(t) = 1$$

which requires $\lim_{t\to\infty} b(t)/a(t) = 0$. Therefore (1.29) becomes for $y > 0$

$$\lim_{t\to\infty} V_G(ty)/a(t) = \left(\frac{1}{-\log H_2}\right)^{\leftarrow}(y) = (y^{1/\alpha} - b)/a.$$

However because $a(t) \in RV_{1/\alpha}$ we also have

$$\lim_{t\to\infty} V_G(ty)/a(t) = \lim_{t\to\infty} (V_G(ty)/a(ty))(a(ty)/a(t)) = a^{-1}(1 - b)y^{1/\alpha}$$

whence for $y > 0$

$$(y^{1/\alpha} - b)/a = a^{-1}(1 - b)y^{1/\alpha}$$

which reduces to

$$b = by^{1/\alpha}.$$

This necessitates $b = 0$. Thus (1.28) and (1.29) reduce to

$$V_F(ty)/a(t) \to y^{1/\alpha}$$
$$V_G(ty)/a(t) \to a^{-1} y^{1/\alpha}$$

and so

$$V_F(t) \sim aV_G(t)$$

and since $V_F \in RV_{1/\alpha}$ we get by inverting and using Proposition 0.8(vi)

$$\frac{1}{1 - F(t)} \sim a^{-\alpha}\left(\frac{1}{1 - G(t)}\right)$$

i.e., $a^\alpha = \lim_{t\to\infty} (1 - F(t))/(1 - G(t))$.

Next suppose $H_1 = \Psi_\alpha$ so that $a(t) \in RV_{-\alpha^{-1}}$. Then (1.29) becomes

$$(V_F(ty) - b(t))/a(t) \to -(y^{-1/\alpha}), \qquad y > 0$$

so that $(V_F(t) - b(t))/a(t) \to -1$. On the other hand recall from Proposition 1.13 that ($x_0 = V_F(\infty)$)

$$x_0 - V_F(t) \sim a(t)$$

so that

$$(b(t) - x_0)/a(t) = \frac{b(t) - V_F(t)}{a(t)} + \frac{V_F(t) - x_0}{a(t)} \to 1 - 1 = 0. \qquad (1.30)$$

From (1.29) we conclude $H_2(x) = \Psi_\alpha(ax + b)$ since $a(t) \in RV_{-\alpha^{-1}}$ and we now show why $b = 0$. Relation (1.29) becomes

$$\frac{V_G(ty) - b(t)}{a(t)} \to \frac{-(y^{-1/\alpha}) - b}{a}$$

and hence using (1.30)

$$\frac{x_0 - V_G(ty)}{a(t)} \to (y^{-1/\alpha} + b)/a \qquad \text{for } y > 0.$$

On the other hand

$$\frac{x_0 - V_G(ty)}{a(t)} = \frac{x_0 - V_G(ty)}{a(ty)} \cdot \frac{a(ty)}{a(t)} \to \left(\frac{1 + b}{a}\right) y^{-1/\alpha} = \left(\frac{y^{-1/\alpha} + b}{a}\right)$$

for $y > 0$ which requires $b = 0$. We now conclude

$$x_0 - V_G(t) \sim a^{-1}a(t)$$

and

$$x_0 - V_F(t) \sim a(t)$$

and therefore

$$(x_0 - V_F(t))/(x_0 - V_G(t)) \to a;$$

i.e.,

$$\frac{1}{x_0 - V_G(t)} \sim a\left(\frac{1}{x_0 - V_F(t)}\right)$$

and hence applying Proposition 0.8(vi) we get

$$\left(\frac{1}{x_0 - V_G}\right)^{\leftarrow}(t) \sim a^{-\alpha}\left(\frac{1}{x_0 - V_F}\right)^{\leftarrow}(t);$$

i.e.,

$$\frac{1}{1 - G(x_0 - t^{-1})} \sim a^{-\alpha}\frac{1}{1 - F(x_0 - t^{-1})}$$

so that

$$(1 - F(x_0 - t^{-1}))/(1 - G(x_0 - t^{-1})) \to a^{-\alpha}$$

as $t \to \infty$ as required.

Now suppose $H_1 = \Lambda(x)$ so that

$$(V_F(ty) - b(t))/a(t) \to \log y, \qquad y > 0. \tag{1.31}$$

Recall $a(t) \in RV_0$ and $b(t) \in \Pi$. Then $H_2(x) = \Lambda(ax + b)$, which entails on the one hand

$$(V_G(ty) - b(t))/a(t) \to ((\log y) - b)/a \tag{1.32}$$

and on the other

$$\frac{V_G(ty) - b(t)}{a(t)} = \frac{V_G(ty) - b(ty)}{a(ty)} \cdot \frac{a(ty)}{a(t)} + \frac{b(ty) - b(t)}{a(t)}$$
$$\to ((\log 1) - b)/a + \log y = -a^{-1}b + \log y.$$

However

$$(\log y - b)/a = -a^{-1}b + \log y$$

means

$$a^{-1}\log y = \log y$$

and hence $a = 1$. From (1.31)

$$(V_F(t) - b(t))/a(t) \to 0$$

and from (1.32) with $b(t)$ replaced by $V_F(t)$ we get

$$(V_G(ty) - V_F(t))/a(t) \to (\log y) - b, \qquad y > 0$$

and inverting we get for $x \in \mathbb{R}$

$$V_G^{\leftarrow}(xa(t) + V_F(t))/t \to e^{x+b}$$

and so if $x = 0$

$$V_G^{\leftarrow}(V_F(t))/t \to e^b.$$

Change variables now to obtain

$$V_G^{\leftarrow}(s)/V_F^{\leftarrow}(s) \to e^b;$$

i.e.,

$$\lim_{s \to x_0} (1 - F(s))/(1 - G(s)) = e^b. \qquad \square$$

EXAMPLE 1 (Cauchy). Let $F'(x) = (\pi(1 + x^2))^{-1}$, $x \in \mathbb{R}$. Then as $x \to \infty$, $F'(x) \sim (\pi x^2)^{-1}$ and so by Karamata's theorem

$$1 - F(x) = \int_x^{\infty} F'(u)du \sim \int_x^{\infty} \pi^{-1} u^{-2} du = (\pi x)^{-1}.$$

Therefore $F \in D(\Phi_1)$. Instead of solving $1 - F(x) = n^{-1}$ we solve $(\pi x)^{-1} = n^{-1}$ to get $a_n = n/\pi$ and so

$$F^n((n/\pi)x) \to \Phi_1(x).$$

EXAMPLE 2 (Normal). $F'(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$, $x \in R$. We have already checked $F$ is a Von Mises function, $F \in D(\Lambda)$, and we know the auxiliary function $f(t)$ satisfies

$$f(t) = \frac{1 - F(t)}{F'(t)} \sim \frac{n(t)/t}{n(t)} = 1/t \qquad \text{as } t \to \infty.$$

We now show $a_n = (2 \log n)^{-1/2}$

$$b_n = (2 \log n)^{1/2} - 1/2(\log\log n + \log 4\pi)/(2 \log n)^{1/2}$$

are acceptable choices of norming constants. Since $1 - F(t) \sim n(t)/t$ (Feller, 1968, page 174) we seek by tail equivalence to solve

$$(2\pi)^{-1/2} b_n^{-1} \exp\{-b_n^2/2\} = n^{-1}$$

and taking $-\log$ of both sides gives

$$(1/2)b_n^2 + \log b_n + 1/2 \log 2\pi = \log n. \tag{1.33}$$

We will construct an expansion of $b_n$ and indicate how many terms are necessary. Since $b_n \to \infty$ we see by dividing left and right sides of (1.33) by $b_n^2$ that as $n \to \infty$

$$b_n \sim (2 \log n)^{1/2}. \tag{1.34}$$

Since $a_n = f(b_n) \sim b_n^{-1}$ we see that an acceptable choice for $a_n$ is

$$a_n = (2\log n)^{-1/2}.$$

This tells us that in an expansion of $b_n$, we may neglect terms which are $o((\log n)^{-1/2})$. For if our expansion of $b_n$ is of the form $b_n = \beta_n + o((\log n)^{-1/2})$ then

$$(b_n - \beta_n)/a_n = o((\log n)^{-1/2})(\log n)^{1/2} \to 0$$

and the convergence to types theorem assures us $\beta_n$ is acceptable.

From (1.34) we see that

$$b_n = (2\log n)^{1/2} + r_n \tag{1.35}$$

where $r_n$ is a remainder which is $o((\log n)^{1/2})$. Now substitute (1.35) into (1.33) and we find

$$\begin{aligned} &1/2r_n^2 + (2\log n)^{1/2}r_n + 1/2\log\log n + 1/2\log 4\pi \\ &\quad + \log(1 + (2\log n)^{-1/2}r_n) = 0. \end{aligned} \tag{1.36}$$

Divide through by $(2\log n)^{1/2}r_n$ and we get

$$\begin{aligned} &\frac{r_n}{2(2\log n)^{1/2}} + 1 + (1/2)\frac{(\log\log n + \log 4\pi)}{r_n(2\log n)^{1/2}} \\ &\quad + \frac{\log(1 + (2\log n)^{-1/2}r_n)}{(2\log n)^{1/2}r_n} = 0. \end{aligned} \tag{1.37}$$

Because $r_n = o((\log n)^{1/2})$ and since the last term is asymptotic to $r_n(2\log n)^{-1/2}/(r_n(2\log n)^{1/2}) = 1/(2\log n) \to 0$ we see that (1.37) is of the form

$$o(1) + 1 + 1/2(\log\log n + \log 4\pi)/(r_n(2\log n)^{1/2}) = 0,$$

i.e.,

$$r_n = -1/2(\log\log n + \log 4\pi)/(2\log n)^{1/2} + s_n \tag{1.38}$$

where $s_n = o(\log\log n/(\log n)^{1/2})$. In fact $s_n = o((\log n)^{-1/2})$, which means we have done enough expanding. To see this observe that (1.36) implies

$$\begin{aligned} (2\log n)^{1/2}r_n + 1/2(\log\log n + \log 4\pi) &= -\log(1 + (2\log n)^{-1/2}r_n) - r_n^2/2 \\ &= -(2\log n)^{-1/2}r_n(1 + o(1)) - r_n^2/2 \to 0 \end{aligned}$$

because of (1.38), and if we substitute (1.38) into the left side of this relation we get

$$(\log n)^{1/2}s_n \to 0$$

as required. Hence we conclude

$$b_n = (2\log n)^{1/2} - 1/2(\log\log n + \log 4\pi)/(2\log n)^{1/2}.$$

Example 3 (Gamma). Suppose $F$ is the gamma distribution with density

$$F'(t) = t^\alpha e^{-t}/\Gamma(\alpha + 1), \qquad t > 0, \alpha > 0.$$

Then

$$F''(t) = (-t^\alpha e^{-t} + \alpha t^{\alpha-1} e^{-t})/\Gamma(\alpha + 1)$$
$$= -F'(t)(1 + \alpha t^{-1}) \sim -F'(t)$$

and furthermore by L'Hospital's rule

$$\lim_{t\to\infty} \frac{1 - F(t)}{F'(t)} = \lim_{t\to\infty} \frac{-F'(t)}{F''(t)} = 1$$

and

$$\lim_{t\to\infty} F''(t)(1 - F(x))/(F'(t))^2 = -1$$

so that $F$ is a Von Mises function, $F \in D(\Lambda)$, and the auxiliary function $f$ satisfies

$$f(t) = \frac{1 - F(t)}{F'(t)} \to 1$$

as $t \to \infty$. Therefore $a_n$ may be taken equal to 1. To find $b_n$ we solve $F'(b_n) = 1/n$ instead of $1 - F(b_n) = 1/n$ since $1 - F(t) \sim F'(t)$. So we have $F'(b_n) = 1/n$ equivalent to

$$b_n^\alpha e^{-b_n}/\Gamma(\alpha + 1) = 1/n;$$

i.e.,

$$b_n - \alpha \log b_n + \log\Gamma(\alpha + 1) = \log n. \tag{1.39}$$

Since $b_n \to \infty$ we see by dividing through that

$$b_n \sim \log n$$

and consequently

$$b_n = \log n + r_n \tag{1.40}$$

where $r_n = o(\log n)$. Substituting (1.4) into (1.39) we obtain

$$\log n + r_n - \alpha \log(\log n + r_n) + \log\Gamma(\alpha + 1) = \log n,$$

i.e.,

$$r_n + \log\Gamma(\alpha + 1) = \alpha \operatorname{loglog} n + \alpha \log(1 + (r_n/\log n)),$$

i.e.,

$$r_n = -\log\Gamma(\alpha + 1) + \alpha \operatorname{loglog} n + o(1).$$

Therefore

$$b_n - (\log n - \log\Gamma(\alpha + 1) + \alpha \operatorname{loglog} n)/a_n = o(1)/a_n \to 0$$

and so

$$b_n = \log n + \alpha \operatorname{loglog} n - \log\Gamma(\alpha + 1)$$

is an acceptable choice.

Note when $\alpha = 0$, $b_n = \log n$, which is well known for the exponential density.

EXERCISES

1.5.1. For two distributions $F$ and $G$ with the same endpoint $x_0$, the ratio $(1 - F(x))/(1 - G(x))$ need not have a limit as $x \to x_0$. Construct examples.

1.5.2. $\{X_n, n \geq 1\}$ is a sequence of iid random variables with common distribution $F$ and $\{\tau_n, n \geq 1\}$ is iid, positive integer valued, and $\{\tau_n\}$ and $\{X_n\}$ are independent. Suppose $E\tau_1 < \infty$ and set

$$S_n = \sum_1^n \tau_j, \qquad \chi_n = \bigvee_{j=S_{n-1}+1}^{S_n} \chi_j, \qquad \overline{M}_n = \bigvee_{j=1}^{n} \chi_j.$$

What is the distribution of $\chi_i$? Show $\{\overline{M}_n\}$ has a limit distribution iff $\{M_n\}$ has one. In fact, there exist norming constants $a_n > 0$, $b_n$, $n \geq 1$ such that

$$P[M_n \leq a_n x + b_n] \to H(x),$$

nondegenerate, iff

$$P[\overline{M}_n \leq a_n x + b_n] \to (H(x))^{E\tau_1} \qquad \text{(Resnick, 1971).}$$

1.5.3. $F$ and $G$ are distributions and $H$ is an extreme value distribution. Suppose

$$F^n(a_n x + b_n) \to H(x)$$

for $a_n > 0$, $b_n \in \mathbb{R}$. Then

$$(FG)^n(a_n x + b_n) := F^n(a_n x + b_n)G^n(a_n x + b_n) \to H(Ax + B)$$

iff

(i) $H = \Phi_\alpha$: $B = 0, 0 < A \leq 1$, and

$$\lim_{x\to\infty} (1 - F(x))/(1 - G(x)) = (A^{-\alpha} - 1)^{-1};$$

(ii) $H = \Psi_\alpha$: $B = 0$, $\infty > A \geq 1$, and

$$\lim_{x\to x_0} (1 - F(x))/(1 - G(x)) = (A^{\alpha} - 1)^{-1};$$

(iii) $H = \Lambda$: $A = 1$, $B < 0$, and

$$\lim_{x\to x_0} (1 - F(x))/(1 - G(x)) = (e^{-B} - 1)^{-1} \qquad \text{(Resnick, 1971).}$$

1.5.4. There is a weaker form of equivalence than tail equivalence. Say $F_1, F_2, \in D(\Lambda)$ are $a$-equivalent if there exists $a_n > 0$, $b_n^{(i)} \in \mathbb{R}$, $i = 1, 2$ such that for $i = 1, 2$

$$F_i^n(a_n x + b_n^{(i)}) \to \Lambda(x);$$

i.e., the same $a_n$ can be used for both distributions but not necessarily the same $b_n$.

(a) $F_1$ and $F_2$ are $a$-equivalent iff

$$\left(\frac{1}{1 - F_1}\right)^{\leftarrow} \overset{\Pi}{\sim} \left(\frac{1}{1 - F_2}\right)^{\leftarrow}$$

(cf. prior to Proposition 0.16).

(b) Let $U_i = 1/(1 - F_i)$ and suppose $F_i$ are Von Mises functions for $i = 1, 2$. Then $F_1$ and $F_2$ are $a$-equivalent iff

$$U_1' \circ U_1^{\leftarrow} \sim U_2' \circ U_2^{\leftarrow}.$$

(c) $F_1$ and $F_2$ are $a$-equivalent iff there exists a positive function $b(x)$ with $\lim_{x \to x_0} b(x) = 1$ and constants $C_1, C_2 > 0$ such that

$$1 - F_1(x) \sim C_1(1 - F_2(P(x)))$$

as $x \to x_0$ where

$$P(x) = C_2 + \int_0^x b(t)dt.$$

Hint: It suffices to consider Von Mises functions. Use part (b).

(d) Check that the standard normal distribution is $a$-equivalent to $F(x) = 1 - \exp\{-x^2\}$. What is a suitable choice of $P$? Use this to compute $a_n$ for the normal distribution (de Haan, 1974a).

1.5.5. The normal distribution has the property

$$P[b_n(M_n - b_n) \le x] \to \Lambda(x). \tag{1.41}$$

Find a characterization of the distributions $F$ with property (1.41). Cf. Exercise 1.1.7.

1.5.6. Suppose $\{X_n, n \ge 1\}$ are iid random variables with common distribution $F(x)$. Set $M_n = \bigvee_{i=1}^n X_i$, $m_n = \bigwedge_{i=1}^n X_i$, and assume there exist $a_n > 0$, $\alpha_n > 0$, $b_n \in \mathbb{R}$, $\beta_n \in \mathbb{R}$ such that

$$P[M_n \le a_n x + b_n] \to G_1(x)$$

and

$$P[-m_n \le \alpha_n x + \beta_n] \to G_2(x)$$

where $G_1$ and $G_2$ are nondegenerate.

(a) Show joint convergence

$$P[M_n \le a_n x + b_n, -m_n \le \alpha_n y + \beta_n] \to G_1(x)G_2(y).$$

(b) Particularize to the case where the common distribution $\{X_n\}$ is the standard normal $N(x)$. Show the range $M_n - m_n$ has a limit distribution which is the second convolution power of $\Lambda(x)$:

$$P[M_n - m_n \le a_n x] \to \Lambda * \Lambda(x).$$

(c) If in (a), $G_1 = G_2 = \Phi_\alpha$, show $M_n - m_n$ has a limit distribution if

$$\lim_{x \to \infty} F(-x)/(1 - F(x)) = \rho \in [0, \infty]$$

exists.

(d) If in (a), $G_1 = G_2 = \Lambda$, show $M_n - m_n$ has a limit distribution if $F(x)$ and $1 - F(-x)$ are $a$-equivalent (cf. 1.5.4) (de Haan, 1974b).

1.5.7. Compute norming constants for the $t$ and $F$ densities. In which domain of attraction are these distributions?

CHAPTER 2

# Quality of Convergence

The previous chapters contain information characterizing possible limit distributions for extremes and also discuss domain of attraction criteria. So if the familiar relation

$$F^n(a_n x + b_n) \to G(x)$$

holds, we know the class of possible G's, what conditions $F$ must satisfy, and how to characterize $a_n$ and $b_n$. The present chapter amplifies our knowledge by describing in various ways how close $F^n(a_n x + b_n)$ is to $G(x)$ and how if approaches $G(x)$ asymptotically. The topics discussed include moment convergence, local limit theory and density convergence, large deviations, and uniform rates of convergence.

## 2.1. Moment Convergence (Von Mises, 1936; Pickands, 1968)

Suppose $X_n$, $n \geq 1$ are iid with common distribution $F$ and $F \in D(G)$ for an extreme value distribution $G$. Then there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that $M_n = \bigvee_{i=1}^n X_i$ satisfies

$$P[a_n^{-1}(M_n - b_n) \leq x] = F^n(a_n x + b_n) \to G(x). \tag{2.1}$$

We ask for which values of $k > 0$ it is true that

$$\lim_{n\to\infty} E(a_n^{-1}(M_n - b_n))^k = \int_{-\infty}^{\infty} x^k G(dx). \tag{2.2}$$

It is well known that convergence of a sequence of random variables does not imply that moments converge (cf. Chung, 1974, pages 94–98). The canonical example is to take $(0, 1)$ as the probability space with Lebesgue measure as the probability. Set $X_n(\omega) = 2^n 1_{(0,n^{-1})}(\omega)$ so that $X_n \to 0$ a.s. but $EX_n = n^{-1}2^n \to \infty$. A condition which controls tail probabilities and thus prevents improbable large values from disturbing moment convergence is needed.

Note that the tail conditions which comprise the domain of attraction criteria are only a control on the right tail. For instance, if $F \in D(\Phi_\alpha)$ then $1 - F(x) \sim x^{-\alpha}L(x)$ as $x \to \infty$. This implies

$$\int_0^\infty x^k F(dx) < \infty \qquad \text{if } k < \alpha$$

(Exercise 1.2.2), but no control is provided over the left tail and it is possible for $\int_{-\infty}^{0} |x|^k F(dx) = \infty$ for any $k > 0$. Similarly $F \in D(\Lambda)$ implies when $x_0 = \infty$ that

$$\int_0^\infty x^k F(dx) < \infty \qquad \text{for all } k > 0$$

(Exercise 1.1.1) but implies nothing about behavior of the left tail.

Thus in investigating (2.1) it is necessary to impose some condition on the left tail of $F$.

**Proposition 2.1.** *For an extreme value distribution $G$, suppose $F \in D(G)$.*
(i) *If $G = \Phi_\alpha$, set $a_n = (1/(1-F))^{\leftarrow}(n), b_n = 0$. If for some integer $0 < k < \alpha$*

$$\int_{-\infty}^{0} |z|^k F(dx) < \infty \tag{2.3}$$

*then*

$$\lim_{n\to\infty} E(M_n/a_n)^k = \int_{-\infty}^{\infty} x^k \Phi_\alpha(dx) = \Gamma(1 - \alpha^{-1}k).$$

(ii) *If $G = \Psi_\alpha$ and $F$ has right end $x_0$ set*

$$a_n = x_0 - (1/(1-F))^{\leftarrow}(n), \qquad b_n = x_0.$$

*If for some integer $k > 0$*

$$\int_{-\infty}^{x_0} |x|^k F(dx) < \infty \tag{2.4}$$

*then*

$$\lim_{n\to\infty} E((M_n - x_0)/a_n)^k = \int_{-\infty}^{0} x^k \Psi_\alpha(dx) = (-1)^k \Gamma(1 + \alpha^{-1}k).$$

(iii) *If $G = \Lambda$ and $F$ has right end $x_0$ with representation (1.5) set $b_n = (1/(1-F))^{\rightarrow}(n), a_n = f(b_n)$. If for some integer $k > 0$*

$$\int_{-\infty}^{0} |x|^k F(dx) < \infty \tag{2.5}$$

*then*

$$\lim_{n\to\infty} E((M_n - b_n)/a_n)^k = \int_{-\infty}^{\infty} x^k \Lambda(dx) = (-1)^k \Gamma^{(k)}(1)$$

*where $\Gamma^{(k)}(1)$ is the $k$th derivative of the gamma function at $x = 1$.*

*Remarks.* (i) Conditions (2.3), (2.4), and (2.5) can be weakened slightly. See Exercise 2.1.1.
(ii) For any norming constants $a_n$, $b_n$ satisfying (2.1) (not just the ones specified in the statement of the proposition), we also have (2.2) satisfied provided the appropriate condition (2.3), (2.4), or (2.5) holds. See Exercise 2.1.2.

We only prove (i) and (iii) in Proposition 2.1. Part (iii) requires that our tool box be equipped with the following inequalities.

**Lemma 2.2.** *Suppose $F \in D(\Lambda)$ with representation* (1.5) *and that $a_n$ and $b_n$ are as specified in part* (iii) *of Proposition* 2.1.
(a) *Given $\varepsilon > 0$, we have for $s > 0$ and all sufficiently large $n$*

$$f(b_n)/f(a_n s + b_n) \geq (1 + \varepsilon s)^{-1} \tag{2.6}$$

*and consequently if $y > 0$ and $n$ is large*

$$1 - F^n(a_n y + b_n)) \leq (1 + \varepsilon)^3 (1 + \varepsilon y)^{-\varepsilon^{-1}}. \tag{2.7}$$

(b) *Recall the meaning of $z_0$ in the representation* (1.5). *Given $\varepsilon$, pick $z_1 \in (z_0, x_0)$ such that $|f'(x)| \leq \varepsilon$ if $x > z_1$. Then for large $n$ and $u \in (a_n^{-1}(z_1 - b_n), 0)$ we have*

$$f(b_n)/f(a_n u + b_n) \geq (1 + \varepsilon |u|)^{-1} \tag{2.8}$$

*and consequently for large $n$ and $s \in (a_n^{-1}(z_1 - b_n), 0)$*

$$F^n(a_n s + b_n) \leq e^{-(1-\varepsilon)^2 (1 + \varepsilon|s|)^{\varepsilon^{-1}}}. \tag{2.9}$$

Proof of Lemma 2.2. (a) For $n$ such that $|f'(t)| \leq \varepsilon$ if $t \geq b_n$ we have for $s > 0$

$$(f(a_n s + b_n)/f(b_n)) - 1 \quad = \int_{b_n}^{a_n s + b_n} (f'(u)/f(b_n)) du$$

and recalling $a_n = f(b_n)$ this is

$$\int_0^s f'(a_n u + b_n) du \leq \varepsilon s.$$

Consequently $f(b_n)/f(a_n s + b_n) \geq (1 + \varepsilon s)^{-1}$ as asserted.
To check (2.7) note that

$$1 - F(b_n) \sim n^{-1}$$

so that for large $n$ and $y > 0$

$$n(1 - F(a_n y + b_n) \leq (1 + \varepsilon)(1 - F(a_n y + b_n))/(1 - F(b_n))$$

and from (1.5) this is

$$(1 + \varepsilon) c(a_n y + b_n) c^{-1}(b_n) e^{-\int_{b_n}^{a_n y + b_n} (1/f(s)) ds}.$$

Since $c(x) \to c > 0$ as $x \uparrow x_0$ we have the preceding

$$\leq (1 + \varepsilon)^2 e^{-\int_0^y (f(b_n)/f(a_n s + b_n)) ds}$$

(recall $a_n = f(b_n)$), and applying (2.6) this is bounded by

$$\le (1 + \varepsilon)^2 e^{-\int_0^y (1+\varepsilon s)^{-1} ds}$$
$$= (1 + \varepsilon)^2 (1 + \varepsilon y)^{-\varepsilon^{-1}}.$$

Therefore

$$1 - F^n(a_n y + b_n) = 1 - \exp\{n(-\log F(a_n y + b_n))\}$$
$$< n(-\log F(a_n y + b_n))$$

and since $-\log F(a_n y + b_n) \sim 1 - F(a_n y + b_n)$ uniformly for $y > 0$ (use $\lim_{z \to 1} (-\log z)/(1 - z) = 1$) the preceding is bounded for large $n$ by

$$\le (1 + \varepsilon) n (1 - F(a_n y + b_n))$$

and the result follows.

(b) As earlier for $n$ large and $u \in (a_n^{-1}(z_1 - b_n), 0)$

$$1 - f(a_n u + b_n)/f(b_n) = \int_{a_n u + b_n}^{b_n} (f'(w)/f(b_n)) dw$$
$$\int_u^0 f'(a_n w + b_n) dw$$

and since $a_n w + b_n > a_n u + b_n > z_1$ the preceding integral is $\ge -\varepsilon |u|$, and thus we have shown

$$1 - f(a_n u + b_n)/f(b_n) \ge -\varepsilon |u|;$$

i.e.,

$$1 + \varepsilon |u| \ge f(a_n u + b_n)/f(b_n)$$

which is equivalent to (2.8). To check (2.9) write for large $n$

$$\begin{aligned} F^n(a_n s + b_n) &= (1 - (1 - F(a_n s + b_n)))^n \\ &\le \exp\{-n(1 - F(a_n s + b_n))\} \\ &\le \exp\{-(1 - \varepsilon)(1 - F(a_n s + b_n))/(1 - F(b_n))\} \\ &= \exp\left\{-(1 - \varepsilon) c(a_n s + b_n) c^{-1}(b_n) \exp\left\{-\int_{b_n}^{a_n s + b_n} (1/f(u)) du\right\}\right\} \end{aligned}$$

and supposing $z_1$ has been chosen so that $c(z_1)/c(b_n) \ge 1 - \varepsilon$ the preceding is bounded by

$$\le \exp\left\{-(1 - \varepsilon)^2 e^{\int_s^0 (f(b_n)/f(a_n u + b_n)) du}\right\}$$

and applying (2.8) we get an upper bound of

$$\le \exp\left\{-(1 - \varepsilon)^2 e^{\int_s^0 (1 + \varepsilon |u|)^{-1} du}\right\} = \exp\left\{-(1 - \varepsilon)^2 (1 + \varepsilon |s|)^{\varepsilon^{-1}}\right\}. \qquad \square$$

PROOF OF PROPOSITION 2.1. If (2.1) holds than it follows from the standard weak convergence theory (Chung, 1974, page 87; Loève, 1963, page 180; cf. the Helly–Bray lemma) that for any $L > 0$

$$\lim_{n\to\infty} E(a_n^{-1}(M_n - b_n))^k 1_{[|a_n^{-1}(M_n-b_n)|\le L]} = \int_{-L}^{L} x^k G(dx).$$

It is enough to show

$$\lim_{L\to\infty} \limsup_{n\to\infty} E|a_n^{-1}(M_n - b_n)|^k 1_{[|a_n^{-1}(M_n-b_n)|> L]} = 0 \tag{2.10}$$

because of the following: Write

$$\begin{aligned}
&\left|E(a_n^{-1}(M_n - b_n))^k - \int_{-\infty}^{\infty} x^k G(dx)\right| \\
&\le |E(a_n^{-1}(M_n - b_n))^k - E(a_n^{-1}(M_n - b_n))^k 1_{[|a_n^{-1}(M_n-b_n)|\le L]}| \\
&+ \left|E(a_n^{-1}(M_n - b_n)^k 1_{[|a_n^{-1}(M_n-b_n)|\le L]} - \int_{-L}^{L} x^k G(dx)\right| \\
&+ \left|\int_{-L}^{L} x^k G(dx) - \int_{-\infty}^{\infty} x^k G(dx)\right|.
\end{aligned} \tag{2.11}$$

If (2.10) holds the right side of (2.11) has $\lim_{L\to\infty} \limsup_{n\to\infty} = 0$ and since the left side of (2.11) does not depend on $L$, the desired result follows. So now we concentrate on showing (2.10).

Set $Y = |a_n^{-1}(M_n - b_n)|$ and we use Fubini's theorem to justify an integration by parts:

$$\begin{aligned}
EY^k 1_{[Y>L]} &= E\int_0^Y ks^{k-1} ds\, 1_{[Y>L]} \\
&= E\int_0^L ks^{k-1} ds\, 1_{[Y>L]} + E\int_L^{\infty} ks^{k-1} 1_{[Y>L,Y>s]} ds \\
&= L^k P[|(a_n^{-1}(M_n - b_n)| > L] + \int_L^{\infty} ks^{k-1} P[Y > s] ds \\
&= A + B.
\end{aligned}$$

Because of (2.1) we have

$$\lim_{L\to\infty} \limsup_{n\to\infty} A = \lim_{L\to\infty} L^k(1 - G(L) + G(-L)).$$

When $G = \Phi_\alpha$ and $k < \alpha$

$$\lim_{L\to\infty} \limsup_{n\to\infty} A = \lim_{L\to\infty} L^k(1 - e^{-L^{-\alpha}}) = \lim_{L\to\infty} L^k L^{-\alpha} = 0$$

and when $G = \Lambda$

$$\lim_{L\to\infty} \limsup_{n\to\infty} A = \lim_{L\to\infty} L^k\{1 - \exp\{-e^{-L}\} + \exp\{-e^L\}\}$$
$$= \lim_{L\to\infty} L^k\{e^{-L} + \exp\{-e^L\}\} = 0$$

and so we must concentrate now on $B$.

Write

$$B = \int_L^\infty ks^{k-1}(1 - F^n(a_n s + b_n))ds + \int_L^\infty ks^{k-1}F^n(-a_n s + b_n)ds$$
$$= B_1 + B_2.$$

Consider $B_1$. If $G = \Phi_\alpha$ then

$$1 - F^n(a_n s) \le 1 - \exp\{n(-\log F(a_n s))\}$$
$$\le n(-\log F(a_n s))$$

and for large $n$ we have the bound

$$\le (1 + \varepsilon)n(1 - F(a_n s)) \le (1 + \varepsilon)^2(1 - F(a_n s))/(1 - F(a_n)).$$

Now apply Proposition 0.8(ii), which tells us that given $\varepsilon > 0$, if $n$ is large and $L > 1$ then

$$(1 - F(a_n s))/(1 - F(a_n)) \le (1 + \varepsilon)s^{-\alpha+\varepsilon}$$

so that in this case

$$\lim_{L\to\infty} \limsup_{n\to\infty} B_1 = \lim_{L\to\infty} (1 + \varepsilon)^3 \int_L^\infty ks^{k-1}s^{-\alpha+\varepsilon}ds = 0$$

provided $k - 1 - \alpha + \varepsilon < -1$, i.e., provided $k < \alpha - \varepsilon$. Since $k$ is assumed less than $\alpha$, an appropriate choice of $\varepsilon$ is available.

Now consider $B_1$ in the case that $G = \Lambda$. In this case apply Lemma 2.2(a) and so

$$B_1 \le (1 + \varepsilon)^3 \int_L^\infty ks^{k-1}(1 + \varepsilon s)^{-\varepsilon^{-1}}ds.$$

The integrand is asymptotically equal to

$$(1 + \varepsilon)^3 k\varepsilon^{-\varepsilon^{-1}}s^{k-1-\varepsilon^{-1}}$$

and choosing $\varepsilon$ small enough, so that

$$k - 1 - \varepsilon^{-1} < -1, \quad \text{or } \varepsilon < k^{-1},$$

we obtain

$$\lim_{L\to\infty} \limsup_{n\to\infty} B_1 \le (\text{const}) \lim_{L\to\infty} \int_L^\infty s^{k-1-\varepsilon^{-1}}ds = 0.$$

This takes care of $B_1$ and now we deal with $B_2$. In case $G = \Phi_\alpha$ we have

$$B_2 = \int_L^\infty ks^{k-1}F^n(-a_n s)ds$$
$$= \int_{-\infty}^{-a_n L} k|s|^{k-1}F^n(s)ds/a_n^k$$
$$\leq F^n(-a_n L)\int_{-\infty}^{0} k|s|^{k-1}F(s)ds/a_n^k$$

and because of (2.3) and the fact that $a_n \to \infty$ we have for all $L > 0$

$$\limsup_{n\to\infty} B_2 = 0.$$

Now we consider $B_2$ for the case $G = \Lambda$. Write $B_2$ as

$$\int_{-\infty}^{-L} k|s|^{k-1}F^n(a_n s + b_n)ds = \int_{-\infty}^{(z_1 - b_n)/a_n} + \int_{(z_1 - b_n)/a_n}^{-L} = B_{21} + B_{22}$$

where $z_2$ is chosen as in Lemma 2.2(b). Note that $(z_1 - b_n)/a_n \to -\infty$ so eventually $(z_1 - b_n)/a_n < -L$. The reason $(z_1 - b_n)/a_n \to -\infty$ is as follows: Since $F \in D(\Lambda)$ we have (Propositon 0.10) that $V = (1/(1-F))^{\leftarrow} \in \Pi$ and hence for any $\varepsilon > 0$

$$\limsup_{n\to\infty} (z_1 - b_n)/a_n \leq \limsup_{n\to\infty} (V(n\varepsilon) - V(n))/a_n = \log \varepsilon$$

by the $\Pi$-variation property. Since $\varepsilon > 0$ is arbitrary we must have

$$\limsup_{n\to\infty} (z_1 - b_n)/a_n = -\infty.$$

For $B_{21}$ we have be setting $y = a_n s + b_n$

$$B_{21} = \int_{-\infty}^{z_1} k|y - b_n|^{k-1}F^n(y)dy/a_n^k$$
$$\leq F^{n-1}(z_1)a_n^{-k}k(\text{const})\int_{-\infty}^{z_1} (|y|^{k-1} + b_n^{k-1})F(y)dy.$$

Both $a_n$ and $b_n$ are slowly varying functions of $n$, and hence since $F^{n-1}(z_1) \to 0$ geometrically fast we get as $n \to \infty$

$$F^{n-1}(z_1)a_n^{-k} \to 0, \qquad F^{n-1}(z_1)a_n^{-k}b_n^{k-1} \to 0.$$

Finally observe

$$\int_{-\infty}^{z_1} |y|^{k-1}F(y)dy < \infty$$

and

$$\int_{-\infty}^{z_1} b_n^{k-1}F(y)dy \leq (\text{const})\int_{-\infty}^{z_1} b_n^{k-1}|y|^{k-1}F(y)dy < \infty$$

by assumption (2.5). This shows $\limsup_{n\to\infty} B_{21} = 0$.

Finally we deal with $B_{22}$: We have

$$B_{22} = \int_{(z_1 - b_n)/a_n}^{-L} k|s|^{k-1} F^n(a_n s + b_n) ds$$

and applying (2.9) $B_{22}$ is bounded by

$$\int_{(z_1 - b_n)/a_n}^{-L} k|s|^{k-1} e^{-(1-\varepsilon)^2(1+\varepsilon|s|)^{\varepsilon^{-1}}} ds$$

and thus

$$\lim_{L\to\infty} \limsup_{n\to\infty} B_{22} \leq \lim_{L\to\infty} (\text{const}) \int_{\infty}^{-L} |s|^{k-1} e^{-(1-\varepsilon)^2(\varepsilon|s|)^{\varepsilon^{-1}}} ds = 0$$

since $|s|^{k-1} e^{-(1-\varepsilon)^2(\varepsilon|s|)^{\varepsilon^{-1}}}$ is integrable on $(-\infty, 0)$. □

The following interesting corollary was pointed out by L. de Haan.

**Corollary 2.3.** *Suppose $F \in D(G)$ so that* (2.1) *holds. If $G = \Phi_\alpha$ suppose $\alpha > 2$. Suppose* (2.3), (2.4), *or* (2.5) *holds in the form*

$$\int_{-\infty}^{x_0} x^2 dF(x) < \infty.$$

*(This condition can be weakened as in Exercise* 2.1.1.) *Then normalization of the maximum using the mean and standard deviation is possible:*

$$P[(M_n - E(M_n))/(\mathrm{Var}(M_n))^{1/2} \leq x]$$

$$\to \begin{cases} \Phi_\alpha((\Gamma(1-2\alpha^{-1}) - \Gamma^2(1-\alpha^{-1}))^{1/2}x + \Gamma(1-\alpha^{-1})) & \text{if } G = \Phi_\alpha \\ \Psi_\alpha((\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1}))^{1/2}x - \Gamma(1+\alpha^{-1})) & \text{if } G = \Psi_\alpha \\ \Lambda((\Gamma^{(2)}(1) - (\Gamma^{(1)}(1))^2)^{1/2}x - \Gamma^{(1)}(1)) & \text{if } G = \Lambda. \end{cases}$$

Proof. If $G = \Lambda$ then from Proposition 2.1

$$(EM_n - b_n)/a_n \to -\Gamma^{(1)}(1) \tag{2.12}$$

and

$$E(a_n^{-1}(M_n - b_n))^2 \to \Gamma^{(2)}(1)$$

so that

$$\begin{aligned} &\Gamma^{(2)}(1) - (\Gamma^{(1)}(1))^2 \\ &= \lim_{n\to\infty} E(a_n^{-1}(M_n - b_n))^2 - (E(a_n^{-1}(M_n - b_n)))^2 \\ &= \lim_{n\to\infty} a_n^{-2}\{EM_n^2 - 2b_n EM_n + b_n^2 - E^2 M_n + 2b_n EM_n - b_n^2\} \\ &= \lim_{n\to\infty} a_n^{-2}(EM_n^2 - E^2 M_n) \\ &= \lim_{n\to\infty} a_n^{-2} \operatorname{Var} M_n. \end{aligned} \tag{2.13}$$

The result now follows from (2.12) and (2.13) and the convergence to types theorem (cf. Proposition 0.2). The arguments for the other cases are the same. □

EXERCISES

2.1.1. Condition (2.3) can be weakened to the following: There exists some integer $n_0$ such that

$$\int_{-\infty}^{0} |x|^k F^{n_0}(dx) < \infty.$$

Check this and the analogous weakenings of (2.4) and (2.5). Give an example of $F$ such that

$$\int_{-\infty}^{0} |x|^k F(dx) = \infty$$

but for $n_0 \geq 2$

$$\int_{-\infty}^{0} |x|^k F^{n_0}(dx) < \infty.$$

2.1.2. If $a_n' \sim a_n, b_n' - b_n = o(a_n)$ and (2.3), (2.4), or (2.5) holds, then

$$\lim_{n\to\infty} E((M_n - b_n')/a_n')^k = \int_{-\infty}^{\infty} x^k G(dx).$$

2.1.3. Moment convergence and relative stability: Suppose $\{M_n\}$ is relatively stable (cf. Exercises 1.19, 0.4.1.2) with $x_0 = \infty$; i.e., suppose $x_0 = \infty$ and there exists $b_n$ such that

$$M_n/b_n \xrightarrow{P} 1.$$

Show if

$$\int_{-\infty}^{0} |x|^k F(dx) < \infty$$

then

$$E(M_n/b_n)^k \to 1$$

as $n \to \infty$ so that

$$M_n/EM_n \xrightarrow{P} 1$$

(Pickands, 1968).

2.1.4. The following discusses when centering and scaling by means and standard deviations, respectively, is possible: Suppose $\{X_n\}$ is a sequence of random variables such that for some $a_n > 0$, $b_n \in \mathbb{R}$

$$P[X_n \leq a_n x + b_n\} =: F_n(a_n x + b_n) \to G(x),$$

nondegenerate. Suppose

$$EX_n, \qquad \operatorname{Var} X_n$$

exist and are finite. Prove there exist $a > 0$, $b \in \mathbb{R}$ such that

$$P[(X_n - EX_n)/(\operatorname{Var} X_n)^{1/2} \le x] \to G(ax + b)$$

iff

$$\lim_{n\to\infty} \operatorname{Var} X_n/(F_n^{\leftarrow}(3/4) - F_n^{\leftarrow}(1/2))^2 = c > 0. \tag{2.14}$$

Hints:

(a) Suppose (2.14). Set

$$Y_n = (X_n - F_n^{\leftarrow}(1/2))/(F_n^{\leftarrow}(3/4) - F_n^{\leftarrow}(1/2)).$$

Why does $Y_n$ converge in distribution? (Cf. the remark following Proposition 0.2.) Then

$$EY_n = (EX_n - F_n^{\leftarrow}(1/2))/(F_n^{\leftarrow}(3/4) - F_n^{\leftarrow}(1/2))$$

$$\operatorname{Var} Y_n = \operatorname{Var} X_n/(F_n^{\leftarrow}(3/4) - F_n^{\leftarrow}(1/2))^2$$

and from 2.14, $\operatorname{Var} Y_n$ is bounded.

(b) Show $EY_n$ is bounded (cf. median inequalities on page 244 of Loève, 1963) and hence $EY_n^2$ is bounded.

(c) Since $\sup_{n\ge 1} EY_n^2 < \infty$, $\{Y_n\}$ is uniformly integrable, and since $\{Y_n\}$ converges in distribution, $\lim_{n\to\infty} EY_n$ exists finite (de Haan, 1970, page 59).

## 2.2. Density Convergence

Suppose (2.1) holds so that

$$P[a_n^{-1}(M_n - b_n) \le x] = F^n(a_n x + b_n) \to G(x) \tag{2.1}$$

for an extreme value distribution $G$. We suppose $F$ has left endpoint $x_l$ and as usual denote the right endpoint by $x_0$ so that

$$x_l = \inf\{x: F(x) > 0\}$$

$$x_0 = \sup\{x: F(x) < 1\}$$

and $-\infty \le x_l \le x_0 \le \infty$. We suppose $F$ is absolutely continuous with density $F'$ and ask when (2.1) implies density convergence

$$g_n(x) := na_n F^{n-1}(a_n x + b_n)F'(a_n x + b_n) \to G'(x). \tag{2.15}$$

We will show that local uniform convergence of $g_n$ to $G'$ is equivalent to the appropriate Von Mises condition. (*Local uniform convergence is* convergence on compact subsets.)

Convergence in various modes of the density $g_n$ has been considered by several authors; see Pickands (1967), Anderson (1971), de Haan and Resnick (1982). These efforts culminate in the nice paper by Sweeting (1985).

We begin with a simple lemma showing norming constants $a_n$ and $b_n$ can be chosen at our convenience without affecting results.

**Lemma 2.4.** *Suppose for given $a_n$, $b_n$ we have*

$$g_n \to G'$$

*locally uniformly. If for $\tilde{a}_n > 0$, $\tilde{b}_n \in \mathbb{R}$*

$$\tilde{a}_n \sim a_n, \qquad (\tilde{b}_n - b_n)/a_n \to 0$$

*then*

$$\tilde{g}_n(x) := n\tilde{a}_n F^n(\tilde{a}_n x + \tilde{b}_n)F'(\tilde{a}_n x + \tilde{b}_n) \to G'(x)$$

*locally uniformly. The assertion remains true if "locally uniformly" is replaced by uniformly in a neighborhood of $\pm\infty$.*

PROOF. It suffices to show that if $x_n \to x \in \mathbb{R}$ then $\tilde{g}_n(x_n) \to G'(x)$ (cf. Section 0.1). However, since

$$\tilde{g}_n(x_n) = (\tilde{a}_n a_n^{-1}) g_n(a_n(\tilde{a}_n a_n^{-1} x_n + (\tilde{b}_n - b_n)/a_n) + b_n)$$

and since $\tilde{a}_n a_n^{-1} x_n + (\tilde{b}_n - b_n)/a_n \to x$, the result follows from the assumed local uniform convergence of $g_n$ to $G'$. □

**Proposition 2.5.** *Suppose $F$ is absolutely continuous with density $F'$ and right end $x_0$. If $F \in D(G)$ and*
(a) $G = \Phi_\alpha$ *then* (2.15) *is true locally uniformly on* $(0, \infty)$ iff (1.19) *holds.*
(b) $G = \Psi_\alpha$ *then* (2.15) *is true locally uniformly on* $(-\infty, 0)$ iff (1.20) *holds.*
(c) $G = \Lambda$ *then* (2.15) *is true locally uniformly on* $\mathbb{R}$ iff (1.21) *holds.*

PROOF. We first prove the result in the case $G = \Lambda$. As in Lemma 2.4 we prove continuous convergence. Assume (1.21) holds and set

$$b_n = (1/(1 - F))^{\leftarrow}(n), \qquad f(t) = \int_t^\infty (1 - F(s))ds/(1 - F(t)), \qquad a_n = f(b_n).$$

Then

$$g_n(x_n) = na_n F^{n-1}(a_n x_n + b_n))F'(a_n x_n + b_n)$$

and since

$$F^{n-1}(a_n x + b_n) = (F^n(a_n x + b_n))^{(n-1)/n} \to \Lambda(x)$$

uniformly on $\mathbb{R}$ (see Section 0.1) we have

$$g_n(x_n) \sim na_n \Lambda(x) F'(a_n x_n + b_n)$$

and so it suffices to check

$$na_n F'(a_n x_n + b_n) \to e^{-x}.$$

Since $a_n x_n + b_n \to x_0$ we have upon setting $f_0(t) = (1 - F(t))/F'(t)$

$$\begin{aligned} na_n F'(a_n x_n + b_n) &\sim n(1 - F(a_n x_n + b_n))f(b_n)/f_0(a_n x_n + b_n) \\ &\sim n(1 - F(a_n x_n + b_n))f(b_n)/f(a_n x_n + b_n) \end{aligned}$$

because (1.21) means $f(t) \sim f_0(t)$. However

$$n(1 - F(a_n x + b_n)) \to e^{-x}$$

locally uniformly since monotone functions are converging to a continuous limit and

$$f(b_n + xa_n)/f(b_n) \to 1$$

locally uniformly by Lemma 1.3. Thus the assertion is checked.

Conversely, if (2.15) holds locally uniformly then since density convergence implies weak convergence it follows that $F \in D(\Lambda)$. Set $V = (1/(1 - F))^{\leftarrow}$ and we have $V \in \Pi$ (Proposition 0.10). We may set $b_n = V(n)$, $a(t) = f(V(t))$ where $f(t) = \int_t^{x_0} (1 - F(s))ds/(1 - F(t))$ and $a(\cdot)$ is the auxiliary function of $V$ and is hence slowly varying. Then (2.15) holds with this choice of $b_n$, $a(n)$, and the convergence is still locally uniform (Lemma 2.4). Since $F^{n-1}(a(n)x + b_n) \to \Lambda(x)$ uniformly, (2.15) implies

$$na(n)F'(a(n)x + b_n) \to e^{-x}$$

locally uniformly and thus as $t \to \infty$

$$[t]a([t])F'(a([t])x + b_{[t]}) \to e^{-x} \tag{2.16}$$

locally uniformly. Set

$$x_t = (V(te^x) - V([t]))/a([t]).$$

Since the $\Pi$-variation property holds locally uniformly (monotone functions are converging to a continuous limit) we get as $t \to \infty$ for $x > 0$

$$x_t = (V([t](e^x t/[t])) - b_{[t]})/a_{[t]} \to \log e^x = x$$

and replacing $x$ in (2.16) by $x_t$ and using local uniform convergence we get

$$\begin{aligned} e^{-x} &= \lim_{t \to \infty} [t]a([t])F'(a([t])x_t + b_{[t]}) \\ &= \lim_{t \to \infty} ta(t)F'(V(te^x)). \end{aligned}$$

(The last step uses $ta(t) \sim [t]a([t])$ which is a consequence of $t \sim [t]$, $ta(t) \in RV_1$ and Proposition 0.8(iii).) Putting $s = te^x$ we get

$$e^{-x} = \lim_{s \to \infty} se^{-x}a(se^{-x})F'(V(s))$$

and since slow variation of $a(\cdot)$ implies $a(se^{-x}) \sim a(s)$ the foregoing is equivalent to

$$1/a(s) \sim sF'(V(s))$$

which is the same as the Von Mises condition (1.21) by Proposition 1.17(d).

Details for the case $G = \Phi_\alpha$ are similar. If (1.19) holds then set $a(t) = V(t) = (1/(1 - F))^{\leftarrow}(t)$ and if $x_n \to x > 0$ then $a(n)x_n \to \infty$ and

$$\begin{aligned} g_n(x_n) &= na(n)F^{n-1}(a(n)x_n)F'(a_n x_n) \\ &\sim \alpha na(n)\Phi_\alpha(x)(1 - F(a_n x_n))/a(n)x_n \end{aligned}$$

since the Von Mises condition is the same as $F'(t) \sim \alpha(1 - F(t))t^{-1}$. The right side here is

$$\alpha\Phi_\alpha(x)n(1 - F(a_n x_n))/x_n \to \alpha\Phi_\alpha(x)x^{-\alpha-1} = G'(x).$$

Conversely if (2.15) holds locally uniformly on $(0, \infty)$, $F \in D(\Phi_\alpha)$ and we can set $b_n = 0$, $a_n = V(n)$ and $V = (1/(1 - F))^{\leftarrow} \in RV_{\alpha^{-1}}$ (Proposition 0.8(v)). Therefore setting $x_t = xV(t)/V([t]) \to x$ we get for $x > 0$

$$\Phi_\alpha(x)\alpha x^{-\alpha-1} = \lim_{n\to\infty} na_n F^{n-1}(a_n x_n)F'(a_n x_n)$$

and therefore

$$\begin{aligned}\alpha x^{-\alpha-1} &= \lim_{t\to\infty} [t]F'(xV(t))V([t]) \\ &= \lim_{t\to\infty} tV(t)F'(xV(t))\end{aligned}$$

and setting $s = xV(t)$ this is

$$\begin{aligned}&= \lim_{s\to\infty} V^{\leftarrow}(sx^{-1})V(V^{\leftarrow}(sx^{-1}))F'(s) \\ &= \lim_{s\to\infty} x^{-\alpha}V^{\leftarrow}(s)sx^{-1}F'(s)\end{aligned}$$

(since $V^{\leftarrow} \in RV_\alpha$) and so

$$\alpha = \lim_{s\to\infty} sF'(s)/(1 - F(s))$$

as desired. The treatment for $\Psi_\alpha$ is left as an exercise. □

We now discuss how to extend this result to get uniform convergence of $g_n$ on all of $\mathbb{R}$. Getting uniform convergence in neighborhoods of $+\infty$ is no problem when $G = \Phi_\alpha$ or $\Lambda$ as it comes almost for free under the Von Mises conditions. When $G = \Psi_\alpha$, some care must be taken in extending local uniform convergence on left, closed neighborhoods of 0.

To get uniform convergence in neighborhoods of $\infty$ it suffices to show (see Section 0.1) that if $x_n \to \infty$ then $g_n(x_n) \to 0$. If $G = \Lambda$ then

$$\begin{aligned}g_n(x_n) &= na_n F^{n-1}(a_n x_n + b_n)F'(a_n x_n + b_n) \\ &\le na_n F'(a_n x_n + b_n).\end{aligned}$$

The Von Mises condition (1.21) requires $F' > 0$ in a left neighborhood of $x_0$ and in this neighborhood $F$ is strictly increasing. Now $a_n = f(b_n) = \int_{b_n}^{x_0}(1 - F(s))/(1 - F(b_n)) \sim (1 - F(b_n))/F'(b_n) = 1/(nF'(b_n))$ (by 1.21), and thus

$$\limsup_{n\to\infty} g_n(x_n) \le \limsup_{n\to\infty} F'(a_n x_n + b_n)/F'(b_n).$$

Recall from Proposition 1.17(d) that $tF'((1/(1 - F))^{\leftarrow}(t)) =: L(t) \in RV_0$ so that

$$F'(y) = L((1/(1 - F(y))))(1 - F(y)) \tag{2.17}$$

in a left neighborhood of $x_0$. Since $n(1 - F(a_n x_n + b_n)) \to 0$ we have for $\varepsilon > 0$

$$\limsup_{n\to\infty} F'(a_n x_n + b_n)/F'(b_n)$$

$$= \limsup_{n\to\infty} \frac{L(n/(n(1 - F(a_n x_n + b_n))))n(1 - F(a_n x_n + b_n))}{L(n)}$$

$$\le \limsup_{n\to\infty} (1 + \varepsilon)(1/(n(1 - F(a_n x_n + b_n))))^{\varepsilon} n(1 - F(a_n x_n + b_n)$$

(by applying Proposition 0.8(iii))

$$= \limsup_{n\to\infty} (1 + \varepsilon)(n(1 - F(a_n x_n + b_n)))^{1-\varepsilon} = 0.$$

If $G = \Phi_\alpha$ the details are simpler: If $x_n \to \infty$ then

$$\limsup_{n\to\infty} g_n(x_n) \le \limsup_{n\to\infty} n a_n F'(a_n x_n)$$

$$= \limsup_{n\to\infty} n(1 - F(a_n x_n))x_n^{-1}(a_n x_n F'(a_n x_n)/(1 - F(a_n x_n)))$$

and supposing (1.19) holds this is

$$\limsup_{n\to\infty} \alpha n(1 - F(a_n x_n))x_n^{-1} = 0$$

since $x_n \to \infty$ and $n(1 - F(a_n x_n)) \to 0$.

Now consider the problem of getting uniform convergence in (2.15) on intervals of the form $[-M, 0]$, $M > 0$, when $G = \Psi_\alpha$. Let $x_n \uparrow 0$ and we seek conditions which guarantee $g_n(x_n) \to \Psi_\alpha'(0)$. Since $\Psi_\alpha'(0) = \infty$ if $0 < \alpha < 1$, we only consider the case $\alpha \ge 1$. The Von Mises condition is (1.20) and the norming constants are given in Proposition 1.13 so that

$$g_n(x_n) = n(x_0 - \gamma_n)F^{n-1}(x_0 + (x_0 - \gamma_n)x_n)F'(x_0 + (x_0 - \gamma_n)x_n).$$

Since $F^{n-1}(x_0 + (x_0 - \gamma_n)x_n) \to \Psi_\alpha(0) = 1$ we must consider when

$$n(x_0 - \gamma_n)F'(x_0 + (x_0 - \gamma_n)x_n) \to \begin{cases} 0 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1. \end{cases} \tag{2.18}$$

The Von Mises condition says $F'(y) \sim \alpha(1 - F(y))/(x_0 - y)$ as $y \uparrow x_0$ and using this (2.18) becomes

$$n(1 - F(x_0 + (x_0 - \gamma_n)x_n))/(-x_n) \to \begin{cases} 0 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1. \end{cases} \tag{2.19}$$

It is convenient to recall that $F \in D(\Psi_\alpha)$ iff

$$U(t) := 1/(1 - F(x_0 - t^{-1})) \in RV_\alpha$$

(Proposition 1.13). Check that

$$U^{\leftarrow}(t) = 1/(x_0 - (1/(1 - F))^{\leftarrow}(t))$$

and thus (2.19) can be recast as

$$\frac{U(U^{\leftarrow}(n))}{|x_n|U(U^{\leftarrow}(n)|x_n|^{-1})} \to \begin{cases} 0 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1 \end{cases}$$

and setting $y_n = |x_n|^{-1}$ we have (2.19) equivalent to the condition:

For all sequences $\{y_n\}$ such that

$$0 < y_n \to \infty: \frac{U(U^{\leftarrow}(n)y_n)}{y_n U(U^{\leftarrow}(n))} \to \begin{cases} \infty & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1 \end{cases} \tag{2.19'}$$

which in turn is equivalent to the more appealing condition:

For any function $y(t)$ such that

$$0 < y(t) \to \infty: \frac{U(ty(t))}{y(t)U(t)} \to \begin{cases} \infty & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1. \end{cases} \tag{2.20}$$

(Check this equivalence!) If $\alpha > 1$ we may readily check that (2.20) is a simple consequence of the inequalities of Proposition 0.8(ii). So we now focus on the case $\alpha = 1$.

Suppose the left-hand derivative $F'(x_0)$ of $F$ at $x_0$ exists, is finite and nonzero. Then as $x \to x_0-$ the Von Mises condition is

$$F'(x) \sim \bar{F}(x)/(x_0 - x)$$

and the right side is

$$\frac{F(x_0) - F(x)}{x_0 - x} \to F'(x_0) \in (0, \infty)$$

and we conclude that as $x \to x_0-$

$$F'(x) \to F'(x_0) < \infty.$$

Thus the left side of (2.18) is

$$\begin{aligned} &n(x_0 - \gamma_n)F'(x_0 - (x_0 - \gamma_n)x_n) \\ &\sim F'(x_0)n(x_0 - \gamma_n) = F'(x_0)n/U^{\leftarrow}(n). \end{aligned}$$

However we are assuming

$$\begin{aligned} F'(x_0) &= \lim_{t\to\infty} \frac{F(x_0) - F(x_0 - t^{-1})}{t^{-1}} \\ &= \lim_{t\to\infty} \frac{t}{U(t)} \end{aligned}$$

and thus by an application of Proposition 0.8(vi)

$$\lim_{t\to\infty} U^{\leftarrow}(t)/t = F'(x_0)$$

and hence (2.18) holds.

Conversely suppose (2.18) or equivalently (2.20) holds for $\alpha = 1$. Writing

$U(t) = tL(t)$ this means we assume:

For any function $y(t)$ such that

$$0 < y(t) \to \infty: \frac{L(ty(t))}{L(t)} \to 1. \tag{2.20'}$$

We wish to conclude that (2.20′) implies the existence of a constant $c$, $0 < c < \infty$, such that

$$\lim_{t\to\infty} L(t) = c.$$

Suppose not. Then either (a) $L(t) \to \infty$, or (b) $L(t) \to 0$, or (c) there exist $0 \le c_1 < c_2 \le \infty$ and sequences $s_n \to \infty$, $t_n \to \infty$ with $L(s_n) \to c_1$, $L(t_n) \to c_2$. To see that (c) leads to a contradiction define for $n \ge 1$

$$k(n) = \inf\{k \ge n: s_k/t_n \ge n\}$$

so that $k(n) \to \infty$ and $y_n := s_{k(n)}/t_n \to \infty$. Then from (2.20′)

$$1 = \lim_{n\to\infty} \frac{L(t_n y_n)}{L(t_n)} = \lim_{n\to\infty} \frac{L(s_{k(n)})}{L(t_n)} = \frac{c_1}{c_2} < 1,$$

a contradiction. To see that (a) leads to a contradiction take $t_n \to \infty$ and define for $n \ge 1$

$$k(n) = \inf\{k \ge n: \frac{t_{k(n)}}{t_n} \ge n \quad \text{and} \quad L(t_{k(n)}) > L^2(t_n)\}$$

so that $y_n := t_{k(n)}/t_n \to \infty$ and

$$\begin{aligned} 1 &= \lim_{n\to\infty} L(t_n y_n)/L(t_n) = \lim_{n\to\infty} L(t_{k(n)})/L(t_n) \\ &\ge \lim_{n\to\infty} L^2(t_n)/L(t_n) = \lim_{n\to\infty} L(t_n) = \infty, \end{aligned}$$

a contradiction. The proof that (b) leads to a contradiction is similar.

Thus $L(t) \to c$ whence

$$c^{-1} = \lim_{t\to\infty} t/U(t) = \lim_{t\to\infty} \frac{F(x_0) - F(x_0 - t^{-1})}{t^{-1}}$$

and so the left-hand derivative of $F$ at $x_0$ exists finite and nonzero.

Uniform convergence on $(0, \infty)$ is readily verified from (2.18) when $F \in D(\Psi_\alpha)$ since the left side of (2.18) is zero for $x_n > 0$.

We now summarize these findings.

**Proposition 2.6.** *Suppose $F$ is absolutely continuous with density $F'(x)$ and the appropriate Von Mises condition* (1.19), (1.20), *or* (1.21) *holds. If $F \in D(\Phi_\alpha)$ then density convergence* (2.15) *on* $(0, \infty)$ *is uniform on neighborhoods of* $\infty$. *If $F \in D(\Lambda)$ then density convergence* (2.15) on $\mathbb{R}$ *is uniform on neighborhoods of* $\infty$. *Suppose $F \in D(\Psi_\alpha)$ and $\alpha \ge 1$. If $\alpha > 1$ density convergence* (2.15) is *uniform* *on intervals of the form* $[-M, 0]$, $M > 0$ *(and hence uniform on all neighborhoods of* $\infty$*). If* $\alpha = 1$ *this is true* iff *the left-hand derivative of* $F$ *at* $x_0$ *exists positive and finite.*

We now discuss how to obtain uniform convergence on all of $\mathbb{R}$. Similar to the situation in Section 2.1, behavior not controlled by right tail conditions must be controlled by imposing extra conditions; *cf.* (2.21) later.

**Proposition 2.7.** *Suppose* $F$ *is absolutely continuous with density* $F'(x)$ *and that one of the Von Mises conditions* (1.19), (1.20), *or* (1.21) *holds. If* $F \in D(\Psi_\alpha)$, *suppose either* $\alpha > 1$ *or* $\alpha = 1$ *and the left-hand derivative of* $F$ *at* $x_0$ *exists positive and finite. Then density convergence* (2.15) *holds uniformly on* $\mathbb{R}$ iff *there exist constants* $B > 0$, $C > 0$ *such that for all* $x$

$$F'(x)F(x)^B \le C. \tag{2.21}$$

*Remark.* If the density $F'$ is bounded then (2.21) is satisfied. Condition (2.21), devised by Sweeting (1985), is designed to allow the density to become unbound in right neighborhoods of $x_l = \inf\{x: F(x) > 0\}$. Condition (2.21) says that the distribution $F^{B+1}(x)$ has a bounded density.

Proof. We start by supposing $F \in D(\Lambda)$, (1.21), and (2.21) hold. Let $x_n \to -\infty$ and we must show $g_n(x_n) \to 0$. There are two possibilities along subsequences $\{n'\}$:

(a) $a_{n'}x_{n'} + b_{n'} \to x_0$ or

(b) $a_{n'}x_{n'} + b_{n'} \le K < x_0$.

In case (a), just as in the developments following (2.17), we have by recalling $a_n \sim (1 - F(b_n))/F'(b_n)$ and using Proposition 0.8(ii) that for large $n'$

$$\begin{aligned} g_{n'}(x_{n'}) &\sim n'(1 - F(b_{n'}))F^{n'-1}(a_{n'}x_{n'} + b_{n'})\frac{L(n'/(n'(1 - F(a_{n'}x_{n'} + b_{n'}))))}{L(n')} \\ &\quad \cdot n'(1 - F(a_{n'}x_{n'} + b_{n'}) \\ &\le (1 + \varepsilon)(n'(1 - F(a_{n'}x_{n'} + b_{n'})))^{1-\varepsilon}F^{n'-1}(a_{n'}x_{n'} + b_{n'}) \end{aligned}$$

and remembering that $F = 1 - (1 - F) \le e^{-(1-F)}$ the foregoing is bounded by

$$\le (1 + \varepsilon)(n'(1 - F))^{1-\varepsilon}\exp\{-((n' - 1)/n')n'(1 - F)\}.$$

If $n'(1 - F) \to \infty$ then this bound obviously approaches zero as desired. To check $n'(1 - F) \to \infty$ note that since $x_{n'} \to -\infty$ we have for any $M \ge 0$ and all large $n'$ that

$$x_{n'} \le -M$$

and hence

$$n'(1 - F(a_{n'}x_{n'} + b_{n'})) \ge n'(1 - F(a_{n'}(-M) + b_{n'}) \to e^M$$

and since $M$ is arbitrary the conclusion is that

$$n'(1 - F(a_{n'}x_{n'} + b_{n'}) \to \infty.$$

In case (b) when $a_{n'}x_{n'} + b_{n'} \le K < \infty$ we have

$$\begin{aligned} g_{n'}(x_{n'}) &= a_{n'}n'F^{n'-1}(a_{n'}x_{n'} + b_{n'})F'(a_{n'}x_{n'} + b_{n'}) \\ &\le n'a_{n'}CF^{n'-1-B}(a_{n'}x_{n'} + b_{n'}) \end{aligned}$$

(by 2.21)

$$\le Cn'a_{n'}F^{n'-1-B}(K) \to 0$$

since $na_n \sim nL(n)$ and $F^{n-1-B}(K) \to 0$ geometrically fast as a consequence of $F(K) < 1$.

Conversely suppose (2.21) fails. Then with $C = 1$ and $B = n - 1$ we see there exists $z_n \in (x_l, x_0)$ such that

$$F'(z_n) \ge F(z_n)^{-n+1}.$$

Set $x_n = a_n^{-1}(z_n - b_n)$ and then

$$\begin{aligned} g_n(x_n) &= na_nF^{n-1}(a_nx_n + b_n)F'(a_nx_n + b_n) \\ &\ge na_nF^{n-1}(z_n)F^{-n+1}(z_n) = na_n \to \infty \end{aligned}$$

so that $g_n$ cannot converge uniformly on $\mathbb{R}$ to $\Lambda'(x)$.

Now we deal with the problem assuming $F \in D(\Phi_\alpha)$. Suppose (1.19) and (2.21) hold and we show uniform convergence of $g_n \to \phi'_\alpha$ on intervals of the form $[0, M]$. Suppose $0 < x_n \to 0$ and again along subsequences $\{n'\}$ there are two cases:

(a) $a_{n'}x_{n'} \to \infty$

(b) $a_{n'}x_{n'} \le K < \infty$.

Case (b) is handled as in the discussion for $\Lambda$ so we focus on (a):

$$\begin{aligned} g_{n'}(x_{n'}) &= n'a_{n'}F^{n'-1}(a_{n'}x_{n'})F'(a_{n'}x_{n'}) \\ &\sim \alpha n'(1 - F(a_{n'}x_{n'}))F^{n'-1}(a_{n'}x_{n'})/x_{n'} \qquad \text{(from (1.19))} \\ &\le \alpha n'(1 - F(a_{n'}x_{n'}))\exp\{-(n' - 1)(1 - F(a_{n'}x_{n'}))\}/x_{n'}. \end{aligned}$$

Now $U = 1/(1 - F) \in RV_\alpha$ so that since $U(a_n)/U(a_nx_n) = U(a_nx_nx_n^{-1})/U(a_nx_n)$ we get from Proposition 0.8(ii) for given $0 < \varepsilon < \alpha$ and large $n$

$$(1 - \varepsilon)(x_n^{-1})^{\alpha-\varepsilon} \le U(a_n)/U(a_nx_n) \le (1 + \varepsilon)(x_n^{-1})^{\alpha+\varepsilon}.$$

Since $n(1 - F(a_nx_n)) \sim (1 - F(a_nx_n))/(1 - F(a_n)) = U(a_n)/U(a_nx_n)$ we have for large $n$

$$(1 - \varepsilon)^2x_n^{-\alpha+\varepsilon} \le n(1 - F(a_nx_n)) \le (1 + \varepsilon)^2x_n^{-\alpha-\varepsilon}.$$

So when $n'$ is large, a bound on $g_{n'}(x_{n'})$ is of the form

$$\begin{aligned} &\alpha(1 + \varepsilon)^2x_{n'}^{-\alpha-\varepsilon}\exp\{-(1 - \varepsilon)^3x_{n'}^{-\alpha+\varepsilon}\}/x_{n'} \\ &= O(x_{n'}^{-\alpha-\varepsilon-1}\exp\{-(1 - \varepsilon)^3x_{n'}^{-\alpha+\varepsilon}\}) \to 0. \end{aligned}$$

To get uniform convergence on $(-\infty, 0)$ is easy: If $x_n \to x < 0$ then the simple argument of case (b) works. The converse is nearly the same as for $\Lambda$ and is omitted.

Checking details for $F \in D(\Psi_\alpha)$ is left as an exercise. □

Convergence of $g_n$ to $G'$ in the $L_p$ metric is considered in de Haan and Resnick (1982) and Sweeting (1985). The first reference also considers local limit theorems when it is not assumed that $F$ has a density.

## Exercises

2.2.1. Suppose $F \in D(\Lambda)$ with auxiliary function $f$. For given $\varepsilon > 0$ there exist $t_0$ such that for $x > 0, t > t_0$

$$(1-\varepsilon)\left[\frac{1-F(t)}{1-F(t+xf(t))}\right]^{-\varepsilon} \le \frac{f(t+xf(t))}{f(t)} \le (1+\varepsilon)\left[\frac{1-F(t)}{1-F(t+xf(t))}\right]^{\varepsilon}$$

and for $x < 0, t + xf(t) > t_0$

$$(1-\varepsilon)\left[\frac{1-F(t+xf(t))}{1-F(t)}\right]^{-\varepsilon} \le \frac{f(t+xf(t))}{f(t)} \le (1+\varepsilon)\left[\frac{1-F(t+xf(t))}{1-F(t)}\right]^{\varepsilon}.$$

Hint: With $U = 1/(1-F)$ we have $f \circ U^{\leftarrow} \in RV_0$. Apply Proposition 0.8(ii) (de Haan and Resnick, 1982).

2.2.2. Give an example of $F \in D(\Lambda)$ satisfying (2.21) but with the density $F'$ unbounded near $x_l$ and $g_n \to G'$ uniformly on $\mathbb{R}$.

2.2.3. Prove the unproven statements in the $\Psi_\alpha$ case.

2.2.4. Suppose the conditions of Proposition 2.7 hold so that $g_n \to G'$ uniformly on $\mathbb{R}$. For any sequence $d_n \to \infty$ and $h > 0$

(a) If $F \in D(\Phi_\alpha)$, then $\lim_{n\to\infty} d_n P[x < a_n^{-1} M_n \le x + d_n^{-1} h] = h\Phi'_\alpha(x)$ uniformly on $\mathbb{R}$.

(b) If $F \in D(\Lambda)$, then $\lim_{n\to\infty} d_n P[x < a_n^{-1}(M_n - b_n) \le x + d_n^{-1} h] = h\Lambda'(x)$ uniformly on $\mathbb{R}$.

(c) If $F \in D(\Psi_\alpha)$, then

$$\lim_{n\to\infty} d_n P[x < (M_n - x_0)/(x_0 - \gamma_n) \le x + d_n^{-1} h] = h\Psi'_\alpha(x)$$

(de Haan and Resnick, 1982).

# 2.3. Large Deviations

This section is based on Anderson (1971, 1976, 1978, 1984), de Haan and Hordijk (1972), and Goldie and Smith (1987).

Since convergence in (2.1) holds uniformly in $\mathbb{R}$, we may write (2.1) as

$$\sup_{x\in\mathbb{R}} |F^n(x) - G(a_n^{-1}(x - b_n))| = d_n \to 0. \tag{2.22}$$

This suggests that in a statistical context, if we do not know $F$, we regard $G(a^{-1}(x - b))$ as the approximate distribution of extremes. This is appealing for two reasons. First, we have at most three parameters to estimate ($a$, $b$, and possibly the shape parameter $\alpha$ appearing in the definition of $\Phi_\alpha$ and $\Psi_\alpha$), and second, domain of attraction restrictions on the tail behavior of $F$ are mild and satisfied by most common densities so that this procedure is fairly robust.

Of course there are two possible sources of difficulties: the statistical estimation of parameters and the approximating of $F^n(x)$ by $G(a_n^{-1}(x - b_n))$. For the latter it would be useful to know something about $d_n$ appearing in (2.22), and this is the subject of Section 2.4. However $d_n$ is not always the best way to measure how close $F^n(a_n x + b_n)$ is to $G(x)$. In problems concerning probabilities of exceeding large values, we care about how closely $1 - G(x)$ approximates $P[M_n > a_n x + b_n]$ for large values of $x$. Since both $1 - G(x)$ and $P[M_n > a_n x + b_n]$ are likely to be very small, these quantities have little influence on $d_n$ and it is better instead to use relative error, which is equivalent to examining how close

$$\frac{P[M_n > a_n x + b_n]}{1 - G(x)} = \frac{1 - F^n(a_n x + b_n)}{1 - G(x)}$$

is to one for large $x$.

We seek $x_n \uparrow \infty$, the convergence to infinity being as fast as possible, such that

$$\lim_{n \to \infty} P[(M_n - b_n)/a_n > y_n]/(1 - G(y_n)) = 1 \tag{2.23}$$

for any sequence $\{y_n\}$ such that $y_n = O(x_n)$ or equivalently for any positive $A$

$$\lim_{n \to \infty} \sup_{x \le A x_n} |P[(M_n - b_n)/a_n > x]/(1 - G(x)) - 1| = 0. \tag{2.23'}$$

We always assume $\{x_n\}$ is strictly increasing. The relations (2.23) or (2.23′) describe those values of $x$ such that $1 - G(x)$ is a good approximation to $P[(M_n - b_n)/a_n > x]$. Obviously the faster we can allow $x_n$ to increase to $\infty$, the better the approximation.

We begin by considering the case $F \in D(\Phi_\alpha)$. In this case

$$1 - F(x) = x^{-\alpha} L(x), \qquad x \to \infty$$

with $L \in RV_0$ and we may take $b_n = 0$, $a_n = (1/(1 - F))^{\leftarrow}(n)$. Then (2.23) can be written as (assume $y_n \to \infty$, the contrary case being covered by uniform convergence in (2.1))

$$\begin{aligned} 1 &= \lim_{n \to \infty} (1 - F^n(a_n y_n))/(1 - \Phi_\alpha(y_n)) \\ &= \lim_{n \to \infty} n(1 - F(a_n y_n))/y_n^{-\alpha} \end{aligned}$$

(since $1 - \phi_\alpha(x) \sim x^{-\alpha}$, $x \to \infty$ and $1 - F^n(a_n y_n) = 1 - e^{-n(-\log F(a_n y_n))} \sim n(-\log F(a_n y_n)) \sim n(1 - F(a_n y_n))$ where the asymptotic equivalences are justified by the fact that $F^n(a_n y_n) \to 1$ implies $n \log F(a_n y_n) \to 0$). Since $n^{-1} \sim 1 - F(a_n)$ the previous expression becomes

$$\begin{aligned} 1 &= \lim_{n\to\infty} y_n^\alpha(1 - F(a_n y_n))/(1 - F(a_n)) \\ &= \lim_{n\to\infty} (a_n y_n)^\alpha(1 - F(a_n y_n))/a_n^\alpha(1 - F(a_n))) \\ &= \lim_{n\to\infty} L(a_n y_n)/L(a_n) \end{aligned} \tag{2.24}$$

and this is equivalent to (2.23). Condition (2.24) is rephrased slightly in the following.

**Proposition 2.8.** (Anderson, 1978). *Suppose* $x^\alpha(1 - F(x)) = L(x) \in RV_0$ *and* $\{x_n\}$ *is strictly increasing,* $x_n \uparrow \infty$. *Then the large deviation property* (2.23) *holds* iff *there exists a non-decreasing function* $\xi(t)$ *with* $\xi(\infty) = \infty$ *such that* $\xi(a_n) = x_n$ *and*

$$\lim_{t\to\infty} L(t\xi^\delta(t))/L(t) = 1 \tag{2.25}$$

*locally uniformly in* $\delta \in [0, \infty]$.

Proof. Suppose (2.25) holds and we verify (2.24) when $y_n \to \infty$, $y_n \le A x_n$. (Again, note if a subsequence of $\{y_n\}$ is bounded, then (2.24) along that subsequence is a direct consequence of uniform convergence; cf. Proposition 0.5.) Since $y_n \le A x_n$, if we set $\delta_n = \log y_n / \log x_n$ then $y_n = x_n^{\delta_n} = \xi^{\delta_n}(a_n)$ and $\{\delta_n\}$ is bounded. The limit in (2.24) is

$$\lim_{n\to\infty} L(a_n \xi^{\delta_n}(a_n))/L(a_n)$$

and this limit is 1 since (2.25) is assumed to hold locally uniformly in $\delta$.

Conversely if (2.24) holds then for all $\{\delta_n\} \subset [0,1]^\infty = \{(u_1, u_2, \ldots): u_i \in [0,1], i = 1, 2, \ldots\}$

$$\lim_{n\to\infty} L(a_n x_n^{\delta_n})/L(a_n) = 1. \tag{2.26}$$

Define

$$\xi(t) = x_n \qquad \text{for } t \in [a_n, a_{n+1})$$

and

$$n(t) = \sup\{n: a_n \le t\}$$

so that

$$a_{n(t)} \le t < a_{n(t)+1}. \tag{2.27}$$

Then since $a_n \sim a_{n+1}$ (because of $a(t) = (1/(1 - F))^\leftarrow(t) \in RV_{\alpha^{-1}}$ and Proposition 0.8(iii)) we get by dividing through by $a_{n(t)}$ in (2.27) that as $t \to \infty$

$$a_{n(t)} \sim t.$$

Furthermore

$$a_{n(t)}\xi^{\delta_{[t]}}(a_{n(t)}) \le t\xi^{\delta_{[t]}}(t) \le a_{n(t)+1}\xi^{\delta_{[t]}}(a_{n(t)})$$

which implies

$$t\xi^{\delta_{[t]}}(t) \sim a_{n(t)}\xi^{\delta_{[t]}}(a_{n(t)})$$

and from (2.26) and Proposition 0.8(iii)

$$\lim_{t\to\infty} L(t\xi^{\delta_{[t]}}(t))/L(t) = 1$$

and since $\{\delta_n\}$ is arbitrarily chosen in $[0,1]^\infty$ (2.25) holds uniformly for $0 \le \delta \le 1$. It is easy to extend this to local uniform convergence on $[0,\infty)$. For instance if $\delta(t) \in [1,2]$ then

$$\lim_{t\to\infty} \frac{L(t\xi^{\delta(t)}(t))}{L(t)} = \lim_{t\to\infty} \frac{L(t\xi(t)\xi^{\delta(t)-1}(t))L(t\xi(t))}{L(t\xi(t))L(t)} = 1. \qquad \square$$

*Remark.* Proposition 2.8 does not assert (2.25) holds for all $\xi$ satisfying $\xi(a_n) = x_n$.

Since (2.25) strengthens the slow variation property, Anderson (1978) aptly termed (2.25) *super slow variation.*

**Definition.** Let $\xi$ be nondecreasing with $\xi(\infty) = \infty$. A slowly varying function is $\xi$-super slowly varying ($\xi$-ssv) if

$$\lim_{t\to\infty} L(t\xi^{\delta}(t))/L(t) = 1 \tag{2.25}$$

locally uniformly for $\delta \in [0,\infty)$.

Actually, uniformity in (2.25) is a consequence of pointwise convergence in $\delta$, provided the function $\xi$ satisfies a growth condition. See Anderson, 1984, Theorem 2, and Goldie and Smith, (1987), Section 2.3.

We now examine sufficient conditions for super slow variation and hence for the large deviation property.

Suppose $L \in RV_0$ has Karamata representation (cf. Corollary 0.7)

$$L(x) = c(x)\exp\left\{\int_1^x t^{-1}\varepsilon(t)dt\right\} \tag{2.28}$$

where $c(x) \to c > 0$, $\varepsilon(x) \to 0$ as $x \to \infty$. Since $\xi(t) \to \infty$ we have

$$L(t\xi^{\delta}(t))/L(t) \sim \exp\left\{\int_t^{t\xi^{\delta}(t)} u^{-1}\varepsilon(u)du\right\}$$

so that the ratio goes to 1 locally uniformly in $\delta$ iff the integral goes to zero locally uniformly in $\delta$. In the integral make the change of variable $z = (\log u - \log t)/\log \xi(t)$ and

$$\int_t^{t\xi^{\delta}(t)} u^{-1}\varepsilon(u)du = \int_0^{\delta} \varepsilon(\exp\{\log t + z\log\xi(t)\})\log\xi(t)dz$$

and this goes to zero locally uniformly in $\delta$ provided

$$\varepsilon(t\xi^z(t))\log\xi(t)\to 0 \tag{2.29}$$

as $t\to\infty$ locally uniformly in $z$. However (2.29) is true iff

$$\varepsilon(t)\log\xi(t)\to 0 \tag{2.30}$$

and this is the desired sufficient condition. (To see the equivalence of (2.29) and (2.30) note by setting $z=0$ in (2.29) we get that (2.29) obviously implies (2.30). Conversely if (2.30) holds, then because $\xi$ is nondecreasing

$$\begin{aligned} 0 &= \lim_{t\to\infty}\varepsilon(t\xi^z(t))\log\xi(t\xi^z(t)) \\ &\geq \lim_{t\to\infty}\varepsilon(t\xi^z(t))\log\xi(t)\geq 0 \end{aligned}$$

and the convergence is locally uniform in $z$.)

As a special case consider what happens if the Von Mises condition (1.19) holds; i.e.,

$$\alpha(t):=tF'(t)/(1-F(t))\to\alpha$$

as $t\to\infty$. Then

$$1-F(t)=c\exp\left\{-\int_1^x t^{-1}\alpha(t)dt\right\}$$

and thus

$$\varepsilon(t)=\alpha(t)-\alpha$$

and sufficient condition (2.30) becomes

$$\log\xi(t)((tF'(t)/(1-F(t)))-\alpha)\to 0$$

as $t\to\infty$.

Another interesting special case, which covers the Cauchy, Pareto, $t$, and $F$ distributions, is when

$$t^\alpha(1-F(t))\to c>0$$

as $t\to\infty$. In this case

$$L(t)=t^\alpha(1-F(t))=c(t)e^0$$

so $\varepsilon(t)\equiv 0$ and all growth rates for $\varepsilon(t)$ are allowed.

Interestingly enough, when $\xi$ satisfies a growth condition, (2.30) is necessary as well.

**Proposition 2.9** (Anderson, 1978; Goldie and Smith, 1987). *Suppose $\xi(t)\to\infty$ and*

$$\log\xi(x\xi(x))=0(\log\xi(x)) \tag{2.31}$$

*as $x\to\infty$. Then $L\in RV_0$ is $\xi-$* ssv iff *$L$ has a Karamata representation of the* *form* (2.28) *with*

$$\varepsilon(t)\log \xi(t) \to 0 \tag{2.30}$$

*as* $t \to \infty$.

*Remark.* In cases where $\xi$ is differentiable, (2.31) is close to saying $\log \xi(e^x)$ has a bounded derivative. See Exercise (2.3.4).

PROOF. You might wish to review Exercises 0.4.3.8 and 0.4.3.11, and Lemma 1.3. We need only prove necessity of (2.30). Define

$$L^{\#}(t) = L(e^t)$$

$$\xi^*(t) = \log \xi(e^t)$$

so that assuming $L$ is $\xi$ − ssv and satisfies (2.25) we get

$$\lim_{s\to\infty} L^{\#}(s + \delta\xi^*(s))/L^{\#}(s) = 1 \tag{2.25$'$}$$

locally uniformly in $\delta$ and the rephrasing of (2.25) and (2.31) becomes

$$\xi^*(x + \xi^*(x))/\xi^*(x)) \le K. \tag{2.31$'$}$$

Pick $c$ such that $\xi^*(t) > 0$ for $t \ge c$. The function

$$\int_c^x (1/\xi^*(u))du$$

for $x \ge c$ is continuous, strictly increasing, and since for $x \ge c$

$$\begin{aligned}\int_x^{x+\xi^*(x)} (1/\xi^*(u))du &= \int_0^1 \xi^*(x)/\xi^*(x + u\xi^*(x))du \\ &\ge 1\xi^*(x)/\xi^*(x + \xi^*(x)) \ge K^{-1} \qquad \text{(from (2.31$'$))}\end{aligned} \tag{2.32}$$

we have

$$\int_c^\infty (1/\xi^*(u))du \ge \int_c^{c+\xi^*(c)} + \int_{c+\xi^*(c)}^{c+\xi^*(c)+\xi^*(c+\xi^*(c))} + \cdots = \infty.$$

Thus the function

$$U(x) = \exp\left\{\int_c^x (1/\xi^*(u))du\right\}$$

is continuous, strictly increasing with $U(\infty) = \infty$, and $U^{\leftarrow}$ is well defined on $[0, \infty)$. We show

$$L^{\#} = L_1 \circ U \tag{2.33}$$

with $L_1 \in RV_0$. Repeat the argument leading to (2.32) to get for $0 \le \delta \le 1$

$$U(x + \delta\xi^*(x))/U(x) \ge e^{\delta K^{-1}} \tag{2.34}$$

for $x \ge c$ and hence for $v \in [1, e^{K^{-1}}]$

$$0 \le (U^{\leftarrow}(vU(x)) - x)/\xi^*(x) \le K \log v \le 1. \tag{2.35}$$

Thus to show $L_1 = L^{\#} \circ U^{\leftarrow} \in RV_0$ we write for $v \in [1, e^{K^{-1}}]$

$$\lim_{t\to\infty} \frac{L_1(tv)}{L_1(t)} = \lim_{t\to\infty} \frac{L^{\#}(U^{\leftarrow}(tv))}{L^{\#}(U^{\leftarrow}(t))}$$

and setting $x = U^{\leftarrow}(t)$ we have this equal to

$$= \lim_{x\to\infty} \frac{L^{\#}(U^{\leftarrow}(vU(x))}{L^{\#}(x)}$$

$$= \lim_{x\to\infty} L^{\#}\left(\left[\frac{U^{\leftarrow}(vU(x)) - x}{\xi^*(x)}\right]\xi^*(x) + x\right)/L^{\#}(x) = 1$$

since

$$0 \le (U^{\leftarrow}(vU(x)) - x)/\xi^*(x) \le 1$$

and the convergence in (2.25′) is uniform for $\delta \in [0, 1]$. This verifies $\lim_{t\to\infty} L_1(tv)/L_1(t) = 1$ for $v \in [1, e^{K^{-1}}]$ and a simple argument extends this to all $v \ge 1$. Hence (2.33) is checked.

Since $L_1 \in RV_0$, it has a Karamata representation

$$L_1(x) = c_1(x) \exp\left\{\int_1^x t^{-1}\varepsilon_1(t)dt\right\}$$

where $c_1(x) \to c_1 > 0$, $\varepsilon_1(x) \to 0$. Thus

$$L^{\#}(x) = c_1(U(x)) \exp\left\{\int_1^{U(x)} t^{-1}\varepsilon_1(t)dt\right\}.$$

Make the change of variable $s = \exp\{U^{\leftarrow}(t)\}$ so that $t = U(\log s)$, $dt = s^{-1}U'(\log s)ds = U(\log s)/(s\xi^*(\log s))ds$, and

$$L(e^x) = L^{\#}(x) \sim (\text{const}) \exp\left\{\int_1^{e^x} (\varepsilon_1(U(\log s))/\log \xi(s))s^{-1}\, ds\right\}$$

and setting

$$\varepsilon(s) = \varepsilon_1(U(\log s))/\log \xi(s)$$

we see that $L$ has a Karamata representation with $\varepsilon(s)$ satisfying

$$\varepsilon(s) \log \xi(s) \to 0$$

as required. □

Now we consider (2.23) for $F \in D(\Lambda)$ supposing throughout the discussion that $x_0 = \infty$. Since $1 - G(y_n) = 1 - \exp\{-e^{-y_n}\} \sim e^{-y_n}$ for $y_n \to \infty$ we have (2.23) equivalent to

$$e^{y_n}(1 - F^n(a_n y_n + b_n)) \sim e^{y_n} n(1 - F(a_n y_n + b_n) \to 1 \tag{2.36}$$

as $n \to \infty$ where we recall that $a_n y_n + b_n \to \infty$ so that $F(a_n y_n + b_n) \to 1$. From Proposition 1.4 we have for $z_0 < x < \infty$

$$1 - F(x) = c(x)\exp\left\{-\int_{z_0}^{x} (1/f(u))du\right\} \tag{1.5}$$

where $\lim_{x\uparrow x_0} c(x) = c > 0$, $\lim_{x\uparrow x_0} f'(x) = 0$, and we may take

$$b_n(1/(1-F))^{\leftarrow}(n), \qquad a_n = f(b_n).$$

Thus (2.36) can be expressed as

$$\begin{aligned}1 &\sim e^{y_n}(1 - F(a_n y_n + b_n))/(1 - F(b_n)) \\ &= e^{y_n}(c(a_n y_n + b_n)/c(b_n))\exp\left\{-\int_{b_n}^{a_n y_n + b_n} (1/f(u))du\right\}\end{aligned} \tag{2.37}$$

and making the change of variable in the integral $v = (u - b_n)/a_n y_n$ the preceding expression becomes

$$\begin{aligned}&e^{y_n}(1 + o(1))\exp\left\{-\int_0^1 y_n f(b_n)/f(b_n + v a_n y_n)dv\right\} \\ &\sim \exp\left\{-\int_0^1 ((f(b_n)/f(b_n + v a_n y_n)) - 1)y_n\, dv\right\}.\end{aligned}$$

Since $y_n \le A x_n$, if we set $\delta_n = y_n/x_n \in [0, A]$ we can write $y_n = \delta_n x_n$. If $\xi(x)$ is nondecreasing satisfying $\xi(\infty) = \infty$ and $\xi(b_n) = x_n$ then we see (2.36) may be rewritten as

$$\xi(b_n)\int_0^{\delta_n} (f(b_n)/f(b_n + v\xi(b_n)f(b_n)) - 1)dv \to 0,$$

and sufficient for this is

$$\xi(x)\left(\frac{f(x)}{f(x + \delta f(x)\xi(x))} - 1\right) \to 0 \tag{2.38}$$

as $x \to \infty$, locally uniformly in $\delta$. Finally we check that (2.38) holds if the de Haan and Hordijk (1972) condition holds, viz

$$\xi^2(x)f'(x) \to 0. \tag{2.39}$$

Note first that

$$\begin{aligned}\left|\frac{f(x + \delta f(x)\xi(x))}{f(x)} - 1\right| &\le \int_x^{x+\delta f(x)\xi(x)} \left|\frac{f'(u)}{f(x)}\right| du \\ &= \int_0^{\delta} |f'(x + uf(x)\xi(x))|\xi(x)du \\ &\le \int_0^{\delta} |f'(x + uf(x)\xi(x))\xi^2(x + uf(x)\xi(x))|\, du\end{aligned}$$

and this approaches zero locally uniformly in $\delta$ because of (2.38). To show that (2.39) implies (2.38) we repeat these steps as follows:

$$\xi(x)\left|\frac{f(x)}{f(x+\delta f(x)\xi(x))}-1\right| = \xi(x)\left|\frac{f(x+\delta f(x)\xi(x))-f(x)}{f(x+\delta f(x)\xi(x))}\right|$$

$$\sim \xi(x)\left|\frac{f(x+\delta f(x)\xi(x))-f(x)}{f(x)}\right| \qquad \text{(by the previous calculation)}$$

$$\le \int_0^\delta |f'(x+uf(x)\xi(x))|\xi^2(x)du$$

$$\le \int_0^\delta |f'(x+uf(x)\xi(x))\xi^2(x+uf(x)\xi(x))|du \to 0$$

locally uniformly in $\delta$ as required.

We now summarize our large deviation results when $F \in D(\Lambda)$.

**Proposition 2.10.** *Suppose $F \in D(\Lambda)$ with $x_0 = \infty$ and $1 - F$ has representation* (1.5). *Set $R(x) = \int_{z_0}^x (1/f(u))du$ for $x > z_0$. Then the large deviation property* (2.23) *holds* iff *there exists a nondecreasing function $\xi(t)$ satisfying $\xi(\infty) = \infty$, $\xi(b_n) = x_n$, and*

$$\lim_{x\to\infty} (R(x+\delta f(x)\xi(x)) - R(x) - \delta\xi(x)) = 0 \tag{2.40}$$

*locally uniformly in $\delta \in [0, \infty)$. Sufficient for* (2.40) *is the de Haan and Hordijk* (1972) *condition*

$$\lim_{x\to\infty} \xi^2(x)f'(x) = 0. \tag{2.39}$$

Proof. We need only check the equivalence of (2.23) with (2.40). Note that (2.23) and (2.37) are equivalent, so taking logarithms in the latter we obtain

$$R(a_n x_n \delta_n + b_n) - R(b_n) - \delta_n x_n \to 0 \tag{2.37'}$$

or, what is the same,

$$R(b_n + \delta f(b_n)\xi(b_n)) - R(b_n) - \delta\xi(b_n) \to 0 \tag{2.37''}$$

locally uniformly in $\delta$, and obviously (2.37″) is implied by (2.40), so we need to check the converse. Supposing (2.37″) we define

$$\xi(x) = x_n \qquad \text{for } x \in [b_n, b_{n+1}).$$

If $x \in [b_n, b_{n+1})$ we may write $x = b_n + \theta(b_{n+1} - b_n)$ for $0 \le \theta < 1$. We assume $F \in D(\Lambda)$ so that $V = (1/(1-F))^{\leftarrow} \in \Pi$ with auxiliary $a$-function $f \circ V$ (Propositions 0.9 and 0.10). Thus for any $\varepsilon > 0$

$$0 \le \limsup_{n\to\infty} (b_{n+1} - b_n)/f(b_n) = \limsup_{n\to\infty} (V(n+1) - V(n))/f \circ V(n)$$

$$\le \limsup_{n\to\infty} (V(n(1+\varepsilon)) - V(n))/f \circ V(n)$$

$$= \log(1+\varepsilon),$$

and hence we conclude

$$\lim_{n\to\infty} (b_{n+1} - b_n)/f(b_n) = 0.$$

Recall

$$\lim_{t\to\infty} f(t + \delta f(t))/f(t) = 1$$

locally uniformily in $\delta$ (Lemma 1.3) and therefore for $x \in [b_n, b_{n+1})$

$$\begin{aligned}\frac{f(x)}{f(b_n)} &= \frac{f(b_n + \theta(b_{n+1} - b_n))}{f(b_n)} \\ &= f\left(b_n + \left(\theta\frac{(b_{n+1} - b_n)}{f(b_n)}\right)f(b_n)\right)/f(b_n) \to 1 \end{aligned} \tag{2.41}$$

as $n \to \infty$. Since $\xi(x)$ is constant on $[b_n, b_{n+1})$ and $R$ is monotone nondecreasing

$$\begin{aligned}&R(x + \delta f(x)\xi(x)) - R(x) - \delta\xi(x) \\ &\geq R\left(b_n + \left(\delta\frac{f(x)}{f(b_n)}\right)f(b_n)\xi(b_n)\right) - R(b_{n+1}) - \delta\xi(b_n) \\ &= R\left(b_n + \left(\delta\frac{f(x)}{f(b_n)}\right)f(b_n)\xi(b_n)\right) - R(b_n) - \delta\xi(b_n) \\ &\quad - (R(b_{n+1}) - R(b_n)).\end{aligned}$$

From the definition of $R$

$$\begin{aligned}R(b_{n+1}) - R(b_n) = &-\log((1 - F(b_{n+1}))/(1 - F(b_n))) \\ &+ \log c(b_{n+1})/c(b_n) \to 0\end{aligned}$$

and we get on applying (2.41) and (2.37″)

$$\liminf_{x\to\infty} (R(x + \delta f(x)\xi(x)) - R(x) - \delta\xi(x)) \geq 0.$$

For an inequality in the reverse direction write for $x \in [b_n, b_{n+1})$

$$\begin{aligned}&R(x + \delta f(x)\xi(x)) - R(x) - \delta\xi(x) \\ &\leq R(b_n + \theta(b_{n+1} - b_n) + \left(\delta\frac{f(x)}{f(b_n)}\right)f(b_n)\xi(b_n)) - R(b_n) - \delta\xi(b_n) \\ &= R\left(b_n + \left(\frac{\theta(b_{n+1} - b_n)}{f(b_n)\xi(b_n)} + \delta\frac{f(x)}{f(b_n)}\right)f(b_n)\xi(b_n)\right) - R(b_n) - \delta\xi(b_n)\end{aligned}$$

and since

$$\frac{\theta(b_{n+1} - b_n)}{f(b_n)\xi(b_n)} + \delta\frac{f(x)}{f(b_n)} \to \delta$$

we get from (2.37″) that

$$\limsup_{x\to\infty} (R(x+\delta f(x)\xi(x)) - R(x) - \delta\xi(x) \le 0$$

and thus we have shown that

$$\lim_{x\to\infty} (R(x+\delta f(x)\xi(x)) - R(x) - \delta\xi(x)) = 0.$$

Checking local uniform convergence is no problem using continuous convergence (Section 0.1), and thus we are done. □

As an example consider the normal distribution $N(x)$ with density

$$n(x) = (2\pi)^{-1/2}e^{-x^2/2}.$$

Then

$$f(t) = (1 - N(t))/n(t) \sim t^{-1}$$

(Feller, 1968, page 175), and since

$$f'(t) = -1 + t(1 - N(t))/n(t)$$

and

$$f''(t) = -t + (1 + t^2)(1 - N(t))/n(t) \ge 0$$

we have $f'(t)$ nondecreasing and an application of Proposition 0.7(b) gives

$$f'(t) \sim -t^{-2}$$

and (2.39) becomes

$$t^{-2}\xi^2(t) \to 0;$$

i.e.,

$$\xi(t) = o(t).$$

Since $b_n \sim (2\log n)^{1/2}$ (Example 2, Section 1.5) we see that the large deviation property holds for

$$x_n = O(\xi(b_n)) = o((\log n)^{1/2}).$$

If Von Mises condition (1.21) holds, then from the proof of Proposition 1.17(a) we see that (2.39) can be formulated

$$\lim_{x\to\infty} \xi^2(x)\left(-1 + \frac{F'(x)\int_x^\infty (1 - F(u))du}{(1 - F(x))^2}\right) = 0$$

and if Von Mises condition (1.24) holds, condition (2.39) becomes

$$\lim_{x\to\infty} \xi^2(x)\left(1 + \frac{(1 - F(x))F''(x)}{(F'(x))^2}\right) = 0.$$

Rewriting (2.40) as

$$\lim_{x\to\infty} \xi(x)\left(\frac{R(x+\delta f(x)\xi(x)) - R(x)}{\xi(x)} - \delta\right) = 0$$

makes it clear that (2.40) describes a rate on how fast

$$\frac{R(x+\delta f(x)\xi(x)) - R(x)}{\xi(x)} \to \delta.$$

The latter condition suggests that

$$R = \pi \circ H$$

where $\pi \in \Pi$, $H \in \Gamma$, and under suitable growth conditions on $\xi(x)$ this should enable one to show that (2.39) is necessary. However this interesting conjecture has yet to be fully verified. See Exercise 2.3.8 for evidence in support of this conjecture.

## Exercises

2.3.1. When $F \in D(\Phi_\alpha)$, if (2.23) holds for a certain choice of normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$, then it holds for every choice; i.e., if (2.23) holds and $a_n \sim \alpha_n$, $b_n - \beta_n = o(a_n)$, then (2.23) holds with $(\alpha_n, \beta_n)$ replacing $(a_n, b_n)$.

2.3.2. Even if the Von Mises condition 1.19 fails, we may set

$$\alpha(t) = (1 - F(t)) \Big/ \int_t^\infty s^{-1}(1 - F(s))ds$$

so that $\alpha(t) \to \alpha$ and

$$1 - F(x) = \left(\alpha(x) \int_1^\infty s^{-1}(1 - F(s))ds\right) \exp\left\{-\int_1^x t^{-1}\alpha(t)dt\right\}.$$

Thus

$$\varepsilon(t) = \alpha(t) - \alpha$$

and condition (2.30) becomes

$$\log \xi(t)\left(\left((1 - F(t)) \Big/ \int_t^\infty s^{-1}(1 - F(s))ds\right) - \alpha\right) \to 0.$$

2.3.3. Derive large deviation results for $F \in D(\Psi_\alpha)$ either *ab* initio or by using the fact that if $F \in D(\Psi_\alpha)$,

$$F_0(x) = \begin{cases} 0 & x < 0 \\ F(x_0 - x^{-1}) & x > 0 \end{cases}$$

is in $D(\Phi_\alpha)$.

2.3.4. (a) If $\xi^*(t) = \log \xi(e^t)$ is the integral of a bounded derivative then 2.31 holds. If $(\xi^*)'(t)$ is nondecreasing, then (2.31) or (2.31′) implies $(\xi^*)'$ is bounded.
(b) Show that any power function of the form $\xi(t) = t^a$, $a > 0$, satisfies (2.31).

2.3.5. If $\xi_1$ grows so fast that

$$\int_1^\infty (1/\xi_1^*(u))du < \infty$$

in violation of (2.31), then (2.30) implies $L(\infty) < \infty$ and $L$ is $\xi_1$-ssv, and in fact $L$ is $\xi$-ssv for any $\xi$ nondecreasing with $\xi(\infty) = \infty$ (Anderson, 1984).

2.3.6. If $\xi_\beta^*(t) = t^\beta$, $\beta > 0$, $L^*(t) = \log_3 t$ then for $0 < \beta \leq 1$, (2.30) and (2.31) hold, and for $\beta > 1$ (2.30) and (2.31), do not hold, $\int_1^\infty (1/\xi_\beta^*(u))du < \infty$ but $L$ is $\xi_\beta$-ssv. So (2.30) can be too strong and the super slow variation property can hold without it (Anderson, 1984).

2.3.7. For which sequences $\{x_n\}$ does the large deviation property hold when $F$ is gamma, Weibull, lognormal?

2.3.8. (a) Observe from the proof of (2.38) that $f'(x)\xi(x) \to 0$ implies

$$f(x)/f(x + \delta f(x)\xi(x)) \to 1$$

locally uniformly. By symmetry, if we suppose $\xi$ is the integral of a density $\xi'$, then $\xi'(x)f(x) \to 0$ implies $\xi(x)/\xi(x + \delta f(x)\xi(x)) \to 1$ locally uniformly. If

$$\lim_{x\to\infty} f'(x)\xi^2(x) = c > 0$$

then show

$$\lim_{n\to\infty} P[a_n^{-1}(M_n - b_n) > \xi(b_n)]/(1 - \Lambda(\xi(b_n))) = e^c$$

(de Haan and Hordijk, 1972).
(b) If $\lim f'(x)\xi(x) = \lim_{x\to\infty} \xi'(x)f(x) = 0$ and (2.40) holds, then locally uniformly

$$\lim_{x\to\infty} \int_0^\delta \xi^2(x + sf(x)\xi(x))f'(x + sf(x)\xi(x))ds = 0.$$

If $f'$ does not change sign and $|f'|$ is nonincreasing

$$\lim_{x\to\infty} \xi^2(x)f'(x) = 0.$$

(c) If $\lim_{x\to\infty} f'(x)\xi(x) = \lim_{x\to\infty} \xi'(x)f(x) = 0$ then

$$H(x) := \exp\left\{\int_1^x (1/(f(u)\xi(u)))du\right\}$$

is in class $\Gamma$ with auxiliary function $f\xi$ and

$$\pi(x) := \int_1^x u^{-1}\xi(H^\leftarrow(u))du$$

is in class $\Pi$ with auxiliary $a$-function $\xi \circ H^\leftarrow$. Furthermore

$$R = \pi \circ H.$$

## 2.4. Uniform Rates of Convergence to Extreme Value Laws

We now survey some results which describe the rate of convergence of

$$d_n = \sup_{x \in \mathbb{R}} |F^n(a_n x + b_n) - G(x)|$$

to zero when $F \in D(G)$. As in the previous section, we are interested in how good an approximation $G(a_n^{-1}(x - b_n))$ is to $F^n(x)$, and $d_n$ is another way to measure the goodness of the approximation.

Such problems have been considered in the literature dating back to Fisher and Tippett (1928). We have found the following references very instructive: Anderson (1971), Davis (1982b), Hall (1979), Hall and Wellner (1979), Cohen (1982a, b), and Smith (1982). In particular, when $F \in D(\Phi_\alpha)$, Smith relates uniform rates of convergence to the concept of slow variation with remainder (Goldie and Smith, 1984).

In contrast to the approach based on slow variation with remainder, our method centers around the representation results for distributions $F \in D(G)$ (Proposition 1.4, Corollary 1.12, Propositions 1.15, 1.17, and 1.18). In this section, in the interest of usability, we are less than completely general and assume that $F(x)$ satisfies Von Mises conditions so that the convenient representations of $F \in D(G)$ exist as described in Propositions 1.15 and 1.18. We feel this is the appropriate level of generality, but in any event our methods can be generalized by making use of the more general representations for $F$. (cf. Cohen, 1982b). We concentrate on $\Phi_\alpha$ and $\Lambda$ and leave results for $\Psi_\alpha$ to the reader.

The representations described in Propositions 1.15 and 1.18 are in terms of $1 - F$. However, in this section, it is preferable to work with $-\log F$ because tighter bounds are then obtained. Since $-\log F(x) = c(x)(1 - F(x)), c(x) \to 1$, there is little involved in obtaining new representations. The cost of working with $-\log F$ results from the obvious fact that more common distributions are described in terms of $1 - F$ and not $-\log F$.

Some of the approaches discussed later arose from discussions with A. Balkema and L. de Haan.

### 2.4.1. Uniform Rates of Convergence to $\Phi_\alpha(x)$

Write $F = \exp\{-e^{-\phi}\}$ and suppose $F$ is differentiable. A Von Mises condition guaranteeing $F \in D(\Phi_\alpha)$ analogous to 1.19 is

$$h(x) := x\phi'(x) - \alpha = \frac{xF'(x)}{F(x)(-\log F(x))} - \alpha \to 0 \tag{2.42}$$

and we suppose there exists a nonincreasing continuous function $g$ and

$$|h(x)| \le g(x) \downarrow 0 \tag{2.43}$$

as $x \to \infty$. Typically we take $g(x) = \sup_{y \ge x} |h(y)|$. Set $\exp\{-\phi(a_n)\} = -\log F(a_n) = n^{-1}$ so that

$$\phi(a_n x) - \phi(a_n) = \int_1^x t^{-1}(\alpha + h(a_n t))dt \to \alpha \log x, \qquad x \to \infty,$$

for $x > 0$ showing that (2.42) is sufficient for $F \in D(\Phi_\alpha)$. The rate of convergence will be given in terms of $g$.

We begin with the following simple lemma:

**Lemma 2.11.** *For* $\alpha_1, \alpha_2 > 0$,

$$\sup_{x>0} |\Phi_{\alpha_1}(x) - \Phi_{\alpha_2}(x)| \le (.2701)|\alpha_1 - \alpha_2|/(\alpha_1 \wedge \alpha_2).$$

PROOF. Observe for $x > 0, \dfrac{\partial}{\partial \beta}\Phi_\beta(x) = \Phi_\beta(x)x^{-\beta}\log x$ and so assuming $\alpha_1 < \alpha_2$

$$\begin{aligned}\sup_{x>0} |\Phi_{\alpha_1}(x) - \Phi_{\alpha_2}(x)| &\le \sup_{x>0} \sup_{\beta \in [\alpha_1, \alpha_2]} |\Phi_\beta(x)x^{-\beta}\log x|\,|\alpha_2 - \alpha_1| \\ &= \sup_{y>0} \sup_{\beta \in [\alpha_1, \alpha_2]} |e^{-y}y \log y|\beta^{-1}|\alpha_2 - \alpha_1| \qquad (\text{set } y = x^{-\beta}) \\ &\le \left(\sup_{y>0} |e^{-y}y\log y|\right)\alpha_1^{-1}|\alpha_2 - \alpha_1|.\end{aligned}$$

The supremum is found numerically to four decimal places. □

It is now easy to obtain the rate of convergence on the interval $[1, \infty)$.

**Proposition 2.12.** *If* (2.42) *holds and* $\exp\{-\phi(a_n)\} = n^{-1}$ *then*

$$\sup_{x \ge 1} |F^n(a_n x) - \Phi_\alpha(x)| \le (.2701)(\alpha - g(a_n))^{-1}g(a_n) = O(g(a_n)).$$

PROOF. For $x \ge 1$

$$\begin{aligned}\phi(a_n x) - \phi(a_n) &= \int_1^x (\alpha + h(a_n t))t^{-1}dt \le \int_1^x (\alpha + g(a_n t))t^{-1}dt \\ &\le (\alpha + g(a_n))\log x.\end{aligned}$$

Obtaining a lower bound in a similar way we finally get

$$(\alpha - g(a_n))\log x \le \phi(a_n x) - \phi(a_n) \le (\alpha + g(a_n))\log x$$

and taking negative exponentials twice gives for $x \ge 1$

$$\Phi_{(\alpha - g(a_n))}(x) \le F^n(a_n x) \le \Phi_{(\alpha + g(a_n))}(x) \tag{2.44}$$

and an application of Lemma 2.1 gives the desired result. □

On the region $(-\infty, 1)$, more care must be taken. We first present a method which works quite generally, and then we show that if more is assumed about $g$ a better bound can be obtained.

If $x < 1$ we find in the same way as (2.44) was obtained that

$$\Phi_{(\alpha+g(a_n x))}(x) \leq F^n(a_n x) \leq \Phi_{(\alpha-g(a_n x))}(x) \tag{2.45}$$

and hence $|F^n(a_n x) - \Phi_\alpha(x)| \leq (.2701)(\alpha - g(a_n x))^{-1} g(a_n x)$ from Lemma 2.11. Now suppose $\{x_n\}$ is a sequence (to be specified) satisfying $x_n \to 0$ and $a_n x_n \to \infty$. Then we conclude that

$$\sup_{x_n \leq x < \infty} |F^n(a_n x) - \Phi_\alpha(x)| \leq (.2701)(\alpha - g(a_n x_n))^{-1} g(a_n x_n).$$

For $x \leq x_n$ observe from (2.45) that

$$F^n(a_n x) \leq F^n(a_n x_n) \leq \Phi_{(\alpha-g(a_n x_n))}(x_n)$$

and for $x_n \leq 1$

$$\Phi_\alpha(x_n) \leq \Phi_{(\alpha-g(a_n x_n))}(x_n)$$

and so the uniform bound becomes

$$\sup_x |F^n(a_n x) - \Phi_\alpha(x)| \leq (.2701)(\alpha - g(a_n x_n))^{-1} g(a_n x_n) \vee \Phi_{(\alpha-g(a_n x_n))}(x_n). \tag{2.46}$$

The way to choose $x_n$ so that the right side of (2.46) is minimized is to pick $x_n$ to satisfy

$$(.2701)(\alpha - g(a_n x_n))^{-1} g(a_n x_n) = \Phi_{(\alpha-g(a_n x_n))}(x_n)$$

or equivalently

$$a_n x_n(-\log((.2701)(\alpha - g(a_n x_n))^{-1} g(a_n x_n)))^{(\alpha-g(a_n x_n))^{-1}} = a_n.$$

To get an expression for $x_n$ it is convenient to switch to a continuous variable. Define

$$a(t) = \left(\frac{1}{-\log F}\right)^{\leftarrow}(t) = \inf\{u\colon 1/(-\log F(u)) \geq t\}$$

and define a nondecreasing function $\rho(t)$ by

$$\rho(a(t)) = a(t)x(t)$$

where $x(t)$ is an unknown function decreasing to zero, while $a(t)x(t)$ increases to $\infty$. Then we have

$$\rho(a(t))\{-\log(.2701)(\alpha - g(\rho(a(t))))^{-1} g(\rho(a(t)))\}^{(\alpha-g(\rho(a(t))))^{-1}} = a(t).$$

Change variables replacing $a(t)$ by $t$. If we let $\rho^{\leftarrow}(t)$ be the inverse of $\rho$ we obtain

$$\rho^{\leftarrow}(t) = t(-\log((.2701)(\alpha - g(t))^{-1} g(t)))^{(\alpha-g(t))^{-1}}. \tag{2.47}$$

It may be difficult to invert this expression, but an asymptotic inversion can usually be performed. Note that it is clear that

$$\lim_{t\to\infty} \uparrow \rho^{\leftarrow}(t)/t = \infty$$

and hence (Proposition 0.8(vi))

$$\lim_{t\to\infty} \downarrow \rho(t)/t = 0$$

so from the definition of $\rho$ we get $x(t) = \rho(a(t))/a(t) \downarrow 0$. Since $\rho(t) \uparrow \infty$ we get $a(t)x(t) \uparrow \infty$ and so $x(t)$ has the desired properties.

We now summarize our findings.

**Proposition 2.13.** *Suppose* (2.42) *holds so that* $F \in D(\Phi_\alpha)$. *Then*

$$\sup_{x\in\mathbb{R}} |F^n(a_n x) - \Phi_\alpha(x)| \le (.2701)(\alpha - g(\rho(a_n)))^{-1} g(\rho(a_n))$$

*where* $\rho$ *is given in terms of its inverse by* (2.47).

*Remark.* One sometimes gets the impression from results in the literature that $O(n^{-1})$ is the best convergence rate possible. This is not the case. Since the function $g$ can be any function converging monotonically to zero, a wide variety of convergence rates are to be expected. For example, suppose $g(t) = e^{-t}$ and we define a distribution $F$ by

$$F(x) = \begin{cases} 0, & x < 1 \\ \exp\{-\exp\{-\int_1^x (1 + e^{-u})u^{-1}du\}\}, & x \ge 1. \end{cases}$$

Then $\rho^{\leftarrow}(t) \sim t(-\log(t))^{(1-g(t))^{-1}} \sim t^2$ so that $\rho(t) \sim t^{1/2}$. Since $\log t = \int_1^{a(t)} (1 + e^{-u})u^{-1}du$ we have $\log a^{\leftarrow}(t) = \int_1^t (1 + e^{-u})u^{-1}du = \log t + c + o(1)$, and thus $a^{\leftarrow}(t) = te^c(1 + o(1))$ and $a(t) = te^{-c}(1 + o(1))$. So an order of convergence $g(\rho(a_n))$ is of the form $\exp\{-kn^{-1/2}\}$, $k > 0$.

It is an interesting fact that, as shown by Rootzen (1984), the convergence rate cannot be faster than exponential without $F$ actually being an extreme value distribution.

The convergence rate on $[1, \infty)$ is $g(a_n)$, but the preceding technique gives the overall rate $g(\rho(a_n))$. When $g$ satisfies growth conditions of regular variation type it is possible to improve the bound from $O(g(\rho(a_n)))$ to $O(g(a_n))$ as is done in Smith (1982). Indeed when $g$ is regularly varying with index $-\beta < 0$ we have

$$\lim_{n\to\infty} \frac{g(\rho(a_n))}{g(a_n))} = \lim_{n\to\infty} \frac{g(a_n(\rho(a_n)/a_n))}{g(a_n)} = \lim_{n\to\infty} (\rho(a_n)/a_n)^{-\beta} = \infty$$

(recall $\rho(t)/t \to 0$ as $t \to \infty$) so that $O(g(a_n))$ is a significant improvement over $O(g(\rho(a_n)))$.

We show that $O(g(a_n))$ is a valid convergence rate under the following assumption: Pick $\delta$ so large that $g(\delta) < \alpha$. Then for $n$ such that $a_n^{-1}\delta < 1$ we assume

$$\frac{g(a_n x)}{g(a_n)} \le x^{-\beta}, \qquad a_n^{-1}\delta \le x \le 1, \tag{2.48}$$

where $\beta > 0$.

*Remark.* There are two circumstances in which (2.48) is easy to verify. The first is when $g$ is of the form $g(x) = x^{-\beta}$, which occurs in several common cases. The other situation is when $g$ is regularly varying, is differentiable, and satisfies $tg'(t)/g(t) \to -\hat{\beta}$, as $t \to \infty$.

We need a variant of Lemma 2.11.

**Lemma 2.14.** *Suppose* (2.48) *holds and* $\delta > 0$ *is chosen so that* $g(\delta) < \alpha$. *We have for n such that* $a_n^{-1}\delta < 1$

$$\sup_{a_n^{-1}\delta \le x \le 1} (|\Phi_{(\alpha - g(a_n x))}(x) - \Phi_\alpha(x)| \vee |\Phi_{(\alpha + g(a_n x))}(x) - \Phi_\alpha(x)| \le c(\alpha, \beta, \delta) g(a_n) \tag{2.49}$$

where

$$c(\alpha, \beta, \delta) = \beta^{-1}\theta \sup_{s \ge 1} \{s^{1+\theta}(\log s)e^{-s}\}$$

and

$$\theta = \beta/(\alpha - g(\delta)).$$

Proof. Since $\Phi_\gamma(x) = \Lambda(\gamma \log x)$ for $x > 0$ and $\Lambda'(y)$ is increasing for $y < 0$ we have

$$\begin{aligned}
\sup_{a_n^{-1}\delta \le x \le 1} |\Phi_\alpha(x) - \Phi_{(\alpha - g(a_n x))}(x)| &\le \sup_{a_n^{-1}\delta \le x \le 1} \int_{\alpha - g(a_n x)}^{\alpha} \Lambda'(\gamma \log x)|\log x|\, d\gamma \\
&\le \sup_{a_n^{-1}\delta \le x \le 1} \Lambda'((\alpha - g(a_n x))\log x)|\log x| g(a_n x) \\
&\le \sup_{a_n^{-1}\delta \le x \le 1} \Lambda'((\alpha - g(\delta))\log x)|\log x| g(a_n x) \\
&\le \sup_{0 < x \le 1} \Lambda'((\alpha - g(\delta))\log x)|\log x| x^{-\beta} g(a_n) \\
&= \sup_{0 < y < 1} \{y^{-1} e^{-y^{-1}} |\log y| y^{-\theta}\theta\beta^{-1}\} g(a_n) \\
&= c(\alpha, \beta, \delta) g(a_n).
\end{aligned}$$

The bound for $|\Phi_{\alpha + g(a_n x)}(x) - \Phi_\alpha(x)|$ is dominated by the one just presented, so we are done. □

*Remark.* The constant $c(\alpha, \beta, \delta)$ must be computed numerically once $g$, $\delta$, and $\beta$ are specified. For example:

| $\theta$ | .25 | .5 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| $\sup_{s \ge 1} \{s^{1+\theta}(\log s)e^{-s}\}$ | .2372 | .2976 | .4928 | 1.6392 | 6.8703 | 35.058 |

**Proposition 2.15.** *Suppose that* (2.48) *holds and* $\delta > 0$ *is chosen so large that* $g(\delta) < \alpha$. *Then for n such that* $a_n^{-1}\delta < 1$

$$\sup_{x \in \mathbb{R}} |F^n(a_n x) - \Phi_\alpha(x)| \le (.2701)(\alpha - g(a_n))^{-1} g(a_n) \vee c(\alpha, \beta, \delta) g(a_n)$$
$$\vee F^n(\delta) \vee \Phi_\alpha(a_n^{-1}\delta) = O(g(a_n)).$$

PROOF. Proposition 2.12 is still applicable. From Equation (2.45) and Lemma 2.14 we have

$$\sup_{a_n^{-1}\delta \le x \le 1} |F^n(a_n x) - \Phi_\alpha(x)| \le c(\alpha, \beta, \delta) g(a_n).$$

Finally

$$\sup_{x \le a_n^{-1}\delta} |F^n(a_n x) - \Phi_\alpha(x)| \le F^n(\delta) \vee \Phi_\alpha(a_n^{-1}\delta). \qquad \square$$

For many distributions it is convenient to work with $1 - F$ rather than $-\log F$. In cases in which the convergence rate is slower than $1/n$, the following result is useful. Set $\bar{F} = 1 - F$.

**Proposition 2.16.** *Suppose* (2.42) *holds and set*

$$B(x) = \sum_{k=1}^{\infty} \frac{\bar{F}^k(x)}{k(k+1)} \qquad \textit{so that}$$

$$\bar{F}(x)/2 \le B(x) \le \tfrac{1}{2}\bar{F}(x)(1 + \bar{F}(x))$$

*and* $B(x) \sim \frac{1}{2}\bar{F}(x)$ *as* $x \to \infty$. *Then* $-F \log F = \bar{F}(1 - B)$ *and hence*

$$\frac{xF'(x)}{F(x)(-\log F(x))} - \alpha = \frac{xF'(x)}{\bar{F}(x)} - \alpha + \frac{xF'(x)B(x)}{\bar{F}(x)(1 - B(x))} \tag{2.50}$$

*so that*

$$\frac{xF'(x)}{F(x)(-\log F(x))} - \alpha = \frac{xF'(x)}{\bar{F}(x)} - \alpha + c(x)\bar{F}(x) \tag{2.51}$$

*where* $c(x) \to \alpha/2$.

Unless $(xF'(x)/\bar{F}(x)) - \alpha$ goes to zero more slowly than $\bar{F}(x)$, the use of this formula will lead to a convergence rate of $O(n^{-1})$. This will be the case, for instance, with the Cauchy distribution. Cf. Exercise 3.4.3.

EXAMPLE (Cf. Smith, 1982, Example 1). Suppose for $x \ge 1$

$$\bar{F}(x) = cx^{-\alpha} + dx^{-\beta-\alpha}$$

where $c > 0, d > 0, 0 < \beta < \alpha, c + d = 1$. We find

$$\frac{xF'(x)}{\bar{F}(x)} = \alpha\left(\frac{1 + d(\alpha + \beta)c^{-1}\alpha^{-1}x^{-\beta}}{1 + dc^{-1}x^{-\beta}}\right)$$

and so

$$\frac{xF'(x)}{\bar{F}(x)} - \alpha = \frac{d\beta c^{-1}x^{-\beta}}{1 + dc^{-1}x^{-\beta}}$$

which implies

$$\left|\frac{xF'(x)}{\bar{F}(x)} - \alpha\right| \le c_1 x^{-\beta}$$

where $c_1 = d\beta c^{-1}$. Set $c_2 = \alpha + d(\alpha + \beta)c^{-1}$ and for $x > 1$ we have

$$\left|\frac{xF'(x)B(x)}{\bar{F}(x)(1 - B(x))}\right| \le c_2 \frac{B(x)}{1 - B(x)} \le \frac{c_2 \frac{1}{2}\bar{F}(x)(1 + \bar{F}(x))}{1 - \frac{1}{2}\bar{F}(x)(1 + \bar{F}(x))}$$

$$\le c_2 \bar{F}(x)/(1 - \bar{F}(x)) \le 2c_2 \bar{F}(x)$$

provided $\bar{F}(x) < \frac{1}{2}$, as will be the case if $x \ge x_0$ and $x_0 = (4c)^{1/\alpha} \vee (4d)^{1/(\alpha+\beta)} \vee 1$ is a suitable and convenient choice.

According to (2.50) we have for $x \ge x_0$

$$\left|\frac{xF'(x)}{F(x)(-\log F(x))} - \alpha\right| \le x^{-\beta}[c_1 + 2c_2(cx^{-(\alpha-\beta)} + dx^{-\alpha})] \le kx^{-\beta} =: g(x)$$

where $k = c_1 + 2c_2(c + d) = c_1 + 2c_2$.

We do not find $a_n$ but instead compute $\alpha_n \le a_n$, which will be more convenient but still give a valid bound $O(g(\alpha_n))$. Recall that $a_n$ is the solution of $-\log F(x) = n^{-1}$. Let $a'_n$ be the solution of $\bar{F}(x) = n^{-1}$. Since $-\log F \ge \bar{F}$ we have $a_n \ge a'_n$. Also $\bar{F}(x) \ge cx^{-\alpha}$ so if we set $\alpha_n = (cn)^{1/\alpha}$ we have $\alpha'_n \le a_n$ and also $\alpha_n \sim a'_n \sim a_n$ as $n \to \infty$.

If we pick $\delta \ge x_0$ we then have for all $n$ such that $\alpha_n > \delta$ (i.e., $n > c^{-1}\delta^{\alpha}$)

$$\sup_x |F^n(a_n x) - \Phi_\alpha(x)|$$

$$\le (.2701)(\alpha - g(\alpha_n))^{-1} g(\alpha_n) \vee c(\alpha, \beta, \delta) g(\alpha_n) \vee F^n(\delta) \vee \Phi_\alpha(\alpha_n^{-1}\delta)$$

where $g(x) = kx^{-\beta}$, $\alpha_n = (cn)^{1/\alpha}$. The order of convergence is $O(n^{-\beta/\alpha})$.

To get a better feel for the method, suppose $\alpha = 1/2$, $\beta = 1/4$, $c = 3/4$, $d = 1/4$. Then we find $x_0 = 9$, $c_1 = .0833$, $c_2 = .75$, $k = 1.5833$. We pick $\delta$ to give a reasonable value for $\theta$ and hence for $c(\alpha, \beta, \delta)$. If $\theta = 1$ then $\delta = 1608.9012$ and $c(\alpha, \beta, \delta) = 1.9712$. The condition $a_n \ge \delta$ requires $n \ge 54$, and on this range the dominant term in the bound is $c(\alpha, \beta, \delta) g(\alpha_n)$, showing the dependence of the bound on $c(\alpha, \beta, \delta)$. Some values for the bounds are given in the following table to four decimal places.

| $n$ | $(.2701)(\alpha - g(\alpha_n))^{-1} g(\alpha_n)$ | $c(\alpha, \beta, \delta) g(\alpha_n)$ | $F^n(\delta)$ | $\Phi_\alpha\left(\frac{\delta}{\alpha_n}\right)$ |
|---|---|---|---|---|
| 54 | .2675 | .4904 | .0393 | .0000 |
| 75 | .1974 | .4161 | .0112 | .0000 |
| 100 | .1557 | .3604 | .0024 | .0000 |
| 150 | .1150 | .2943 | .0000 | .0000 |
| 300 | .0723 | .2081 | .0000 | .0000 |
| 500 | .0528 | .1612 | .0000 | .0000 |
| 1,000 | .0353 | .1140 | .0000 | .0000 |
| 5,000 | .0148 | .0510 | .0000 | .0000 |
| 10,000 | .0102 | .0360 | .0000 | .0000 |

### 2.4.2. Uniform Rates of Convergence to $\Lambda(x)$

Again write $F = \exp\{-e^{-\phi}\}$ and suppose that $F$ is twice differentiable. The Von Mises condition analogous to (1.24) guaranteeing $F \in D(\Lambda)$ is

$$h(x) = (1/\phi'(x))' = -\log F(x) - \left\{\frac{-F(x)F''(x)\log F(x)}{(F'(x))^2} + 1\right\} \to 0 \quad (2.52)$$

as $x \to x_0 := \sup\{y\colon F(y) < 1\}$. There exists a nonincreasing function $g$ with

$$|h(x)| \leq g(x) \downarrow 0 \quad (2.53)$$

as $x \to x_0$. Set $f(x) = 1/\phi'(x)$ and define $b_n$ by $F(b_n) = \exp\{-n^{-1}\}$ so $\phi(b_n) = \log n$ and define $a_n$ by $a_n = f(b_n) = n^{-1}F(b_n)/F'(b_n)$. Observe that

$$\phi(a_n x + b_n) - \phi(b_n) = \int_0^x \frac{f(b_n)}{f(a_n v + b_n)}\,dv.$$

Since $f' = h \to 0$ we have by Lemma 1.3 that

$$\frac{f(b_n + f(b_n)v)}{f(b_n)} \to 1$$

locally uniformly in $v$ as $n \to \infty$, and thus we see that

$$\phi(a_n x + b_n) - \phi(b_n) \to x$$

as $n \to \infty$, showing $F \in D(\Lambda)$. The function $g$ will again yield the rate of convergence. The following estimates are basic to our approach.

**Proposition 2.17.** *For a positive real number $g$ define the distribution functions*

$$F(g, x) = \begin{cases} 0 & \text{if } x < -g^{-1} \\ \exp\{-(1 + gx)^{-g^{-1}}\} & \text{if } x > -g^{-1} \end{cases}$$

$$F(-g, x) = \begin{cases} \exp\{-(1 - gx)^{g^{-1}}\} & \text{if } x < g^{-1} \\ 1 & \text{if } x > g^{-1}. \end{cases}$$

*Then for* $0 < g < 1$

$$\sup_{x \in R} |F(\pm g, x) - \Lambda(x)| \leq e^{-1}g \approx .3679g.$$

*Remark.* It is possible the constant $e^{-1}$ can be improved by using techniques of Hall and Wellner (1979). Hall and Wellner have proved

$$\sup_{x \geq 0} |(1 - n^{-1}x)^n 1_{[0,n]}(x) - e^{-x}| \leq (2 + n^{-1})e^{-2}n^{-1} = 0(n^{-1}).$$

Proof. The method follows Ailam (1968) and Hall and Wellner (1979). We consider only the bound on $F(-g, x) - \Lambda(x)$, the other case being similar. We have

$$\sup_{x\in\mathbb{R}} |F(-g,x) - \Lambda(x)| = \sup_{x<0} |F - \Lambda| \vee \sup_{0\le x\le g^{-1}} |F - \Lambda| \vee \sup_{x>g^{-1}} |F - \Lambda|$$
$$= A \vee B \vee C.$$

Now

$$C = \sup_{x\ge g^{-1}} |1 - e^{-e^{-x}}| = 1 - \exp\{-e^{-g^{-1}}\} \le e^{-g^{-1}}.$$

Note that $e^{-g^{-1}} \le e^{-1}g$ for $0 < g < 1$ since $xe^{-x}$ is decreasing on $[1, \infty)$. For $A$ we have

$$A = \sup_{x<0} |\exp\{-(1 - gx)^{g^{-1}}\} - \exp\{-e^{-x}\}|$$
$$= \sup_{y>0} (\exp\{-(1 + y)^{g^{-1}}\} - \exp\{-e^{y/g}\}) = \sup_{y>0} q(y).$$

Check that the supremum of $q(y)$ can be found by solving $q'(y) = 0$ for the nonzero root. Since $q'(y) = 0$ gives

$$\exp\{-(1 + y)^{g^{-1}}\}(1 + y)^{g^{-1}-1} = \exp\{-e^{y/g}\}e^{y/g}$$

we have

$$A \le \sup_{y>0} (\exp\{-e^{y/g}\}e^{y/g}(1 + y)^{1-g^{-1}} - \exp\{-e^{y/g}\})$$
$$= \sup_{y>0} (e^{y/g}\exp\{-e^{y/g}\}((1 + y)^{1-g^{-1}} - e^{-y/g}))$$
$$\le e^{-1} \sup_{y>0} ((1 + y)^{1-g^{-1}} - e^{-y/g}) = e^{-1} \sup_{y>0} \bar{q}(y)$$

since $\sup_{y>0} ye^{-y} = e^{-1}$. Again check that $\sup \bar{q}(y)$ is achieved at the nonzero root of $\bar{q}'(y) = 0$. The equation $\bar{q}'(y) = 0$ yields

$$e^{-y/g} = (1 - g)(1 + y)^{-g^{-1}}$$

so

$$\sup_{y>0} \bar{q}(y) \le \sup_{y>0} ((1 + y)^{1-g^{-1}} - (1 - g)(1 + y)^{-g^{-1}})$$
$$= \sup_{y>0} ((1 + y)^{-g^{-1}}(y + g)) = g$$

since the supremum is achieved at $y = 0$. The bound for $B$ is obtained by a similar but simpler argument than the one used on $A$ and is omitted. □

On the region $[0, \infty)$ we have the following result.

**Proposition 2.18.** *If* (2.52) *and* (2.53) *hold and* $a_n$ *and* $b_n$ *are as specified after* (2.53) *then*

$$\sup_{x\ge 0} |F^n(a_n x + b_n) - \Lambda(x)| \le e^{-1} g(b_n).$$

PROOF. Recalling that $f = 1/\phi'$ we have for $v > 0$

$$\left|\frac{f(b_n + a_n v) - f(b_n)}{f(b_n)}\right| \leq \int_{b_n}^{b_n + a_n v} \frac{|f'(u)|du}{f(b_n)} \leq \frac{g(b_n)a_n v}{f(b_n)} = g(b_n)v.$$

Therefore for $v > 0$

$$1 - g(b_n)v \leq \frac{f(b_n + a_n v)}{f(b_n)} \leq 1 + g(b_n)v$$

and taking reciprocals we have, assuming that $g(b_n)v < 1$, that

$$\frac{1}{1 + g(b_n)v} \leq \frac{f(b_n)}{f(b_n + a_n v)} \leq \frac{1}{1 - g(b_n)v}.$$

For $x$ such that $x > 0$ and $g(b_n)x < 1$ we get by integrating

$$\begin{aligned} -\log(-\log F(g(b_n), x)) &\leq \phi(a_n x + b_n) - \phi(b_n) \\ &\leq -\log(-\log F(-g(b_n), x)). \end{aligned}$$

Taking negative exponentials twice the following is true for $x > 0$:

$$F(g(b_n), x) \leq F^n(a_n x + b_n) \leq F(-g(b_n), x). \tag{2.54}$$

The desired result follows by means of Proposition 2.17. □

We now obtain a bound on the region $(-\infty, 0)$ which will be generally applicable. For $x < 0$, the analogue of (2.54) is

$$F(g(a_n x + b_n), x) \leq F^n(a_n x + b_n) \leq F(-g(a_n x + b_n), x) \tag{2.55}$$

and so

$$|F^n(a_n x + b_n) - \Lambda(x)| \leq e^{-1} g(a_n x + b_n) \tag{2.56}$$

by an appeal to Proposition 2.17. Let $\{x_n\}$ satisfy

$$x_n \downarrow -\infty \text{ and } a_n x_n + b_n \to \infty. \tag{2.57}$$

Combining (2.56) and Proposition 2.18 gives

$$\sup_{x \geq x_n} |F^n(a_n x + b_n) - \Lambda(x)| \leq e^{-1} g(a_n x_n + b_n),$$

and using (2.55) we obtain

$$\begin{aligned} &\sup_x |F^n(a_n x + b_n) - \Lambda(x)| \\ &\quad \leq e^{-1} g(a_n x_n + b_n) \vee F(-g(a_n x_n + b_n), x_n) \vee \Lambda(x_n). \end{aligned}$$

It is easy to check that

$$F(-g(a_n x_n + b_n), x_n) \geq \Lambda(x_n)$$

so that the uniform bound becomes

$$e^{-1} g(a_n x_n + b_n) \vee F(-g(a_n x_n + b_n), x_n).$$

At this point we see the bound is minimized if we pick $\{x_n\}$ to satisfy

$$e^{-1}g(a_n x_n + b_n) = F(-g(a_n x_n + b_n), x_n). \tag{2.58}$$

It is necessary to check whether this choice of $\{x_n\}$ satisfies (2.57). Suppose to get a contradiction, that $\{x_n\}$ does not converge to $-\infty$ so that for a subsequence $\{n'\}$ and a number $K$, $x_{n'} \geq K$. Then $a_{n'}x_{n'} + b_{n'} \to \infty$ and the left side of (2.58) converges to zero as $n' \to \infty$. However, the right side is of the order of $\Lambda(x_{n'})$, which does not converge to 0, and this gives the desired contradiction. Next suppose $a_n x_n + b_n$ does not converge to $\infty$ so that for a subsequence $\{n'\}$ and $M < \infty$ we have $a_{n'}x_{n'} + b_{n'} \leq M$. Then $g(a_{n'}x_{n'} + b_{n'}) \geq g(M) > 0$ and $g(a_{n'}x_{n'} + b_{n'})x_{n'} \to -\infty$. So the left side of (2.58) does not converge to zero but the right side does, again giving a contradiction.

We summarize our findings.

**Proposition 2.19.** *Suppose* (2.52) *and* (2.53) *hold so that* $F \in D(\Lambda)$ *and suppose* $a_n$, $b_n$ *are chosen as specified after* (2.53). *Then with* $\{x_n\}$ *chosen as in* (2.58) *we have*

$$\sup_x |F^n(a_n x + b_n) - \Lambda(x)| \leq e^{-1}g(a_n x_n + b_n).$$

This bound may be compared with the more attractive bound $e^{-1}g(b_n)$ valid for $x \in [0, \infty)$. When $g$ satisfies conditions of regular variation type we may extend the bound $g(b_n)$ to cover all $x \in \mathbb{R}$. If $|f'|$ is regularly varying then $|f'(x)| \sim g(x) = \sup_{y \geq x}|f'(y)|$ as $x \to x_0$ (cf. Exercise 0.4.2.11), and so by Karamata's theorem we hope

$$\frac{f(x)}{xg(x)} \to \text{constant}$$

as $x \to x_0$. So with the regular variation case in mind we assume there exists $k \in (0, \infty)$ such that for $n \geq n_0$

$$\frac{f(b_n)}{b_n g(b_n)} \leq k \tag{2.59}$$

and for $c < k^{-1}$ and $n \geq n_0$, $\beta > 0$, $\gamma > 0$

$$g(b_n(1 - ck)) \leq \gamma(1 - ck)^{-\beta}g(b_n). \tag{2.60}$$

**Proposition 2.20.** *If* (2.59) *and* (2.60) *hold then for* $n \geq n_0$

$$\sup_{x \in \mathbb{R}} |F^n(a_n x + b_n) - \Lambda(x)|$$

$$\leq F^n\left(a_n\left(\frac{c}{g(b_n)}\right) + b_n\right) \vee \Lambda\left(\frac{c}{g(b_n)}\right) \vee (\gamma(1 - kc)^{-\beta} \vee 1)e^{-1}g(b_n) = O(g(b_n)).$$

PROOF. From (2.56)

$$\sup_{-c/g(b_n) \leq x \leq 0} |F^n(a_n x + b_n) - \Lambda(x)| \leq \sup_{-c/g(b_n) \leq x \leq 0} e^{-1}g(a_n x + b_n)$$

and by monotonicity of $g$ this is

$$\leq e^{-1}g(a_n(-c/g(b_n)) + b_n) = e^{-1}g\left(b_n\left(1 - \frac{cf(b_n)}{b_n g(b_n)}\right)\right)$$

and using (2.59) and the fact that $g$ is nonincreasing we have the preceding bounded by

$$\leq e^{-1}g(b_n(1 - ck))$$

and from (2.60) this is

$$\leq \gamma e^{-1}g(b_n)(1 - ck)^{-\beta}.$$

Combining this with Proposition 2.18 gives the result. □

*Remarks.* From (2.56) and (2.60) we see that

$$F^n(a_n(-c/g(b_n)) + b_n) \leq \Lambda(-c/g(b_n)) + \gamma e^{-1}g(b_n)(1 - ck)^{-\beta}$$

for $n \geq n_0$. Hence the order of the bound in Proposition 2.20 is $O(g(b_n))$. If $g$ is regularly varying, we have that $g(b_n)$ is a slowly varying function of $n$ since $b_n$ is slowly varying (cf. Proposition 0.8(iv)) and hence the bound converges to zero at a slow rate.

As in Section 2.4.1, we often prefer to work with $\bar{F} := 1 - F$ rather than $-\log F$. Recall $B(x) = \sum_{k=1}^{\infty} \bar{F}^k(x)/(k(k+1))$ and $B(x) \sim \bar{F}(x)/2$.

**Proposition 2.21.** *Set $\rho(x) = \bar{F}(x)/F'(x)$. Then since $f = 1/\phi' = \rho(1 - B)$*

$$\left(\frac{1}{\phi'(x)}\right)' = \rho'(x)(1 - B(x)) + \sum_{k=1}^{\infty} \bar{F}^k(x)/(k+1)$$

$$= \rho'(x)c_1(x) + \frac{1}{2}\bar{F}(x)(1 + c_2(x)) \tag{2.61}$$

*where $c_1(x) \to 1$, $c_2(x) \to 0$.*

*Note*

$$\sum_1^{\infty} \bar{F}^k/(k+1) = \frac{1}{2}\bar{F} + \bar{F}\sum_{k=2}^{\infty} \bar{F}^{k-1}/(k+1)$$

$$\leq \frac{1}{2}\bar{F} + \frac{1}{3}\bar{F}(\bar{F}/F) \leq \bar{F} \tag{2.62}$$

*provided $\bar{F} \leq 3/5$.*

Example (Weibull Distribution). Suppose for $x > 0$, $\beta > 0$, $\beta \neq 1$

$$\bar{F}(x) = \exp\{-x^{\beta}\}.$$

Then $F'(x) = \bar{F}(x)\beta x^{\beta-1}$ and

$$\rho(x) = \frac{\bar{F}(x)}{F'(x)} = \frac{\bar{F}(x)}{\bar{F}(x)\beta x^{\beta-1}} = \beta^{-1}x^{-(\beta-1)}$$

$$\rho'(x) = -\beta^{-1}(\beta-1)x^{-\beta}.$$

Using (2.61) and (2.62) gives

$$\left|\left(\frac{1}{\phi'(x)}\right)'\right| \le |\beta-1|\beta^{-1}x^{-\beta} + e^{-x^\beta} =: g(x)$$

for $x$ such that $\bar{F}(x) \le 3/5$; i.e., $x \ge (\log 5/3)^{1/\beta}$. We have $f(x) = \beta^{-1}x^{-(\beta-1)}(1 - B(x))$ and so

$$\frac{f(x)}{xg(x)} \le \frac{\beta^{-1}x^{-(\beta-1)}}{x|\beta-1|\beta^{-1}x^{-\beta}} = \frac{1}{|\beta-1|} =: k.$$

For $s \ge s_0 \ge 1$ (where $s_0$ is the solution of $s_0^\beta e^{-s_0^\beta} = \delta$) we have $\exp\{-s^\beta\} \le \delta s^{-\beta}$ and therefore

$$g(x) \le (|\beta-1|\beta^{-1} + \delta)x^{-\beta}.$$

So for $x \ge (s_0 \vee (\log 5/3)^{1/\beta})/(1-c)$

$$\frac{g(x(1-c))}{g(x)} \le \left(\frac{\delta + \beta^{-1}|\beta-1|}{\beta^{-1}|\beta-1|}\right)(1-c)^{-\beta}$$

and so we get $\gamma = (\delta + \beta^{-1}|\beta-1|)/(\beta^{-1}|\beta-1|)$.

For concreteness, suppose $\beta = 2$. Then $k = 1$, $b_n = (-\log(1 - e^{-n^{-1}}))^{1/2}$. $a_n = f(b_n) \ge \frac{1}{2}b_n^{-1}(1 - \frac{1}{2}e^{-b_n^2}(1 + e^{-b_n^2})) =: \alpha_n$. A moderate value of $c$ must be chosen, otherwise the very slow decrease of $g(b_n)$ will prevent $\Lambda(-c/g(b_n))$ from being small for reasonable sample sizes. We choose $c = .1$, $s_0 = 1.75$ so that $\delta = .1432$, $\gamma = 1.2864$, and the bound in Proposition 2.21 is valid for $n \ge 44$. Some typical values are given in the table.

| $n$ | $e^{-1}\gamma(1-c)^{-2}g(b_n)$ | $\Lambda(-c/g(b_n))$ | $F^n(\alpha_n(-c/g(b_n)) + b_n)$ |
|---|---|---|---|
| 44 | .0901 | .1477 | .1548 |
| 75 | .0753 | .1139 | .1220 |
| 100 | .0692 | .0976 | .1060 |
| 250 | .0552 | .0561 | .0644 |
| 500 | .0482 | .0346 | .0418 |
| 1,000 | .0429 | .0201 | .0258 |
| 10,000 | .0318 | .0019 | .0032 |

In Proposition 2.20 we considered the situation where $f'(x) \to 0$ roughly like a negative power of $x$. We consider now what happens when $f'$ decays to zero roughly like an exponential function. More precisely we suppose $0 \le f'(x) \downarrow 0$ and $1/f' \in \Gamma$. Thinking of $f'$ as a distribution tail (as in Proposition 1.18) we get from Proposition 1.19

$$\lim_{x\to\infty} f'(x)\int_x^\infty f(u)du/f^2(x) = 1. \tag{2.63}$$

Also from Proposition 1.9 we get that the auxiliary function of $1/f'$ can be taken as $f/f'$ so that locally uniformly in $x \in \mathbb{R}$ we have

$$\lim_{t\to\infty} f'(t + x(f(t)/f'(t)))/f'(t) = e^{-x}. \tag{2.64}$$

With this case in mind, we state the final result.

**Proposition 2.22.** *Suppose $F \in D(\Lambda)$ and for $\varepsilon > 0$, $c > 0$, and $n \geq n_0$ we have*

$$g(b_n - ca_n/g(b_n)) \leq e^{c+\varepsilon} g(b_n). \tag{2.65}$$

*Then for $n \geq n_0$*

$$\sup_{x\in\mathbb{R}} |F^n(a_n x + b_n) - \Lambda(x)|$$
$$\leq F^n(a_n(-c/g(b_n)) + b_n) \vee \Lambda(-c/g(b_n)) \vee e^{c+\varepsilon-1} g(b_n).$$

The proof is virtually identical to the proof of Proposition 2.20 and is left as an exercise. Again the order of convergence is $0(g(b_n))$. If $0 \leq f'(x) \downarrow 0$ and (2.63) holds then

$$g(b_n) \sim f^2(b_n)\Big/\int_{b_n}^{\infty} f(u)du.$$

Note if we change variables $u = b(s) := \phi^{\leftarrow}(\log s)$ then

$$\int_{b(n)}^{\infty} f(u)du = \int_n^{\infty} f(b(s))b'(s)ds$$

and since

$$b'(s) = 1/\{\phi'(\phi^{\leftarrow}(\log s))s\} = 1/\{\phi'(b(s))s\}$$
$$= f(b(s))/s =: a(s)/s$$

we have

$$g(b_n) \sim a^2(n)\Big/\int_n^{\infty} a^2(s)s^{-1}\,ds.$$

According to Proposition 0.11(a), $\int_x^\infty a^2(s)s^{-1}\,ds =: \pi(x)$ being the integral of a $-1$-varying function is $\Pi$-varying with auxiliary function $a^2(\cdot)$. So $g(b_n) \to 0$ like the reciprocal of a $\Pi$-varying function divided by its auxiliary function. Both $\pi$ and $a^2(\cdot)$ are slowly varying so again the convergence rate is rather slow.

## Exercises

2.4.1. If $g$ is differentiable and satisfies

$$\lim_{t\to\infty} tg'(t)/g(t) = -\beta < 0$$

show (2.48) is satisfied. What is a suitable $\beta$?

2.4.2. If $1 - F(x) = \begin{cases} 1 & x < 1 \\ x^{-\alpha} & x \geq 1 \end{cases}$ set $a_n = n^{1/\alpha}$ and show

$$\sup_{x \in \mathbb{R}} |F^n(a_n x) - \Phi_\alpha(x)| \leq O(n^{-1})$$

and using Hall and Wellner (1979)

$$O(n^{-1}) = (2 + n^{-1})e^{-2}n^{-1}.$$

The same calculations give a rate of convergence for $1 - F(x) = e^{-x}$, $x > 0$.

2.4.3. Suppose $F$ concentrates on $[0, \infty)$ and the density $F'(x)$ is of the form

$$F'(x) = c(x)x^{-\beta}, \qquad x > 0$$

for $\beta > 1$, where $\lim_{x \to \infty} c(x) = c > 0$. Suppose

$$|c(x) - c| \leq g(x) \downarrow 0$$

as $x \to \infty$. Check

$$\left| \frac{xF'(x)}{1 - F(x)} - (\beta - 1) \right| \leq (\beta - 1)\frac{2g(x)}{c - g(x)} = O(g(x))$$

so that Proposition 2.16 leads to a convergence rate no better than $O(g(a_n) \vee n^{-1})$. Illustrate this by analyzing the Cauchy, $t$, and $F$ densities.

2.4.4. Check the algebra in Proposition 2.21.

2.4.5. Prove Proposition 2.22.

2.4.6. Check that for the normal distribution, the bound is of order $O(1/\log n)$. Check that (2.59) and (2.60) are satisfied and a suitable $g$ is

$$g(x) = x^{-2} + 1 - N(x).$$

See Hall (1979) for a more precise result. Find the order of the bound for the gamma distribution.

2.4.7. (a) Show $d_n \geq d_1^n$.
(b) Consequently if $d_n = O(\theta^n)$ for all $\theta > 0$ then $d_1 = 0$ and $F$ is an extreme value distribution
(c) If $G$ is an extreme value distribution define $F$ for fixed $\theta \in (0, 1)$ by

$$F(x) = \begin{cases} G(x) & x > G^{\leftarrow}(\theta) \\ \theta & G^{\leftarrow}(\theta) > x \geq G^{\leftarrow}(\theta) - 1 \\ 0 & G^{\leftarrow}(\theta) - 1 > x. \end{cases}$$

Show

$$0 < d_n \leq \theta^n.$$

Thus despite the remarks following Propositions 2.20 and 2.22, the convergence rate can be exponentially fast.
(d) However the convergence rate can be arbitrarily slow. Let

$$F^{\#}(x) = 1 - e^{-x^+}$$

and suppose $\{\theta_n\}$ is any sequence whatever satisfying $\theta_n \downarrow 0$. Define

$$\xi_n = 4e^e\theta_{[\exp\{n+1\}]}$$

and set for large $n$

$$F(x) = \begin{cases} F^{\#}(n + \xi_n) & x \in [n, n + \xi_n) \\ F^{\#}(x) & \text{otherwise.} \end{cases}$$

Show $\bar{F} \sim \bar{F}^{\#}$ so that $F$ and $F^{\#}$ are tail equivalent and $F \in D(\Lambda)$. With $a_n = 1$, $b_n = \log n$, show $d_n \geq \theta_n$ for all large $n$. (Hint: Show for large $n$

$$F^n([\log n]) - F^n([\log n]-) \geq e^{-e}\xi_{[\log n]}$$

(Rootzen, 1984).)

CHAPTER 3

# Point Processes

For a thorough understanding of many structural results in extreme value theory, knowledge of point processes is desirable. We present a brief account of those parts of the theory that are useful for understanding the behavior of extremes. Some skimming may be advisable. Parts of this account are fashioned after Neveu (1976). An additional excellent reference is Kallenberg (1983).

## 3.1. Fundamentals

A *point process* is a random distribution of points in space. How can we make a model for this?

We begin by specifying some notation. The state space where the points live will be denoted by $E$. It is convenient to suppose $E$ is locally compact with a countable basis. (By this we understand that $E$ is Hausdorff, every $x \in E$ has a compact neighborhood, and there exist open $G_n$, $n \geq 1$ such that any open $G$ can be written $G = \bigcup_{\alpha \in I} G_\alpha$ for $I$ a finite or countable index set.) For us, $E$ will always be a subset of a compactified Euclidean space of finite dimension, and the reader with little background or interest in topology is urged to consider $E$ in this way. Let $\mathscr{E}$ be the Borel $\sigma$-algebra of subsets of $E$, i.e., the $\sigma$-algebra generated by the open sets.

For $x \in E$, define the measure $\varepsilon_x$ on $\mathscr{E}$ by

$$\varepsilon_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \qquad \text{for } A \in \mathscr{E}.$$

A *point measure* on $E$ is a measure $m$ of the following form: Let $\{x_i, i \geq 1\}$ be a countable collection of (not necessarily distinct) points of $E$. Then

$$m := \sum_{i=1}^{\infty} \varepsilon_{x_i}$$

and if $K \in \mathscr{E}$ is compact then $m(K) < \infty$ (i.e., $m$ is *Radon* meaning the measure of compact sets is always finite).

Let $S_m = \{x \in E: m(\{x\}) \neq 0\}$ so that $S_m$ is the set of points charged by $m$, i.e., the distinct points of $\{x_i, i \geq 1\}$. We may check that $S_m$ is the support of $m$; i.e., $S_m$ is the smallest closed set $F$ such that $m(F^c) = 0$. If $x \in S_m$, call $m(\{x\})$ the *multiplicity* of $x$ and call $m$ *simple* if $m(\{x\}) \leq 1$ for all $x \in E$.

Designate by $M_p(E)$ the space of all point measures defined on $E$ and define a $\sigma$-algebra $\mathscr{M}_p(E)$ of subsets of $M_p(E)$ to be the smallest $\sigma$-algebra containing all sets of the form $\{m \in M_p(E): m(F) \in B\}$ for $F \in \mathscr{E}$, $B \in \mathscr{B}([0, \infty])$. (Since $m(F)$ has range $\{0, 1, \ldots, \infty\}$, it is excessive to take $B \in \mathscr{B}([0, \infty])$, but for generalizations to random measures later, this is the most convenient formulation.) Alternatively, $\mathscr{M}_p(E)$ is the smallest $\sigma$-algebra making all evaluation maps $m \to m(F)$ (from $M_p(E) \to [0, \infty]$) measurable for all $F \in \mathscr{E}$.

A *point process* on $E$ is a measurable map, call it $N$, from a probability space $(\Omega, \mathscr{A}, P) \to (M_p(E), \mathscr{M}_p(E))$; i.e., a point process is a random element of $M_p(E)$. The probability law, denoted by $P_N$ of the point process $N$, is the measure $P \circ N^{-1} = P[N \in \cdot\,]$ on $\mathscr{M}_p(E)$.

So if we pick $\omega$, then $N(\omega, \cdot)$ is a point measure and $N(\omega, F)$ is the number of points in $F$ for the realization $\omega$.

Just from the definition of a point process one would think it difficult to verify that a map $N: \Omega \to M_p(E)$ is a point process. The following is a more palatable criterion in that it says $N$ is a point process iff $N(F)$ is an (extended real valued) random variable for each $F \in \mathscr{E}$.

**Proposition 3.1.** *$N$ is a point process* iff *the map $\omega \to N(\omega, F)$ is measurable from $(\Omega, \mathscr{A}) \to ([0, \infty], \mathscr{B}([0, \infty])$ for every $F \in \mathscr{E}$.*

PROOF: NECESSITY. If $N$ is a point process, then $\omega \to N(\omega, \cdot)$ is measurable from $(\Omega, \mathscr{A}) \to (M_p(E), \mathscr{M}_p(E))$, and $m \to m(F)$ is measurable from $(M_p(E)\mathscr{M}_p(E)) \to ([0, \infty], \mathscr{B}([0, \infty]))$ by the definition of $\mathscr{M}_p(E)$. Therefore $\omega \to N(\omega, F)$ is measurable since it is the composition of these two measurable maps.

PROOF: SUFFICIENCY. Suppose $\omega \to N(\omega, F)$ is measurable, i.e., $\{\omega: N(\omega, F) \in B\} \in \mathscr{A}$ for $B \in \mathscr{B}([0, \infty])$ and $F \in \mathscr{E}$. Define

$$\mathscr{G} = (A \in \mathscr{M}_p(E): N^{-1}A \in \mathscr{A}\}.$$

It is easy to check that $\mathscr{G}$ is a $\sigma$-algebra. Note that $\mathscr{G}$ contains all sets of the form $\{m: m(F) \in B\}$ since

$$N^{-1}\{m: m(F) \in B\} = \{\omega: N(\omega, F) \in B\} \in \mathscr{A}$$

by assumption. Hence

$$\mathscr{G} \supset \sigma\{\{m: m(F) \in B\}, F \in \mathscr{E}, B \in \mathscr{B}([0, \infty])\} = \mathscr{M}_p(E). \qquad \square$$

The notation $\sigma(A_\alpha, \alpha \in I)$ is standard and means the smallest $\sigma$-algebra containing the collection $A_\alpha$, $\alpha \in I$.

Further simplifying criteria are given later.

*Reminders.* See, for example, Billingsley (1979) for the following: If $\mathscr{T} \subset \mathscr{E}$, call $\mathscr{T}$ a $\Pi$-system if $\mathscr{T}$ is closed under finite intersections; i.e., if $A, B \in \mathscr{T}$ then $A \cap B \in \mathscr{T}$. If $\mathscr{J} \subset \mathscr{E}$, call $\mathscr{J}$ a $\lambda$-system ($\sigma$-additive class) if

(i) $E \in \mathscr{J}$,
(ii) If $A, B \in \mathscr{J}$, and $A \supset B$ then $A - B \in \mathscr{J}$,
(iii) If $A_n \in \mathscr{J}$ and $A_n \subset A_{n+1}$ then

$$\lim_{n\to\infty} \uparrow A_n \in \mathscr{J}.$$

The following is very useful:

**Dynkin's Theorem.** *If $\mathscr{T}$ is a $\Pi$-system and $\mathscr{J}$ is a $\lambda$-system than $\mathscr{J} \supset \mathscr{T}$ implies $\mathscr{J} \supset \sigma(\mathscr{T})$.*

*A useful corollary: If two probability measures are equal on a $\Pi$-system which generates the $\sigma$-algebra then they are equal on the $\sigma$-algebra. Cf. Exercise* 3.1.3.

To check that $N$ is a point process, one does not have to check that $\omega \to N(\omega, F)$ is measurable for all $F$ but just for $F$ in a restricted class, say for bounded rectangles in case $E$ is Euclidean.

**Proposition 3.2.** *Suppose $\mathscr{T}$ are relatively compact subsets in $\mathscr{E}$ satisfying*

(i) *$\mathscr{T}$ is a $\Pi$-system.*
(ii) *$\sigma(\mathscr{T}) = \mathscr{E}$.*
(iii) *Either* (a) *there exist $E_n \in \mathscr{T}$, $E_n \uparrow E$ or* (b) *there exist $\{E_n\}$, a partition of $E$, with $\sum E_j = E$ and $E_n \in \mathscr{T}$.*

*Then $N$ is a point process on $(\Omega, \mathscr{A})$ in $(E, \mathscr{E})$* iff *$\omega \to N(\omega, I)$ is measurable from $\Omega \to [0, \infty)$ for each $I \in \mathscr{T}$.*

PROOF. Suppose $\omega \to N(\omega, I)$ is measurable for all $I \in \mathscr{T}$ and define for $n$ fixed

$$\mathscr{G} = \{F \in \mathscr{E} : \omega \to N(F \cap E_n) \text{ is measurable from } \Omega \to [0, \infty]\}.$$

Suppose (iii(a)) holds; the proof under (iii(b)) is similar. We note the following properties of $\mathscr{G}$:

1. $\mathscr{G} \supset \mathscr{T}$ because if $F \in \mathscr{T}$ then $F \cap E_n \in \mathscr{T}$ since $\mathscr{T}$ is a $\Pi$-system and thus $\omega \to N(F \cap E_n)$ is measurable.
2. $\mathscr{G} \supset E$ since $\omega \to N(E_n)$ is measurable.
3. $\mathscr{G}$ is closed under proper differences: If $F_1, F_2 \in \mathscr{G}$, $F_1 \supset F_2$ then

$$N((F_1 - F_2) \cap E_n) = N(F_1 \cap E_n) - N(F_2 \cap E_n). \tag{3.1}$$

Note $N(F_1 \cap E_n) \le N(E_n) < \infty$ since $E_n \in \mathscr{T}$ and is relatively compact and all point measures $m$ have the property $m(K) < \infty$ if $K$ is compact. Therefore the difference in (3.1) is that of two finite measurable functions and hence is measurable.

4. $\mathscr{G}$ is closed under nondecreasing limits.

Properties 2, 3, and 4 indicate that $\mathscr{G}$ is a $\lambda$-system. Also $\mathscr{G} \supset \mathscr{T}$ so $\mathscr{G} \supset \sigma(\mathscr{T}) = \mathscr{E}$ by Dynkin's theorem. So for any $F \in \mathscr{E}$, $\omega \to N(F \cap E_n)$ is measurable. Measurability is preserved by taking limits so let $n \to \infty$ (which sends $E_n \uparrow E$) to get $\omega \to N(F)$ is measurable for any $F$. □

**Corollary 3.3.** *Let $\mathscr{T}$ satisfy the hypotheses of Proposition* 3.2 *and set*

$$\mathscr{J} = \{\{m: m(I_j) = n_j, 1 \le j \le k\}, k = 1,2,\ldots, I_j \in \mathscr{T} \text{ and } n_j \ge 0 \text{ integers}\}. \tag{3.2}$$

*Then $\sigma(\mathscr{J}) = \mathscr{M}_p(E)$ and $\mathscr{J}$ forms a* $\Pi$*-system.*

Proof. In the previous proposition, set $\Omega = M_p(E)$. Then by definition

$$\mathscr{M}_p(E) := \sigma\{\{m: m(F) \in A\}, F \in \mathscr{E}, A \in \mathscr{B}([0,\infty])$$

and Proposition 3.2 assures us that also

$$\mathscr{M}_p(E) = \sigma\{\{m: m(I) \in A\}, I \in \mathscr{T}, A \in \mathscr{B}([0,\infty]\}.$$

Since $m(\cdot)$ has range $\{0, 1, \ldots, \infty\}$, it is clear the right side in the preceding line is also $\sigma(\mathscr{J})$. □

If $Q$ is a probability measure on $(M_p(E), \mathscr{M}_p(E))$ then defining $N(m, \cdot) = m(\cdot)$ gives a point process with law $Q$, called the *canonical point process.*

For $F \in \mathscr{E}$, $N(F)$ is a random variable. So for $F_1, \ldots, F_k \in \mathscr{E}$. $(N(F_i), i \le k))$ is a random vector. The set of finite dimensional distributions of such random vectors determines the law $P \circ N^{-1} = P_N$ as is proved next.

**Proposition 3.4.** *Let $N$ be a point process in $(E, \mathscr{E})$ and suppose $\mathscr{T}$ satisfies the hypotheses of Proposition* 3.2. *Define the mass functions*

$$P_{I_1,\ldots,I_k}(n_1,\ldots,n_k) = P[N(I_j) = n_j, 1 \le j \le k]$$

*for $I_i \in \mathscr{T}$, $n_i \ge 0$ integers, $1 \le i \le k$. Then $P_n$ is uniquely determined by knowledge of*

$$\{P_{I_1,\ldots,I_k}, k = 1,2,\ldots; I_j \in \mathscr{T}\}.$$

The proof is made apparent if we formulate the result in an alternate way: If $P$ and $Q$ are probability measures on $\mathscr{M}_p(E)$ and $P = Q$ on $\mathscr{J}$ (defined by (3.2)) then $P \equiv Q$. (From Corollary 3.3, $P \equiv Q$ on a $\Pi$-system generating the full $\sigma$-algebra and hence everywhere.)

This uniqueness result does not say much about how to construct point processes. A construction of a Poisson process is given later.

Let $(\Omega, \mathscr{A}, P)$ be a probability space and let $(E_i, \mathscr{E}_i)$ be state spaces. If $N_i: \Omega \to M_p(E_i)$, $i \ge 1$ are point processes we say $N_i$, $i \ge 1$ are independent if the induced $\sigma$-algebras

$$N_i^{-1}(\mathscr{M}_p(E_i)), \qquad i \ge 1$$

are independent. In particular if $F_i \in \mathscr{E}_1$, $1 \le i \le k$, and $G_j \in \mathscr{E}_2$, $1 \le j \le l$ then the vectors

$$(N_1(F_i), i \le k) \quad \text{and} \quad (N_2(G_j), j \le l)$$

are independent vectors. (The converse holds as well; cf. Exercise 3.1.5.)

The intensity or mean measure of a point process $N$ is the measure $\mu$ defined as

$$\mu(F) = EN(F) = \int_\Omega N(\omega, F)P(d\omega)$$

$$\doteq \int_{M_p(E)} m(F)P_N(dm)$$

for $F \in \mathscr{E}$. (Check that $\mu$ really is a measure; Exercise 3.1.4.) Warning: $\mu$ need not be Radon.

Suppose $f: (E, \mathscr{E}) \to ([0, \infty], \mathscr{B}([0, \infty]))$ is measurable. Recall there exist simple $f_n$ with $0 \le f_n \uparrow f$ and $f_n$ is of the form

$$f_n = \sum_1^{k_n} c_i^{(n)} 1_{A_i^{(n)}}, \qquad A_i^{(n)} \in \mathscr{E} \quad \text{and} \quad \{A_i^{(n)}, i \le k_n\} \text{ disjoint.}$$

Define, as usual, for $\omega \in \Omega$

$$N(\omega, f) = \int_E f(x)N(\omega, dx) \le \infty. \tag{3.3}$$

This is a random variable since by monotone convergence

$$N(\omega, f) = \lim_{n\to\infty} \uparrow N(\omega, f_n)$$

and each

$$N(\omega, f_n) = \sum_1^{k_n} c_i^{(n)} N(\omega, A_i^{(n)})$$

is a random variable. (If $N$ is canonical so that $\Omega = M_p(E)$ and $N(m, \cdot) = m(\cdot)$ then this argument shows that $m \to m(f) := \int_E f(x)m(dx)$ is measurable from $M_p(E) \to [0, \infty]$.) Furthermore

$$EN(\omega, f) = \mu(f) := \int_E f d\mu$$

since

$$EN(f) = \lim_{n\to\infty} \uparrow EN(f_n) = \lim_{n\to\infty} \uparrow E\sum_1^{k_n} c_i^{(n)} N(\omega, A_i^{(n)})$$

$$= \lim_{n\to\infty} \uparrow \sum_1^{k_n} c_i^{(n)} \mu(A_i^{(n)}) = \lim_{n\to\infty} \uparrow \int_E f_n(x)\mu(dx)$$

$$= \int_E \left(\lim_{n\to\infty} \uparrow f_n(x)\right)\mu(dx) = \int f(x)\mu(dx).$$

If $f \in L_1(\mu)$ but $f$ is not necessarily non-negative we may still define $N(f)$ by $N(f) = N(f^+) - N(f^-)$ since $N(f^\pm) < \infty$, a.s.

EXERCISES

3.1.1. Show $R^\infty$ and $C[0,1]$ are not locally compact. (Here

$$R^\infty = \{(x_1, x_2, \ldots): x_i \in \mathbb{R}, i = 1, 2, \ldots\}$$

and $C[0,1]$ = all real valued continuous functions on $[0,1]$.)

3.1.2. Verify that the $\sigma$-algebra generated by the sets $\{m \in M_p(E): m(F) \in B\}$, $F \in \mathscr{E}$, $B \in \mathscr{B}([0,\infty])$ is the same as the $\sigma$-algebra generated by the maps

$$m \to m(F)$$

for $F \in \mathscr{E}$.

3.1.3. Verify using Dynkin's theorem that if two probability measures are equal on a $\Pi$-system which generates the $\sigma$-algebra then they are equal on the $\sigma$-algebra. (The collection of sets where the two probability measures are equal is a $\lambda$-system containing the $\pi$-system.)

3.1.4. Check that the intensity measure of a point process is indeed a measure. Give an example to show it need not be Radon.

3.1.5. Suppose $N_1$, $N_2$ are point processes defined on $(\Omega, \mathscr{A}, P)$ with state spaces $E_1$, $E_2$, respectively. Then independence of

$$(N_1(F_i), 1 \le i \le k) \quad \text{and} \quad (N_2(G_j), 1 \le j \le l)$$

for any $k, l, F_1, \ldots, F_k \in \mathscr{E}_1, G_1, \ldots, G_l \in \mathscr{E}_2$ implies $N_1$, $N_2$ independent. Generalize to the case that the $F$'s and $G$'s are selected from subclasses of $\mathscr{E}_1$ and $\mathscr{E}_2$, respectively.

3.1.6. Suppose $N$ is a point process with state space $\mathbb{R}^k$. If $\tau$ is a random vector in $\mathbb{R}^k$ show $N(\cdot + \tau)$ is a point process. Note the value of $N(\cdot + \tau)$ on set $F$ for realization $\omega$ is $N(\omega, F + \tau(\omega))$ where $F + \mathbf{t} = \{\mathbf{x} + \mathbf{t}, \mathbf{x} \in F\}$.

3.1.7. A common way of specifying a point process is as follows: Let $\{X_n, n \ge 1\}$ be random elements of $E$ defined on $(\Omega, \mathscr{A})$. Show

$$N = \sum_{i=1}^{\infty} \varepsilon_{X_i}$$

is a point process.

3.1.8. Suppose $\mathscr{T}$ is a $\Pi$-system of subsets of $E$ and $\mathscr{H}$ is a linear function space of real valued functions on $E$ satisfying
(a) $1 \in \mathscr{H}$ and $1_A \in \mathscr{H}$ for all $A \in \mathscr{T}$,
(b) if $0 \le f_j \le f_{j+1} \in \mathscr{H}$ and $f = \sup f_j$ is finite, then $f \in \mathscr{H}$. Prove $\mathscr{H}$ contains all functions which are $\sigma(\mathscr{T})$ measurable (Jagers, 1974).

## 3.2. Laplace Functionals

Let $Q$ be a probability measure on $(M_p(E), \mathscr{M}_p(E))$. The *Laplace transform* of $Q$ is the map $\psi$ which takes non-negative Borel functions on $E$ into $[0, \infty)$ defined by

$$\psi(f) := \int_{M_p(E)} \left( \exp\left\{ -\int_E f(x)m(dx) \right\} \right) Q(dm).$$

If $N: (\Omega, \mathscr{A}) \to (M_p(E), \mathscr{M}_p(E))$ is a point process, the *Laplace functional* of $N$ is the Laplace transform of the law of $N$:

$$\psi_N(f) := E\exp\{-N(f)\} = \int_\Omega \exp\{-N(\omega, f)\} P(d\omega)$$

$$= \int_{M_p(E)} \left( \exp\left\{ -\int_E f(x)m(dx) \right) P_N(dm).$$

**Proposition 3.5.** *The Laplace transform $\psi$ of $Q$ uniquely determines $Q$. The Laplace functional $\psi_N$ of $N$ uniquely determines the law of $N$.*

Proof. For $k \geq 1$ and $F_1, \ldots, F_k \in \mathscr{E}$, and $\lambda_i \geq 0$, $i = 1, \ldots, k$ define $f: E \to [0, \infty)$ by

$$f(x) = \sum_{i=1}^{k} \lambda_i 1_{F_i}(x).$$

Then

$$N(\omega, f) = \int f(x) N(\omega, dx) = \sum_{i=1}^{k} \lambda_i N(\omega, F_i)$$

and

$$\psi_N(f) = E\exp\left\{ -\sum_{i=1}^{k} \lambda_i N(F_i) \right\}$$

which is the joint Laplace transform of the random vector $(N(F_i), i \leq k)$. Using the uniqueness theorem for Laplace transforms of random vectors we see $\psi_N$ uniquely determines the law of $(N(F_i), i \leq k)$ for any $F_i, \ldots, F_k \in \mathscr{E}$. The proof is completed by an appeal to Proposition 3.4. $\square$

Laplace functionals are useful for studying weak convergence of point processes. We will compute some after discussing Poisson processes.

A final comment: Moments of $N$ can be determined from $\psi_N$. For example, for $f \geq 0$ and measurable

$$\mu(f) = EN(f) = \lim_{t \downarrow 0} t^{-1}(1 - \psi_N(tf).$$

To check this observe that

$$\lim_{t \downarrow 0} \uparrow t^{-1}(1 - e^{-tN(f)}) = N(f)$$

and taking expectations and using monotone convergence give the result.

## 3.3. Poisson Processes

### 3.3.1. Definition and Construction

Given a Radon measure $\mu$ on $\mathscr{E}$, a point process $N$ is called a *Poisson process* or *Poisson random measure* (PRM) with mean measure $\mu$ if $N$ satisfies
(a) For any $F \in \mathscr{E}$, and any non-negative integer $k$

$$P[N(F) = k] = \begin{cases} \exp\{-\mu(F)\}(\mu(F))^k/k! & \text{if } \mu(F) < \infty \\ 0 & \text{if } \mu(F) = \infty, \end{cases}$$

and
(b) For any $k \geq 1$, if $F_1, \ldots, F_k$ are mutually disjoint sets in $\mathscr{E}$ then

$$N(F_i), \quad i \leq k$$

are independent random variables.

It follows from (a) that if $\mu(F) = \infty$ then $N(F) = \infty$ a.s. and that $\mu$ is the intensity of $N$. As a shorthand for "Poisson process with mean measure $\mu$" we will sometimes write PRM($\mu$).

**Proposition 3.6.** (i) PRM($\mu$) *exists! Its law is uniquely determined by* (a) *and* (b) *in the previous definition.*
(ii) *The Laplace functional of* PRM($\mu$) *is given* (*for* $f \geq 0$, *measurable*) *by*

$$\Psi_N(f) = \exp\left\{-\int_E (1 - e^{-f(x)})\mu(dx)\right\}, \tag{3.4}$$

*and conversely a point process with Laplace functional of the form* (3.4) *must be* PRM($\mu$).

PROOF. We begin by proving that any point process satisfying (a) and (b) has a Laplace functional (3.4), and conversely any point process with Laplace functional (3.4) satisfies (a) and (b). Then we give a construction of a point process satisfying (a) and (b). Since the distribution of a point process is uniquely determined by the Laplace functional, we will be done.

So suppose $N$ is a point process for which (a) and (b) hold. If $c > 0$, $F \in \mathscr{E}$, $f(x) = c1_F(x)$ then $N(f) = cN(F)$ and from (a)

$$\begin{aligned} \psi_N(f) &= E\exp\{-N(f)\} = E\exp\{-cN(F)\} \\ &= \exp\{(e^{-c} - 1)\mu(F)\} \qquad \text{(since } N(F) \text{ is a Poisson random variable)} \\ &= \exp\left\{-\int_E (1 - e^{-f(x)})\,\mu(dx)\right\} \end{aligned}$$

which is the form given in (3.4). Next suppose $c_i \geq 0$, $F_1, \ldots, F_k$ are disjoint in $\mathscr{E}$. Then $N(F_1), \ldots, N(F_k)$ are independent and if $f(x) = \sum_{i=1}^k c_i 1_{F_i}(x)$ we have

$$\begin{aligned}
\psi_N(f) &= E\exp\{-N(f)\} = E\exp\left\{-\sum_{i=1}^{k} c_i N(F_i)\right\} \\
&= \prod_{i=1}^{k} E\exp\{-c_i N(F_i)\} \qquad \text{from independence} \\
&= \prod_{i=1}^{k} \exp\left\{-\int_E (1 - e^{-c_i 1_{F_i}(x)})\mu(dx)\right\} \qquad \text{from the previous step} \\
&= \exp\left\{-\int_E \sum_{i=1}^{k} (1 - e^{-c_i 1_{F_i}(x)})\mu(dx)\right\} \\
&= \exp\left\{-\int_E (1 - e^{-\sum_{i=1}^{k} c_i 1_{F_i}(x)})\mu(dx)\right\} \\
&= \exp\left\{-\int_E (1 - e^{-f(x)})\mu(dx)\right\}
\end{aligned}$$

which again gives (3.4). Now for general $f \geq 0$, measurable, there exist simple $f_n$ of the form just considered, $f_n = \sum_{i=1}^{k_n} c_i^{(n)} 1_{F_i^{(n)}}$ with $c_i^{(n)} \geq 0$, $\{F_i^{(n)}, 1 \leq i \leq k_n\}$ disjoint, and $0 \leq f_n \uparrow f$. By monotone convergence

$$N(f_n) \uparrow N(f)$$

for all $\omega$. Since $e^{-N(g)} \leq 1$ for any measurable $g \geq 0$ we have by dominated convergence as $n \to \infty$

$$\psi_N(f_n) = E\exp\{-N(f_n)\} \to E\exp\{-N(f) = \psi_N(f).$$

On the other hand (3.4) holds for $f_n$ so

$$\psi_N(f_n) = \exp\left\{-\int_E (1 - e^{-f_n})d\mu\right\}.$$

If $f_n \uparrow f$ then also $1 - e^{-f_n} \uparrow 1 - e^{-f}$ and monotone convergence applies to give

$$\int_E (1 - e^{-f_n})d\mu \uparrow \int_E (1 - e^{-f})d\mu$$

and so we conclude (3.4) holds for $f$ as required.

Conversely, suppose a point process $N$ has Laplace functional given by (3.4). Setting $f = \lambda 1_F$, $F \in \mathscr{E}$ gives

$$Ee^{-N(f)} = Ee^{-\lambda N(F)} = \exp\{-(1 - e^{-\lambda})\mu(F)\},$$

which is the Laplace transform of a Poisson random variable with parameter $\mu(F)$. So (a) holds in the definition of PRM. Likewise for $F_1, \ldots, F_k$ disjoint in $\mathscr{E}$ and with $f = \sum_{i=1}^{k} \lambda_i 1_{F_i}$, $\lambda_i \geq 0$ we get

$$
\begin{aligned}
Ee^{-N(f)} &= Ee^{-\sum_1^k \lambda_i N(F_i)} \\
&= \exp\left\{-\int_E (1 - e^{-\sum_1^k \lambda_i 1_{F_i}}) d\mu\right\} \\
&= \exp\left\{-\int_E \sum_1^k (1 - e^{-\lambda_i 1_{F_i}}) d\mu\right\} = \prod_1^k \exp\{-(1 - e^{-\lambda_i})\mu(F_i)\} \\
&= \prod_1^k Ee^{-\lambda N(F_i)}
\end{aligned}
$$

and so $(N(F_i), \ldots, N(F_k))$ are independent; this verifies (b) in the definition of PRM.

We now focus on the construction of PRM$(\mu)$. Suppose initially that the given measure $\mu$ is finite ($\mu(E) < \infty$) so we can write $\mu = c\nu$ where $\nu$ is a probability measure. Construct a probability space which supports independent random elements

$$\tau, X_1, X_2, \ldots$$

where $\tau$ is a Poisson random variable with parameter $c > 0$ and $\{X_j, j \geq 1\}$ are iid random elements of $E$ with distribution $\nu$; i.e., $P[X_1 \in F] = \nu(F)$, $F \in \mathscr{E}$. (Note, we may take $\Omega = N \times E \times E \times \ldots$, where $N$ is the non-negative integers. Give $\Omega$ the product $\sigma$-algebra and product measure.) Now define $N^*$ on $\Omega$ by

$$
\begin{aligned}
N^* &= \sum_1^\tau \varepsilon_{X_i} \qquad \text{on } [\tau > 0] \\
&= 0 \qquad \text{on } [\tau = 0].
\end{aligned}
$$

We first verify that $N^*$ is a point process and for this its suffices to check whether $N^*(F)$ is a random variable for any $F \in \mathscr{E}$. For $k \geq 1$

$$[N^*(F) = k] = \sum_{l=k}^\infty \left(\left[\sum_{i=1}^l 1_F(X_i) = k\right] \cap [\tau = l]\right)$$

and so $[N^*(F) = k]$ is measurable. A similar argument works for $k = 0$. In fact $N^*(F)$ is a Poisson distributed random variable. For $k \geq 1$:

$$P[N^*(F) = k] = \sum_{l=k}^\infty P\left[\sum_{i=1}^l 1_F(X_i) = k\right] P[\tau = l].$$

The first probability on the right is binomial so

$$
\begin{aligned}
P[N^*(F) &= \sum_{l=k}^\infty \binom{l}{k} (\nu(F))^k (1 - \nu(F))^{l-k} e^{-c} c^l / l! \\
&= \sum_{l=k}^\infty \frac{(c(1 - \nu(F)))^{l-k}}{(l-k)!} \frac{e^{-c}(c\nu(F))^k}{k!} = e^{c(1-\nu(F))} \frac{e^{-c}(c\nu(F))^k}{k!} \\
&= e^{-c\nu(F)} (c\nu(F))^k / k!.
\end{aligned}
$$

So $N^*$ is a point process and for fixed $F$, $N^*(F)$ has a Poisson distribution. We now verify that the independence property (b) holds for $N^*$. To accomplish this, let $F_0, \ldots, F_k$ be a measurable partition of $E(F_i \in \mathscr{E}, F_i \cap F_j = \phi, i \neq j, \sum_{i=1}^k F_i = E)$. Suppose $n_i, i = 0, \ldots, k$ are non-negative integers and $\sum_{i=1}^k n_i = n$ and for $n \geq 1$ (similar procedures for $n = 0$)

$$\begin{aligned} P[N^*(F_0) &= n_0, \ldots, N^*(F_k) = n_k] \\ &= P[N^*(F_i) = n_i, \qquad i = 0, \ldots, k; \qquad \tau = n] \\ &= P\left[\sum_{i=1}^n \mathscr{E}_{X_i}(F_0) = n_0, \ldots, \sum_{i=1}^n \mathscr{E}_{X_i}(F_k) = n_k\right] P[\tau = n] \end{aligned}$$

and recognizing the first probability as a multinomial we get

$$\frac{n!}{\prod_{i=0}^k n_i!} \prod_{i=0}^k (\nu(F_i))^{n_i} \frac{e^{-c} c^n}{n!}$$

and since $1 = \sum_{i=0}^k \nu(F_i)$ and $n = \sum_{i=0}^k n_i$ the preceding is

$$\prod_{i=0}^k e^{-c\nu(F_i)} (c\nu(F_i))^{n_i}/n_i! = \prod_{i=0}^k P[N^*(F_i) = n_i].$$

Now suppose $F_1, \ldots, F_k$ are any disjoint sets in $\mathscr{E}$. Set $F_0 = E - \sum_{i=1}^k F_i$ so that $F_0, \ldots, F_k$ is a partition of $E$. For any non-negative integers $n_1, \ldots, n_k$

$$\begin{aligned} P[N^*(F_1) &= n_1, \ldots, N^*(F_k) = n_k] \\ &= \sum_{n_0=0}^\infty P[N^*(F_0) = n_0, N^*(F_1) = n_1, \ldots, N^*(F_k) = n_k] \\ &= \sum_{n_0=0}^\infty P[N^*(F_0) = n_0] \prod_{i=1}^k P[N^*(F_i) = n_i] \\ &= \prod_{i=1}^k P[N^*(F_i) = n_i], \end{aligned}$$

which is property (b) in the definition of PRM.

It remains to indicate the construction when $\mu(E) = \infty$. In this case we decompose $\mu$ as $\mu = \sum_1^\infty \mu_k$ and we do this as follows: Take a partition $\{F_k, k \geq 1\}$ of $E$ by relatively compact sets of $\mathscr{E}$ and define $\mu_k = \mu(\cdot \cap F_k)$ so that $\mu_k$ concentrates on $F_k$. Since $\mu_k$ is finite ($\mu_k(E) = \mu(E \cap F_k) = \mu(F_k) < \infty$ since $F_k$ is relatively compact and $\mu$ is Radon) we know how to construct PRM$(\mu_k)$. Let $N_k^*$ be such a process and we may suppose $N_k^*$, $k \geq 1$ are independent. Then we assert $N^* := \sum_k N_k$ is the desired PRM$(\mu)$ and we prove this assertion via Laplace functionals. Let $f \geq 0$ be measurable. Then

$$\begin{aligned} \psi_{N^*}(f) &= E \exp\left\{-\sum_k N_k^*(f)\right\} \\ &= \lim_{r \to \infty} E \exp\left\{-\sum_{k=1}^r N_k^*(f)\right\} \end{aligned}$$

$$= \lim_{r\to\infty} \prod_{k=1}^{r} E\exp\{-N_k^*(f)\}$$

$$= \lim_{r\to\infty} \prod_{k=1}^{r} \exp\left\{-\int_E (1-e^{-f})d\mu_k\right\}$$

(since $N_k$ is $\mathrm{PRM}(\mu_k)$)

$$= \lim_{r\to\infty} \exp\left\{-\int_E (1-e^{-f(x)})\left(\sum_1^r \mu_k\right)(dx)\right\}$$

$$= \exp\left\{-\int_E (1-e^{-f(x)})\left(\sum_1^\infty \mu_k\right)(dx)\right\}$$

$$= \exp\left\{-\int_E (1-e^{-f})d\mu\right\}$$

as required since $\sum \mu_k = \mu$. □

### 3.3.2. Transformations of Poisson Processes

There are several transformations of Poisson processes that are enormously useful in limit theory. Much of this material was learned from Cinlar (1976). Additional applications are contained in Resnick (1986).

We first show that mapping the points of a Poisson process yields a new Poisson process.

**Proposition 3.7.** *Let $E_i$, $i = 1, 2$ be two locally compact spaces with countable bases. Let $\mathscr{E}_i$, $i = 1, 2$ be the associated σ-fields. Let $T$: $(E_1, \mathscr{E}_1) \to (E_2, \mathscr{E}_2)$ be measurable. If $N$ is* $\mathrm{PRM}(\mu)$ *on $E_1$ then*

$$\tilde{N} := N \circ T^{-1} \qquad \text{is } \mathrm{PRM}(\tilde{\mu} = \mu \circ T^{-1}) \text{ on } E_2.$$

*If we have a representation*

$$N = \sum_i \varepsilon_{X_i}$$

*then*

$$\tilde{N} = N \circ T^{-1} = \sum_i \varepsilon_{TX_i}.$$

Proof. Let $f_2$: $E_2 \to [0, \infty)$ be measurable. Then

$$\psi_{\tilde{N}}(f_2) = E\exp\{-\tilde{N}(f_2)\} = E\exp\left\{-\int_{E_2} f_2(x_2) N \circ T^{-1}(\omega, dx_2)\right\}$$

$$= E\exp\left\{-\int_{E_1} f_2(Tx_1) N(\omega, dx_1)\right\}$$

by the transformation theorem for integrals. Now $f_2 \circ T$ is non-negative and measurable on $E_1$ and $N$ is PRM($\mu$) on $E_1$ so by (3.4)

$$\begin{aligned}\psi_{\tilde{N}}(f_2) &= \exp\left\{-\int_{E_1}(1-e^{-f_2\circ T})d\mu\right\}\\ &= \exp\left\{-\int_{E_2}(1-e^{-f_2(x)})\mu\circ T^{-1}(dx)\right\},\end{aligned}$$

which is the Laplace functional of PRM($\mu \circ T^{-1}$) on $E_2$. □

The next result shows that starting from PRM, we may construct a new PRM whose points live in a higher dimensional space.

**Proposition 3.8.** *Let $(E_i, \mathscr{E}_i)$ be two state spaces as in the previous proposition. Suppose*

$$N_1 = \sum_i \varepsilon_{X_i}$$

*is* PRM($\mu$) *on $(E_1, \mathscr{E}_1)$ and let $K\colon E_1 \times \mathscr{E}_2 \to [0,1]$ be a transition function from $E_1 \to E_2$; i.e., $K(\cdot, F_2)$ is $\mathscr{E}_1$-measurable for every $F_2 \in \mathscr{E}_2$ and $K(x,\cdot)$ is a probability measure on $\mathscr{E}_2$ for each $x \in E_1$. Let $\{J_i\}$ be $E_2$-valued random elements which are conditionally independent given $\{X_i\}$:*

$$P[J_i \in F_2 \mid \{X_n\}, \{J_\alpha, \alpha \neq i\}] = K(X_i, F_2) \tag{3.5}$$

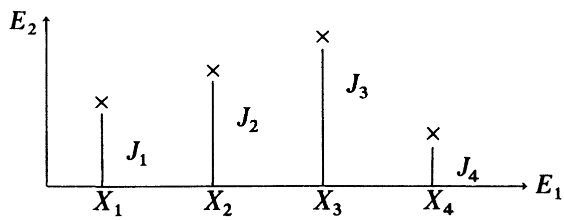
*for any $i$ and $F_2 \in \mathscr{E}_2$. Then*

$$N^* := \sum_i \varepsilon_{(X_i, J_i)}$$

*is* PRM *on $E_1 \times E_2$ with mean measure*

$$\mu^*(dx, dy) = \mu(dx)K(x, dy).$$

*Special Case*: $\{J_i\}$ are iid and independent of $(X_i)$. Suppose $\{J_i\}$ have common distribution $F$. Then $\sum \varepsilon_{(X_i, J_i)}$ is PRM with mean measure which is the product $\mu(dx)F(dy)$.




Before the proof, we need two lemmas which interpret (3.5).

**Lemma 3.9.** *Suppose $f\colon E_1 \times E_2 \to [0, \infty)$ is bounded. Then a.s.*

$$E(f(X_i, J_i) \mid \{X_n\}) = \int_{E_2} f(X_i, y)K(X_i, dy). \tag{3.6}$$

PROOF. Let

$$\mathscr{H} = \{f : f \geq 0, \text{ bounded}, \mathscr{E}_1 \times \mathscr{E}_2 \text{ measurable, and } f \text{ satisfies 3.6}\}.$$

First note $\mathscr{H}$ contains functions of the form $f(x, y) = f_1(x)f_2(y)$ where $f_i \geq 0$, bounded, and $\mathscr{E}_i$ measurable, $i = 1, 2$. To check this note that

$$\begin{aligned} E(f(X_i, J_i)|\{X_n\}) &= E(f_1(X_i)f_2(J_i)|\{X_n\}) \\ &= f_1(X_i)E(f_2(J_i)|\{X_n\}) \end{aligned}$$

and applying (3.5) this is

$$\begin{aligned} &= f_1(X_i)\int_{E_2} f_2(y)K(X_i, dy) \\ &= \int_{E_2} f_1(X_i)f_2(y)K(X_i, dy) = \int_{E_2} f(X_i, y)K(X_i, dy). \end{aligned}$$

This means that if $A_i \in \mathscr{E}_i$, $i = 1, 2$ then

$$I_{A_1 \times A_2} \in \mathscr{H}.$$

Now let $\mathscr{C} = \{G \in \mathscr{E}_1 \times \mathscr{E}_2 : 1_G \in \mathscr{H}\}$ and observe that

(i) $E_1 \times E_2 \in \mathscr{C}$;
(ii) $\mathscr{C}$ is closed under proper differences,
(iii) $\mathscr{C}$ is closed under nondecreasing limits.

Hence $\mathscr{C}$ is a $\lambda$-system containing the $\pi$-system of rectangles $A_1 \times A_2$, $A_i \in \mathscr{E}_i$, $i = 1, 2$ and so by Dynkin's theorem $\mathscr{C} \supset \sigma\{\text{rectangles}\} = \mathscr{E}_1 \times \mathscr{E}_2$. Now it is easy to check that if $G_1, \ldots, G_k$ are disjoint in $\mathscr{E}_1 \times \mathscr{E}_2$ and $c_i \geq 0$, $i = 1, \ldots, k$ then

$$\sum_1^k c_i 1_{G_i} \in \mathscr{H}. \tag{3.7}$$

Finally any $f \geq 0$, bounded and $\mathscr{E}_1 \times \mathscr{E}_2$ measurable can be written as the monotone limit of functions of the form (3.7). The monotone convergence theorem then shows $f \in \mathscr{H}$ as desired. $\square$

**Lemma 3.10.** *If $g: E_1 \times E_2 \to [0, 1]$ is $\mathscr{E}_1 \times \mathscr{E}_2$ measurable then a.s.*

$$E\left(\prod_{i=1}^{\infty} g(X_i, J_i)|\{X_n\}\right) = \prod_{i=1}^{\infty} E(g(X_i, J_i)|\{X_n\}).$$

The proof of this result can be accomplished in a manner similar to the proof of Lemma 3.9 and is left as an exercise.

PROOF OF PROPOSITION 3.8. We proceed by using Laplace functionals. Let $f \geq 0$ be $\mathscr{E}_1 \times \mathscr{E}_2$ measurable. Then

$$\Psi_{N^*}(f) = \mathrm{E}\exp\{-N^*(f)\} = E\exp\left\{-\sum_1^\infty f(X_i, J_i)\right\}$$
$$= E\left(\prod_1^\infty e^{-f(X_i,J_i)}\right) = E\left(E\left(\prod_1^\infty e^{-f(X_i,J_i)}\Big|\{X_n\}\right)\right)$$

and via Lemma 3.10 this equals

$$E\left(\prod_1^\infty E(e^{-f(X_i,J_i)}|\{X_n\})\right)$$
$$= E\left(\prod_1^\infty \int_{E_2} e^{-f(X_i,y)}K(X_i, dy)\right) \qquad \text{(from (3.6))}.$$

Now set $\theta(X_i) = \int_{E_2} e^{-f(X_i,y)}K(X_i, dy)$ so that $0 \le \theta \le 1$ and

$$\Psi_{N^*}(f) = E\prod_1^\infty \theta(X_i) = E\left(\exp\left\{-\sum_1^\infty(-\log\theta(X_i))\right\}\right)$$
$$= E\exp\left\{-\int_{E_1}(-\log\theta(x))N_1(dx)\right\}$$
$$= \Psi_{N_1}(-\log\theta)$$

and because $N_1$ is PRM$(\mu)$ we get from (3.4) that

$$\Psi_{N^*}(f) = \exp\left\{-\int_{E_1}(1 - e^{-(-\log\theta(x))})\mu(dx)\right\}$$
$$= \exp\left\{-\int_{E_1}(1 - \theta(x))\mu(dx)\right\}$$
$$= \exp\left\{-\int_{E_1}\left(1 - \int_{E_2} e^{-f(x,y)}K(x, dy)\right)\mu(dx)\right\}$$

and by Fubini's theorem we get

$$= \exp\left\{-\int_{E_1\times E_2}(1 - e^{-f(x,y)})\mu(dx)K(x, dy)\right\}$$
$$= \exp\left\{-\int_{E_1\times E_2}(1 - e^{-f(x,y)})\mu_*(dx, dy)\right\},$$

which is the Laplace functional of PRM$(\mu_*)$, and this completes the proof. □

Example. Let $\sum_{i=1}^\infty \varepsilon_{X_i}$ be PRM$(\mu)$ on $\mathbb{R}^d$. Displace each point $X_i$ by an independent amount $J_i$ where $\{J_i\}$ are iid with common distribution $F$ and $\{X_i\}$ is independent of $\{J_i\}$. The resulting point process is PRM$(\mu * F)$ where $*$ denotes convolution. To see this observe that by Proposition 3.8

$$N^* = \sum \varepsilon_{(X_i, J_i)}$$

is PRM on $\mathbb{R}^d \times \mathbb{R}^d$ with mean measure $\mu \times F$. Define $T: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by $T(x, y) = x + y$. Then by Proposition 3.7

$$N^* \circ T^{-1} = \sum_i \varepsilon_{T(X_i, J_i)} = \sum_i \varepsilon_{X_i + J_i}$$

is PRM$((\mu \times F) \circ T^{-1})$ and for $B \in \mathscr{B}(\mathbb{R}^d)$

$$\mu \times F \circ T^{-1}(B) = \mu \times F\{(x,j): x + j \in B\} = \mu * F(B)$$

where $*$ denotes convolution.

Observe that if $\mu$ is Lebesgue measure on $\mathbb{R}^d$ then

$$\begin{aligned} \mu * F(B) &= \int_{\mathbb{R}^d} F(dy)\mu\{x: x + y \in B\} \\ &= \int_{\mathbb{R}^d} F(dy)\mu(B) = \mu(B). \end{aligned}$$

When $\mu$ is Lebesgue measure we usually call the process homogeneous especially if the dimension $d = 1$. So independent displacement of the points of a homogeneous Poisson process results in a homogeneous Poisson process.

### Exercises

3.3.1. Prove Proposition 3.7 directly without using Laplace functionals. Check that $\tilde{N}$ satisfies (a) and (b) in Section 3.3.1.

3.3.2. Prove Lemma 3.10.

3.3.3. Let $N$ be PRM$(\mu)$ on $E$. Show that $N$ is simple a.s. iff $\mu$ is diffuse (atomless) on $E$. Show $N(\cdot \cap F)$ is PRM where $F \in \mathscr{E}$. What is the mean measure?

3.3.4. Let $N$ be PRM$(\mu(dt) = t^{-1}dt)$ on $(0, \infty)$. Express $N$ as a time changed homogeneous process.

3.3.5. Let $N$ be a nonhomogeneous Poisson process on $\mathbb{R}$ with mean measure

$$\mu(B) = \int_B \lambda(s)ds$$

where $\lambda$ is locally integrable. Express $N$ as a function of a homogeneous PRM. Of particular interest later will be the cases
(a) $E = (0, \infty]$, $\quad \lambda(s) = \alpha s^{-\alpha-1}$, $\quad \alpha > 0$
(b) $E = (-\infty, \infty]$, $\quad \lambda(s) = e^{-s}$, $\quad -\infty < s < \infty$.

3.3.6. $M/G/\infty$ Queue: Calls arrive to a telephone exchange according to a homogeneous Poisson process at times $\{X_i\}$ $(X_i \in (-\infty, \infty))$. Lengths of calls are iid random variables $(J_i)$ with common distribution $F$. Times when calls terminate form a homogeneous Poisson process if $\{J_i\}$ and $\{X_i\}$ are independent.

3.3.7. (a) Let $\{E_i, i \geq 1\}$ be iid exponential random variables on $[0, \infty)$: $P[E_i > x] = e^{-x}$, $x > 0$. Let $\Gamma_n = \sum_{i=1}^n E_i$. Show that $\sum_n \varepsilon_{\Gamma_n}$ is a homogeneous Poisson process on $[0, \infty)$.

(b) Use Proposition (3.8) and (a) to construct homogeneous PRM on $[0,\infty) \times [0,\infty)$.

(c) If $\sum_i \varepsilon_{(X_i, J_i)}$ is homogeneous PRM on $[0,\infty) \times [0,1)$ then $\sum_i \varepsilon_{X_i}$ is PRM on $[0,\infty)$ and $\{J_i\}$ is iid uniform on $[0,1)$.

3.3.8. Let $\{\Gamma_n\}$ be as in the previous proposition and suppose $\{U_n, n \geq 1\}$ are iid, uniform on $[0,1]$. Suppose $\nu$ is a measure on $\mathbb{R}$ with $Q(x) := \nu(x,\infty) < \infty$ for all $x \in \mathbb{R}$. Define for $y > 0$

$$Q^{\leftarrow}(y) = (1/Q)^{\leftarrow}(y^{-1}).$$

Show

$$\sum_i \varepsilon_{(U_i, Q^{\leftarrow}(\Gamma_i))}$$

is PRM on $[0,1] \times \mathbb{R}$ with mean measure $du \times \nu(dx)$.

3.3.9. Two point process $N_1$ and $N_2$ on $E$ are equal in distribution ($N_1 \stackrel{d}{=} N_2$) iff for each $f \geq 0$ bounded and measurable we have

$$N_1(f) \stackrel{d}{=} N_2(f)$$

as random variables. (It is only necessary to check $N_1(f) = N_2(f)$ for $f \in C_k^+(E)$; cf. Section 3.4.)

3.3.10. Suppose $N_i$ are PRM$(\mu_i)$ with domain $(\Omega, \mathscr{A})$ and state space $(E, \mathscr{E})$, $i = 1, 2, \ldots$. If $\sum \mu_i$ is Radon, $\sum N_i$ is PRM$(\sum \mu_i)$.

## 3.4. Vague Convergence

Weak convergence of point processes is a basic tool in the study of stochastic process behavior of extremes and records. In order to discuss weak convergence of point processes we need a notion of convergence in $M_p(E)$, and in fact we will show how to make $M_p(E)$ into a complete, separable metric space. There is little extra cost if we discuss these issues in the context of random measures.

Let $(E, \mathscr{E})$ be a state space as before and let $\rho$ be a metric on $E$ which makes $E$ a complete, separable metric space. We have need for additional notation which parallels and amplifies what was introduced in Section 3.1.

Let $C_K(E)$ be the continuous, real valued functions on $E$ with compact support so that $f \in C_K(E)$ means there exists a compact set $K$ and the continuous function satisfies $f(x) = 0$ for $x \in K^c$. $C_K^+(E)$ is the subset of $C_K(E)$ consisting of continuous, non-negative functions with compact support. Let $M_+(E)$ be all non-negative Radon measures on $(E, \mathscr{E})$ and define $\mathscr{M}_+(E)$ to be the smallest $\sigma$-algebra of subsets of $M_+(E)$ making the maps $m \to m(f) = \int_E f\,dm$ from $M_+(E) \to \mathbb{R}$ measurable for all $f \in C_K^+(E)$.

Some equivalent descriptions of $\mathscr{M}_+$ are

$$\begin{aligned}\mathscr{M}_+(E) &= \sigma\{\{m \in M_+(E) \colon m(f) \in B\}, f \in C_K^+(E), B \in \mathscr{B}([0,\infty])\}\\ &= \sigma\{m \to m(G), G \text{ open, relatively compact}\}\\ &= \sigma\{m \to m(G), G \in \mathscr{E}\}\end{aligned}$$

and so on. The monotone class arguments needed to verify the equivalent descriptions of $M_+$ will be omitted.

Since point measures are the primary objects of interest, it is natural to wonder whether $M_p(E)$ is measurable; i.e., is $M_p(E) \in \mathscr{M}_+(E)$? This is certainly the case, and a method to check this is outlined in Problem 3.4.2. Alternatively we will see that if $M_+(E)$ is topologized with the vague topology, $\mathscr{M}_+(E)$ coincides with the Borel $\sigma$-algebra (the $\sigma$-algebra generated by the vaguely open sets) and $M_p(E)$ is vaguely closed and hence measurable.

## *Random Measures*

$\xi$ is a random measure if it is a measurable map from a probability space $(\Omega, \mathscr{A}, P)$ into $(M_+(E), \mathscr{M}_+(E))$. If $\xi$ takes all its values in $M_p(E)$ then $\xi$ is a point process. The Laplace functional of $\xi$ is the map $\Psi_\xi$ from positive, $\mathscr{E}$-measurable functions to $[0, 1]$ defined by

$$\Psi_\xi(f) = E\exp\{-\xi(f)\} = \int_\Omega \exp\left\{-\int_E f(x)\xi(\omega, dx)\right\} P(d\omega)$$

$$= \int_{M_+(E)} \exp\left\{-\int_E f(x)m(dx)\right\} P_\xi(dm)$$

where $P_\xi = P \circ \xi^{-1}$. The Laplace functional restricted to $C_K^+(E)$ uniquely determines the distribution $P_\xi$ of the process $\xi$ (Exercise 3.4.3.; cf. Lemma 3.11).

Two random measures $\xi_1$, $\xi_2$ on $(\Omega, \mathscr{A}, P)$ have the same distribution if $P_{\xi_1} = P_{\xi_2}$. In this case we sometimes write $\xi_1 \stackrel{d}{=} \xi_2$. Check $\xi_1 \stackrel{d}{=} \xi_2$ iff for any $k$ and $f_i \in C_K^+(E)$, $i \le k$ we have

$$(\xi_1(f_i), i \le k) \stackrel{d}{=} (\xi_2(f_i), i \le k)) \text{ in } \mathbb{R}^k.$$

In order to discuss weak convergence of random measures and point processes we must topologize $M_+(E)$. Ideally this topology will be metrizable as a complete separable metric space to facilitate link-ups with standard accounts of weak convergence theory such as that of Billingsley (1968).

For $\mu_n$, $\mu \in M_+(E)$ we say $\mu_n$ converges vaguely to $\mu$ (written $\mu_n \xrightarrow{v} \mu$) if $\mu_n(f) \to \mu(f)$ for all $f \in C_K^+(E)$. A topology on $M_+$ giving this notion of convergence is obtained as follows: A sub-base for this topology consists of sets of the form

$$\{\mu \in M_+ : s < \mu(f) < t\}$$

for some $f \in C_K^+$ and $s < t$. Finite intersections of such sets (example, $\{\mu \in M_+ : \mu(f_i) \in (s_i, t_i), i = 1, \ldots, k\}$) form a basis, and open sets are obtained by taking unions of basis sets. A basic neighborhood of $\mu \in M_+$ is a set of the form

$$\{v \in M_+ : |v(f_i) - \mu(f_i)| < \varepsilon, i = 1, \ldots, k\}.$$

For a topological space such as $M_+(E)$, a natural $\sigma$-algebra is the Borel $\sigma$-algebra $\mathscr{B}(M_+(E))$, which is just the $\sigma$-algebra generated by open sets.

Since we have already discussed a $\sigma$-algebra, namely $\mathscr{M}_+(E)$, for $M_+(E)$ it is important to know the relation between the two. However

$$\mathscr{M}_+(E) = \mathscr{B}(M_+(E))$$

so that either description of the measure structure can be used, depending on convenience. Cf. Jagers, 1974, or Exercise 3.4.5.

Developments thus far seem to be emphasizing integrals over sets. From your knowledge of weak convergence of probability measures on $\mathbb{R}$ it is plausible that $\{\mu(f), f \in C_K^+(E)\}$ holds the same information as $\{\mu(A), A \in \mathscr{E}\}$. Integrals are frequently more convenient to deal with (at least theoretically). We interpret $\mu(f)$ as the value of $\mu$ at coordinate $f$. Compare this with a function $x(\cdot) = \{x(t), t \geq 0\}$ of a real variable where $x(t)$ is the value of $x(\cdot)$ at coordinate $t$.

Theoretical justification for looking at integrals rather than sets is given by the next result, which is a variant of Urysohn's lemma (e.g., Simmons, 1963, page 135) as presented in Kallenberg (1983).

**Lemma 3.11.** (a) *Let $K$ be compact. There exist compact $K_n \downarrow K$ and a nonincreasing sequence $\{f_n\}$, $f_n \in C_K^+(E)$ and*

$$1_K \leq f_n \leq 1_{K_n} \downarrow 1_K.$$

(b) *Let $G$ be open, relatively compact. There exist open, relatively compact $G_n \uparrow G$ and a nondecreasing sequence $\{f_n\}$, $f_n \in C_K^+(E)$ and $1_G \geq f_n \geq 1_{G_n} \uparrow 1_G$.*

Proof. We use the fact that if $E$ is locally compact with countable base, then it is metrizable as a complete, separable metric space. Call the metric $\rho$. For $B \subset E$ let $B^-$ be the closure and $B^\circ$ be the interior of $B$.

(a) Let $\{B_k\}$ be open, relatively compact and $B_k \uparrow E$. If $K$ is compact, $\{B_k\}$ is an open cover of $K$ and hence there exists $k_0$ such that $K \subset B_{k_0}$. Furthermore $\rho(K, B_{k_0}^c) > 0$ since if $\rho(K, B_{k_0}^c) = 0$, $K$ and $B_{k_0}^c$ both being closed, there would be $x \in K \cap B_{k_0}^c$ and this would contradict $K \subset B_{k_0}$. Suppose $\rho(K, B_{k_0}^c) > \varepsilon > 0$ and define

$$K^\delta = \{x \in E: \rho(x, K) \leq \delta\}$$

to be the $\delta$-swelling of $K$. For $n$ such that $n^{-1} \leq \varepsilon$ set $K_n = K^{1/n}$ and note that $K_n$ is closed and

$$K_n \subset B_{k_0} \subset B_{k_0}^-.$$

Since $B_{k_0}$ is relatively compact, $K_n$ is compact. Also $K_n \downarrow K$.

Define $f_n(x) = 1 - (n\rho(x, K) \wedge 1)$ so that $0 \leq f_n \leq 1$, $f_n \in C_K^+(E)$ and $\{f_n\}$ is nonincreasing. To check $1_{K_n} \geq f_n \geq 1_K$ observe first that on $K$, $1 = 1_{K_n} = f_n = 1_K$ so everywhere $f_n \geq 1_K$. To verify $1_{K_n} \geq f_n$ observe that

(i) If $x \in K_n$, then $1_{K_n} = 1 \geq f_n(x)$ and
(ii) If $x \in K_n^c$ then by definition of $K_n$, $\rho(x, K) > n^{-1}$ and $f_n(x) = 1 - (n\rho(x, K_n) \wedge 1) = 1 - 1 = 0$.

(b) Every open set in $E$ is an $F_\sigma$ and since $G$ is assumed relatively compact there exist compact $K_n \uparrow G$ from which it is easy to construct open, relatively compact $G_n \uparrow G$ (cf. Cohn, 1980, page 198) such that $G_n^- \subset G_{n+1}$. This last property forces $\rho(G_n, G^c) > 0$ since if not, then

$$0 = \rho(G_n, G^c) \geq \rho(G_n^-, G^c)$$

implies that there exists

$$x \in G_n^- \cap G^c \subset G_{n+1} \cap G^c = \varnothing,$$

a contradiction. Now set

$$g_n(x) = 1 - \left(\frac{\rho(x, G_n)}{\rho(G^c, G_n)}\right) \wedge 1$$

so that $0 \leq g_n \leq 1$. If $x \in G$, then for large $n$, $x \in G_n$ and hence $\rho(x, G_n) = 0$ and $g_n(x) = 1$. If $x \in G^c$ then for all $n$

$$\rho(x, G_n) \geq \inf_{y \in G^c} \rho(y, G_n) =: \rho(G^c, G_n)$$

so that $\rho(x, G_n)/\rho(G^c, G_n) \geq 1$ and $g_n(x) = 0$. If we define $f_n = \bigvee_{i=1}^n g_i$ then $\{f_n\}$ is nondecreasing and $1_G \geq f_n \geq 1_{G_n}$. $\square$

*Remark.* The proof of (a) shows that if $K$ is compact, then for all small $\delta > 0$, $K^\delta$ is compact.

With Lemma 3.11 under our belts, we can now give some interpretations of vague convergence in terms of sets.

**Proposition 3.12.** *Let $\mu, \mu_1, \mu_2, \ldots$ be in $M_+(E)$. The following are equivalent:*

(i) $\mu_n \xrightarrow{v} \mu$;
(ii) $\mu_n(B) \to \mu(B)$ *for all relatively compact $B$ for which $\mu(\partial B) = 0$; i.e., the boundary of $B$ has $\mu$ measure* 0;
(iii) $\limsup_{n\to\infty} \mu_n(K) \leq \mu(K)$ *and* $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *for all compact $K$ and all open, relatively compact $G$.*

Proof. (i) $\to$ (iii): If $K$ is compact then by Lemma 3.11 there exist compact $K_n$, $f_n \in C_K^+$, and

$$1_K \leq f_n \leq 1_{K_n} \downarrow 1_K.$$

Then for $m$ fixed

$$\limsup_{n\to\infty} \mu_n(K) \leq \limsup_{n\to\infty} \mu_n(f_m) = \mu(f_m).$$

Since $f_m \leq 1_{K_{m_0}}$ for $m \geq m_0$ and $1_{K_{m_0}}$ is $\mu$-integrable we find by dominated convergence that letting $m \to \infty$ gives $\limsup_{n\to\infty} \mu_n(K) \leq \mu(K)$ as desired. The second result for open, relatively compact $G$ is proved similarly by using Lemma 3.11(b).

(iii) → (ii): Suppose $B$ is relatively compact with $\mu(\partial B) = 0$. Then $B^\circ \subset B \subset B^-$, and by applying (iii) we get

$$\mu(B^\circ) \le \liminf_{n\to\infty} \mu_n(B^\circ) \le \liminf_{n\to\infty} \mu_n(B) \le \limsup_{n\to\infty} \mu_n(B) \le \lim_{n\to\infty} \mu_n(B^-) \le \mu(B^-)$$

and since $\mu(\partial B) = 0$ implies $\mu(B^\circ) = \mu(B) = \mu(B^-)$ (ii) follows.

(ii) → (i): Let $f \in C_K^+(E)$ and suppose $F$ is the support of $f$, so that $F$ is compact. We must show that $\mu_n(f) \to \mu(f)$.

A simple helpful preliminary is the following: Let $(E_i, \mathscr{E}_i)$, $i = 1, 2$ be two metric spaces and suppose $T: E_1 \to E_2$ is continuous. Then if $A_2 \in \mathscr{E}_2$

$$\partial(T^{-1}A_2) \subset T^{-1}(\partial A_2). \tag{3.8}$$

(Note on the left, $\partial$ is an operator on $E_1$ but on the right $\partial$ operates in $E_2$.) You may check this as Exercise 3.4.8.

Let $\hat{f}$ be the restriction of $f$ to $F$. Define

$$\Gamma_n := \{\gamma > 0: \mu\{x \in F: \hat{f}(x) = \gamma\} > n^{-1}\}$$

and $\Gamma := \bigcup_n \Gamma_n$. Now the sets $\{x \in F: \hat{f}(x) = \gamma\}$ are disjoint for different values of $\gamma$ so that $\Gamma_n$ is finite and hence $\Gamma$ is countable (remember that $\mu(F) < \infty$). Note that $\Gamma = \{\gamma > 0: \mu \circ \hat{f}^{-1}\{\gamma\} > 0\}$ is the set of atoms in $(0, \infty)$ of the measure $\mu \circ \hat{f}^{-1}$ on $\mathbb{R}_+$.

There exists $\beta > 0$, such that on $F$, $0 \le \hat{f} \le \beta$. Given $\varepsilon > 0$, there exist $\alpha_i \in \Gamma^c$, $0 \le i \le k$ with

$$0 = \alpha_0 < \alpha_1 < \cdots < \alpha_k = \beta$$

and

$$\sup_{1 \le i \le k} (\alpha_i - \alpha_{i-1}) \le \varepsilon. \tag{3.9}$$

On $F$,

$$\sum_1^k \alpha_{i-1} 1_{(\alpha_{i-1}, \alpha_i]}(\hat{f}(x)) \le \hat{f}(x) \le \sum_1^k \alpha_i 1_{(\alpha_{i-1}, \alpha_i]}(\hat{f}(x)) \tag{3.10}$$

and also $\{x \in F: \hat{f}(x) \in (\alpha_{i-1}, \alpha_i]\}$ is relatively compact since it is a subset of compact $F$. Finally observe that by (3.8)

$$\begin{aligned}\mu(\partial \hat{f}^{-1}(\alpha_{i-1}, \alpha_i]) &\le \mu(\hat{f}^{-1}(\partial(\alpha_{i-1}, \alpha_i]))\\ &= \mu(\hat{f}^{-1}\{\alpha_{i-1}, \alpha_i\}) = \mu(\hat{f}^{-1}\{\alpha_i\}) + \mu(\hat{f}^{-1}\{\alpha_{i-1}\}) = 0\end{aligned}$$

since $\alpha_{i-1}, \alpha_i \in \Gamma$.

From (3.10) we have

$$\sum_1^k \alpha_{i-1}\mu[\hat{f} \in (\alpha_{i-1}, \alpha_i]] \le \mu(f) = \int_F \hat{f}\, d\mu \le \sum_1^k \alpha_i \mu[\hat{f} \in (\alpha_{i-1}, \alpha_i]] \tag{3.11}$$

and the difference in the extremes of the inequalities is

$$\sum_1^k (\alpha_i - \alpha_{i-1})\mu[\hat{f} \in (\alpha_{i-1}, \alpha_i]] \le \varepsilon\mu(F) \tag{3.12}$$

from (3.9). Therefore

$$\begin{aligned}
\limsup_{n\to\infty} \mu_n(f) &= \limsup_{n\to\infty} \int_F \hat{f} d\mu_n \\
&\le \limsup_{n\to\infty} \int_F \sum_1^k \alpha_i 1_{[\hat{f}\in(\alpha_{i-1},\alpha_i]]}(x)\mu_n(dx) \qquad \text{from (3.10)} \\
&= \limsup_{n\to\infty} \sum_1^k \alpha_i \mu_n\{x: \hat{f}(x)\in(\alpha_{i-1},\alpha_i]\} \\
&= \sum_1^k \alpha_i \mu\{x: \hat{f}(x)\in(\alpha_{i-1},\alpha_i]\} \qquad \text{(applying (ii))} \\
&\le \mu(f) + \varepsilon\mu(F) \qquad \text{from (3.11) and (3.12).}
\end{aligned}$$

Similarly

$$\liminf_{n\to\infty} \mu_n(f) \ge \mu(f) - \varepsilon\mu(F)$$

and we find that $\mu_n(f) \to \mu(f)$; i.e., (i) holds. □

Vague convergence of point measures, $m_n \xrightarrow{v} m$, has the following interpretation in terms of convergence of the points of $m_n$ to the points of $m$.

**Proposition 3.13.** *Suppose $m_n$, $m \in M_p(E)$ and $m_n \xrightarrow{v} m$. For $K$ compact and satisfying $m(\partial K) = 0$ we have for $n \ge n(K)$ a labeling of the points of $m_n$ and $m$ in $K$ such that*

$$m_n(\cdot \cap K) = \sum_{i=1}^{p} \varepsilon_{x_i^{(n)}}, \qquad m(\cdot \cap K) = \sum_{i=1}^{p} \varepsilon_{x_i}$$

*and in $E^p$*

$$(x_i^{(n)}, 1 \le i \le p) \to (x_i, 1 \le i \le p) \qquad \text{as } n \to \infty.$$

*(Of course, $E^p$ has the product topology so that convergence of vectors in $E^p$ means componentwise convergence.)*

Proof. We may write

$$m(\cdot \cap K) = \sum_1^s c_r \varepsilon_{y_r}$$

where $y_1, \dots, y_s$ are the atoms of $m$ in $K$ (in fact, in $K^0$, since $m(\partial K) = 0$) and $c_1, \dots, c_s$ are integers giving multiplicities.

For each $y_r$ choose a neighborhood $G_r \subset K^0$, $G_1, \dots, G_s$ disjoint and $m(\partial G_r) = 0$. Then $\lim_{n\to\infty} m_n(G_r) = m(G_r)$, $1 \le r \le s$, and for $n$ sufficiently large, $n \ge n(K)$ say, $m_n(G_r) = m(G_r) = c_r$, $1 \le r \le s$, and $m_n(K) = m(K)$. Labeling points properly now gives the result. □

Next is a proposition which assures us that topological or metric information about $M_+(E)$ may be transferred to $M_p(E)$.

**Proposition 3.14.** *$M_p(E)$ is vaguely closed in $M_+(E)$.*

The following lemma precedes the proof.

**Lemma 3.15.** *Suppose $K$ is compact and $\mu \in M_+$. There exist $\delta_n \downarrow 0$ and $K^{\delta_n} \downarrow K$ and $\mu(\partial K^{\delta_n}) = 0$.*

PROOF OF LEMMA. By the remark following Proposition 3.11, $K^\delta$ is compact for all sufficiently small $\delta$, $\delta \le \delta_0$ say. Now $\mu(K^{\delta_0}) < \infty$ since $\mu$ is Radon. Also

$$\partial K^\delta \subset \{x\colon \rho(x, K) = \delta\}$$

and therefore

$$\{\partial K^\delta, 0 < \delta \le \delta_0\}$$

is a disjoint family of sets. It follows that

$$\{\delta \in (0, \delta_0]\colon \mu(\partial K^\delta) > n^{-1}\}$$

is finite since otherwise $\mu K^{\delta_0} = \infty$. Therefore

$$\{\delta \in (0, \delta_0]\colon \mu(\partial K^\delta) > 0\}$$

is countable. □

PROOF OF PROPOSITION 3.14. Suppose $\mu_n \in M_p(E)$, $\mu \in M_+(E)$, and $\mu_n \xrightarrow{v} \mu$. We show that $\mu \in M_p(E)$. Let $G_n$ be relatively compact, $G_n \uparrow E$, and $\mu(\partial G_n) = 0$. (If your favorite covering sequence happens to have $\mu$-mass on a boundary, Lemma 3.15 assures us that by swelling a bit we can find an equally serviceable covering set with no mass on its boundary.) Define

$$\mathscr{B}_\mu = \{B \in \mathscr{E}\colon B \text{ is relatively compact, } \mu(\partial B) = 0\}.$$

Observe that $\mathscr{B}_\mu$ is a $\Pi$-system since for $B_i \in \mathscr{E}$, $i = 1, 2$

$$\partial(B_1 \cap B_2) \subset \partial B_1 \cup \partial B_2.$$

Let

$$\mathscr{G}_n = \{B \in \mathscr{E}\colon \mu(B \cap G_n) \text{ is a non-negative integer}\}.$$

By Proposition 3.12

$$\mu_n(B) \to \mu(B), \qquad B \in \mathscr{B}_\mu, \tag{3.13}$$

so that $\mu(B)$ is a non-negative integer and therefore since $B \in \mathscr{B}_\mu$ implies $B \cap G_n \in \mathscr{B}_\mu$ we have $\mathscr{G}_n \supset \mathscr{B}_\mu$. One readily checks that $\mathscr{G}_n$ is a $\lambda$-system and hence by Dynkin's theorem

$$\mathscr{G}_n \supset \sigma(\mathscr{B}_\mu) = \mathscr{E}.$$

(Why is $\sigma(\mathscr{B}_\mu) = \mathscr{E}$? If $K$ is compact, Lemma 3.15 assures us there are $K_n \downarrow K$, $K_n \in \mathscr{B}_\mu$. Therefore $\sigma(\mathscr{B}_\mu) \supset \sigma$(compact subsets of $E$) $= \mathscr{E}$.) We conclude that $\mu(A \cap G_n)$ is integer valued for any $A \in \mathscr{E}$ and any $n$. Hence $\mu$ is a point measure. $\square$

Now we consider a criterion for relative compactness in $M_p$ or $M_+$.

**Proposition 3.16.** *A subset $M$ of $M_p(E)$ or $M_+(E)$ is vaguely relatively compact*

$$\text{iff } \sup_{\mu \in M} \mu(f) < \infty \qquad \textit{for each } f \in C_K^+(E)$$

$$\text{iff } \sup_{\mu \in M} \mu(B) < \infty \qquad \textit{for each relatively compact } B \in \mathscr{E}.$$

Proof. Since $M_p(E)$ is a closed subset of $M_+(E)$, it suffices to prove the result for $M_+(E)$. We prove only the first equivalence; the second one is left as an exercise.

We observe first that for each $f \in C_K^+(E)$

$$\sup_{\mu \in M} \mu(f) = \sup_{\mu \in M^-} \mu(f).$$

This is almost obvious since the map $T_f\mu = \mu(f)$ is continuous; here are details. Note that there are $\mu_n \in M^-$ such that $\mu_n(f) \to \sup_{\mu \in M^-} \mu(f)$. By the definition of the vague topology there exists for each $n$, $\nu_n \in M$ such that

$$|\mu_n(f) - \nu_n(f)| < 2^{-n}.$$

Therefore

$$\sup_{\mu \in M} \mu(f) \geq \nu_n(f) \to \sup_{\mu \in M^-} \mu(f) \geq \sup_{\mu \in M} \mu(f)$$

and the result follows.

If $M^-$ is compact then for $f \in C_K^+(E)$ define $T_f\colon M_+ \to [0, \infty)$ by $T_f\mu = \mu(f)$. The map $T_f$ is continuous on $M_+$ by definition of the vague topology, and hence the image of $M^-$ under $T_f$ is compact; i.e., $\{T_f\mu, \mu \in M^-\} = \{\mu(f), \mu \in M^-\}$ is compact in $[0, \infty)$. Compact sets on $[0, \infty)$ are bounded so $\sup_{\mu \in M^-} \mu(f) < \infty$.

For the converse, suppose that for each $f \in C_K^+(E)$, $\sup_{\mu \in M^-} \mu(f) < \infty$. Then

$$I_f := [0, \sup_{\mu \in M^-} \mu(f)]$$

is a compact subset of $[0, \infty)$ and hence by Tychonoff's theorem (Simmons, 1963, page 119)

$$I := \prod_{f \in C_K^+(E)} I_f$$

is a compact subset of $\mathbb{R}^{C_K^+(E)} = \prod_{f \in C_K^+(E)} \mathbb{R}$ with product topology.

Identify $\mu \in M^-$ with $\{\mu(f): f \in C_K^+\} \in I$. This defines a map $T$ from $M^-$ into $I$, and the topologies on $M^-$ and on $T(M^-)$ coincide since $\mu_n \to \mu$ in $M^-$ means $\mu_n(f) \to \mu(f)$ for all $f \in C_K^+(E)$ and the latter statement is just componentwise convergence in the product topology on $I$. Thus $M^-$ and $T(M^-)$ are homeomorphic. Since $M^-$ is closed, $T(M^-)$ is a closed subset of $I$ and hence $T(M^-)$ is compact. Thus so too is $M^-$ a compact set. □

We now metrize $M_p(E)$ and $M_+(E)$.

**Proposition 3.17.** *The vague topology on $M_p$ or $M_+$ is metrizable as a complete, separable metric space.*

PROOF. From Proposition 3.14 it is enough to consider $M_+$. The idea of the proof is to find a countable collection $\{h_l\}$ in $C_K^+(E)$ such that if $\mu_n$, $\mu \in M_+(E)$ we have $\mu_n \xrightarrow{v} \mu$ iff $\mu_n(h_l) \to \mu(h_l)$ for all $l$. An explicit construction (cf. Kallenberg (1983)) is as follows.

Let $\{G_i, i \geq 1\}$ be a countable base of relatively compact sets and suppose without loss of generality that $\{G_i\}$ is closed under finite unions and finite intersections. (If the $G$-family you started with does not have this property, switch to finite unions of sets of the form $\bigcap_{i \in I} G_i$ where $I$ is a finite subset of the non-negative integers.) From the approximation Lemma 3.11 there exist $f_{i,n}, g_{i,n} \in C_K^+(E)$ and

$$\lim_{n\to\infty} \uparrow f_{i,n} = 1_{G_i}, \qquad \lim_{n\to\infty} \downarrow g_{i,n} = 1_{G_i^-}.$$

Enumerate $\{f_{i,n}, g_{i,n}, i \geq 1, n \geq 1\}$ as $\{h_1, h_2, \ldots\}$.

Any $\mu \in M_+(E)$ is uniquely determined by $\{\mu(h_l), l \geq 1\}$. For any $i$, $\mu(G_i) = \lim_{n\to\infty} \uparrow \mu(f_{i,n})$ and hence $\{\mu(h_l)\}$ determines $\{\mu(G_i)\}$. Since $\{G_i\}$ is a $\Pi$-system generating the $\sigma$-algebra $\mathscr{E}$, $\mu$ is determined everywhere.

Next we observe that for $\mu_n$, $\mu \in M_+(E)$, $\mu_n \xrightarrow{v} \mu$ iff

$$\text{For each } l, \text{ there exists a finite positive constant } c_l \text{ and } \mu_n(h_l) \to c_l. \text{ In this case } c_l = \mu(h_l). \tag{3.14}$$

To verify this we suppose that (3.14) holds and as a first step to showing $\mu_n \xrightarrow{v} \mu$ we show that $\{\mu_n\}$ is relatively compact in $M_+(E)$. We proceed by means of the criterion in Proposition 3.16. Take any $f \in C_K^+(E)$ and suppose the compact set $K$ is the support of $f$. Compactness implies there is a finite subset $I$ of the integers and

$$K \subset \bigcup_{i \in I} G_i.$$

Since $\{G_i\}$ is closed under finite unions, for some $i_0$, $G_{i_0} = \bigcup_{i \in I} G_i$ and $K \subset G_{i_0}$. Therefore if $\|f\| = \sup\{f(x): x \in E\}$ we have

$$f \leq \|f\| 1_K \leq \|f\| 1_{G_{i_0}} \leq \|f\| 1_{G_{i_0}^-} \leq \|f\| g_{i_0,k}$$

for any fixed $k$ and so

$$\sup_n \mu_n(f) \leq \|f\| \sup_n \mu_n(g_{i_0,k}) < \infty$$

since $\{\mu_n(h_l)\}$ is a convergent sequence for each $l$ and hence bounded. So Proposition 3.16 implies that $\{\mu_n\}$ is relatively compact.

Since $\{\mu_n\}$ is relatively compact, there exists a subsequence $\{n'\}$ and $\mu \in M_+(E)$ such that $\mu_{n'} \xrightarrow{v} \mu$ as $n' \to \infty$. So we have on the one hand

$$\mu_{n'}(h_l) \to c_l$$

and on the other

$$\mu_{n'}(h_l) \to \mu(h_l).$$

Any limit point $\mu$ thus has the property $\mu(h_l) = c_l$ and since any measure in $\mu_+(E)$ is uniquely determined by its values on $\{h_l\}$ we conclude that all limit points of $\{\mu_n\}$ must be the same and hence $\mu_n \xrightarrow{v} \mu$ as desired.

Since $\mu_n(h_l) \to \mu(h_l)$ for all $l$ is enough for $\mu_n \xrightarrow{v} \mu$, the following is a satisfactory metric on $M_+(E)$ for the vague topology: For $\mu$, $\mu' \in M_+(E)$ define

$$d(\mu, \mu') = \sum_{i=1}^{\infty} 2^{-i}(1 - \exp\{-|\mu(h_i) - \mu'(h_i)|\}).$$

The metric is complete: If $d(\mu_n, \mu_m) \to 0$ then for each $l$ we have

$$|\mu_n(h_l) - \mu_m(h_l)| \to 0$$

and so $\{\mu_n(h_l), n \geq 1\}$ is a Cauchy sequence of real numbers. Consequently there exists $c_l$: $\lim_{n\to\infty} \mu_n(h_l)$ and from (3.14) we conclude there exists $\mu \in M_+(E)$ with $\mu_n \xrightarrow{v} \mu$; i.e., $d(\mu_n, \mu) \to 0$. The metric is also separable: A countable base is

$$\{\{\mu \in M_+ \colon \mu(h_l) \in (r_i, r_j)\}, l = 1, 2, \ldots, r_i, r_j \text{ rational}\}. \qquad \square$$

Applications of weak convergence theory, discussed later, require a knowledge of what functionals are continuous. The following is very useful.

**Proposition 3.18.** *Suppose that E, E′ are two spaces which are locally compact with countable bases.*

*Suppose $T\colon E \to E'$ is continuous and satisfies*

$$T^{-1}(K') \text{ is compact in } E \text{ for every compact } K' \text{ in } E'. \tag{3.15}$$

*Then $\hat{T}\colon M_+(E) \to M_+(E')$ defined by*

$$\hat{T}\mu = \mu \circ T^{-1}$$

*is continuous.*

*Note $\hat{T}$ restricted to $M_p(E)$ is of the form*

$$\hat{T}\left(\sum \varepsilon_{x_i}\right) = \sum \varepsilon_{Tx_i}$$

*so that a continuous function on the points which also satisfies* (3.15) *induces a continuous function on the point measures.*

PROOF. For $\mu_n$, $\mu \in M_+(E)$ suppose $\mu_n \xrightarrow{v} \mu$. If $f \in C_K^+(E')$ then

$$\hat{T}\mu_n(f) = \mu_n \circ T^{-1}(f) = \mu_n(f \circ T).$$

Because of (3.15) we have $f \circ T \in C_K^+(E)$ since the support of $f \circ T$ is $T^{-1}$ (support of $f$) which is compact in $E$. Therefore $\mu_n \xrightarrow{v} \mu$ implies

$$\hat{T}\mu_n(f) = \mu_n(f \circ T) \to \mu(f \circ T) = \mu \circ T^{-1}(f)$$

and thus

$$\hat{T}\mu_n \xrightarrow{v} \hat{T}\mu. \qquad \square$$

*Remark.* For continuous $T$, the compactness requirement (3.15) is satisfied whenever either $T$ is a homeomorphism or $E$ is compact. Sometimes neither is the case, and if one is desperate to apply Proposition 3.18 an approximation procedure which restricts attention to a compact subset of $E$ must be applied.

## EXERCISES

3.4.1. Verify that the different descriptions of $\mathscr{M}_+(E)$ are equivalent.

3.4.2. Let $G_j, j \geq 1$ be a countable base of the topology of $E$ such that $G_j$ is open and relatively compact. (For example, if $E = \mathbb{R}^d$ then $\{G_j\}$ can be taken to be open rectangles with rational vertices.)

(a) The only positive Radon measures on $E$ which take non-negative integer values on $G_j$, for each $j$, are the point measures.
Hint: If $m$ is such a measure, then for all $x \in E$ there is an open neighborhood $V_x \in \{G_j\}$ of $x$ such that $m\{x\} = m(V_x)$. For compact $K$, a finite collection $S = \{x_1, \ldots, x_k\}$ of points of $K$ exists such that $m\{x\} \geq 1$, $x \in S$, and if $A \cap S = \phi$, $A$ closed and $A \subset K$, then $m(A) = 0$. Write

$$m(\cdot \cap K) = \sum_{x \in S} m\{x\}\varepsilon_x.$$

(b) Check $M_p(E) = \bigcap_{j=1}^{\infty} \{m \in M_+(E)\colon m(G_j) \text{ is non-negative integer valued}\}$ so that $M_p(E) \in \mathscr{M}_+(E)$ and

$$\begin{aligned} \mathscr{M}_p(E) &= \mathscr{M}_+(E) \cap M_p(E) \\ &:= \{A \cap M_p(E)\colon A \in \mathscr{M}_+(E)\} \end{aligned}$$

(Neveu, 1976).

3.4.3. Use Lemma 3.4.3 to prove that the Laplace functional of a random measure $\xi$ restricted to $C_K^+(E)$ uniquely determines the distribution of $\xi$.

3.4.4. Two random measures $\xi_1$ and $\xi_2$ are equal in distribution if for every $k$ and $f_1, \ldots, f_k \in C_K^+(E)$ we have in $\mathbb{R}^k$

$$(\xi_1(f_i), 1 \le i \le k) = (\xi_2(f_i), 1 \le i \le k).$$

Formulate and prove an analogous result with sets replacing the $f$'s.

3.4.5. Prove $\mathscr{M}_+(E) = \mathscr{B}(M_+(E))$. Hint: Since $m \to m(f)$ is $\mathscr{M}_+(E)$ measurable, every basis set of $M_+(E)$ is $\mathscr{M}_+(E)$ measurable. Since $M_+(E)$ has a countable basis, $\mathscr{B}(M_+(E)) \subset \mathscr{M}_+(E)$. For the converse, $m \to m(f)$ is continuous and hence $m \to m(f)$ is $\mathscr{B}(M_+(E))$ measurable for each $f \in C_K^+(E)$. So $m \to m(G)$ is $\mathscr{B}(M_+(E))$ measurable for each relatively compact, open $G$ (use Lemma 3.11) and hence $\mathscr{M}_+(E) \subset \mathscr{B}(M_+(E))$.

3.4.6. Check (3.8).

3.4.7. Check $\sigma$(compact subsets of $E$) $= \mathscr{E}$.

3.4.8. Prove the second equivalence in Proposition 3.16.

3.4.9. We know $M_+(E)$ is a complete, separable metric space. Is it locally compact?

3.4.10. Show that the following transformations are vaguely continuous:
(i) $T_1\colon M_+(E) \times M_+(E) \to M_+(E)$
$T_1(\mu, \nu) = \mu + \nu$
(ii) $T_2\colon M_+(E) \times (0, \infty) \to M_+(E)$
$T_2(\mu, \lambda) = \lambda\mu$
(Kallenberg, 1983).

3.4.11. If $K \in \mathscr{E}$ is compact, show that $\{\mu \in M_+(E)\colon \mu(K) < t\}$ is open in $M_+(E)$. If $G$ is open, relatively compact show $\{\mu \in M_+(E)\colon \mu(G) > t\}$ is open in $M_+(E)$. Cf. Proposition 3.12 (Kallenberg, 1983).

3.4.12. Suppose $x_n, x \in E$, $c_n \ge 0$, $c > 0$. Then

$$c_n \varepsilon_{x_n} \xrightarrow{v} c\varepsilon_x$$

iff $c_n \to c$ and $x_n \to x$ (Kallenberg, 1983).

3.4.13. If $E = (0, \infty)^2$, is (3.15) satisfied for the following $T$:

$$T(x, y) = x + y$$

$$T(x, y) = xy?$$

If $E = (0, \infty)$ how about $Tx = 2x$?

3.4.14. Let $m_n = \sum_{i=1}^{\infty} n^{-1}\varepsilon_{i/n}$ be a discrete version of Lebesque measure $m$ on $[0, \infty]$. Show $m_n \xrightarrow{v} m$.

## 3.5. Weak Convergence of Point Processes and Random Measures

We first review some facts from the theory of weak convergence in metric spaces (cf. Billingsley, 1968).

Let $S$ be a complete, separable metric space with metric $d$ and let $\mathscr{S}$ be the Borel $\sigma$-algebra of subsets of $S$ generated by open sets. Suppose $(\Omega, \mathscr{A}, \mathbf{P})$ is a probability space. A random element $X$ in $S$ is a measurable map from such a space $(\Omega, \mathscr{A})$ into $(S, \mathscr{S})$. The most common examples are

| | |
|---|---|
| $S = \mathbb{R}$ | $X =$ random variable |
| $= \mathbb{R}^d$ | $=$ random vector |
| $= C[0, \infty)$, the space of real valued, continuous functions on $[0, \infty)$ | $=$ random process with continuous paths |
| $= D[0, \infty)$, the space of real valued, right continuous functions on $[0, \infty)$ with finite left limits existing on $(0, \infty)$ | $=$ random process with jump discontinuities |
| $= M_p(E)$ | $=$ stochastic point process |
| $= M_+(E)$ | $=$ random measure. |

Given a sequence $\{X_n, n \geq 0\}$ of random elements, there is a corresponding sequence of distributions

$$P_n = \mathbf{P} \circ X_n^{-1} \quad \text{on} \quad \mathscr{S}, \qquad n \geq 0.$$

Then $X_n$ converges weakly to $X_0$ (written $X_n \Rightarrow X_0$ or $P_n \Rightarrow P_0$) if whenever $f \in C(S)$, the class of bounded, continuous real valued functions on $S$, we have

$$\mathbf{E}f(X_n) = \int_S f(x) P_n(dx) \to Ef(X_0) = \int_S f(x) P_0(dx).$$

This is equivalent (Billingsley, 1968, page 12; cf. Proposition 3.12) to

$$\lim_{n\to\infty} \mathbf{P}[X_n \in A] = \mathbf{P}[X_0 \in A] \qquad \text{for all } A \in \mathscr{S} \tag{3.16}$$

such that $\mathbf{P}[X_0 \in \partial A] = 0$ or

$$\limsup_{n\to\infty} \mathbf{P}[X_n \in F] \leq \mathbf{P}[X_0 \in F] \qquad \text{for all closed } F \in \mathscr{S} \quad \text{or} \tag{3.17}$$

$$\liminf_{n\to\infty} \mathbf{P}[X_n \in G] \geq \mathbf{P}[X_0 \in G] \qquad \text{for all open } G \in \mathscr{S}. \tag{3.18}$$

A nice way to think about weak convergence is from *Skorohod's theorem* (Billingsley, 1971; cf. Proposition 0.2): $X_n \Rightarrow X_0$ iff there exist random elements $\{X_n^*, n \geq 0\}$ on the uniform probability space $([0,1], \mathscr{B}[0,1], m)$ where $m$ is Lebesgue measure such that

$$X_n \stackrel{\mathrm{d}}{=} X_n^* \qquad \text{for each } n \geq 0$$

and

$$X_n^* \to X_0^* \qquad \text{a.s.}$$

The second statement means

$$m\left\{t\in[0,1]: \lim_{n\to\infty} d(X_n^*(t), X_0^*(t)) = 0\right\} = 1.$$

The power of weak convergence theory comes from the fact that once a basic convergence result has been proved, many corollaries emerge with little effort. The arguments usually involve continuity. Suppose $(S_i, d_i)$, $i = 1, 2$, are metric spaces and $h\colon S_1 \to S_2$ is continuous. If $X_n$, $n \geq 0$ are random elements in $(S_1, \mathscr{S}_1)$ and $X_n \Rightarrow X_0$ then $h(X_n) \Rightarrow h(X_0)$ in $(S_2, \mathscr{S}_2)$. To check this is easy: Let $f_2 \in C(S_2)$ and we must show that $Ef_2(h(X_n) \to Ef_2(h(X_0))$. But $f_2(h(X_n)) = f_2 \circ h(X_n)$ and since $f_2 \circ h \in C(S_1)$ the result follows from $X_n \Rightarrow X_0$.

In fact $h$ need not be continuous everywhere.

**Continuous Mapping Theorem** (Billingsley, 1968, page 30). *Let $(S_i, d_i)$, $i = 1, 2$ be two metric spaces and suppose $X_n$, $n \geq 0$ are random elements of $(S_1, \mathscr{S}_1)$* and $X_n \Rightarrow X_0$. *If $h\colon S_1 \to S_2$ satisfies*

$$\mathbf{P}[X_0 \in D_h] = \mathbf{P}[X_0 \in \{s_1 \in S_1\colon h \text{ is discontinuous at } s_1\}] = 0$$

*then*

$$h(X_n) \Rightarrow h(X_0) \qquad \text{in } S_2.$$

This is an immediate consequence of Skorohod's theorem. Alternatively proceed from first principles as follows. Let $F_2 \in \mathscr{S}_2$ be closed and we must show that

$$\limsup_{n\to\infty} \mathbf{P}[h(X_n) \in F_2] \leq \mathbf{P}[X_0 \in F_2].$$

But

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}[h(X_n) \in F_2] &= \limsup_{n\to\infty} \mathbf{P}[X_n \in h^{-1}(F_2)] \\ &\leq \limsup_{n\to\infty} \mathbf{P}[X_n \in (h^{-1}(F_2))^-] \leq \mathbf{P}|X_0 \in (h^{-1}(F_2))^-].\end{aligned}$$

However, $(h^{-1}(F_2))^- \subset D_h \cup h^{-1}(F_2)$ since if $s_1 \in (h^{-1}(F_2))^- \cap D_h^c$ there exist $s_n \in h^{-1}(F_2)$ and $s_n \to s_1$ implying $h(s_n) \to h(s_1)$. Since $h(s_n) \in F_2$ and $F_2$ is closed we get $h(s_1) \in F_2$ and therefore $s_1 \in h^{-1}(F_2)$. Finally

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}[h(X_n) \in F_2] &\leq \mathbf{P}[X_0 \in D_h \cup h^{-1}(F_2)] = \mathbf{P}[X_0 \in h^{-1}(F_2)] \\ &= \mathbf{P}[h(X_0) \in F_2]\end{aligned}$$

and so weak convergence follows by equivalence (b) subsequent to the definition of weak convergence.

Often to prove weak convergence subsequence arguments are used and the following is useful. A family $\Pi$ of probability measures on a complete, separable metric space is relatively compact if every sequence $\{P_n\} \subset \Pi$ contains a weakly convergent subsequence. Relative compactness is theoretically useful but hard to check in practice so we need a workable criterion. Call the family $\Pi$ tight (and by abuse of language we will refer to the corresponding random elements also as a tight family) if for any $\varepsilon$, there exists a compact $K_\varepsilon \in \mathscr{S}$ such that

$$P(K_\varepsilon) > 1 - \varepsilon \qquad \text{for all } P \in \Pi.$$

This is the sort of condition that precludes probability mass from escaping from the state space. Prohorov's theorem (Billingsley, 1968 or Williams, 1979) assures us that when $S$ is separable and complete, tightness of $\Pi$ is the same as relative compactness.

We now are prepared to give a usable criterion for weak convergence in $M_+(E)$ or $M_p(E)$.

**Proposition 3.19.** *Let $P_n, n \geq 0$ be probability measures on $M_+(E)$. Then $P_n \Rightarrow P_0$* iff *Laplace functionals converge, i.e.,* iff *for any $f \in C_K^+(E)$, $\Psi_{P_n}(f) \to \Psi_{P_0}(f)$. Equivalently if $\xi_n, n \geq 0$ are random measures (i.e., random elements in $M_+(E)$), then $\xi_n \Rightarrow \xi_0$* iff *$\Psi_{\xi_n}(f) \to \Psi_{\xi_0}(f)$, for all $f \in C_K^+(E)$.*

Proof. Suppose $\xi_n \Rightarrow \xi_0$ in $M_+(E)$. For $f \in C_K^+(E)$ define the continuous map $T_f$: $M_+(E) \to [0, \infty)$ by $T_f \mu = \mu(f)$. The continuous mapping theorem assures us that $\xi_n \Rightarrow \xi_0$ entails $T_f \xi_n = \xi_n(f) \Rightarrow T_f \xi_0 = \xi_0(f)$. By the dominated convergence theorem

$$\Psi_{\xi_n}(f) = E\exp\{-\xi_n(f)\} \to E\exp\{-\xi_0(f)\} = \Psi_{\xi_0}(f).$$

For the converse we need the following lemma. □

**Lemma 3.20.** *$\{\xi_n\}$ is tight in $M_+(E)$* iff *for any $f \in C_K^+(E)$ we have $\{\xi_n(f)\}$ tight in $\mathbb{R}$.*

Proof. Suppose $\{\xi_n(f)\}$ is tight in $\mathbb{R}$ for any $f \in C_K^+(E)$. Pick $g_i \in C_K^+(E)$ with $g_i \uparrow 1$. Since $\{\xi_n(g_i), n \geq 1\}$ is tight, for any $\varepsilon$ there exists $c_i$ large enough that for all $n$

$$\mathbf{P}[\xi_n(g_i) > c_i] \leq \varepsilon/2^{i+1}.$$

The set $M = \bigcap_{i \geq 1}\{\mu \in M_+(E): \mu(g_i) \leq c_i\}$ is relatively compact by Proposition 3.16 since for any $f \in C_K^+(E)$

$$\sup_{\mu \in M} \mu(f) < \infty.$$

To see this, note that given $f$, there exists some $i_0$ and a constant $K_0$ such that $f \leq K_0 g_{i_0}$ and hence

$$\sup_{\mu \in M} \mu(f) \leq K_0 \sup_{\mu \in M} \mu(g_{i_0}) \leq K_0 c_{i_0}.$$

Furthermore for all $n$

$$\mathbf{P}[\xi_n \notin M^-] \le \mathbf{P}[\xi_n \notin M] = \mathbf{P}\left\{\bigcup_{i\ge 1} [\xi_n(g_i) > c_i]\right\} \le \sum_i \mathbf{P}[\xi_n(g_i) > c_i]$$

$$\le \sum_1^\infty \varepsilon/2^{i+1} = \varepsilon,$$

which shows that $\{\xi_n\}$ is tight. The converse is an easy exercise. □

Now for the converse to Proposition 3.19. Suppose $\Psi_{\xi_n}(f) \to \Psi_{\xi_0}(f)$, for all $f \in C_K^+(E)$. Then for any $\lambda > 0$, replacing $f$ by $\lambda f$ gives

$$E\exp\{-\lambda\xi_n(f)\} = E\exp\{-\xi_n(\lambda f)\} = \Psi_{\xi_n}(\lambda f)$$
$$\to \Psi_{\xi_0}(\lambda f) = E\exp\{-\xi_0(\lambda f)\} = E\exp\{-\lambda\xi_0(f)\}.$$

Thus the Laplace transform of the random variable $\xi_n(f)$ converges to the transform of $\xi_0(f)$ and hence $\xi_n(f) \Rightarrow \xi_0(f)$ in $\mathbb{R}$. A convergent sequence is certainly relatively compact and hence, by Prohorov, $\{\xi_n(f)\}$ is tight. From the lemma $\{\xi_n\}$ is tight in $M_+(E)$ and hence relatively compact in $M_+(E)$. So given any subsequence $\{n''\} \subset \{n\}$ there exists a further subsequence $\{n'\} \subset \{n''\}$ and for some random measure $\xi \in M_+(E)$

$$\xi_{n'} \Rightarrow \xi.$$

From the first half of this proposition we get $\Psi_{\xi_{n'}}(f) \Rightarrow \Psi_\xi(f)$ for every $f \in C_K^+(E)$. By assumption $\Psi_{\xi_n}(f) \to \Psi_{\xi_0}(f)$ and hence $\Psi_\xi(f) = \Psi_{\xi_0}(f)$. This means $\xi \stackrel{d}{=} \xi_0$ and so all weak subsequential limits of $\{\xi_n\}$ are equal in distribution to $\xi_0$ and hence $\xi_n \Rightarrow \xi_0$ in $M_+(E)$ as desired.

Typical of the applications of Proposition 3.19 is the next result, which, though rather simple, has far-reaching implications and provides the link between regular variation and point processes to be discussed in the next chapter.

**Proposition 3.21.** *For each n suppose $\{X_{n,j}, j \ge 1\}$ are iid random elements of $(E, \mathscr{E})$ and $\mu$ is a Radon measure on $(E, \mathscr{E})$. Define $\xi_n := \sum_{j=1}^\infty \varepsilon_{(jn^{-1}, X_{n,j})}$ and suppose $\xi$ is* PRM *on $[0,\infty) \times E$ with mean measure $dt \times d\mu$. Then $\xi_n \Rightarrow \xi$ in* $M_p([0,\infty) \times E)$ iff

$$nP[X_{n,1} \in \cdot\,] \xrightarrow{v} \mu \qquad \text{on } E. \tag{3.19}$$

Proof. As a warm-up, we prove the following simpler result: If $N$ is PRM($\mu$) on $E$ then

$$N_n = \sum_{j=1}^n \varepsilon_{X_{n,j}} \Rightarrow N \qquad \text{in } M_p(E)$$

iff (3.19) holds. To see this we show convergence of Laplace functionals so if $f \in C_K^+(E)$ we have

$$\begin{aligned}\psi_{N_n}(f) &= E\exp\{-N_n(f)\} = E\exp\left\{-\sum_{j=1}^{n} f(X_{n,j})\right\}\\ &= (E\exp\{-f(X_{n,1})\})^n\\ &= \left(1 - \frac{\int_E(1-e^{-f(x)})nP[X_{n,1}\in dx]}{n}\right)^n\end{aligned}$$

and this converges to

$$\exp\left\{-\int_E (1-e^{-f(x)})\mu(dx)\right\},$$

the Laplace functional of PRM$(\mu)$ (cf. (3.4)) iff (3.19) holds.

This illustrates the method in a simple context, and we now concentrate on showing the full result of Proposition 3.21. For $f \in C_K^+([0,\infty) \times E)$, the Laplace functional of $\xi_n$ is

$$\begin{aligned}\Psi_{\xi_n}(f) = E\exp\{-\xi_n(f)\} &= E\exp\left\{-\sum_k f(kn^{-1}, X_{n,k})\right\}\\ &= \prod_k\left(1 - \int_E (1-e^{-f(kn^{-1},x)})\mathbf{P}[X_{n,1}\in dx]\right)\end{aligned}$$

and $\Psi_{\xi_n}(f) \to \Psi_\xi(f)$ iff

$$\begin{aligned}&-\log\Psi_{\xi_n}(f)\\ &= -\sum_k \log\left(1 - \int_E (1-e^{-f(kn^{-1},x)})\mathbf{P}[X_{n,1}\in dx]\right) \to -\log\Psi_\xi(f).\end{aligned}$$

Suppose (3.19) holds. Define $\lambda_n$ by

$$\lambda_n(ds,dx) = \sum_k \varepsilon_{kn^{-1}}(ds)\mathbf{P}[X_{n,1}\in dx]$$

so that by (3.19)

$$\lambda_n(ds,dx) \xrightarrow{v} ds\mu(dx).$$

Then

$$\begin{aligned}&\sum_k \int_E (1-e^{-f(kn^{-1},x)})\mathbf{P}[X_{n,1}\in dx]\\ &= \iint_{[0,\infty)\times E}(1-e^{-f})d\lambda_n \to \iint (1-e^{-f(s,x)})ds\mu(dx) \qquad (3.20)\end{aligned}$$

as $n \to \infty$. Furthermore if $K$ is the compact support of $f$ in $[0,\infty) \times E$, there is some compact $A \subset E$ such that

$$\sup_{k\geq 1}\int_E (1-e^{-f(kn^{-1},x)})\mathbf{P}[X_{n,1}\in dx] \leq \mathbf{P}[X_{n,1}\in A] \to 0 \qquad (3.21)$$

as $n \to \infty$ by (3.19).

From the elementary expansion

$$\log(1+z) = z(1+\varepsilon(z)), \qquad |\varepsilon(z)| \le |z| \quad \text{if} \quad |z| \le 1/2$$

we have

$$\left| -\log \Psi_{\xi_n}(f) - \sum_k \int_E (1 - e^{-f(kn^{-1},x)}) \mathbf{P}[X_{n,1} \in dx] \right|$$

$$\le \sum_{k=1}^{n} \left( \int_E (1 - e^{-f(kn^{-1},x)}) \mathbf{P}[x_{n,1} \in dx] \right)^2$$

(for $n$ sufficiently large)

$$\le \left( \sup_{k \ge 1} \int_E (1 - e^{-f(kn^{-1},x)}) \mathbf{P}[X_{n,1} \in dx] \right)$$

$$\times \sum_{k=1}^{\infty} \int_E (1 - e^{-f(kn^{-1},x)}) \mathbf{P}[X_{n,1} \in dx] \to 0$$

as $n \to \infty$ by 3.20 and (3.21). Therefore if (3.19) holds $\Psi_{\xi_n}(f) \to \Psi_\xi(f)$ by (3.20) since the right side of (3.20) is $-\log$ of the Laplace functional of PRM with mean measure $ds \times d\mu$.

Conversely if we know $\Psi_{\xi_n}(f) \to \Psi_\xi(f)$ for all $f \in C_K^+([0,\infty) \times E)$, set $f(s,x) = 1_{[0,1]}(s) g(x)$ where $g \in C_K^+$ and we get

$$\Psi_{\xi_n}(f) = E \exp\left\{ -\sum_{k=1}^{n} g(X_{n,k}) \right\} \to \exp\left\{ -\int_E (1 - e^{-g}) d\mu \right\}. \tag{3.22}$$

Note this $f$ is not in $C_k^+([0,\infty) \times E)$ but by writing

$$h_l^-(s) g(x) \le f(s,x) \le h_l^+(s) g(x)$$

where $0 \le h_l^-(s) \uparrow 1_{[0,1]}(s)$, $h_l^+(s) \downarrow 1_{[0,1]}(s)$ as $l \to \infty$ for $h_l^\pm \in C_K^+([0,\infty))$ we get (3.22) by a standard approximation argument. Now (3.22) says $\sum_{k=1}^n \varepsilon_{X_{n,k}}$ converges weakly to PRM$(\mu)$ and so (3.19) follows by the discussion in the warm-up. □

Propositions 3.19 and 3.21 will be adequate for almost all our needs in the next chapter. However, for many limit theorems involving point processes, particularly in extreme value theory of dependent stationary sequences (cf. Leadbetter, Lindgren, Rootzen, 1983), the following striking result of Kallenberg has been very useful. (Cf. Kallenberg, 1983; Jagers, 1974.)

Call a point process $\xi$ simple if its distribution concentrates on the simple point measures of $M_p(E)$. This means

$$\mathbf{P}[\xi(\{x\}) \le 1 \text{ for all } x \in E] = 1.$$

**Proposition 3.22.** *Suppose $\xi$ is a simple point process on E and $\mathscr{T}$ is a basis of relatively compact open sets such that $\mathscr{T}$ is closed under finite unions and* *intersections and for* $I \in \mathscr{T}$

$$\mathbf{P}[\xi(\partial I) = 0] = 1.$$

*If* $\xi_n$, $n \geq 1$, *are point processes on E and for all* $I \in \mathscr{T}$

$$\lim_{n\to\infty} \mathbf{P}[\xi_n(I) = 0] = \mathbf{P}[\xi(I) = 0] \tag{3.23}$$

*and*

$$\lim_{n\to\infty} \mathbf{E}\xi_n(I) = \mathbf{E}\xi(I) < \infty \tag{3.24}$$

*then*

$$\xi_n \Rightarrow \xi$$

*in* $M_p(E)$.

*Remark.* Typically if $E$ is Euclidean, $\mathscr{T}$ consists of finite unions of bounded rectangles.

Before the proof we need to develop a uniqueness result which essentially says that the distribution of a simple point process $\xi$ is uniquely determined by knowledge of $\mathbf{P}[\xi(I) = 0]$, $I \in \mathscr{T}$.

For $m \in M_p(E)$, if $S$ is the support of $m$ we may write

$$m = \sum_{y_r \in S} c_r \varepsilon_{y_r}$$

where $c_r$ are non-negative integers, $c_r \geq 1$, and $c_r$ represents the weight or multiplicity given location $y_r$. Define $T^*$: $M_p(E) \to M_p(E)$ by

$$T^*m = \sum_{y_r \in S} \varepsilon_{y_r} =: m^*, \tag{3.25}$$

so that by construction $m^*$ is simple. Later we will show that $T^*$ is measurable (in fact more will be shown). Now the uniqueness result will be stated and proved.

**Proposition 3.23.** *Suppose* $\xi_i$, $i = 1, 2$ *are point processes on E, and* $\mathscr{T}$ *is a basis of open, relatively compact sets closed under finite unions and intersections. Then* $T^*$ *defined by* (3.25) *is measurable from* $(M_p(E), \sigma(\{m \in M_p(E)\colon m(I) = 0\}, I \in \mathscr{T}))$ into $((M_p(E), \mathscr{M}_p(E))$ *and if*

$$\mathbf{P}[\xi_1(I) = 0] = \mathbf{P}(\xi_2(I) = 0] \tag{3.26}$$

*for all* $I \in \mathscr{T}$, *then*

$$\xi_1^* \stackrel{d}{=} \xi_2^*$$

*in* $M_p(E)$, *where* $\xi_i^* = T^*\xi_i$, $i = 1, 2$.

Proof of Proposition 3.23. Set

$$\mathscr{C} = \{\{m \in M_p(E)\colon m(I) = 0\}, I \in \mathscr{T}\}.$$

First of all observe that $\mathscr{C}$ is a $\pi$-system. It is closed under finite intersections because

$$\{m: m(I_1) = 0\} \cap \{m: m(I_2) = 0\} = \{m: m(I_1 \cup I_2) = 0\}$$

and since $I_1, I_2 \in \mathscr{T}$ implies $I_1 \cup I_2 \in \mathscr{T}$ by assumption, this last set is in $\mathscr{C}$.

Let $P_i := \mathbf{P} \circ \xi_i^{-1}$ be the distribution of $\xi_i$. Assumption (3.24) says $P_1 = P_2$ on the $\pi$-system $\mathscr{C}$ and hence by Dynkin's theorem we have $P_1 = P_2$ on $\sigma(\mathscr{C})$ (cf. Exercise 3.1.3).

We now must check $T^*$ is a measurable map from

$$(M_p(E), \sigma(\mathscr{C})) \to (M_p(E), \mathscr{M}_p(E))$$

and as in Proposition 3.2 it suffices to check for each $I \in \mathscr{T}$ that

$$T_1^*: m \to m^*(I)$$

is measurable from $(M_p(E), \sigma(\mathscr{C})) \to \{0, 1, \ldots\}$, this being much easier than checking the measurability of $T^*$ directly. Since $I$ is relatively compact, it is possible for each $n$ to cover $I$ by a finite number of sets $A_{n,j} \in I$, $1 \le j \le k_n$ such that the diameter of $A_{n,j} = \sup_{x,y \in A_{n,j}} \rho(x, y)$ is less than $n^{-1}$. Furthermore we may suppose the $\{A_{n,j}; 1 \le j \le n\}_{n \ge 1}$ family is nested so that $A_{n+1,j} \subset A_{n,i}$ for some $i$. Thus

$$T_1^* m = m^*(I) = \lim_{n\to\infty} \uparrow \sum_{j=1}^{k_n} (m(A_{n,j}) \wedge 1).$$

Defining $T_2^* m = m(A_{n,j}) \wedge 1$ we have

$$(T_2^*)^{-1}(\{0\}) = \{m: m(A_{n,j}) = 0\} \in \sigma(\mathscr{C})$$

so that $T_2^*$ and hence $T_1^*$ is $\sigma(\mathscr{C})$ measurable as required.

It is now easy to prove $\xi_1^* = \xi_2^*$. We have for $B \in M_p(E)$:

$$\mathbf{P}[\xi_1^* \in B] = \mathbf{P}[T^* \xi_1 \in B] = \mathbf{P}[\xi_1 \in (T^*)^{-1}(B)] = P_1((T^*)^{-1}(B)).$$

Since $(T^*)^{-1}(B) \in \sigma(\mathscr{C})$ and $P_1 = P_2$ on $\sigma(\mathscr{C})$, the preceding equals

$$P_2((T^*)^{-1}(B)) = \mathbf{P}[\xi_2^* \in B]. \qquad \square$$

Proof of Proposition 3.22. Condition (3.24) implies that $\{\xi_n\}$ is tight since for $I \in \mathscr{T}$ we have by Chebychev's inequality

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbf{P}[\xi_n(I) > t]$$

$$\le \lim_{t\to\infty} \limsup_{n\to\infty} t^{-1} \mathbf{E}\xi_n(I) = \lim_{t\to\infty} t^{-1} \mathbf{E}\xi(I) = 0$$

and therefore by covering any compact $K$ with a finite number of $I$'s from $\mathscr{T}$ we get

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbf{P}[\xi_n(K) > t] = 0$$

which is equivalent to tightness. (Cf. Lemma 3.20 and Exercise 3.5.2.)

From tightness we have for any subsequence $\{n''\}$ that there exists a further subsequence $\{n'\} \subset \{n''\}$ such that $\{\xi_{n'}\}$ converges weakly to a limit which we call $\eta$. From Proposition 3.14 $\eta$ must be a point process. We need to know two facts about $\eta$:

$$\eta \text{ is simple} \tag{3.27}$$

and

$$\mathbf{P}[\xi(\partial A) = 0] = 1 \qquad \text{implies } \mathbf{P}[\eta(\partial A) = 0] = 1 \tag{3.28}$$

for any relatively compact $A$. Accept these two facts as true temporarily. From (3.28), for $I \in \mathscr{T}$, the map $m \to m(I)$ is a.s. continuous with respect to $\eta$ (Proposition 3.12(ii)) and thus

$$\xi_{n'} \Rightarrow \eta$$

in $M_p(E)$ entails for $I \in \mathscr{T}$

$$\xi_{n'}(I) \Rightarrow \eta(I)$$

in $\mathbb{R}$ so that

$$\mathbf{P}[\xi_{n'}(I) = 0] \to \mathbf{P}[\eta(I) = 0].$$

On the other hand, we have assumed

$$\mathbf{P}[\xi_n(I) = 0] \to \mathbf{P}[\xi(I) = 0]$$

so that

$$\mathbf{P}[\xi(I) = 0] = \mathbf{P}[\eta(I) = 0].$$

Hence

$$\begin{aligned} \xi &= \xi^* && (\xi^* \text{ is assumed simple}) \\ &\stackrel{d}{=} \eta^* && (\text{Proposition 3.23}) \\ &= \eta && (3.27). \end{aligned}$$

So any subsequential limit $\eta$ of $\{\xi_n\}$ has the property $\eta \stackrel{d}{=} \xi$ and hence $\xi_n \Rightarrow \xi$. So it only remains to prove (3.27) and (3.28).

We check (3.28) first and this will be proved if we show for any compact $K$

$$\mathbf{P}[\eta(K) = 0] \geq \mathbf{P}[\xi(K) = 0] \tag{3.29}$$

since $\partial A$ is compact for $A$ relatively compact. From the approximation Lemma 3.11, there exist $f_j \in C_K^+(E)$ and compacta $K_j$ such that as $j \to \infty$

$$1_K \leq f_j \leq 1_{K_j} \downarrow 1_K.$$

Therefore

$$\begin{aligned} \mathbf{P}[\eta(K) = 0] &\geq \mathbf{P}[\eta(f_j) = 0] \\ &= \mathbf{P}[\eta(f_j) \leq 0]. \end{aligned}$$

Since $\xi_{n'}(f_j) \Rightarrow \eta(f_j)$ and $\{t: t \le 0\}$ is closed we get from (3.17) that the preceding is not less than

$$\ge \limsup_{n' \to \infty} \mathbf{P}[\xi_{n'}(f_j) = 0]$$

$$\ge \limsup_{n' \to \infty} \mathbf{P}[\xi_{n'}(K_j) = 0]$$

and because $\mathscr{T}$ contains a basis, there exist $I_j \in \mathscr{T}$ such that $K_j \subset I_j \downarrow K$. The previous limsup thus has a lower bound of

$$\ge \limsup_{n' \to \infty} \mathbf{P}[\xi_{n'}(I_j) = 0]$$

$$= \mathbf{P}[\xi(I_j) = 0] \qquad \text{(from 3.23)}$$

and since $\xi(I_j) \downarrow \xi(K)$ the proof of (3.29) is complete.

Now concentrate on (3.27). To show $\eta$ is simple we pick a relatively compact $I \in \mathscr{T}$ and show

$$\mathbf{P}[\eta \text{ has a multiple point in } I] = 0. \tag{3.30}$$

If (3.30) is true, we can cover $E$ with a countable collection of relatively compact sets from $\mathscr{T}$ and easily replace $I$ with $E$ in (3.30).

Now the probability on the left of (3.30) is

$$\begin{aligned} \mathbf{P}[\eta(I) > \eta^*(I)] &= \mathbf{P}[|\eta(I) - \eta^*(I)| > 1/2] \\ &\le 2(\mathbf{E}\eta(I) - \mathbf{E}\eta^*(I)). \end{aligned}$$

However $\xi_{n''} \Rightarrow \eta$ and the already proven (3.28) imply as before that

$$\mathbf{P}[\xi(I) = 0] = \mathbf{P}[\eta(I) = 0]$$

for all $I \in \mathscr{T}$ so that by Proposition 3.23 we have $\xi \overset{d}{=} \eta^*$. Thus for $I \in \mathscr{T}$

$$\mathbf{E}\xi(I) = \mathbf{E}\eta^*(I) \le \mathbf{E}\eta(I)$$

and by Fatou's lemma this is

$$\le \liminf_{n' \to \infty} \mathbf{E}\xi_{n'}(I) \le \limsup_{n' \to \infty} \mathbf{E}\xi_{n'}(I)$$

$$= \mathbf{E}\xi(I)$$

this last step following by assumption (3.24). We therefore conclude $E\eta(I) = E\eta^*(I)$, whence

$$\mathbf{P}[\eta(I) > \eta^*(I)] = 0$$

as desired.

## Exercises

3.5.1. For random measures $\xi_n$, $\xi$ we have

$$\xi_n \Rightarrow \xi \qquad \text{in } \mathscr{M}_+(E)$$

iff for all $f \in C_K^+(E)$

$$\xi_n(f) \Rightarrow \xi(f) \qquad \text{in } \mathbb{R}.$$

3.5.2. The sequence of random measures $\{\xi_n\}$ is tight in $\mathscr{M}_+(E)$ iff for every relatively compact $B \in \mathscr{E}$

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbf{P}[\xi_n(B) > t] = 0.$$

3.5.3. Let $X_n$, $n \geq 0$ be random elements of $(S, \mathscr{S})$ and suppose $X_n \Rightarrow X_0$ in $S$. Suppose $h_n$, $n \geq 0$ are real valued, uniformly bounded, and continuous with $h_n \to h_0$ locally uniformly. Show $h_n(X_n) \Rightarrow h(X_0)$ in $\mathbb{R}$.

3.5.4. In the converse to Proposition 3.21 prove the map

$$\sum_k \varepsilon_{(t_k, j_k)} \to \sum_k 1_{[0,1]}(t_k)\varepsilon_{j_k}$$

from $M_p([0, \infty) \times E) \to M_p(E)$ is a.s. continuous.

3.5.5. If $\xi$ is PRM$(\mu)$ and $\mu$ is atomless, what is a suitable $\mathscr{T}$ for Proposition 3.22? (Cf. Exercise 3.3.3.)

3.5.6. Prove that if $\xi$ is a random measure then

$$\mathscr{B} = \{A \in \mathscr{E}: \mathbf{P}[\xi(\partial A) = 0] = 1\}$$

contains a topological base. Thus the difference in hypotheses for $\mathscr{T}$ between Propositions 3.22 and 3.23 is not theoretically significant (cf. Lemma 3.15) (Kallenberg, 1983, pages 32–33).

3.5.7. Let $\{X_{k,n}, 1 \leq k \leq n, n \geq 1\}$ be random elements of $(E, \mathscr{E})$ and suppose for each $n$, $\{X_{k,n}, 1 \leq k \leq n\}$ is iid. For $0 < a_n \uparrow \infty$ and some $\mu \in \mathscr{M}_+(E)$ show

$$a_n^{-1}\xi_n := a_n^{-1} \sum_{k=1}^n \varepsilon_{X_{k,n}} \Rightarrow \mu$$

in $\mathscr{M}_+(E)$ iff

$$\mu_n := a_n^{-1} n\mathbf{P}[X_{1,n} \in \cdot\,] \xrightarrow{v} \mu \text{ (Resnick, 1986).}$$

# CHAPTER 4

# Records and Extremal Processes

Many natural questions about extremes need a stochastic process context for precise formulation.Imagine observing $X_1, X_2, \ldots$ at a rate of one per unit of time. Suppose while observing we compute the maxima $M_1, M_2, \ldots$. We may ask how often or at what frequency does the maximum change. A change in the maximum means a record was observed, a record being a value larger than previous values. So the previous question is equivalent to asking how often records occur. Also, at what indices do records occur and do the actual record values have any pattern?

These and related questions requiring a stochastic process point of view are dealt with in this chapter for an iid sequence $\{X_n, n \geq 1\}$. The careful study of point processes in the previous chapter will pay rich dividends, as many processes which explain the time varying behavior of extremes are based on simple point processes.

In Section 4.1 the structure of records times is explored. The Markov character of $\{M_n\}$ is discussed. The indices $\{L(j), j \geq 1\}$ where the process $\{M_n\}$ jumps are called *record times* and the values $\{M_{L(j)}, j \geq 1\} = \{X_{L(j)}, j \geq 1\}$ are the *record values*, i.e., those values larger than previous ones. When the underlying distribution is continuous, the record values form a Poisson process and the record times are asymptotically Poisson in a sense to be made precise. There is a very elegant structure to record value processes, and ideally this brief description will whet your appetite for the more detailed results to come.

The stochastic process orientation is temporarily suspended in Section 4.2, where we return to the analytic arena to classify the class of limit laws for records and the domains of attraction of these limit laws. Surprisingly (at first glance) these limit laws are different from the Gnedenko classes given in Theorem 0.3.

In Section 4.3 we introduce a class of continuous time stochastic processes called *extremal processes*. Any sequence of maxima $\{M_n, n \geq 1\}$ of iid random variables can be embedded in a continuous time extremal process $\{Y(t), t > 0\}$ so that

$$\{M_n, n \geq 1\} \stackrel{d}{=} \{Y(n), n \geq 1\}$$

in $\mathbb{R}^\infty$. Sometimes structural properties stand out more clearly in continuous time than in discrete time, and thus our pattern of investigation could be to discover some fact about $\{Y(t), t > 0\}$ and then see how it applies to $\{M_n\}$. Also it will be shown that if the underlying distribution is in a domain of attraction, the sequence of maxima converges (in a powerful stochastic process sense) to a limiting extremal process and thus many properties of extremal processes will be true asymptotically for $\{M_n\}$ through the magic of the invariance principle.

Thus extremal processes are important because of finite time structural results arising out of embeddings and because of asymptotic theory and weak convergence applications. Both uses of extremal processes require detailed knowledge of the properties of extremal processes, and these results are given in Section 4.3. In Section 4.4 the basic weak convergence results are given, and many applications from the continuous mapping theorem (see Section 3.5) are discussed.

Section 4.4.1 is a technical section which discusses how to metrize $D(0, \infty)$ (the class of functions on $(0, \infty)$ which are right continuous with finite left limits), which is the natural space where extremal processes live. You are advised to skim this section, especially if your knowledge of weak convergence at the level of Billingsley (1968), say, is good.

The weak convergence techniques in Section 4.4 are based on the point process method: In order to prove that $(M_n - b_n)/a_n$ converges in a stochastic process sense, we prove that the sequence of point processes with points $\{(k/n, (X_k - b_n)/a_n), k \geq 1\}$ converges as $n \to \infty$ to a limiting point process and then apply a suitable functional to get weak convergence of extremes. This is a powerful technique and in Section 4.5 we apply it to an important class of stationary dependent variables, namely the infinite order moving averages

$$\left\{X_n = \sum_{j=-\infty}^{\infty} c_j Z_{n-j}, n \geq 1\right\}$$

where $\{c_j\}$ are real constants and $\{Z_n, -\infty < n < \infty\}$ are iid with a distribution whose tails are regularly varying.

In the last section we discuss recent results about the process of $k$-records, i.e., those observations which have relative rank $k$ upon being observed. If we observe over $n$ time units the collection of $k$-record processes for $1 \leq k \leq l$, we obviously have more information than if we just note records, and in restricted circumstances the $k$-record process can be of inferential use. For different $k$, the point processes based on $k$ records are iid, a surprising and very attractive result. In Section 4.6 we also discuss behavior of records when the underlying distribution is not assumed continuous.

The importance of extremal processes became apparent with the simultaneous appearance of the two articles by Dwass (1964) and Lamperti (1964); the Dwass paper was followed by Dwass (1966, 1974). See also Tiago de Oliveira (1968). The study of record times was stimulated by Renyi (1962) and Dwass (1960). The importance of Poisson processes was emphasized in Pickands (1971), Shorrock (1972, 1973, 1974, 1975), Resnick (1974, 1975), and Weissman (1975a, b, c, 1976). The use of embeddings and various strong approximations was discussed in Shorrock (loc. cit.), Resnick (loc. cit), Pickands (1971), and Deheuvels (1981, 1982a, 1983). Useful surveys are Resnick and Rubinovitch (1973), Resnick (1983, 1986), de Haan (1984a), Goldie (1983), and Vervaat (1973a).

## 4.1. Structure of Records

Let us recall some relevant facts from the theory of Markov processes (Breiman, 1968, Chapter 15). Let $\{Y(t), t \in T\}$ be a Markov process with state space $\mathbb{R}$. The index set $T$ will be either $(0, \infty)$ or $\{1, 2, \ldots\}$. Suppose there are stationary transition probabilities and there is a family of Markov kernels $K_t(x, B)$ for $t \in T$, $x \in \mathbb{R}$, $B \in \mathscr{B}(\mathbb{R})$ and

$$P\{Y(t+\tau) \in B \mid Y(s), s \in T, s \le t\} = K_\tau(Y(t), B)$$

for all $t \in T$, $t + \tau \in T$, $\tau > 0$, $B \in \mathscr{B}(\mathbb{R})$. The finite dimensional distributions of the process $Y$ have the form

$$\begin{aligned} &P[Y(t_0) \in dy_0, Y(t_1) \in dy_1, \ldots, Y(t_k) \in dy_k] \\ &\quad = \pi_0(dy_0) K_{t_1 - t_0}(y_0, dy_1) \ldots K_{t_k - t_{k-1}}(y_{k-1}, dy_k) \end{aligned} \tag{4.1}$$

for $t_i \in T$, $i = 0, \ldots, k$, and $\pi_0$ is the distribution of $Y(t_0)$. For our purposes we will take $Y$ to be a pure jump process. This means there is a sequence (possibly doubly infinite for the case $T = (0, \infty)$) of jump times $\{\tau_n\}$ and $\tau_n < \tau_{n+1}$, $\{\tau_n\}$ has no cluster point, and $Y(t)$ is constant for $\tau_n \le t < \tau_{n+1}$. In case $T = (0, \infty)$ we assume $Y$ has right continuous paths. The times between jumps when $T = (0, \infty)$ are conditionally exponentially distributed: Given that the process is in state $x$, the holding time has an exponential distribution with parameter $\lambda(x)$. Given that the process is finished holding in state $x$, it moves to a new state in set $B$ with probability $\Pi(x, B)$. The relation of $K_t(\cdot, \cdot)$ to $\lambda(\cdot)$ and $\Pi(\cdot, \cdot)$ is

$$K_t(x, \{x\}) = 1 - \lambda(x)t + o(t), \qquad t \to 0 \tag{4.2}$$

$$K_t(x, B) = (\lambda(x)t + o(t))\Pi(x, B) \qquad t \to 0, x \in B^c. \tag{4.3}$$

When $T = \{1, 2, \ldots\}$, the holding time distributions are conditionally geometrically distributed.

The process $\{Y(\tau_n)\}$ is Markov with discrete index set, and it has stationary transition probabilities $\Pi(x, B)$:

$$P[Y(\tau_{n+1}) \in B \mid Y(\tau_n) = x] = \Pi(x, B).$$

Observe that $\{Y(\tau_n)\}$ represents the succession of states visited by the process $\{Y(t), t \in T\}$. Given $\{Y(\tau_n)\}$, the variables $\{\tau_{n+1} - \tau_n\}$ are conditionally independent and exponentially distributed in the case $T = (0, \infty)$:

$$P[\tau_{m+1} - \tau_m > x \mid \{Y(\tau_n)\}, \{\tau_{k+1} - \tau_k, k \ne m\}] = e^{-\lambda(Y(\tau_m))x}. \tag{4.4}$$

In the case $T = \{1, 2, \dots\}$ the variables $\{\tau_{n+1} - \tau_n\}$ are conditionally independent and geometrically distributed.

Now let $\{X_n, n \geq 1\}$ be iid with common distribution function $F$. For $n \geq 1$ set $M_n = \bigvee_{i=1}^n X_i$ and consider the stochastic process $\{M_n, n \geq 1\}$. Let us write down the finite dimensional distributions of this process. In the bivariate case we have for $x_1 < x_2$, $t_1 < t_2$, and $t_1, t_2$ non-negative integers

$$P\left[M_{t_1} \leq x_1, M_{t_2} \leq x_2\right] = P\left[M_{t_1} \leq x_1, \bigvee_{j=t_1+1}^{t_2} X_j \leq x_2\right]$$
$$= F^{t_1}(x_1)F^{t_2-t_1}(x_2).$$

In case $x_1 > x_2$,

$$P[M_{t_1} \leq x_1, M_{t_2} \leq x_2] = F^{t_2}(x_2]$$

so in general

$$P[M_{t_1} \leq x_1, M_{t_2} \leq x_2] = F^{t_1}(x_1 \wedge x_2)F^{t_2-t_1}(x_2).$$

Following this pattern we get for the $k$-variate case

$$P[M_{t_1} \leq x_1, M_{t_2} \leq x_2, \dots, M_{t_k} \leq x_k]$$
$$F^{t_1}\left(\bigwedge_{i=1}^k x_i\right)F^{t_2-t_1}\left(\bigwedge_{i=2}^k x_i\right)\dots F^{t_k-t_{k-1}}(x_k) \tag{4.5}$$

for $t_1 < t_2 < \cdots < t_k$, $k \geq 1$, and $t_i$ non-negative integers.

Comparing (4.5) and (4.1) we see that $\{M_n\}$ is a Markov process with stationary transition probabilities. The transition kernels are given by

$$K_t(x, (-\infty, z]) = \begin{cases} F^t(z) & z \geq x \\ 0 & z < x. \end{cases} \tag{4.6}$$

(This is checked by substituting (4.6) into (4.1) to verify (4.5).) The paths of $\{M_n\}$ are constant except for those indices $n$ such that $M_n > M_{n-1}$, which are call *record times*. Define $L(1) = 1$ and inductively for $n \geq 1$

$$L(n+1) = \inf\{j > L(n): M_j > M_{L(n)}\}.$$

So the record times $\{L(n), n \geq 1\}$ are the times when the Markov process $\{M_n\}$ jumps. The succession of states visited $\{X_{L(n)}, n \geq 1\} = \{M_{L(n)}, n \geq 1\}$ is called the *record values* and constitute the embedded Markov process of states visited.

**Proposition 4.1.** (i) $\{X_{L(n)}, n \geq 1\}$ *is a Markov process with stationary transition probabilities and for* $n \geq 1$

$$\Pi(x, (y, \infty)) = P[X_{L(n+1)} > y \mid X_{L(n)} = x]$$
$$\begin{cases} (1 - F(y))/(1 - F(x)) & y > x \\ 1 & y \leq x. \end{cases}$$

(ii) *If* $F(x) = 1 - e^{-x}, x > 0$, *then* $\{X_{L(n)}, n \geq 1\}$ *are the points of homogeneous*

PRM *on* $(0, \infty)$. *This means*

$$\{X_{L(n)}, n \geq 1\} \stackrel{d}{=} \{\Gamma_n, n \geq 1\}$$

*in* $R^\infty$ *where* $\Gamma_n = E_1 + \cdots + E_n$ *and* $\{E_j, j \geq 1\}$ *are iid*, $P[E_j > x] = e^{-x}, x > 0$.
(iii) *Suppose* $R(t) = -\log(1 - F(t))$ *and set* $x_l = \inf\{y: F(y) > 0\}$, $x_0 = \sup\{y: F(y) < 1\}$. *Suppose* $F$ *is continuous. Then* $R: (x_l, x_0) \to (0, \infty)$ *so that* $R^{\leftarrow}: (0, \infty) \to (x_l, x_0)$. *Then* $\{X_{L(n)}, n \geq 1\}$ *are the points of PRM on* $(x_l, x_0)$ *with mean measure*

$$R(a, b] = R(b) - R(a)$$

*for* $x_l < a \leq b < x_0$.
(iv) *If* $F$ *is continuous*, $\{X_{L(k)}, L(k+1) - L(k), k \geq 1\}$ *are the points of a two dimensional* PRM *on* $(x_l, x_0) \times \{1, 2, 3, \ldots\}$ *with mean measure*

$$\mu^*((a, b] \times \{j\}) = (F^j(b) - F^j(a))/j$$

*for* $x_l < a < b < x_0, j \geq 1$.

PROOF. (i) Since $\{X_{L(n)}\}$ is the embedded jump chain of $\{M_n\}$ we already know that $\{X_{L(n)}\}$ is Markov with stationary transitions. To compute $\Pi(x, dy)$ observe for $y > x$ and $1 - F(x) > 0$

$$\begin{aligned}
\Pi(x, (y, \infty)) &= P[X_{L(2)} > y \mid X_{L(1)} = x] \\
&= \sum_{n=2}^{\infty} P[X_{L(2)} > y, L(2) = n \mid X_{L(1)} = x] \\
&= \sum_{n=2}^{\infty} P\left[\bigvee_{j=2}^{n-1} X_j \leq x, X_n > y \mid X_1 = x\right] \\
&= \sum_{n=2}^{\infty} F^{n-2}(x)(1 - F(y)) = (1 - F(y))/(1 - F(x)).
\end{aligned}$$

(ii) In case $F(x) = 1 - e^{-x}$ we have for $y > x$

$$\begin{aligned}
\pi(x, (y, \infty)) &= (1 - F(y))/(1 - F(x)) = \exp\{-(y - x)\} \\
&= P[\Gamma_{n+1} > y \mid \Gamma_n = x].
\end{aligned}$$

Since also $X_{L(1)} = X_1 \stackrel{d}{=} \Gamma_1$, we have

$$\{X_{L(n)}, n \geq 1\} \stackrel{d}{=} \{\Gamma_j, j \geq 1\},$$

since two Markov sequences with stationary transition probabilities are equal in distribution if their initial distributions and transition kernels coincide.
(iii) If $X$ is a random variable with distribution $F$ then the variant of the probability integral transform discussed in 0.2 gives $R^{\leftarrow}(E_1) \stackrel{d}{=} X_1$ and similarly

$$\left\{\bigvee_{i=1}^{n} R^{\leftarrow}(E_i), n \geq 1\right\} \stackrel{d}{=} \{M_n, n \geq 1\}.$$

Since $R^{\leftarrow}$ is nondecreasing

$$\left\{\bigvee_{i=1}^{n} R^{\leftarrow}(E_i), n \geq 1\right\} = \left\{R^{\leftarrow}\left(\bigvee_{i=1}^{n} E_i\right), n \geq 1\right\}.$$

Furthermore $R$ continuous makes $R^{\leftarrow}$ strictly increasing and hence

$$\{R^{\leftarrow}(E_{L(n)}), n \geq 1\} \stackrel{d}{=} \{X_{L(n)}, n \geq 1\} \tag{4.7}$$

where notation has been slightly abused; the $L(n)$ on the left refers to record times of $\{E_j, j \geq 1\}$, and the $L(n)$ on the right refers to record times of $\{X_j, j \geq 1\}$. (Observe that without the assumption that $R^{\leftarrow}$ is strictly increasing, there would be the potential that intervals of constancy of $R^{\leftarrow}$ would cause records of $\{E_j\}$ to be ignored in the left side of (4.7).)

Notice that $\{R^{\leftarrow}(E_{L(n)})\}$ are the points $\{E_{L(n)}\}$ of homogeneous PRM transformed. A glance at Proposition 3.7 assures us that $\{R^{\leftarrow}(E_{L(n)})\}$ are the points of PRM with mean measure $m \circ (R^{\leftarrow})^{-1}$ where $m$ is Lebesgue measure. For $x_l < a \leq b < x_0$

$$m \circ (R^{\leftarrow})^{-1}(a, b] = m\{s\colon a < R^{\leftarrow}(s) \leq b\}$$

and applying 0.6(c) this is

$$= m\{s\colon R(a) < s \leq R(b)\} = R(b) - R(a).$$

(iv) We have from the discrete time analog of (4.4) that

$$\begin{aligned} P[L(n+1) - L(n) = j \mid \{X_{L(j)}, j \geq 0\}, \{L(i+1) - L(i), i \neq n\}] \\ = P[L(n+1) - L(n) = j \mid X_{L(n)}] \\ = F(X_{L(n)})^{j-1}(1 - F(X_{L(n)})) \end{aligned} \tag{4.8}$$

for $j \geq 1$. From (4.8), the fact that $\sum_{n=1}^{\infty} \varepsilon_{X_{L(n)}}$ is PRM($R$), and Proposition 3.8 we conclude that

$$\sum_{n=1}^{\infty} \varepsilon_{(X_{L(n)}, L(n+1) - L(n))}$$

is PRM on $(x_l, x_0) \times \{1, 2, \ldots\}$ with mean measure $\mu^*$ given by

$$\begin{aligned} \mu^*((a, b) \times \{j\}) &= \int_{(a,b]} R(dx) F(x)^{j-1}(1 - F(x)) \\ &= \int_{(F(a), F(b)]} y^{j-1} dy = (F^j(b) - F^j(a))/j. \qquad \square \end{aligned}$$

**Corollary 4.2.** *Suppose $F$ is continuous. Define for $x_l < t < x_0$*

$$\eta(t) = \inf\{n\colon M_n > t\}.$$

*Then $\eta(t)$ is a process with independent increments and for $x_l < a < b < x_0$*

$$P[\eta(b) = k] = F^{k-1}(b)(1 - F(b)) \qquad k \geq 1$$

$$P[\eta(b) - \eta(a) = 0] = (1 - F(b))/(1 - F(a))$$

*and for* $n \geq 1$

$$P[\eta(b) - \eta(a) = n] = \left(\frac{1 - F(b)}{1 - F(a)}\right) F(a, b] F^{n-1}(b).$$

PROOF. Observe that

$$\eta(b) - \eta(a) = \#\{n: M_n \in (a, b]\} = \sum_{k=1}^{\infty} (L(k+1) - L(k) \varepsilon_{X_{L(k)}}(a, b]$$

and if $N^* = \sum_{k=1}^{\infty} \varepsilon_{(X_{L(k)}, L(k+1) - L(k))}$ then

$$\eta(b) - \eta(a) = \sum_{j=1}^{\infty} jN^*((a, b] \times \{j\}).$$

Since $N^*$ is PRM the independent increment property of $\eta$ follows from that of $N^*$; cf. property (b), Section 3.3.1.

Next we have that for $k \geq 1$

$$[\eta(b) \leq k] = [M_k > b]$$

so that

$$P[\eta(b) \leq k)] = 1 - F^k(b)$$

and

$$\begin{aligned} P[\eta(b) = k) &= (1 - F^k(b)) - (1 - F^{k-1}(b)) \\ &= F^{k-1}(b) - F^k(b) = F^{k-1}(b)(1 - F(b)). \end{aligned}$$

Therefore, taking generating functions gives for $0 \leq s \leq 1$

$$Es^{\eta(b)} = \sum_{k=1}^{\infty} s^k F^{k-1}(b)(1 - F(b)) = s(1 - F(b))/(1 - sF(b)).$$

From the independent increment property

$$Es^{\eta(b)} = E(s^{(\eta(b) - \eta(a))} s^{\eta(a)}) = Es^{(\eta(b) - \eta(a))} Es^{\eta(a)}$$

and thus

$$\begin{aligned} Es^{(\eta(b) - \eta(a))} = Es^{\eta(b)}/Es^{\eta(a)} &= \left(\frac{1 - F(b)}{1 - F(a)}\right) \frac{1 - sF(a)}{1 - sF(b)} \\ &= ((1 - F(b))/(1 - F(a)))(1 - sF(a)) \sum_{k=0}^{\infty} s^k F^k(b) \\ &= \left(\frac{1 - F(b)}{1 - F(a)}\right) \left\{ \sum_{k=0}^{\infty} s^k F^k(b) - \sum_{k=0}^{\infty} s^{k+1} F^k(b) F(a) \right\} \\ &= \left(\frac{1 - F(b)}{1 - F(a)}\right) \left\{ 1 + \sum_{k=1}^{\infty} s^k (F^k(b) - F^{k-1}(b) F(a)) \right\} \\ &= \left(\frac{1 - F(b)}{1 - F(a)}\right) \left\{ 1 + \sum_{k=1}^{\infty} s^k F^{k-1}(b) F(a, b] \right\} \end{aligned}$$

and the desired formulas are obtainable as coefficients of $s^k$. □

We now study the sequence $\{L(n), n \geq 1\}$ and the point process generated by this sequence. The next pearl is a basic structural result. See Dwass (1960, 1964) and Renyi (1962).

**Proposition 4.3.** *Let $\{X_n, n \geq 1\}$ be iid with common continuous distribution $F(x)$. Let $R_n$ be the relative rank of $X_n$ among $X_1, \ldots, X_n$; i.e., $R_n = \sum_{i=1}^{n} 1_{[X_i \geq X_n]}$. Thus*

$$R_n = 1 \text{ iff } X_n = M_n,$$

$$= 2 \text{ iff } X_n \textit{ is the second largest among } X_1, \ldots, X_n,$$

*and so on.*

(i) *$\{R_n, n \geq 1\}$ is a sequence of independent random variables with*

$$P[R_n = k] = n^{-1}, \qquad 1 \leq k \leq n.$$

(ii) *The events*

$$A_j = [X_j \textit{ is a record}] = [R_j = 1], \qquad j \geq 1$$

*are independent and*

$$PA_j = j^{-1}.$$

Proof. The second result follows directly from the first, which is checked as follows: Since ties among the observations occur with probability 0 ($F$ is assumed continuous) each of $n!$ orderings $X_{i_1} < \cdots < X_{i_n}$ is equally likely ($i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$). There is a one to one correspondence between each such ordering and a realization of $R_1, \ldots, R_n$. For example when $n = 3$

$$X_3 < X_1 < X_2$$

corresponds to the realization

$$R_1 = 1, \qquad R_2 = 1, \qquad R_3 = 3.$$

So each realization of $R_1, \ldots, R_n$ has equal probability $1/n!$; i.e.,

$$P[R_1 = r_1, \ldots, R_n = r_n] = 1/n!$$

for $r_i \in \{1, \ldots, i\}$, $i = 1, \ldots, n$. To get the mass function of $R_n$ we sum over $r_1, \ldots, r_{n-1}$ remembering that $r_i$ has $i$ possible values. Hence

$$P[R_n = r_n] = \sum_{r_1, \ldots, r_{n-1}} P[R_1 = r_1, \ldots, R_n = r_n]$$

$$\sum_{r_1, \ldots, r_{n-1}} 1/n! = (1 \cdot 2 \cdot \ldots \cdot (n-1))/n! = 1/n.$$

Hence $P[R_1 = r_1, \ldots, R_n = r_n] = \prod_{i=1}^{n} P[R_i = r_i]$ showing $\{R_n, n \geq 1\}$ is a sequence of independent random variables. □

Define now the point process $\mu$ on $(0, \infty)$ by

$$\mu(\cdot) = \sum_{n=1}^{\infty} \varepsilon_{L(n)} = \sum_{j=1}^{\infty} 1_{A_j} \varepsilon_j(\cdot)$$

so that $\mu([1,n])$, the number of records in the first $n$ observations, is a sum of independent Bernoulli random variables. This fact can be used to obtain asymptotic behavior of $\mu[1,n]$ or $L(n)$ as done by Renyi (1962). We give one argument in this classical vein but prefer to await discussion of extremal processes for a fuller treatment. Also it is noteworthy that whereas $\{X_{L(n)}\}$ are the points of PRM, $\mu$ is only asymptotically Poisson, a statement which will be made precise later. This is explored both as a corollary of Proposition 4.3 and also later in connection with extremal processes.

**Corollary 4.4.** *If $F$ is continuous then*

$$\mu([1,n])/\log n \to 1 \text{ a.s.}$$

PROOF. Observe that $E1_{A_j} = 1/j$ and

$$\operatorname{Var} 1_{A_j} = E1_{A_j}^2 - (E1_{A_j})^2 = 1/j - 1/j^2 = (j-1)/j^2.$$

Since $\mu[1,n] = \sum_{j=1}^n 1_{A_j}$, consider the series $\sum_1^\infty ((1_{A_j} - j^{-1})/\log j)$. We have

$$\sum_1^\infty \operatorname{Var}((1_{A_j} - j^{-1})/\log j) = \sum_1^\infty (\operatorname{Var} 1_{A_j})/(\log j)^2 = \sum_1^\infty (j-1)/(j\log j)^2.$$

Since $(j-1)/(j\log j)^2 \sim 1/j(\log j)^2$ is summable (approximate by

$$\int_1^\infty x^{-1}(\log x)^{-2}dx < \infty)$$

we have by the Kolmogorov convergence criterion (Feller, 1971, p. 243) that

$$\sum_1^\infty ((1_{A_j} - j^{-1})/(\log j) \text{ converges almost surely.}$$

Applying the classical Kronecker lemma we get

$$\frac{\sum_1^n (1_{A_j} - j^{-1})}{\log n} = \frac{\mu[1,n]}{\log n} - 1 + o(1) \to 0 \qquad \text{a.s.}$$

since $\sum_1^n j^{-1} \sim \log n$ as $n \to \infty$. □

The next result says the point process of record times is asymptotically Poisson.

**Corollary 4.5.** *Let $F$ be continuous. Define the point processes $\mu_n$ on $(0,\infty)$ by*

$$\mu_n(\cdot) = \sum_{j=1}^\infty \varepsilon_{L(j)}(n\cdot) = \sum_{j=1}^\infty \varepsilon_{n^{-1}L(j)}(\cdot) = \sum_{i=1}^\infty 1_{A_i}\varepsilon_{n^{-1}i}(\cdot).$$

*Let $\mu_\infty$ be PRM on $(0,\infty)$ with mean measure of $(a,b] = \log(b/a)$. Then*

$$\mu_n \Rightarrow \mu_\infty$$

*in $M_p((0,\infty))$.*

**Proof.** We show $\Psi_{\mu_n}(f) \to \Psi_{\mu_\infty}(f)$, for all $f \in C_K^+((0,\infty))$. We have that

$$\begin{aligned}
\Psi_{\mu_n}(f) &= E\exp\left\{-\int_{(0,\infty)} f(x) \sum_{i=1}^{\infty} 1_{A_i}\varepsilon_{n^{-1}i}(dx)\right\} \\
&= E\exp\left\{-\sum_{i=1}^{\infty} f(n^{-1}i)1_{A_i}\right\} \\
&= \prod_{i=1}^{\infty} E\exp\{-f(n^{-1}i)1_{A_i}\} \\
&= \prod_{i=1}^{\infty} (e^{-f(n^{-1}i)}i^{-1} + (1 - i^{-1})) \\
&= \prod_{i=1}^{\infty} (1 - i^{-1}(1 - e^{-f(n^{-1}i)})).
\end{aligned}$$

As an alternative to taking logarithms and expanding as in Proposition 3.21 we use the elementary inequalities

$$\left|\prod_{i=1}^{q} a_i - \prod_{i=1}^{q} b_i\right| \le \sum_{i=1}^{q} |a_i - b_i| \tag{4.9}$$

valid for $|a_i| \le 1$ and $|b_i| \le 1$, $1 \le i \le q$, and

$$|e^{-x} - 1 + x| \le x^2/2 \tag{4.10}$$

valid for $x > 0$. (For (4.9) suppose $q = 2$, write $a_1a_2 - b_1b_2 = (a_1 - b_1)a_2 + b_1(a_2 - b_2)$, take absolute values, and use induction. For 4.10 write $|1 - e^{-x} - x| = |\int_0^x (e^{-u} - 1)du| \le \int_0^x |e^{-u} - 1|du \le \int_0^x u\,du$.) Set $x_{i,n} = i^{-1}(1 - \exp\{-f(i/n)\})$ and we have

$$\begin{aligned}
\left|\Psi_{\mu_n}(f) - \prod_{i=1}^{\infty} \exp\{-x_{i,n}\}\right| &= \left|\prod_{i=1}^{\infty} (1 - x_{i,n}) - \prod_{i=1}^{\infty} \exp\{-x_{i,n}\}\right| \\
&\le \sum_{i=1}^{\infty} |e^{-x_{i,n}} - 1 + x_{i,n}| \le \frac{1}{2}\sum_{i=1}^{\infty} (x_{i,n})^2 \\
&= \sum_{i=1}^{\infty} \frac{1}{2}(i^{-1}(1 - \exp\{-f(i/n)\}))^2.
\end{aligned} \tag{4.11}$$

Now

$$(i^{-1}(1 - e^{-f(n^{-1}i)}))^2 \le i^{-2}$$

which is summable, and for fixed $i$,

$$\lim_{n\to\infty} (i^{-1}(1 - e^{-f(n^{-1}i)}))^2 = 0$$

because $f \in C_K^+((0,\infty))$ has support in some interval $[a, b]$, $a > 0$. Therefore by dominated convergence, the series in (4.11) goes to zero as $n \to \infty$.

This means that if we set

$$m_n = \sum_{i=1}^{\infty} n^{-1}\varepsilon_{n^{-1}i}(\cdot)$$

then

$$\begin{aligned}\Psi_{\mu_n}(f) &= o(1) + \exp\left\{-\sum_{i=1}^{\infty} i^{-1}(1 - \exp\{-f(i/n)\})\right\}\\ &= o(1) + \exp\left\{-\sum_{i=1}^{\infty} \frac{1 - \exp\{-f(i/n)\}}{i/n} n^{-1}\right\}\\ &= o(1) + \exp\left\{-\int_{(0,\infty)} \frac{1 - \exp\{-f(x)\}}{x} m_n(dx)\right\}.\end{aligned}$$

Observe that $m_n \xrightarrow{v} m$, Lebesgue measure (cf. Exercise 3.4.14) since for $f \in C_K^+((0,\infty))$

$$m_n(f) = \sum_{i=1}^{\infty} f(i/n)n^{-1}$$

which we recognize as a Riemann approximating sum to an integral and thus

$$m_n(f) \to \int_{(0,\infty)} f(x)dx = m(f).$$

It therefore follows from vague convergence that

$$\int_{(0,\infty)} \frac{(1 - e^{-f(x)})}{x} m_n(dx) \to \int_{(0,\infty)} (1 - e^{-f(x)})\frac{dx}{x}$$

since $(1 - e^{-f(x)})/x \in C_K^+((0,\infty))$ and $m_n \xrightarrow{v} m$. Our conclusion is

$$\lim_{n\to\infty} \Psi_{\mu_n}(f) = \exp\left\{-\int_{(0,\infty)} (1 - e^{-f(x)})x^{-1}dx\right\}$$

as required for $\mu_n \Rightarrow \mu_\infty$. □

## Exercises

4.1.1. (a) If $\{E_n, n \geq 1\}$ are iid with $P[E_1 > x] = e^{-x}$, $x > 0$ show using the Borel-Cantelli lemma that

$$\limsup_{n\to\infty} E_n/\log n = 1 \text{ a.s.}$$

(b) Show

$$\lim_{n\to\infty} E_{L(n)}/n = 1 \text{ a.s.}$$

(c) Show

$$\lim_{n\to\infty} X_{L(n)}/R^{\leftarrow}(n) = 1 \text{ a.s.}$$

provided for all $t \in \mathbb{R}$

$$\lim_{s\to\infty} R^{\leftarrow}(s + t(s\log\log s)^{1/2})/R^{\leftarrow}(s) = 1.$$

The last condition is equivalent to

$$\lim_{t\to\infty} (R(tx) - R(t))/(2R(t)\log\log R(t))^{1/2} = \infty.$$

Hint: Use the representation (4.7) and the law of iterated logarithm for partial sums.

(d) What is the norming constant in (c) when the underlying distribution $F$ is normal?

4.1.2. For two random variables $X$ and $Y$ write $X >^P Y$ ($X$ is stochastically larger than $Y$) if $P[X > x] \geq P[Y > x]$. Show if $X > Y$ then there exists $X^*$ and $Y^*$ defined on the uniform probability space with

$$X \stackrel{d}{=} X^*, \; Y \stackrel{d}{=} Y^*$$

and $X^* \geq Y^*$ a.s. (Use the probability integral transform.)

Construct examples where

(a) $X_{L(n+1)} - X_{L(n)} >^P X_{L(n)} - X_{L(n-1)}$

and

(b) $X_{L(n+1)} - X_{L(n)} {}^P\!< X_{L(n)} - X_{L(n-1)}$

(Try the Weibull distribution.)
(Duane Boes, unpublished comments)

4.1.3. When is $\eta(t) = \inf\{n: M_n > t\}$ stochastically continuous? (Stochastic continuity means if $t \to t_0$ then $\eta(t) \xrightarrow{P} \eta(t_0)$. If $\eta$ is *not* stochastically continuous it has fixed discontinuities.)

4.1.4. Define the point process

$$O(\cdot) = \sum \varepsilon_{M_k}$$

where $\{X_n, n \geq 1\}$ are iid from a continuous *df* $F(x)$ and $M_n = \bigvee_{i=1}^n X_i$. Compute the Laplace functional of $O$. Hint: Write

$$O = \sum (L(k+1) - L(k))\varepsilon_{X_{L(k)}}.$$

Relate this to the $\{\eta(t)\}$ process (de Haan and Resnick, 1982).

4.1.5. Kendall's tau: A natural distance between two orderings of the objects $\{a_1, \ldots, a_n\}$ is

$$T = \#(\text{discordant pairs})$$

where

$(a_i, a_j)$ is a discordant pair if $a_i$ precedes $a_j$ in one order but not the other.

For instance if the ranks of $a_1, \ldots, a_4$ are

1st order 2, 1, 4, 3
2nd order 2, 4, 1, 3

then $T = 5$. To get a permutation distribution for $T$ suppose one order is

$a_1, \ldots, a_n$ ("natural order") and the other is a random reordering of the natural one. Let

$$Y_{j-1} = \#\{\text{objects among } a_1, \ldots, a_{j-1} \text{ which } a_j \text{ precedes in second ordering}\}.$$

Then $Y_{j-1}$ takes values 0, 1, ..., $j-1$ each with probability $1/j$ and $\{Y_i\}$ are independent. Check

$$\operatorname{Var} Y_i = \frac{i}{12}\{i+2\}, \qquad EY_i = i/2, \qquad ET = \frac{n}{4}(n-1),$$

$$\operatorname{Var} T = n(n-1)(2n+5)/72.$$

The rank statistic is $\tau = 1 - 4T/(n(n-1))$ and $\tau$ is approximately

$$N\left(0, \frac{2(2n+5)}{9n(n-1)}\right).$$

4.1.6. For any $k > 0$ show

$$\mu_\infty(\cdot) \stackrel{d}{=} \mu_\infty(k(\cdot)).$$

Hint: What is the mean measure of $\mu_\infty(k(\cdot))$?

4.1.7. Prove Proposition 4.1(iii) directly from Proposition 4.1(i) by computing the Laplace functional of $\sum_{n=1}^\infty \varepsilon_{X_{L(n)}}$. Hint: Let $N(x)$ be the number of records in $(-\infty, x]$. Using Proposition 4.1(i), what is

$$P[X_{L(1)} \in dx_1, \ldots, X_{L(n)} \in dx_n;\ N(x) = n]$$

for $x_1 < \cdots < x_n$. Use this to compute

$$E \exp\left\{\sum_{k=1}^{N(x)} f(X_{L(i)})\right\}$$

and then let $x \to \infty$.

4.1.8. Let $\{X_n, n \geq 1\}$ be iid with common distribution $F(x)$. Prove that the total number of records is finite iff the right endpoint $x_0$ of $F$ is an atom: $F(x_0-) < 1$. (Recall $x_0 = \sup\{x: F(x) < 1\}$) (Shorrock, 1970).

## 4.2. Limit Laws for Records

Here we consider the class of possible limit laws for $X_{L(n)}$ and also domain of attraction criteria. It is striking that neither the limit laws nor the domains are the same as for maxima.

Suppose $F$ is continuous and return to relation (4.7). Since the records $\{E_{L(n)}\}$ from an iid exponentially distributed sequence are equal in distribution to $\{\Gamma_n\}$, the points of a homogeneous Poisson process on $[0, \infty)$, we may write

$$\{X_{L(n)}, n \geq 1\} \stackrel{d}{=} \{R^{\leftarrow}(\Gamma_n), n \geq 1\}. \tag{4.12}$$

Suppose next that there exist $\alpha_n > 0$, $\beta_n \in \mathbb{R}$, such that for some nondegenerate distribution $G$

$$P[(X_{L(n)} - \beta_n)/\alpha_n \leq x] \to G(x). \tag{4.13}$$

Therefore from (4.12) and (4.13)

$$\begin{aligned}P[X_{L(n)} \le \alpha_n x + \beta_n] &= P[R^{\leftarrow}(\Gamma_n) \le \alpha_n x + \beta_n]\\ &= P[\Gamma_n \le R(\alpha_n x + \beta_n)]\\ &= P[(\Gamma_n - n)/n^{1/2} \le (R(\alpha_n x + \beta_n) - n)/n^{1/2}]\\ &\to G(x). \end{aligned} \tag{4.14}$$

However from the central limit theorem

$$\lim_{n\to\infty} P[(\Gamma_n - n)/n^{1/2} \le x] = N(x)$$

uniformly for all $x$, where $N$ is the standard normal distribution. Therefore (4.14) can hold iff there is a nondecreasing function $g(x)$ with more than one point of increase such that

$$(R(\alpha_n x + \beta_n) - n)/n^{1/2} \to g(x), \tag{4.15}$$

weakly. In this case $G$ is of the form

$$G(x) = N(g(x)).$$

Convert (4.15) into a convergence with a continuous parameter by setting $\alpha(t) = \alpha_{[t]}$, $\beta(t) = \beta_{[t]}$. Then

$$\begin{aligned}&(R(\alpha(t)x + \beta(t)) - t)/t^{1/2}\\ &\quad\le (R(\alpha_{[t]}x + \beta_{[t]}) - [t])/[t]^{1/2} \to g(x)\end{aligned}$$

and

$$\begin{aligned}&(R(\alpha(t)x + \beta(t)) - t)/t^{1/2}\\ &\quad\ge (R(\alpha_{[t]}x + \beta_{[t]}) - ([t]+1))/([t]+1)^{1/2}\\ &\quad= \frac{(R(\alpha_{[t]}x + \beta_{[t]}) - [t])}{[t]^{1/2}}\left(\frac{[t]}{[t]+1}\right)^{1/2} - \frac{1}{([t]+1)^{1/2}}\\ &\quad\to g(x) \qquad \text{as } t \to \infty\end{aligned}$$

and therefore

$$(R(\alpha(t)x + \beta(t)) - t)/t^{1/2} \to g(x). \tag{4.16}$$

If in (4.16) we divide by $t$ instead of $t^{1/2}$ we may conclude

$$R(\alpha(t)x + \beta(t)) \sim t$$

as $t \to \infty$, for all $x$ such that $g(x)$ is finite. Hence

$$\begin{aligned}&(R(\alpha(t)x + \beta(t)) - t)/t^{1/2}\\ &\quad= R^{1/2}(\alpha(t)x + \beta(t)) - t^{1/2})(R^{1/2}(\alpha(t)x + \beta(t)) + t^{1/2})/t^{1/2}\\ &\quad\sim 2(R^{1/2}(\alpha(t)x + \beta(t)) - t^{1/2})\end{aligned}$$

as $t \to \infty$ and so (4.16) is equivalent to

$$R^{1/2}(\alpha(t)x + \beta(t)) - t^{1/2} \to \tfrac{1}{2}g(x). \tag{4.17}$$

Define a distribution function $H$, called the *associated distribution*, by

$$1 - H(x) = \exp\{-R^{1/2}(x)\}.$$

Exponentiating in (4.17) gives

$$e^{t^{1/2}}(1 - H(\alpha(t)x + \beta(t))) \to e^{-g(x)/2}$$

and letting $s = e^{t^{1/2}}$ we get $s \to \infty$

$$s(1 - H(\alpha((\log s)^2)x + \beta((\log s)^2)) \to e^{-g(x)/2}$$

or equivalently as $s \to \infty$

$$H^s(\alpha((\log s)^2)x + \beta((\log s)^2)) \to \exp\{-e^{-g(x)/2}\}.$$

From Theorem 0.3 we conclude that $\exp\{-e^{-g(x)/2}\}$ must be an extreme value distribution and hence up to affine shifts we have either

$$g(x)/2 = x \qquad x \in \mathbb{R}$$

or

$$= \alpha \log x \qquad x > 0$$

or

$$= -\alpha \log(-x), \qquad x < 0$$

where $\alpha > 0$.

**Proposition 4.6.** (a) *The class of limit laws for record values is of the form*

$$N(-\log(-\log B(x))$$

*where $B(x)$ is an extreme value distribution and $N(x)$ is the standard normal distribution. More explicitly the limit laws are of the following types:*

(i) $N(x)$

(ii) $N_{1,\alpha}(x) = \begin{cases} 0 & x < 0 \\ N(\alpha \log x) & x \geq 0 \end{cases}$

(iii) $N_{2,\alpha}(x) = \begin{cases} N(-\alpha \log(-x)) & x < 0 \\ 1 & x \geq 0 \end{cases}$

*where $\alpha > 0$.*

(b) *A limit law $G$ in* (4.13) *exists for $\{X_{L(n)}\}$* iff $H = 1 - \exp\{-R^{1/2}\}$ *is in the domain of attraction of an extreme value distribution. In fact $G =$*

(i) $N(x)$ iff $H \in D(\Lambda)$. *In this case may take*

$$\alpha_n = R^{\leftarrow}(n + \sqrt{n}) - R^{\leftarrow}(n),\ \beta_n = R^{\leftarrow}(n).$$

(ii) $N_{1,\alpha}(x)$ iff $H \in D(\Phi_{\alpha/2})$. *In this case we may take*

$$\alpha_n = R^{\leftarrow}(n),\ \beta_n = 0.$$

(iii) $N_{2,\alpha}(x)$ iff $H \in D(\Psi_{\alpha/2})$. In this case we may take

$$\alpha_n = x_0 - R^{\leftarrow}(n),\ \beta_n = x_0$$

*where* $x_0$ is the (*necessarily*) *finite right end of* $F(x)$.

PROOF. We only comment briefly on (b), the rest being clear. For instance in case (ii) we have

$$P[(X_{L(n)} - \beta_n)/\alpha_n \le x] \to N(-\log(-\log \Phi_\alpha(x)))$$
$$= N(g(x)) = N_{1,\alpha}(x)$$

iff

$$H^s(\alpha((\log s)^2)x + \beta((\log s)^2)) \to \exp\{-e^{-g(x)/2}\} = \Phi_{\alpha/2}(x).$$

From Theorem 1.11 we may set $\beta(s) = 0$ and

$$\alpha((\log s)^2) = \left(\frac{1}{1-H}\right)^{\leftarrow}(s).$$

However $1 - H = \exp\{-R^{1/2}\}$ so

$$\left(\frac{1}{1-H}\right)^{\leftarrow}(s) = R^{\leftarrow}((\log s)^2)$$

and hence

$$\alpha(t) = R^{\leftarrow}(t)$$

as asserted. □

EXERCISES. Throughout, suppose $F$ is continuous.

4.2.1. (a) $G = N_{1\alpha}$ iff

$$R^{1/2}(x) = c(x) + \int_1^x a(t)/t\,dt \quad \text{where} \quad c(x) \to c \in \mathbb{R}$$

and

$$a(t) \to \alpha/2 \quad \text{as} \quad t \to \infty$$

iff

$$\lim_{n\to\infty} \frac{R(tx) - R(x)}{R^{1/2}(t)} = \alpha \log x.$$

(b) Suppose $x_0 < \infty$. Then $G = N_{2,\alpha}$ iff

$$R^{1/2}(x_0 - x^{-1}) = c(x) + \int_1^x t^{-1}a(t)dt$$

where $c(x) \to c \in \mathbb{R}$, $a(t) \to \alpha/2$, $t \to \infty$
iff

$$\lim_{t \downarrow 0} \frac{R(x_0 - tx) - R(x_0 - t)}{R^{1/2}(x_0 - t)} = -\alpha \log x \quad \text{for} \quad t > 0.$$

(c) $G = N$ iff

$$\lim_{s \to \infty} \frac{R^{\leftarrow}(s + x\sqrt{s}) - R^{\leftarrow}(s)}{R^{\leftarrow}(s + \sqrt{s}) - R^{\leftarrow}(s)} = x$$

iff

$$R^{\leftarrow}((\log s)^2) \in \Pi$$

iff

$$R^{1/2}(x) = c(x) + \int_{z_0}^{x} \left(\frac{1}{f(s)}\right) ds$$

where $f$ is absolutely continuous with density $f'(x) \to 0$ as $x \to x_0$ and $c(x) \to c \in \mathbb{R}$ as $x \to x_0$.

4.2.2. What happens to weak limits of records when

$$1 - F(x) = \exp\{-(x/(1-x))\}, \qquad 0 \le x \le 1?$$

4.2.3. If $G = N_{1,\alpha}$ show

$$R(x) \sim (\tfrac{1}{2}\alpha \log x)^2. \tag{4.18}$$

If $R(x) = (\frac{1}{2}\alpha \log x)^2$ show records have limit distribution $N_{1,\alpha}$. Show (4.18) is necessary for $G = N_{1\alpha}$ but not sufficient by considering

$$R^{1/2}(x) = \int_1^x \frac{1}{2} \alpha(1 - \cos(\log t)) t^{-1} dt.$$

4.2.4. Call $\{X_{L(n)}\}$ relatively stable in probability if there exists $B_n > 0$ such that

$$X_{L(n)}/B_n \xrightarrow{P} 1$$

as $n \to \infty$. Show that the following conditions are equivalent and any one of them implies $\{X_{L(n)}\}$ is relatively stable (assume $x_0 = \infty$):

(i) $X_{L(n)}/R^{\leftarrow}(n) \xrightarrow{P} 1$

(ii) $\displaystyle\lim_{x \to \infty} \frac{R(tx) - R(x)}{R^{1/2}(x)} = \infty$

(iii) $R^{\leftarrow}((\log x)^2) \in RV_0$

(iv) $\displaystyle\lim_{x \to \infty} R^{\leftarrow}(x + cx^{1/2})/R^{\leftarrow}(x) = 1$
for $c \in \mathbb{R}$.

(v) Maxima of iid random variables from the associated distribution $H(x) = 1 - \exp\{-R^{1/2}(x)\}$ are relatively stable in probability; i.e., $1 - H$ is rapidly varying.

4.2.5. Suppose $x_0 = \infty$.
(a) If $X_{L(n)}/R^{\leftarrow}(n) \xrightarrow{P} 1$ then $M_n/R^{\leftarrow}(\log n) \xrightarrow{P} 1$.

(b) If $R(x) = (\log x)^2$, then $M_n/R^{\leftarrow}(\log n) \xrightarrow{P} 1$ but $\{X_{L(n)}\}$ is not relatively stable in probability.
(c) If (4.13) holds with $G = N$ then $\{X_{L(n)}\}$ is stable in probability.

4.2.6. If (4.13) holds write $R \in DR(G)$.
(a) If $R \in DR(G)$ and for some extreme value distribution $G_1$, we have $F \in D(G_1)$, then $G_1 = \Lambda$.
(b) Let

$$R_\alpha(x) = (\tfrac{1}{2}\beta \log x)^\alpha$$

for $x > e$, $\alpha > 0$, $\beta > 0$ and call the corresponding distribution $F_\alpha$. Check
(i) For $\alpha > 2$, $R_\alpha \in DR(N)$, $F_\alpha \in D(\Lambda)$.
(ii) For $\alpha = 2$, $R_2 \in DR(N_{1,\alpha})$, $F_2 \in D(\Lambda)$.
(iii) For $1 < \alpha < 2$, $R_\alpha \notin DR(N_{1,\gamma})$ for any $\gamma$ and $F_\alpha \in D(\Lambda)$. Also $R_\alpha \in DR(N)$.
(iv) For $\alpha = 1$, $R_\alpha \notin D(N_{1,\gamma})$ for any $\gamma$, $F_1 \in D(\Phi_{\beta/2})$, $R_\alpha \notin D(N)$.
Conclude

$$D(\Lambda) \cap DR(N) \neq \varnothing$$

$$D(\Lambda) \cap DR(N_{1,\beta}) \neq \varnothing \qquad \text{for any } \beta > 0$$

$$D(\Lambda) \cap DR(N_{2,\beta}) \neq \varnothing \qquad \text{for any } \beta > 0.$$

However $\{M_n\}$ can have a limit distribution but not $\{X_{L(n)}\}$.

4.2.7. Let $R^{1/2}(x) = x^{1/2} + \frac{1}{2}x^{-1/2} \sin x$, $x \geq 1$. Then $H = 1 - \exp\{-R^{1/2}\} \in D(\Lambda)$ and hence $R \in DR(N)$. But $F = 1 - \exp(-R\}$ is not in $D(\Lambda)$.

4.2.8. Let $H = 1 - \exp\{-R^{1/2}\}$ and suppose $F = 1 - \exp\{-R\}$ has a differentiable density.
(a) If $H$ satisfies (1.19) then $R \in DR(N_{1,2\alpha})$ and $F \in D(\Lambda)$.
(b) If $H$ satisfies (1.20) then $R \in DR(N_{2,2\alpha})$ and $F \in D(\Lambda)$.
(c) If $H$ satisfies (1.24) then $R \in DR(N)$ and $F \in D(\Lambda)$.

## 4.3. Extremal Processes

For the study of the stochastic behavior of maxima and records, extremal processes are a useful tool.

Let $F$ be a distribution function on $\mathbb{R}$ and define a family of finite dimensional distributions $F_{t_1,\ldots,t_k}(x_1,\ldots,x_k)$ for $k \geq 1$, $t_i > 0$, $x_i \in \mathbb{R}$, $i = 1,\ldots,k$ by

$$F_{t_1,\ldots,t_k}(x_1,\ldots,x_k)$$
$$F^{t_1}\left(\bigwedge_{i=1}^{k} x_i\right) F^{t_2-t_1}\left(\bigwedge_{i=2}^{k} x_i\right) \ldots F^{t_k-t_{k-1}}(x_k). \tag{4.19}$$

If we compare (4.19) with (4.5) we see the formulas are the same except that in (4.5) the $t_i$'s are restricted to be positive integers. The family (4.19) forms a consistent family of finite dimensional distributions so that by the Kolmogorov extension theorem there exists a continuous time stochastic process $Y = \{Y(t), t > 0\}$ with these finite dimensional distributions. Such a process is called an *extremal process* generated by $F$ or an extremal-$F$ process.

The following is a constructive approach to extremal processes: Suppose $x_l$ and $x_0$ are the left and right endpoints of $F$. Let $N = \sum_k \varepsilon_{(t_k, j_k)}$ be PRM on $(0, \infty) \times [x_l, x_0]$ (if either $x_l$ or $x_0$ is infinite, change a square bracket to a parenthesis) with mean measure ($x_l < a < b < x_0$)

$$E\{N(0,t] \times (a,b]\} = t(-\log F(a) - (-\log F(b))).$$

Set

$$Y(t) = \sup\{j_k \colon t_k \le t\} \tag{4.20}$$

for $t > 0$. Then this $Y$ process has (4.19) as its finite dimensional distributions. This may be readily checked in a manner similar to the development of (4.5). In particular

$$\begin{aligned} P[Y(t) \le t] &= P[N((0,t] \times (x,\infty)) = 0] \\ &= \exp\{-EN((0,t] \times (x,\infty))\} = F^t(x). \end{aligned}$$

The reason for the interest in extremal processes is that a sequence of maxima of iid random variables can be embedded in an extremal process. If $\{X_n, n \ge 1\}$ are iid random variables with common distribution $F$ and $M_n = \bigvee_{i=1}^n X_i$ then as random elements of $R^\infty$

$$\{M_n, n \ge 1\} \stackrel{d}{=} \{Y(n), n \ge 1\}. \tag{4.21}$$

This is readily checked by noting that if we restrict the $t_i$'s in (4.19) to be integers, we get (4.5). The sequence $\{M_n, n \ge 1\}$ may be considered embedded in $Y$ in the sense of (4.21); we can always switch spaces to get a sequence distributionally equivalent to $\{M_n\}$ which is embedded in $Y$.

Here are some elementary properties of $Y$:

**Proposition 4.7.**
(i) *$Y$ is stochastically continuous.*
(ii) *There is a version in $D(0, \infty)$, the space of right continuous functions on $(0, \infty)$, with finite limits existing from the left.*
(iii) *$Y$ has nondecreasing paths and almost surely*

$$\lim_{t \to \infty} \uparrow Y(t) = x_0, \qquad \lim_{t \to 0} \downarrow Y(t) = x_l.$$

(iv) *$Y$ is a Markov jump process with*

$$P[Y(t+s) \le x \mid Y(s) = y] = \begin{cases} F^t(x) & x \ge y \\ 0 & x < y \end{cases}$$

*for $t > 0$, $s > 0$. Set $Q(x) = -\log F(x)$. The parameter of the exponential holding time in state $x$ is $Q(x)$, and given that a jump is due to occur the process jumps from $x$ to $(-\infty, y]$ with probability*

$$\begin{cases} 1 - (Q(y)/Q(x)) & \text{if } y > x \\ 0 & \text{if } y \le x. \end{cases}$$

Proof. (ii) Consider $Y$ defined by (4.20). It follows from the definition that $Y$ has nondecreasing paths and hence each path has finite limits from the left. We now check that $Y(t,\omega)$ is right continuous for almost all $\omega$. Suppose initially $x_l$ is not an atom of $F$; a small modification is necessary to the discussion if $x_l$ is a finite atom. We show there exist $\omega$-sets $\Lambda_i$, $i = 2$ such that $P\Lambda_i = 1$ and

$$\omega \in \Lambda_1 \quad \text{implies} \quad Y(t,\omega) > x_l \quad \text{for all} \quad t > 0 \tag{4.22}$$

and

$$\omega \in \Lambda_2 \quad \text{implies} \quad N(\omega,(0,t] \times (x,\infty)) < \infty \quad \text{for all} \quad x > x_l \quad \text{and all } t > 0. \tag{4.23}$$

For the first assertion note that for any $t > 0$

$$EN((0,t] \times (x_l,\infty)) = tQ(x_l) = \infty$$

so that

$$P[N((0,t] \times (x_l,\infty)) = \infty] = 1.$$

If $\{t_i\}$ is countable and dense in $(0,\infty)$

$$\begin{aligned} &P[N((0,t] \times (x_l,\infty)) = \infty \text{ for all } t > 0] \\ &= P\left(\bigcap_i [N((0,t_i] \times (x_l,\infty)) = \infty]\right) \\ &= \lim_{t_i \to 0} P[N((0,t_i] \times (x_l,\infty)) = \infty] = 1. \end{aligned}$$

This is equivalent to (4.22). For (4.23) observe that for fixed $t > 0$, $x > x_l$

$$EN((0,t] \times (x,\infty)) = tQ(x) < \infty$$

and so

$$P[N((0,t] \times (x,\infty)) < \infty] = 1.$$

Therefore

$$\begin{aligned} &P[N((0,t] \times (x,\infty)) < \infty \text{ for all } x > x_l] \\ &= \lim_{x \downarrow x_l} P[N((0,t] \times (x,\infty)) < \infty] = 1 \end{aligned}$$

and

$$\begin{aligned} &P[N((0,t] \times (x,\infty)) < \infty \text{ for all } x > x_l, t > 0] \\ &= \lim_{t \to \infty} P[N((0,t] \times (x,\infty)) < \infty \text{ for all } x > x_l] = 1. \end{aligned}$$

Now for $\omega \in \Lambda_1 \cap \Lambda_2$ and any $t$

$$N(\omega,(0,2t] \times (Y(t,\omega),\infty)) < \infty$$

and hence the points of realization $\omega$ falling in $(0,2t] \times (Y(t,\omega),\infty)$ cannot cluster. So there exists $\delta = \delta(t,\omega)$ and

$$N(\omega,(t,t+\delta] \times (Y(t,\omega),\infty) = 0.$$

Hence $Y(s,\omega) = Y(t,\omega)$ for $t \le s \le t+\delta$ showing right continuity at $t$ when $\omega \in \Lambda_1 \cap \Lambda_2$.

If $x_l$ is an atom, then $Q(x_l) < \infty$ so that

$$P[N((0,t] \times (x_l,\infty)) < \infty \text{ for all } t > 0] = 1$$

and hence points of $(0,2t] \times (x_l,\infty)$ do not cluster. The proof can be completed as before.

(i) Since we may take almost all paths in $D(0,\infty)$, $Y$ must be stochastically continuous from the right; i.e., as $s \downarrow t$, $Y(s) \xrightarrow{P} Y(t)$ since the convergence is in fact almost sure. If $Y$ is not stochastically continuous from the left then there is $t_0$ and as $s \uparrow t_0$, $Y(s)$ does not converge in probability to $Y(t_0)$. Since $Y(s) \to Y(t_0-)$ a.s. this means

$$0 < P[Y(t_0-) < Y(t_0)].$$

But

$$\begin{aligned} P[Y(t_0-) < Y(t_0)] &\le P[N(\{t_0\} \times (x_l,\infty)) > 0] \\ &\le EN(\{t_0\} \times (x_l,\infty)) = 0 \end{aligned}$$

since the Lebesgue measure of $\{t_0\}$ is zero.

(iii) Since $P[Y(t) \le x_0] = F^t(x_0) = 1$ and for $M < x_0$.

$$P[Y(t) \le M] = F^t(M) \to 0$$

as $t \to \infty$ we have

$$Y(t) \xrightarrow{P} x_0.$$

However $Y$ has nondecreasing paths so convergence in probability is the same as a.s. convergence by the subsequence characterization of convergence in probability. The convergence to $x_l$ as $t \to 0$ is handled similarly.

(iv) These results parallel those of Proposition 4.1 in discrete time. $Y$ is Markov with the given transition probability because of the form of the finite-dimensional distributions. The form of the holding time parameter may be obtained from infinitesimal conditions or by observing

$$\begin{aligned} P[Y(t+s) = Y(s) | Y(s) = x] &= P[N((s,t+s] \times (x,\infty)) = 0] \\ &= \exp\{-E(N(s,t+s] \times (x,\infty))\} = e^{-tQ(x)} \end{aligned}$$

so the holding time in $x$ is at least $t$ with exponential probability $e^{-tQ(x)}$. To compute the jump distribution $\Pi(x,(y,\infty))$ for $y > x$ note from (4.3) as $t \to 0$

$$\begin{aligned} t^{-1}P[Y(t+s) > y | Y(s) = x] &\to \lambda(x)\Pi(x,(y,\infty)) \\ &= Q(x)\Pi(x,(y,\infty)). \end{aligned}$$

On the other hand

$$t^{-1}P[Y(t+s) > y \mid Y(s) = x] = t^{-1}(1 - F^t(y))$$
$$= t^{-1}(1 - e^{-tQ(y)}) \to Q(y)$$

and so

$$\Pi(x, (y, \infty)) = Q(y)/Q(x). \qquad \square$$

Now let $\{\tau_n, -\infty < n < \infty\}$ be the jump times of $Y$ so that $\{Y(\tau_n)\}$, the range of $Y$, is a discrete indexed Markov process and by (iv) earlier we have

$$P[Y(\tau_{n+1}) > y \mid Y(\tau_n) = x] = Q(y)/Q(x).$$

Note that if $Q(x) = e^{-x}$, i.e., if $F(x) = \Lambda(x)$, then

$$P[Y(\tau_{n+1}) > y \mid Y(\tau_n) = x] = e^{-(y-x)}$$
$$= P[\Gamma_{n+1} > y \mid \Gamma_n = x]$$

where $\{\Gamma_n\}$ was defined in Proposition 4.1. Therefore $\{Y(\tau_n)\}$ is homogeneous PRM on $\mathbb{R}$. Let $S(x) = -\log Q(x) = -\log(-\log F(x))$. The following parallels Proposition 4.1 and Corollary 4.2.

**Proposition 4.8.** (i) *If $F = \Lambda$, then $\{Y(\tau_n)\}$ are the points of homogeneous* PRM *on $\mathbb{R}$.*
(ii) *If $F$ is continuous, then $\{Y(\tau_n)\}$ are the points of* PRM *on $(x_l, x_0)$ with mean measure $S$.*
(iii) *If $F$ is continuous*

$$\sum_{n=-\infty}^{\infty} \varepsilon_{(Y(\tau_n), \tau_{n+1}-\tau_n)}$$

*is* PRM *on $(x_l, x_0) \times (0, \infty)$ with mean measure*

$$\mu^*((a, b] \times (t, \infty)) = \int_{Q(b)}^{Q(a)} y^{-1}e^{-ty}\,dy$$

*for $x_l < a < b < x_0$, $t > 0$.*
(iv) *If $F$ is continuous $\{Y^{\leftarrow}(x), x_l < x < x_0\}$ is a process with independent increments and exponential marginals:*

$$P[Y^{\leftarrow}(x) \le t] = P[x \le Y(t)] = 1 - e^{-Q(x)t}.$$

Proof. (ii) If $Y_\Lambda$ is extremal-$\Lambda$ then $S^{\leftarrow}(Y_\Lambda)$ is extremal-$F$ since $S^{\leftarrow} \circ Y$ has finite dimensional distributions given by 4.19. For example, for $k = 1$ we have

$$P[S^{\leftarrow}(Y_\Lambda(t)) \le x] = P[Y_\Lambda(t) \le S(x)]$$
$$= \Lambda^t(S(x)) = F^t(x).$$

If $F$ is continuous then $S^{\leftarrow}$ is strictly increasing and $\{S^{\leftarrow}(Y_\Lambda(\tau_n))\} \stackrel{d}{=} \{Y(\tau_n)\}$

where the process $Y$ on the right is extremal-$F$. The result follows by the transformation theory for Poisson processes as in Proposition 4.1.
(iii) This follows as in Proposition 4.1 except that by Proposition 3.8

$$\mu^*((a,b] \times (t,\infty)) = \int_{(a,b]} S(dx)e^{-tQ(x)}$$
$$= \int_{Q(b)}^{Q(a)} e^{-ty}y^{-1}\,dy.$$

(iv) This follows as in Corollary 4.2 since

$$Y^{\leftarrow}(b) - Y^{\leftarrow}(a) = \sum_n (\tau_{n+1} - \tau_n)\varepsilon_{Y(\tau_n)}(a,b]$$
$$= \int_{(0,\infty)} t \sum_n \varepsilon_{(Y(\tau_n),\tau_{n+1}-\tau_n)}((a,b] \times dt). \qquad \square$$

We next discuss the point process of jump times.

**Proposition 4.9.** *Suppose $F$ is continuous. Then*

$$\mu_\infty := \sum_n \varepsilon_{\tau_n}$$

*is* PRM *on* $(0,\infty)$ *with mean measure of* $(a,b]$ *equal to* $\log(b/a)$, $0 < a < b$.

It will be enough to show that $\mu_\infty(\cdot \cap (0,K])$ is PRM on $(0,K]$ where $K$ is arbitrary. Toward this goal we prove the next lemma (cf. Exercise 3.3.8).

**Lemma 4.10.** *If $Q(x) = -\log F(x)$, set for $y > 0$*

$$Q^{\leftarrow}(y) = (1/Q)^{\leftarrow}(y^{-1})$$
$$= \inf\{s\colon Q(s) \le y\}.$$

*Suppose $\{E_i, i \ge 1\}$ is* iid *with $P[E_i > x] = e^{-x}$, $x > 0$, and set $\Gamma_n = E_1 + \cdots + E_n$. Let $\{U_i, i \ge 1\}$ be* iid *uniformly distributed on $(0,K]$ and suppose $\{U_i\}$ and $\{E_i\}$ are independent. Then*

$$N^{\#} := \sum_{i=1}^{\infty} \varepsilon_{(U_i, Q^{\leftarrow}(\Gamma_i/K))}$$

*is PRM on $(0,K] \times (x_l, x_0)$ with mean measure of $(0,t] \times (a,x_0)$ equal to $tQ(a)$, $t > 0$, $x_l < a < x_0$; i.e.,*

$$N^{\#} \stackrel{d}{=} N \quad \textit{restricted to } (0,K] \times (x_l \times x_0)$$

*(where $N$ is used in the construction* (4.20)*).*

**Proof.** As with (0.6) we have

$$Q^{\leftarrow}(y) > t \quad \text{iff} \quad Q(t) < y. \tag{4.24}$$

Since $\sum_n \varepsilon_{K^{-1}\Gamma_n}$ is PRM on $(0, \infty)$ with mean measure $\mathrm{Km}(\cdot)$ ($m$ is Lebesgue measure) we have by Proposition 3.7 that

$$\sum_n \varepsilon_{Q^{\leftarrow}(\Gamma_n/K)}$$

is PRM with mean measure $Km \circ (Q^{\leftarrow})^{-1}$. However for $x_l < a < x_0$

$$\begin{aligned} Km \circ (Q^{\leftarrow})^{-1}(a, \infty) &= Km\{s > 0: Q^{\leftarrow}(s) > a\} \\ &= Km\{s: 0 < s < Q(a)\} \qquad \text{by (4.24)} \\ &= KQ(a). \end{aligned}$$

By Proposition 3.7, $N^{\#}$ is PRM on $(0, K] \times (x_l, x_0)$ and the mean measure of $(0, t] \times (a, x_0)$ is $(0 < t < K, x_l < a < x_0)$

$$K^{-1}tKQ(a) = tQ(a) = E(N(0, t] \times (a, \infty)). \qquad \square$$

Proof. Now for the proof of Proposition 4.9: Define

$$Y^{\#}(t) = \sup\{Q^{\leftarrow}(\Gamma_i/K): U_i \le t\}$$

on $(0, K]$ so that $Y^{\#} \stackrel{d}{=} Y$ and instead of analyzing the jump times of $Y$ we analyze these of $Y^{\#}$. Since $Q^{\leftarrow}(\Gamma_1/K) > Q^{\leftarrow}(\Gamma_2/K) > \cdots$, if we define for $0 < t < K$

$$T(t) = \inf\{i \ge 1: U_i \le t\}$$

then

$$Y^{\#}(t) = Q^{\leftarrow}(\Gamma_{T(t)}/K).$$

Since $F$ continuous makes $Q^{\leftarrow}$ strictly decreasing, the jump times of $\{Y^{\#}(t)\}$ and those of $\{T(t)\}$ coincide. However observe that for $0 < t \le K$

$$\begin{aligned} T(t) &= \inf\{n \ge 1: U_n \le t\} \\ &= \inf\{n \ge 1: U_n^{-1} \ge t^{-1}\} \\ &= \inf\left\{n \ge 1: \bigvee_{i=1}^{n} U_i^{-1} \ge t^{-1}\right\} = \eta(t^{-1}) \end{aligned}$$

where following Corollary 4.2 we set

$$\eta(s) = \inf\left\{n \ge 1: \bigvee_{i=1}^{n} U_i^{-1} \ge s\right\}.$$

Now observe the

$$\begin{aligned} &\text{jump times of } T(t), 0 < t \le K \\ &\quad = \text{jump times of } \eta(s), K^{-1} \le s < \infty \\ &\quad = \text{records of } \left\{\bigvee_{j=1}^{n} U_j^{-1}\right\} \text{ in } [K^{-1}, \infty). \end{aligned}$$

From Proposition 4.1 the records of $\{\bigvee_{j=1}^{n} U_j^{-1}\}$ form a Poisson process on $(K^{-1}, \infty)$ and the mean measure of $(a, b](K^{-1} < a < b)$ is

$$\begin{aligned} &-\log P[U_1^{-1} > b] - (-\log P(U_1^{-1} > a]) \\ &= -\log P[U_1 \le b^{-1}] - (-\log P\{U_1 \le a^{-1}]) \\ &= -\log(b^{-1}/K) - (-\log(a^{-1}/K)) = \log(b/a). \qquad \square \end{aligned}$$

Consider now $\{X_n, n \ge 1\}$ iid from a continuous distribution $F(x)$ and as usual set $M_n = \bigvee_{i=1}^{n} X_i$. If $Y$ is extremal-$F$ we have

$$\{M_n, n \ge 1\} \stackrel{d}{=} \{Y(n), n \ge 1\}$$

so to study the record structure of $\{X_n\}$ we may as well suppose $\{M_n\}$ is embedded in $Y$; i.e., we study functionals of $\{Y(n)\}$ instead of $\{M_n\}$. So, for instance, with this point of view

$$\mu := \sum_j \varepsilon_{L(j)} = \sum_i 1_{[\mu_\infty(i-1,i]>0]}\varepsilon_i(\cdot)$$

can be considered as a functional of $Y$.

Since we may consider both $\mu$ and $\mu_\infty$ defined on the same space, we may hope to compare them $\omega$ by $\omega$. What is the relation of $\mu$ to $\mu_\infty$? Observe that

$$\begin{aligned} [\mu(n-1, n] = 1] &= [\text{record at } n] \\ &= [\mu_\infty(n-1, n] > 0] \end{aligned}$$

and thus $\mu_\infty$ counts jumps that $\mu$ misses since $\mu$ only checks to see whether or not $Y(n) > Y(n-1)$ but is not sensitive to all jumps of $Y$ in $(n-1, n]$. If $\mu_\infty(n-1, n] > 1$ for infinitely many $n$, then $\mu$ and $\mu_\infty$ will not be related in a useful way. Fortunately this is not the case.

**Proposition 4.11.** *For $F$ continuous, we have*

$$P[\mu_\infty(n, n+1] > 1 \text{ i.o.}] = 0.$$

PROOF. From the Borel–Cantelli lemma it suffices to show

$$\sum_{n=1}^{\infty} P[\mu_\infty(n, n+1] > 1] < \infty.$$

Since $\mu_\infty(n, n+1]$ is a Poisson random variable we have

$$\begin{aligned} \sum_n P[\mu_\infty(n, n+1] > 1] &= \sum_n (1 - \exp\{-\log(n^{-1}(n+1))\} \\ &\quad - \log(n^{-1}(n+1))\exp\{-\log(n^{-1}(n+1))\}) \\ &\le \sum_{n=1}^{\infty} \log(n^{-1}(n+1))(1 - \exp\{-\log(n^{-1}(n+1))\}) \\ &\le \sum_{n=1}^{\infty} (\log(n^{-1}(n+1)))^2. \end{aligned}$$

Since $(\log(n^{-1}(n+1)))^2 \sim n^{-2}$ as $n \to \infty$, the desired convergence of the series follows by a comparison argument. $\square$

A conclusion from Proposition 4.11 is that for almost all $\omega$, there exists $n_0(\omega)$ such that if $n \geq n_0(\omega)$ then

$$\mu(\omega, (n, n+1]) = \mu_\infty(\omega, (n, n+1]). \tag{4.25}$$

We now use this to prove Corollary 4.5 again.

**Corollary 4.5.** *If $F$ is continuous and $\mu$ and $\mu_\infty$ are defined on the same space we have*

$$\mu_n = \mu(n\cdot) \Rightarrow \mu_\infty$$

in $M_p((0, \infty))$.

PROOF. Let $f \in C_K^+((0, \infty))$. Suppose $f(x) = 0$ for $x \in [\delta, \delta^{-1}]^c$. We need to show in $\mathbb{R}$

$$\int f(x)\mu_n(dx) \Rightarrow \int f(x)\mu_\infty(dx)$$

(cf. Proposition 3.19).

Now

$$\begin{aligned} \mu_n(f) &= \sum_{j=1}^{\infty} f(n^{-1}L(j)) \\ &= \sum_{i=1}^{\infty} f(n^{-1}i)1_{[Y(i) > Y(i-1)]} \\ &= \sum_{i:n^{-1}i \in [\delta, \delta^{-1}]} f(n^{-1}i)1_{[\mu_\infty(i-1,i] > 0]}. \end{aligned}$$

For $n \geq (n_0(\omega)+1)\delta^{-1}$, if $i \geq n\delta$ then we have by (4.25) $\mu_\infty(\omega, (i-1, i]) = 1_{[\mu_\infty(i-1,i] > 0]}(\omega) = 1_{[\mu_\infty(i-1,i] = 1]}(\omega)$ and for such $n$

$$\begin{aligned} \mu_n(f, \omega) &= \sum_{i:n^{-1}i \in [\delta, \delta^{-1}]} \int_{(i-1,i]} f(n^{-1}i)\mu_\infty(\omega, dx) \\ &= \int g_n(x)\mu_\infty(\omega, dx) \end{aligned}$$

where

$$g_n(x) = \sum_{i=1}^{\infty} f(n^{-1}i)1_{(i-1,i]}(x).$$

The preceding shows that $\mu_n(f) - \mu_\infty(g_n) \to 0$ a.s., and therefore it suffices to show $\mu_\infty(g_n) \Rightarrow \mu_\infty(f)$.

Observe

$$g_n(nx) = \sum_{i=1}^{\infty} f(n^{-1}i) 1_{(n^{-1}(i-1), n^{-1}i]}(x) \to f(x)$$

as $n \to \infty$ and for $n$ large

$$g_n(nx) \le \sup_{y>0} f(y) 1_K(x)$$

where $K$ is compact and $K \supset [\delta^{-1}, \delta]$.

Therefore since $\mu_\infty(\cdot) \stackrel{d}{=} \mu_\infty(n(\cdot))$ (Exercise 4.1.6) we have

$$\begin{aligned}\mu_\infty(g_n) &= \int_{(0,\infty)} g_n(x)\mu_\infty(dx) = \int_{(0,\infty)} g_n(ny)\mu_\infty(ndy)\\ &\stackrel{d}{=} \int_{(0,\infty)} g_n(ny)\mu_\infty(dy) \to \int_{(0,\infty)} f(y)\mu_\infty(dy)\\ &= \mu_\infty(f),\end{aligned}$$

the convergence following by the dominated convergence theorem. □

We now use these ideas to study the asymptotic behavior of $\mu(1,n]$ and $L(n)$. Note first that the structure of $\mu_\infty$ is quite simple since a time change renders it homogeneous; i.e., $\{\mu_\infty(1,e^t], t > 0\}$ is a homogeneous PRM. (It is PRM by Proposition 3.7 and homogeneous since $E\mu_\infty(1,e^t] = \log(e^t/1) = t$.) For a homogeneous Poisson process, the following are standard:

Strong law of large numbers: $\mu_\infty(1,e^t] \sim t, \quad t \to \infty;$

Central limit theorem: $\dfrac{\mu_\infty(1,e^t] - t}{t^{1/2}} \Rightarrow N$

where $N$ is standard normal;

Iterated logarithm theorem: $\limsup_{t\to\infty} \dfrac{\mu_\infty(1,e^t] - t}{(2t \log\log t)^{1/2}} = 1$ a.s.

$$\liminf_{t\to\infty} \frac{\mu_\infty(1,e^t] - t}{(2t \log\log t)^{1/2}} = -1 \text{ a.s.}$$

Changing variables we get

$$\mu_\infty(1,t] \sim \log t, \qquad t \to \infty; \tag{4.26}$$

$$\frac{\mu_\infty(1,t] - \log t}{(\log t)^{1/2}} \Rightarrow N, \qquad t \to \infty; \tag{4.27}$$

$$\limsup_{t\to\infty} \frac{\mu_\infty(1,t] - \log t}{(2 \log t \log\log\log t)^{1/2}} = 1 \text{ a.s.} \tag{4.28}$$

$$\liminf_{t\to\infty} \frac{\mu_\infty(1,t] - \log t}{(2 \log t \log\log\log t)^{1/2}} = -1 \text{ a.s.}$$

From Proposition 4.11 and (4.25) we get that for large enough $n$, $n \geq n_0(\omega)$ say,

$$\mu_\infty(\omega, [1, n]) = j(\omega) + \mu(\omega, [1, n]) \tag{4.29}$$

where $j(\omega)$ is a finite integer valued variable representing the jumps that $\mu_\infty$ sees which are missed by $\mu$. Hence if $\alpha_n \to \infty$

$$(\mu_\infty(1, n] - \mu(1, n])/\alpha_n \to 0 \text{ a.s.}$$

and we get the following.

**Proposition 4.12.** *If we replace $t$ by $n$ and $\mu_\infty(1, t]$ by $\mu(1, n]$ = number of records in the first $n$ observations then* (4.26), (4.27), *and* (4.28) *all hold.*

Applying Proposition 3.7 yet again we see that the points of $\mu_\infty$ which are greater than 1, which we label $\tau_1 < \tau_2 < \cdots$, can be represented as

$$\{\tau_n, n \geq 1\} = \{\exp\{\Gamma_n\}, n \geq 1\}$$

where as before $\Gamma_n$ is a sum $E_1 + \cdots + E_n$ of $n$ iid exponentially distributed random variables. Hence for $\log e^{\Gamma_n} = \Gamma_n$ we get a strong law, central limit theorem and law of the iterated logarithm. Referring to (4.29) and Proposition 4.11 we see that $\{L(n)\}$ and $\{e^{\Gamma_n}\}$ are related for $n \geq n_0(\omega)$ by

$$L(n, \omega) = \exp\{\Gamma_{n+j(\omega)}(\omega)\} + \delta_n(\omega) = \tau_{n+j(\omega)}(\omega) + \delta_n(\omega) \tag{4.30}$$

where $|\delta_n| \leq 1$. This implies

$$\limsup_{n\to\infty} \left|\frac{\log L(n, \omega) - \Gamma_n(\omega)}{\log n}\right| \leq j(\omega) \tag{4.31}$$

a.s. This is checked as follows: From (4.30)

$$\begin{aligned}\log L(n, \omega) - \Gamma_n(\omega) &= \log(\exp\{\Gamma_{n+j(\omega)}(\omega)\} + \delta_n(\omega)) - \Gamma_n(\omega)\\ &= \Gamma_{n+j(\omega)}(\omega) - \Gamma_n + o(1)\end{aligned}$$

and hence

$$\begin{aligned}&\limsup_{n\to\infty} \left|\frac{\log L(n, \omega) - \Gamma_n(\omega)}{\log n}\right|\\ &= \limsup_{n\to\infty} \left|\frac{\sum_{i=n+1}^{n+j(\omega)} E_i}{\log n}\right| \leq j(\omega) \text{ a.s.}\end{aligned}$$

since for any $i \geq 1$

$$\limsup_{n\to\infty} \frac{E_{n+i}}{\log n} = 1, \text{ a.s.}$$

(cf. Exercise 4.1.1(a)). The relation (4.31) says asymptotic behavior of $\Gamma_n$ can be transferred to $L(n)$:

**Proposition 4.13.** *If $F$ is continuous then we have*
(a) $\log L(n) \sim n$ *as* $n \to \infty$;
(b) $(\log L(n) - n)/\sqrt{n} \Rightarrow N$ *as* $n \to \infty$;
(c) $\limsup_{n\to\infty} (\log L(n) - n)/\sqrt{2n \log\log n} = 1$ a.s.

$$\liminf_{n\to\infty} (\log L(n) - n)/\sqrt{2n \log\log n} = -1 \text{ a.s.}$$

We now show that the same asymptotic behavior holds for the interrecord times $\{L(n) - L(n-1), n \geq 2\}$ as for $\{L(n)\}$. We first investigate $\{\tau_n - \tau_{n-1}\}$ and show

$$\limsup_{n\to\infty} |\log(\tau_n - \tau_{n-1}) - \log \tau_n|/\log n = 1 \tag{4.32}$$

a.s. Recalling that $\tau_n = e^{\Gamma_n}$ where $\Gamma_n = E_1 + \cdots + E_n$, we get

$$\begin{aligned} |\log(\tau_n - \tau_{n-1}) - \log \tau_n| &= -\log(1 - \tau_n^{-1}\tau_{n-1}) \\ &= -\log(1 - e^{-E_n}). \end{aligned}$$

Now it is readily checked that

$$\{-\log(1 - e^{-E_n})\} \stackrel{d}{=} \{E_n\}$$

in $\mathbb{R}^\infty$ and since $\limsup_{n\to\infty} E_n/\log n = 1$ a.s. (Exercise 4.1.12), (4.32) follows.

We may now prove the next result on the asymptotic behavior of interrecord times.

**Proposition 4.14.** *$F$ is continuous. We have that the results of Proposition* 4.13 *hold with $L(n) - L(n-1)$ everywhere replacing $L(n)$. Also*

$$\limsup_{n\to\infty} |\log(L(n) - L(n-1)) - \log L(n)|/\log n = 1 \text{ a.s.} \tag{4.33}$$

Proof. It suffices to prove (4.33). From (4.30)

$$\begin{aligned} &|\log(L(n) - L(n-1)) - \log L(n)|/\log n \\ &= |\log(\tau_{n+j} - \tau_{n+j-1} + \delta_n - \delta_{n-1}) - \log(\tau_{n+j} + \delta_n)|/\log n \\ &= |\log(\tau_{n+j} - \tau_{n+j-1}) - \log \tau_{n+j}|/\log n + o(1). \end{aligned}$$

The last step is verified by noting first of all that since $|\delta_n| \leq 1$ we have

$$\log(\tau_{n+j} + \delta_n) = \log \tau_{n+j} + o(1)$$

as $n \to \infty$, and furthermore we can show

$$\begin{aligned} \log(\tau_{n+j} - \tau_{n+j-1} + \delta_n - \delta_{n-1})/\log n = \log(\tau_{n+j} - \tau_{n+j-1})/\log n \\ + o(1) \end{aligned}$$

as follows: From (4.32) it follows that

$$(\log(\tau_n - \tau_{n-1}) - \log \tau_n)/n \to 0,$$

whence because $\tau_n = \exp\{\Gamma_n\}$

$$n^{-1}(\log(\tau_n - \tau_{n-1}) - 1 \to 0$$

and

$$\log(\tau_n - \tau_{n-1}) \sim n.$$

Thus for any fixed $i$ we have $\tau_{n+i} - \tau_{n+i-1} \to \infty$ so that

$$\begin{aligned}&\log(1 + (\delta_n - \delta_{n-1})/(\tau_{n+j} - \tau_{n+j-1}))/\log n \\ &\quad \sim (\delta_n - \delta_{n-1})/((\tau_{n+j} - \tau_{n+j-1})\log n) \to 0 \text{ a.s.}\end{aligned}$$

Therefore

$$\begin{aligned}&\limsup_{n\to\infty} |\log(L(n) - L(n-1)) - \log L(n)|/\log n \\ &\quad = \limsup_{n\to\infty} |\log(\tau_{n+j} - \tau_{n+j-1}) - \log \tau_{n+j}|/\log n = 1\end{aligned}$$

from (4.32). □

We end this section with a chart comparing the various descriptive quantities of interest in discrete time, with the analogous quantities in continuous time.

A comparison of $\{M_n, n \geq 1\}$ and $\{Y(t), t > \infty\}$:

| | |
|---|---|
| 1. $\{M_n, n \geq 1\}$, underlying distribution $F$, is a Markov process in discrete time. | 1. $\{Y(t), t > 0\}$, extremal-$F$, is a Markov process in continuous time. |
| 2. For $k \geq 1$, $1 \leq t_1 < t_2 < \cdots < t_k$ and $t_i$ integers, $P[M_{t_i} \leq x_i, i = 1, \ldots, k]$ $= F^{t_1}\left(\bigwedge_1^k x_i\right) F^{t_2 - t_1}\left(\bigwedge_2^k x_i\right)$ $\cdots \times F^{t_k - t_{k-1}}(x_k)$. | 2. For $k \geq 1$, $0 < t_1 < \cdots < t_k$ $P[Y(t_i) \leq x_i, i = 1, \ldots, k]$ $= F^{t_1}\left(\bigwedge_1^k x_i\right) F^{t_2 - t_1}\left(\bigwedge_2^k x_i\right)$ $\cdots \times F^{t_k - t_{k-1}}(x_k)$ and $\{Y(n), n \geq 1\} \stackrel{d}{=} \{M_n, n \geq 1\}$ in $\mathbb{R}^\infty$. |
| 3. $P[M_{n+t} \leq z \mid M_n = x]$ $= \begin{cases} F^t(z), & z \geq x \\ 0, & z < x \end{cases}$ for $t > 0$, integer. | 3. $P[Y(s+t) \leq z \mid Y(s) = x]$ $= \begin{cases} F^t(z), & z \geq x \\ 0, & z < x \end{cases}$ for $t > 0$. |
| 4. $\{M_n\}$ is a Markov jump process with $(L(n), n \geq 0\}$ as jump times. | 4. $Y$ is a Markov jump process with $\{\tau_n\}$ as jump times. |
| 5. If $F$ is continuous $\sum_0^\infty \varepsilon_{M_{L(n)}} = \sum_0^\infty \varepsilon_{X_{L(n)}}$ is PRM($R$), $R = -\log(1 - F)$. | 5. If $F$ is continuous $\sum_{-\infty}^\infty \varepsilon_{Y(\tau_n)}$ is PRM($S$), $S = -\log Q = -\log(-\log F)$. |
| 6. $\sum \varepsilon_{X_{L(n)}}$ is homogeneous PRM on $(0, \infty)$ if $F(x) = 1 - e^{-x}$, $x > 0$. | 6. $\sum_{-\infty}^\infty \varepsilon_{Y(\tau_n)}$ is homogeneous PRM on $\mathbb{R}$ if $F = \Lambda$. |

7. $P[X_{L(n+1)} > y \mid X_{L(n)} = x] = \begin{cases} (1 - F(y))/(1 - F(x)), & y \ge x \\ 1, & y < x. \end{cases}$
8. $P[L(n+1) - L(n) = k \mid X_{L(n)} = x] = F^{k-1}(x)(1 - F(x)),\ k \ge 1.$
9. Let $\mu = \sum_0^\infty \varepsilon_{L(n)}$. If $F$ is continuous, $\mu$ has independent increments (recall $\{[X_k$ is a record$], k \ge 1\}$ is a sequence of independent events). For $F$ continuous
$$\mu_n := \sum_0^\infty \varepsilon_{L(j)}(n\,\cdot) = \sum_0^\infty \varepsilon_{L(j)/n}(\cdot) \Rightarrow \mu_\infty(\cdot),$$
in $M_p((0, \infty))$.

7. $P[Y(\tau_{n+1}) > y \mid Y(\tau_n) = x] = \begin{cases} Q(y)/Q(x), & y \ge x \\ 1, & y < x. \end{cases}$
8. $P[\tau_{n+1} - \tau_n > t \mid Y(\tau_n) = x] = e^{-Q(x)t},\ t > 0.$
9. Let $\mu_\infty = \sum_{-\infty}^\infty \varepsilon_{\tau_n}$. If $F$ is continuous, $\mu_\infty$ is PRM $(t^{-1}\,dt)$.

## Exercises

4.3.1. When $F$ is continuous show that $\mu_\infty$ is PRM $(t^{-1}\,dt)$ by the following procedure.
(a) When $\{M_n\}$ comes from an underlying continuous distribution
$$\{1_{[M_j > M_{j-1}]}, j \ge 2\}$$
is a sequence of independent random variables.
(b) Let $Y$ be extremal-$F$ and set
$$\mu^{(n)}(\cdot) = \#\text{jumps of } \{Y(i2^{-n}), i \ge 1\} \text{ in } (\cdot).$$
Show $\mu^{(n)}(\cdot)$ has independent increments and for $a < b$, $\mu^{(n)}(a, b] \to \mu_\infty(a, b]$, whence $\mu_\infty$ has independent increments.

4.3.2. In Corollary 4.5, does $\mu_n \to \mu_\infty$ almost surely?

4.3.3. If $\{E_n, n \ge 1)$ are iid, $P[E_i > x] = e^{-x^+}$, show
(a) $E_{L(n)}/n \to 1$ a.s.
(b) $\bigvee_{i=1}^n E_i/\log n \to 1$ a.s.
Hint: $\bigvee_{i=1}^n E_i = E_{L(\mu[1,n])}$.
(c) Under the condition of Exercise 4.1.1(c), show that if $\{X_n, n \ge 1\}$ are iid with common continuous distribution $F$ then
$$X_{L(n)}/R^{\leftarrow}(n) \to 1 \text{ a.s.}$$
and this in turn implies
$$M_n/R^{\leftarrow}(\log n) \to 1 \text{ a.s.}$$
(The most general condition for a.s. stability of $\{M_n\}$ is discussed in Barndorff-Nielsen (1963) and Resnick and Tomkins (1973).)

4.3.4. If $\{Y(t), t > 0\}$ is extremal-$\Lambda$ then $\{S^{\leftarrow}(Y(t)), t > 0)$ is extremal-$F$, where $S(x) = -\log(-\log F(x))$.

4.3.5. (a) If $\{Y(t), t > 0\}$ is extremal-$\Lambda$, so is

$$\{-\log Y^{\leftarrow}(-\log t), t > 0\}.$$

Note that the second process is not right continuous (Robbins and Siegmund, 1971; Resnick, 1974).

(b) If $Y$ is extremal-$\Lambda$, then

$$\lim_{t \to \infty} Y(t)/\log t = 1 \text{ a.s.}$$

and

$$\lim_{t \to 0} Y(t)/\log t = 1 \text{ a.s.}$$

From the result in (a), these two limiting results are the same result.

(c) Use (a), Proposition 4.8, and Exercise 4.3.4 to prove Proposition 4.9.

4.3.6. (a) Suppose $\{X(t), t > 0\}$ is a Lévy process, i.e., a process with stationary, independent increments. If the Levy measure of the process is $\nu$ then

$$\sup_{0 < s \le t} (X(s) - X(s-)) \wedge 0$$

is extremal-$F$ where for $x > 0$, $F(x) = \exp\{-\nu(x, \infty)\}$ (Dwass, 1966; Resnick and Rubinovitch, 1973).

(b) Suppose $\{X(t), t \ge 0\}$ is homogeneous Poisson, rate 1. What is $Y$? What is $\nu$? Why does Proposition 4.9 fail?

4.3.7. Suppose $Y$ is extremal-$F$ and $F$ is not necessarily continuous.

(a) If $F(x_0) > 0$

$$\begin{aligned} P[Y \text{ hits } x_0] &:= P\left(\bigcup_{t>0} [Y(t) = x_0]\right) \\ &= (Q(x_0-) - Q(x_0))/Q(x_0-) \end{aligned}$$

where $Q = -\log F$.

(b) From (a), $P[Y \text{ hits } x_0] > 0$ iff $x_0$ is an atom of $F$. Furthermore

$$\{[Y \text{ hits } x], x \in (\mathscr{C}(F))^c\}$$

is a family of mutually independent events (recall that $\mathscr{C}(F)$ is the continuity set of $F$).

(c) More generally show that

$$\{Y^{\leftarrow}(x), x_l < x < x_0\}$$

is a process with independent increments and one-dimensional exponential marginals.

4.3.8. Suppose $Y$ is extremal-$F$ and $F$ is continuous. Prove for $0 < a < b$

$$1_{[\mu_\infty(a,b] > 0]} \text{ and } Y(b)$$

are independent (Ballerini and Resnick, 1987).

4.3.9. Suppose $Y$ is extremal-$\Lambda$ and $(X_j, -\infty < j < \infty\}$ is iid with common distribution $\Lambda$. Define for $c > 0$

$$M_n^* = \bigvee_{j=-\infty}^{n} (X_j + cj).$$

(a) Show in $\mathbb{R}^\infty$

$$\{M_n^*, n \geq 1\} \stackrel{d}{=} \{Y(p^{-1}e^{cn}), n \geq 1\}$$

where $p = \int_{-\infty}^{\infty} \prod_{j=1}^{\infty} \Lambda(x + cj)\Lambda(dx) = 1 - e^{-c}$.

(b) The sequence

$$\{1_n^*, n \geq 1\} := \{1_{[M_n^* > M_{n-1}^*]}, n \geq 1\}$$

is iid Bernoulli with

$$P[1_n^* = 1] = p = 1 - e^{-c}.$$

(c) $M_n^*$ and $1_n^*$ are independent for each $n$ (use Exercise 4.3.8) (Smith and Miller, 1984; Ballerini and Resnick, 1987).

4.3.10. If $Y$ is extremal-$F$, prove for any $k$, $0 < t_1 < \cdots < t_k$

$$\{Y(t_i), 1 \leq i \leq k\} \stackrel{d}{=} \left(U_1, U_1 \vee U_2, \ldots, \bigvee_{i=1}^{n} U_i\right)$$

where $U_1, \ldots, U_n$ are independent and $P[U_i \leq x] = F^{t_i - t_{i-1}}(x)$, $i = 2, \ldots, k$. Use this to prove $Y$ is Markov (Dwass, 1964).

4.3.11. Let $\{X_n, n \geq 1\}$ be iid random variables with common, continuous distribution $F$. As usual, let $\{L(n), n \geq 1)$ be the record times.

(a) Let $\{E_n, n \geq 1\}$ be iid with $P[E_1 > x] = e^{-x^+}$. Define a sequence $\{L^*(n), n \geq 1\}$ by

$$L^*(1) = 1, \qquad L^*(n) = [L^*(n-1)e^{E_n}] + 1, \qquad n \geq 2$$

where square brackets denote greatest integer function. Prove in $R^\infty$

$$\{L(n), n \geq 1\} \stackrel{d}{=} \{L^*(n), n \geq 1\}$$

(Williams, 1973).

(b) Define

$$T(n) = \inf\{j : j \geq L(n)/L(n-1), j \text{ an integer}\}.$$

Using (a) show $\{T(n)\}$ are iid and

$$P[T(n) = j] = (j(j-1))^{-1}, \qquad j \geq 2$$

(Galambos and Seneta, 1975; Westcott, 1977).

(c) Check

$$e^{E_n} - 1 < (L(n) - L(n-1))/L(n-1) \leq e^{E_n} - 1 + (n-1)^{-1}$$

and so rederive Proposition 4.14 (Westcott, 1977).

4.3.12. (a) Let $\{L(n), n \geq 1\}$ be the record times from a continuous distribution. Define $H(x)$ by

$$1 - H(x) = 1/[x], \qquad x \geq 1$$

and let $\{\xi_n, n \geq 0\}$, $\xi_0 = 1$ be the record value sequence from $H$. Prove in $\mathbb{R}^\infty$

$$\{L(n), n \geq 1\} \stackrel{d}{=} \{\xi_n, n \geq 0\}.$$

(Recall that the record value sequence is a Markov chain.)

(b) Prove in $\mathbb{R}^\infty$

$$\{L(k)/L(k-1), k \geq n\} \Rightarrow \{W_k, k \geq 1\} \text{ where } \{W_k\} \text{ are iid and}$$

$$P[W_k > x] = x^{-1}, \qquad x \geq 1.$$

Proceed via Exercise 4.3.11 or *ab* initio (Shorrock, 1972).

(c) If $\{Y(t), t > 0\}$ is extremal with a continuous distribution and jump times $\{\tau_n\}$ show that $\{\tau_n\}$ is equal in distribution to the range of the extremal process governed by $\Phi_1(x) = \exp\{-x^{-1}\}$.

(d) For the set-up as in (a) and (b) show that

$$\lim_{n\to\infty} P[(L(n+1) - L(n))/(L(n)) - L(n-1)) > a]$$
$$= a^{-1}\log(1+a), \qquad a > 0$$

(Shorrock, 1972).

4.3.13. (a) If $E$ is exponentially distributed and $Z \in (0,1)$ is independent of $E$ check

$$P([E/\log Z^{-1}] + 1 \geq r | Z) = Z^{r-1}.$$

(b) If $\{X_{L(n)}, n \geq 1\}$ is the record value sequence from the uniform distribution on (0, 1), check

$$\{X_{L(n)}, n \geq 1\} \stackrel{d}{=} \{1 - e^{-\Gamma_n}, n \geq 1\}$$

where as usual $(E_n, n \geq 1\}$ are iid,

$$P[E_1 > x] = e^{-x^+} \quad \text{and} \quad \Gamma_n = E_1 + \cdots + E_n, \qquad n \geq 1.$$

(c) From (4.8) show when $F$ is $U(0,1)$:

$$\{(L(n+1) - L(n), X_{L(n)}), n \geq 1\}$$
$$\stackrel{d}{=} \{[(E_n/(-\log(1 - e^{-\Gamma_n^*}))] + 1, 1 - e^{-\Gamma_n^*}), n \geq 1\}$$

where $(E_n, n \geq 1\}$, $\{E_n^*, n \geq 1\}$ are each iid unit exponential sequences, independent of each other, and $\Gamma_n^* = E_1^* + \cdots + E_n^*$.

(d) Prove whenever $F$ is continuous

$$\{L(n), n \geq 1\} \stackrel{d}{=} \left\{1 + \sum_{k=1}^{n-1} ([E_k/(-\log(1 - e^{-\Gamma_k^*}))] + 1), n \geq 1\right\}$$

in $\mathbb{R}^\infty$ (Deheuvels, 1981).

4.3.14. Suppose $\{Y(t), t > 0\}$ is extremal-$F$. Assume $F$ is continuous with left endpoint $x_l$ and right endpoint $x_0$.

(a) Check for $t \in (x_l, x_0)$

$$Y^{\leftarrow}(t) = \sum_{k: Y(\tau_k) \leq t} (\tau_{k+1} - \tau_k).$$

(b) Suppose $\xi = \sum_k \varepsilon_{Z_k}$ is homogeneous PRM on $\mathbb{R}$ and $\{E_k, -\infty < k < \infty\}$ is iid, $P[E_1 > x] = e^{-x^+}$. If $S = -\log(-\log F)$ show

$$\sum_{k=-\infty}^{\infty} \varepsilon_{(S^{\leftarrow}(Z_k), E_k \exp\{Z_k\})} \stackrel{d}{=} \sum_{k=-\infty}^{\infty} \varepsilon_{(Y(\tau_k), \tau_{k+1}-\tau_k)}$$

on $M_p((x_l, x_0) \times (0, \infty))$ by appealing to Proposition 4.8(iii), Proposition 3.7, and Proposition 3.8.

(c) Hence show

$$Y^{\leftarrow}(t) \stackrel{d}{=} \sum_{k: Z_k \leq S(t)} E_k e^{Z_k}, \qquad x_l < t < x_0$$

in the sense of equality of finite dimensional distributions (Deheuvels, 1981).

## 4.4. Weak Convergence to Extremal Processes

### 4.4.1. Skorohod Spaces

Extremal processes (as well as Levy processes and other Markov processes) live in the space $D(0, \infty)$, the set of real functions on $(0, \infty)$ which are right continuous with finite left limits existing everywhere. In order to discuss weak convergence of extremal processes intelligently, we need to study properties of $D(0, \infty)$ and, in particular, impose a metric which will make $D(0, \infty)$ a complete, separable metric space.

The concept of weak convergences of probability measures on a space $S$ (cf. Billingsley, 1968 and Section 3.5) is very much dependent on the choice of metric as this governs continuity concepts as well as which subsets of $S$ belong to the Borel $\sigma$-fields. The usefulness of the theory depends on wise choice of a metric. If a metric on $S$ makes too many functions continuous, it will be difficult to prove weak convergence of a sequence of probability measures. If the chosen metric makes too few functions continuous, then applications of the continuous mapping theorem will be scarce and the resulting weak convergence theory will not be particularly useful.

The topology of local uniform convergence on $(0, \infty)$ presents problems (Billingsley, 1968, page 150), and a consensus of opinion is that the Skorohod topology is a good compromise.

We begin by discussing $D[a, b]$, the functions which are right continuous on $[a, b)$ and have finite left limits on $(a, b]$. This treatment follows Billingsley (1968).

**Lemma 4.15.** *A criterion for $x \in D[a, b]$ is that for any $\varepsilon > 0$, there exist $r \geq 1$ and times $t_0, \ldots, t_r$, $a = t_0 < \cdots < t_r = b$, and*

$$\sup\{|x(s) - x(t)|: t_{i-1} \leq s, t < t_i\} < \varepsilon \tag{4.34}$$

*for $i = 1, \ldots, r$.*

*By avoiding bad points we make the variation of $x$ small.*

Proof. Given $x \in D[a, b]$ and $\varepsilon > 0$. Let

$$\begin{aligned} \tau &= \sup\{t \geq a: [a, t) \quad \text{can be decomposed as described in (4.34)}\} \\ &= \sup G, \end{aligned}$$

where $G$ is the set previously described in the braces. Then $\tau > a$ since $x(a) = x(a+)$ makes the variation of $x$ in a right neighborhood of "$a$" controllable. Similarly $\tau \in G$ because $x(\tau-)$ exists and hence variation in a left neighborhood of $\tau$ is controllable. If $\tau < b$ we would get a contradiction since right continuity at $\tau$ means there is a right neighborhood of $\tau$ for which variation is smaller than $\varepsilon$. □

From the lemma we have the following properties of functions in $D[a, b]$:

1. The set $\{t \in [a, b]: |x(t) - x(t-)| > \varepsilon\}$ is finite (in fact has cardinality at most $r$), and therefore the number of discontinuities of $x$ is at most countable.
2. The function $x \in D[a, b]$ is bounded on $[a, b]$ since
$$\sup\{|x(t)|: t \in [a, b]\} \leq \sup_{1 \leq i \leq r-1} (|x(t_{i-1})| + \varepsilon) \vee |x(b)|.$$
Therefore, functions in $D(0, \infty)$ are locally bounded.
3. Let
$$x_\varepsilon(t) = \sum_{i=1}^{r} x(t_{i-1}) 1_{[t_{i-1}, t_i)}(t) + x(b) 1_{\{b\}}(t).$$
The $\sup_{a \leq t \leq b} |x_\varepsilon(t) - x(t)| < \varepsilon$ and so $x \in D[a, b]$ can be uniformly approximated to any desired accuracy by a simple function. Hence $x$ is Borel measurable.

Now we define a metric. The uniform metric says that two functions $x$ and $y$ are close if their graphs are uniformly close. The Skorohod metric is not so strict; it allows uniformly small deformations of time before comparing the graphs. The time deformations are achieved by homeomorphisms $\lambda \in \Lambda_{a,b}$ where

$$\Lambda_{a,b} = \{\lambda: [a, b] \to [a, b]: \lambda(a) = a, \lambda(b) = b, \lambda \text{ is continuous and strictly increasing}\}.$$

Then define for $x, y \in D[a, b]$

$$\begin{aligned} d_{a,b}(x, y) &= \inf\left\{\varepsilon > 0: \text{there exists } \lambda \in \Lambda_{a,b} \text{ such that } \sup_{a \leq t \leq b} |\lambda(t) - t| \leq \varepsilon, \sup_{a \leq t \leq b} |x(t) - y(\lambda(t))| \leq \varepsilon\right\} \\ &= \inf_{\lambda \in \Lambda_{a,b}} \left(\sup_{a \leq t \leq b} |\lambda(t) - t|\right) \vee \left(\sup_{a \leq t \leq b} |x(t) - y(\lambda(t))|\right). \end{aligned}$$

Then $d_{a,b}$ is a metric generating the Skorohod topology on $D[a, b]$.

For notational simplicity let $e(t) = t$,

$$\|\lambda - e\|_{a,b} = \sup_{a \leq t \leq b} |\lambda(t) - t|, \qquad \|x - y \circ \lambda\|_{a,b} = \sup_{a \leq t \leq b} |x(t) - y(\lambda(t))|.$$

What does convergence mean in this metric? Given $x_n \in D[a,b]$, $n \geq 0$ we have $d_{a,b}(x_n, x_0) \to 0$ iff there exist $\lambda_n \in \Lambda_{a,b}$ such that

$$\|\lambda_n - e\|_{a,b} \to 0, \qquad \|x_n \circ \lambda_n - x_0\|_{a,b} \to 0.$$

Note that if we take $\lambda = e$ in the definition of $d_{a,b}$ we get

$$d_{a,b}(x, y) \leq \|x - y\|_{a,b}$$

so that uniform convergence is more stringent than Skorohod convergence since uniform convergence implies $d_{a,b}$-convergence. The converse is false: Take $x_0(t) = 1_{[a,(a+b)/2]}(t)$, $x_n(t) = 1_{[a,(a+b)/2+n^{-1}]}(t)$. Then for $n \geq 1$

$$\|x_n - x_0\|_{a,b} = 1$$

so there is no hope of uniform convergence. However

$$d_{a,b}(x_n, x_0) \leq n^{-1}$$

since a homeomorphism mapping $[a,(a+b)/2 + n^{-1}]$ onto $[a,(a+b)/2]$ would cause the graphs to match exactly.

This example also shows that $d_{a,b}$-convergence does not imply pointwise convergence everywhere; note $x_n(\frac{1}{2}(a+b)) = 1$, $x_0(\frac{1}{2}(a+b)) = 0$ so $x_n(t)$ does not converge to $x_0(t)$ when $t = \frac{1}{2}(a+b)$. However it is true that $d_{a,b}$-convergence implies convergence at points $t$ which are continuity points of $x_0$. For suppose $t$ is such a point. Then if $\lambda_n \in \Lambda_{a,b}$ and $\|x_n \circ \lambda_n - x_0\| \to 0$, $\|\lambda_n - e\|_{a,b} \to 0$ we get

$$\begin{aligned}|x_n(t) - x_0(t)| &\leq |x_n(t) - x_0(\lambda_n^{\leftarrow}(t))| + |x_0(\lambda_n^{\leftarrow}(t)) - x_0(t)| \\ &\leq \|x_n \circ \lambda_n - x_0\|_{a,b} + |x_0(\lambda_n^{\leftarrow}(t)) - x_0(t)| \\ &= o(1) + |x_0(\lambda_n^{\leftarrow}(t)) - x_0(t)|.\end{aligned}$$

Since $\|\lambda_n^{\leftarrow} - e\|_{a,b} = \|\lambda_n - e\|_{a,b} \to 0$ we have $\lambda_n^{\leftarrow}(t) \to t$ and since $x_0$ is continuous at $t$, $|x_0(\lambda_n^{\leftarrow}(t)) - x_0(t)| \to 0$.

A slight variant of this argument shows that if $d_{a,b}(x_n, x_0) \to 0$ and $x_0$ is continuous on $[a,b]$ then $\|x_n - x_0\|_{a,b} \to 0$. This follows since, as in the preceding argument,

$$\|x_n - x_0\|_{a,b} \leq \|x_n \circ \lambda_n - x_0\|_{a,b} + \|x_0 \circ \lambda_n^{\leftarrow} - x_0\|_{a,b}.$$

Since $x_0$ is uniformly continuous on $[a,b]$, the result follows. Hence Skorohod convergence coincides with unform convergence when the limit is continuous on $[a,b]$.

With this metric, the space $D[a,b]$ is separable. A countable dense set of simple functions can be constructed with the help of Lemma 4.15. However, the space is not complete. Let $x_n = 1_{[(a+b)/2,(a+b)/2+1/n]}$ and let $x_0 \equiv 0$ on $[a,b]$. Then for all $\lambda \in \Lambda_{a,b}$ we have

$$\|x_n \circ \lambda - x_0\|_{a,b} = \|x_n \circ \lambda\|_{a,b} = 1$$

so that $d_{a,b}(x_n, x_0)$ does not converge to zero. On the other hand $d(x_n, x_m) =$

$|n^{-1} - m^{-1}| \to 0$ as $n, m \to \infty$ so that $\{x_n\}$ is Cauchy. Since $x_0$ is the only potential limit ($\lim x_n(t)) = 0$ a.e.) we sadly conclude that $d_{a,b}$ is not complete.

This is a minor irritation since there is an equivalent "slope" metric $d_0$ cooked up by Billingsley (cf. Billingsley, 1968) which makes $D[a, b]$ complete. The Cauchy sequence exhibited previously is no longer Cauchy with respect to $d_0$. Thus reassured, we will by and large continue to work with $d_{a,b}$. Before proceeding however we make some comments about the phenomenon that a space may be complete with respect to one of a pair of equivalent metrics but not the other.

Let $S$ be a set with two metrics, $\rho_1$ and $\rho_2$. The metrics $\rho_1$ and $\rho_2$ are equivalent if $(S, \rho_1)$ and $(S, \rho_2)$ are homeomorphic, i.e., if there is a bicontinuous bijection between the two spaces. In this case, the two spaces have the same open sets and a sequence converges with respect to $\rho_1$ if and only if it converges with respect to $\rho_2$. However if the homeomorphism is only bicontinuous but not uniformly bicontinuous, it is possible to have a sequence which is Cauchy in $(S, \rho_1)$ but not in $(S, \rho_2)$.

As an example define on $\mathbb{R}$

$$t(x) = x/(1 + |x|)$$

and define for $x, y \in \mathbb{R}$

$$\rho_1(x, y) = |t(x) - t(y)|$$

$$\rho_2(x, y) = |x - y|.$$

Note that $\rho_1$ measures distance by homeomorphically sending $x$ and $y$ into the interval $(-1, 1)$. As is well known, $(\mathbb{R}, \rho_2)$ is complete. However $(\mathbb{R}, \rho_1)$ is not. The sequence $\{n\}$ is not Cauchy in $(\mathbb{R}, \rho_2)$ but it is in $(\mathbb{R}, \rho_1)$ since

$$\rho_1(n, m) = \left|\frac{n}{n+1} - \frac{m}{1+m}\right| \to 1 - 1 = 0$$

as $n, m \to \infty$.

Following Whitt (1980) (see also Stone, 1983, and Lindvall, 1973) we now construct a metric $d$ on $D(0, \infty)$ such that for $x_n \in D(0, \infty)$, $n \geq 0$ we have

$$d(x_n, x_0) \to 0$$

iff for all $0 < a < b$, $a, b \in \mathscr{C}(x_0) = \{t > 0\colon x_0 \text{ is continuous at } t\}$

$$d_{a,b}(r_{a,b}x_n, r_{a,b}x_0) \to 0.$$

Here $r_{a,b}\colon D(0, \infty) \to D[a, b]$ is defined by $r_{a,b}x(t) = x(t)$, $a \leq t \leq b$ so $r_{a,b}x$ is just the restriction of $x$ to $[a, b]$. Thus convergence in $D(0, \infty)$ will be reduced to convergence in the more familiar space $D[a, b]$.

Define for $x, y \in D(0, \infty)$

$$d(x, y) = \int_0^1 ds \int_{t=1}^{\infty} e^{-t}(d_{s,t}(r_{s,t}x, r_{s,t}y) \wedge 1)dt. \tag{4.35}$$

We will show in the next lemma that this definiton gives the desired notion of convergence.

**Lemma 4.16.** (a) *The integrals in* (4.35) *exist and define a metric.*
(b) *If* $x_n$, $n \geq 0$ *are functions in* $D(0, \infty)$ *and* $0 < a < b < c$ *with* $b \in \mathscr{C}(x_0)$ *then*

$$d_{a,c}(r_{a,c}x_n, r_{a,c}x_0) \to 0$$

iff

$$d_{a,b}(r_{a,b}x_n, r_{a,b}x_0) \to 0 \quad \textit{and} \quad d_{b,c}(r_{b,c}x_n, r_{b,c}x_0) \to 0.$$

(c) *If* $x_n \in D(0, \infty)$, $n \geq 0$ *then*

$$d(x_n, x_0) \to 0$$

iff *for all* $0 < s < t$, $s$ *and* $t \in \mathscr{C}(x_0)$,

$$d_{s,t}(r_{s,t}x_n, r_{s,t}x_0) \to 0.$$

Proof. (a) For $x$, $y \in D(0, \infty)$ we have for fixed $s > 0$ that $d_{s,t}(r_{s,t}x, r_{s,t}y)$ is continuous at $t \in \mathscr{C}(x) \cap \mathscr{C}(y)$, $t > s$, and for fixed $t$ we have $d_{s,t}(r_{s,t}x, r_{s,t}y)$ is continuous at $s \in \mathscr{C}(x) \cap \mathscr{C}(y)$, $s < t$. To get an idea of how this is proved we show right continuity in $t$. Given $t \in \mathscr{C}(x) \cap \mathscr{C}(y)$ there exists for any $\varepsilon$, $\lambda \in \Lambda_{s,t}$ such that

$$\begin{aligned} \|\lambda - e\|_{s,t} \vee \|x - y \circ \lambda\|_{s,t} - \varepsilon &\leq d_{s,t}(r_{s,t}x, r_{s,t}y) \\ &\leq \|\lambda - e\|_{s,t} \vee \|x - y \circ \lambda\|_{s,t}. \end{aligned} \tag{4.36}$$

Furthermore by continuity there exists $0 < h < \varepsilon$ such that for $0 < \eta < h$ we have

$$|x(t) - x(t \pm \eta)| \vee |y(t) - y(t \pm \eta)| < \varepsilon.$$

Define

$$\lambda'(u) = \begin{cases} \lambda(u) & s \leq u \leq t \\ u & t \leq u \leq t + h \end{cases}$$

so that $\lambda' \in \Lambda_{s,t+h}$. Therefore

$$\begin{aligned} d_{s,t+h} := d_{s,t+h}(r_{s,t+h}x, r_{s,t+h}y) &\leq \|\lambda' - e\|_{s,t+h} \vee \|x - y \circ \lambda'\|_{s,t+h} \\ &\leq \{\|\lambda - e\|_{s,t} \vee \|x - y \circ \lambda\|_{s,t}\} \vee \|\lambda - e\|_{t,t+h} \vee \|x - y\|_{t,t+h} \end{aligned}$$

and from (4.36) applied to the expression in the braces we have the preceding bounded by

$$\leq (d_{s,t} + \varepsilon) \vee \left( \sup_{t \leq u \leq t+h} |x(u) - x(t)| + |x(t) - y(t)| \right.$$

$$\left. + \sup_{t \leq u \leq t+h} |y(t) - y(u)| \right) \leq (d_{s,t} + \varepsilon) \vee (\varepsilon + (d_{s,t} + \varepsilon) + \varepsilon);$$

i.e.,

$$d_{s,t+h} \le d_{s,t} + 3\varepsilon. \tag{4.37}$$

A reverse inequality is obtained as follows: Given $\varepsilon$ there is some $\lambda \in \Lambda_{s,t+h}$ (not the same $\lambda$ as previously) such that

$$\begin{aligned} d_{s,t+h} &:= d_{s,t+h}(r_{s,t+h}x, r_{s,t+h}y) \\ &\ge \|\lambda - e\|_{s,t+h} \vee \|x - y \circ \lambda\|_{s,t+h} - \varepsilon. \end{aligned} \tag{4.38}$$

There are two cases to consider: (1) $\lambda(t) \ge t$ and (2) $\lambda(t) \le t$. We only consider case (1) and for this case define for small $\delta \le h$

$$\lambda'(u) = \begin{cases} \lambda(u) & \text{for } s \le u \le \lambda^{\leftarrow}(t-\delta) \\ \delta(t - \lambda^{\leftarrow}(t-\delta))^{-1}(u - \lambda^{\leftarrow}(t-\delta)) + t - \delta & \text{for } \lambda^{\leftarrow}(t-\delta) \le u \le t \end{cases}$$

so that the graph of $\lambda'$ is linear between the points $(\lambda^{\leftarrow}(t-\delta), t-\delta)$ and $(t,t)$. This definition makes $\lambda' \in \Lambda_{s,t}$. Now

$$\begin{aligned} \|\lambda' - e\|_{s,t} &= \|\lambda - e\|_{s,\lambda^{\leftarrow}(t-\delta)} \vee \|\lambda' - e\|_{\lambda^{\leftarrow}(t-\delta),t} \\ &\le \|\lambda - e\|_{s,t+h} \vee |\lambda^{\leftarrow}(t-\delta) - (t-\delta)| \\ &\le \|\lambda - e\|_{s,t+h} \end{aligned}$$

and therefore

$$\|\lambda - e\|_{s,t+h} \ge \|\lambda' - e\|_{s,t} \ge d_{s,t}. \tag{4.39}$$

Likewise

$$\begin{aligned} \|x - y \circ \lambda'\|_{s,t} &= \|x - y \circ \lambda\|_{s,\lambda^{\leftarrow}(t-\delta)} \\ &\quad \vee \sup_{\lambda^{\leftarrow}(t-\delta) \le u \le t} |x(u) - y(\lambda'(u))| \\ &\le \|x - y \circ \lambda\|_{s,t+h} \vee \sup_{\lambda^{\leftarrow}(t-\delta) \le u \le t} \{|x(u) - y(t)| \\ &\qquad + |y(t) - y(\lambda'(u))|\}. \end{aligned}$$

Now

$$\sup_{\lambda^{\leftarrow}(t-\delta) \le u \le t} |y(t) - y(\lambda'(u))| = \sup_{t-\delta \le v \le t} |y(t) - y(v)| \le \varepsilon$$

and since $\lambda \in \Lambda_{s,t+h}$ implies $\lambda(t+h) = t+h$ we have

$$\begin{aligned} \sup_{\lambda^{\leftarrow}(t-\delta) \le u \le t} |x(u) - y(t)| &\le \sup_{\lambda^{\leftarrow}(t-\delta) \le u \le \lambda^{\leftarrow}(t+h)} |x(u) - y(t)| \\ &= \sup_{t-\delta \le v \le t+h} |x(\lambda^{\leftarrow}(v)) - y(t)| \\ &\le \sup_{t-\delta \le v \le t+h} (|x(\lambda^{\leftarrow}(v)) - y(v)| + |y(v) - y(t)|) \\ &\le \sup_{s \le v \le t+h} |x(\lambda^{\leftarrow}(v)) - y(v)| + \varepsilon \\ &= \|x - y \circ \lambda\|_{s,t+h} + \varepsilon; \end{aligned}$$

i.e.,

$$\|x - y \circ \lambda'\|_{s,t} \le \|x - y \circ \lambda\|_{s,t+h} + \varepsilon.$$

Thus

$$\|x - y \circ \lambda\|_{s,t+h} \ge \|x - y \circ \lambda'\|_{s,t} - \varepsilon \ge d_{s,t} - \varepsilon. \tag{4.40}$$

We conclude from (4.38), (4.39), and (4.40) that

$$\begin{aligned} d_{s,t+h} &\ge d_{s,t} \vee (d_{s,t} - \varepsilon) - \varepsilon \\ &= d_{s,t} - \varepsilon, \end{aligned}$$

which coupled with (4.37) yields the sandwich

$$d_{s,t} - \varepsilon \le d_{s,t+h} \le d_{s,t} + 3\varepsilon$$

giving the desired right continuity in $t$.

It is now relatively straightforward to verify the existence of the integrals in (4.35). For fixed $s$, $e^{-t}(d_{s,t}(r_{s,t}x, r_{s,t}y) \wedge 1)$ is a.e. continuous and bounded and hence Riemann integrable. Also by dominated convergence, $\int_{t=1}^{\infty} e^{-t}(d_{s,t}(r_{s,t}x, r_{s,t}y) \wedge 1)dt$ is a.e. continuous in $s$; it is also bounded and hence Riemann integrable.

The verification that (4.35) defines a metric is routine and is left as Exercise 4.4.1.2.

(b) If $r_{a,b}x_n \to r_{a,b}x_0$ and $r_{b,c}x_n \to r_{b,c}x_0$ then there exist $\lambda'_n \in \Lambda_{a,b}$, $\lambda''_n \in \Lambda_{b,c}$, and

$$\|x_n - x_0 \circ \lambda'_n\|_{a,b} \vee \|\lambda'_n - e\|_{a,b} \vee \|x_n - x_0 \circ \lambda''_n\|_{b,c} \vee \|\lambda''_n - e\|_{b,c} \to 0.$$

Define

$$\lambda_n(u) = \begin{cases} \lambda'_n(u), & a \le u \le b \\ \lambda''_n(u), & b \le u \le c \end{cases}$$

so that $\lambda_n \in \Lambda_{a,c}$. Then clearly

$$\|x_n - x_0 \circ \lambda_n\|_{a,c} \vee \|\lambda_n - e\|_{a,c} \to 0$$

giving $r_{a,c}x_n \to r_{a,c}x_0$. Note that this direction did not require continuity of $x_0$ at $b$.

Conversely suppose that $r_{a,c}x_n \to r_{a,c}x_0$ so that there exist $\lambda_n \in \Lambda_{a,c}$ such that

$$\|\lambda_n - e\|_{a,c} \vee \|x_n - x_0 \circ \lambda_n\|_{a,c} \to 0. \tag{4.41}$$

Modify $\lambda_n$ to get $\lambda'_n$ with the following properties: The $\lambda'_n$ must fix $b$ and still satisfy (4.41) and $\lambda'_n \in \Lambda_{a,c}$. How do we construct $\lambda'_n$? Since $\|\lambda_n - e\|_{a,c} \to 0$ we can obtain $\lambda'_n$ from $\lambda_n$ by modifying $\lambda_n$ only on a neighborhood $(b - \varepsilon_n, b + \varepsilon_n)$ of $b$, where $0 < \varepsilon_n \to 0$. On $(b - \varepsilon_n, b + \varepsilon_n)$ push the graph of $\lambda_n$ closer to the graph of $e$ so that $\|\lambda'_n - e\|_{a,c} \le \|\lambda_n - e\|_{a,c} \to 0$. For the second expression in (4.41) we will have

$$\begin{aligned}\|x_n - x_0 \circ \lambda_n'\|_{a,c} &\le \|x_n - x_0 \circ \lambda_n\|_{a,c} + \|x_0 \circ \lambda_n - x_0 \circ \lambda_n'\| \\ &\le o(1) + \sup_{b-\varepsilon_n \le u,v \le b+\varepsilon_n} |x_0(u) - x_0(v)| \to 0\end{aligned}$$

since $b \in \mathscr{C}(x_0)$.

The construction of $\lambda_n'$ gives $r_{a,b}\lambda_n' \in \Lambda_{a,b}$, $r_{b,c}\lambda_n' \in \Lambda_{b,c}$ and

$$\|\lambda_n' - e\|_{a,b} \vee \|x_n - x_0 \circ \lambda_n'\|_{a,b} \to 0$$

$$\|\lambda_n' - e\|_{b,c} \vee \|x_n - x_0 \circ \lambda_n'\|_{b,c} \to 0,$$

proving the result.

(c) If $d_{s,t}(r_{s,t}x_n, r_{s,t}x_0) \to 0$ as $n \to \infty$ for all $s < t$, $s$ and $t \in \mathscr{C}(x_0)$, then $d_{s,t}(r_{s,t}x_n, r_{s,t}x_0) \to 0$ almost everywhere with respect to the measure

$$ds\, 1_{(0,1]}(s) e^{-t}\, dt\, 1_{(1,\infty)}(t)$$

and by dominated convergence

$$d(x_n, x_0) = \iint_{(s,t) \in (0,1] \times (1,\infty)} (1 \wedge d_{s,t}(r_{s,t}x_n, r_{s,t}x_0)) e^{-t}\, dt\, ds \to 0$$

as required. Conversely suppose that $d(x_n, x_0) \to 0$. For the purpose of getting a contradiction, suppose there exist $0 < s < t$, $s$, $t \in \mathscr{C}(x_0)$ and

$$\liminf_{n\to\infty} d_{s,t}^{(n)} := \liminf_{n\to\infty} d_{s,t}(r_{s,t}x_n, r_{s,t}x_0) > 0.$$

Then for $0 < u < s$

$$\liminf_{n\to\infty} d_{u,t}^{(n)} > 0$$

since otherwise along some subsequence $\{n'\}$ say

$$d_{u,t}^{(n')} \to 0$$

implying by (b) that $d_{u,s}^{(n')} \to 0$, $d_{s,t}^{(n')} \to 0$, a contradiction. Similarly for $v > t > s > u$

$$\liminf_{n\to\infty} d_{u,v}^{(n)} > 0.$$

Thus by Fatou

$$\begin{aligned}&\liminf_{n\to\infty} d(x_n, x_0) \\ &\quad \ge \liminf_{n\to\infty} \int_{u=0}^{s\wedge 1} \int_{v=t\vee 1}^{\infty} (d_{u,v}^{(n)} \wedge 1)\, du\, e^{-v}\, dv \\ &\quad \ge \int_{u=0}^{s\wedge 1} \int_{v=t\vee 1}^{\infty} \liminf_{n\to\infty} (d_{u,v}^{(n)} \wedge 1)\, du\, e^{-v}\, dv \\ &\quad > 0\end{aligned}$$

which is a contradiction. □

If each $d_{s,t}$ used in (4.35) is complete in $D[s,t]$ (for example, if $d_{s,t}$ is the slope metric) then $(D(0,\infty), d)$ is a complete separable metric space.

In practice, to prove Skorohod convergence in $D(0,\infty)$, we pick a typical interval $[a,b]$ and prove convergence in $D[a,b]$. The same is true for weak convergence as Proposition 4.17 later shows.

We first consider projection maps. For $t_1, \ldots, t_k \in (0,\infty)$ define $\pi_{t_1,\ldots,t_k}$: $D(0,\infty) \to \mathbb{R}^k$ by

$$\pi_{t_1,\ldots,t_k}(x) = (x(t_1), \ldots, x(t_k)).$$

Observe that for $t > 0$, $\pi_t$ is continuous at $x$ if $x$ is continuous at $t$. This follows since if $x$ is continuous at $t$ and $x_n \in D(0,\infty)$ with $d(x_n, x) \to 0$ then $x_n(t) = \pi_t x_n \to x(t) = \pi_t x$. Likewise if $t_1, \ldots, t_k$ are continuity points of $x \in D(0,\infty)$ then $\pi_{t_1,\ldots,t_k}$ is continuous at $x$. This has a nice interpretation for processes. As in Billingsley, 1968, page 124, for a probability measure $P$ on $D(0,\infty)$ let

$$T_P = \{t \in (0,\infty): P\{x \in D(0,\infty): x(t) = x(t-)\} = 1\}.$$

The complement of $T_P$ is at most countable (Billingsley, 1968, page 124). If $\{X(t), t > 0\}$ is a stochastic process on $(\Omega, \mathscr{A}, \mathbf{P})$ with all paths in $D(0,\infty)$ then $P = \mathbf{P} \circ X^{-1}$ and

$$\begin{aligned} T_X := T_P &= \{t > 0: \mathbf{P}[X(t) = X(t-)] = 1\} \\ &= \{t > 0: \mathbf{P}[X \text{ is continuous at } t] = 1\} \end{aligned}$$

so that $T_P^c$ is the set of fixed discontinuities of $X$. If in addition $X$ is stochastically continuous (as is always the case for extremal processes, but not always for inverses of extremal processes) then $T_P = (0,\infty)$. (Recall that for a process $X$ in $D(0,\infty)$ stochastic continuity at $t$ implies $X(t-) = X(t)$ a.s.)

**Lemma 4.17.** *Suppose $X_n$, $n \geq 0$ are random elements of $D(0,\infty)$.*
(a) *If $0 < a < b$, $a$ and $b \in T_{X_0}$, then $r_{a,b}$ is* a.s. *continuous with respect to $X_0$. Thus if*

$$X_n \Rightarrow X_0$$

*in $D(0,\infty)$, then*

$$r_{a,b} X_n \Rightarrow r_{a,b} X_0$$

*in $D[a,b]$.*
(b) *If $t_1, \ldots, t_k \in T_{X_0}$, then $\pi_{t_1,\ldots,t_k}$ is* a.s. *continuous with respect to $X_0$. Thus if*

$$X_n \Rightarrow X_0$$

*in $D(0,\infty)$, then*

$$(X_n(t_i), i \leq k) \Rightarrow (X_0(t_i), i \leq k)$$

*in $\mathbb{R}^k$.*

Proof. (a) The a.s. continuity of $r_{a,b}$ is just a rephrasing of Lemma 1.16(c). Finish with the continuous mapping theorem. The proof of (b) is similar upon recalling that $\pi_{t_1,\dots,t_k}$ is continuous at $x \in D(0,\infty)$ if $x$ is continuous at $t_1, \dots, t_k$. □

The open sets of $D(0,\infty)$ generate the Borel $\sigma$-algebra denoted $\mathscr{D}$. Another natural $\sigma$-algebra to consider is the one generated by finite dimensional sets. Suppose $T \subset (0,\infty)$ is dense. A finite dimensional set is a set of the form

$$\{x \in D(0,\infty): (x(t_1),\dots,x(t_k)) \in H\} = \pi^{-1}_{t_1,\dots,t_k}(H)$$

where $H \in \mathscr{B}(\mathbb{R}^k)$, $0 < t_1 < \cdots < t_k$. It is a fundamental fact that

$$\mathscr{D} = \sigma\{\pi^{-1}_{t_1,\dots,t_k}(H), H \in \mathscr{B}(\mathbb{R}^k), t_i \in T, i = 1,\dots,k; k \geq 1\}$$

(Billingsley, 1968; Whitt, 1980, page 73; Lindvall, 1973, page 117). The important consequence of this fact is that if two random elements $X$, $Y$, of $D(0,\infty)$ have the property that for any $k \geq 1$, $t_1, \dots, t_k \in T$

$$(X(t_1),\dots,X(t_k)) \overset{d}{=} (Y(t_1),\dots,Y(t_k)) \qquad \text{in } \mathbb{R}^k$$

then

$$X \overset{d}{=} Y$$

in $D(0,\infty)$. (The distributions of $X$ and $Y$ agree on the $\Pi$-system of finite dimensional sets generating $\mathscr{D}$ and hence agree everywhere.) Thus equality of finite dimensional distributions implies that $X$ and $Y$ are distributionally indistinguishable.

We now state the natural criterion for weak convergence in $D(0,\infty)$, which reduces the problem to weak convergence in $D[a,b]$.

**Proposition 4.18.** *If $\{X_n, n \geq 0\}$ are random elements of $D(0,\infty)$ then*

$$X_n \Rightarrow X_0$$

*in $D(0,\infty)$ iff for each $a$ and $b \in T_{X_0}$ with $0 < a < b$ we have*

$$r_{a,b}X_n \Rightarrow r_{a,b}X_0$$

*in $D[a,b]$.*

*Remark.* For notational simplicity, we will drop $r_{a,b}$ when indicating weak convergence in $D[a,b]$.

Proof (Whitt, 1980). Because of the previous lemma we need only show that weak convergence in $D[a,b]$ for $0 < a < b$, $a$, $b \in T_{X_0}$ implies weak convergence in $D(0,\infty)$. If $F \subset D(0,\infty)$ is closed, weak convergence in $D(0,\infty)$ is equivalent to (cf. (3.16), Billingsley, 1968, Theorem 2.1)

$$\limsup_{n\to\infty} \mathbf{P}[X_n \in F] \leq \mathbf{P}[X_0 \in F]. \tag{4.42}$$

We assume without loss of generality, that for any $x \in F$, $x$ is continuous at all $t \in T_{X_0}$.

The way to relate weak convergence in $D[a, b]$ to convergence in $D(0, \infty)$ is through the following mechanism: For $0 < a < b$, $a, b \in T_{X_0}$, define

$$H_{a,b} = r_{a,b}^{-1}((r_{a,b}F)^{-})$$

where the bar indicates closure in $D[a, b]$. We prove a succession of facts about the relation of $\{H_{a,b}; 0 < a < b; a \text{ and } b \in T_{X_0}\}$ to $F$. Once these are stated and checked, it will be easy to verify (4.42). We have first

$$F \subset H_{a,b} \tag{4.43}$$

for each $a, b \in T_{X_0}$. For if $x \in F$ then obviously $r_{a,b}x \in (r_{a,b}F)^{-}$, which restated gives $x \in r_{a,b}^{-1}((r_{a,b}F)^{-}) = H_{a,b}$.

Next if $a, b, c, d \in T_{X_0}$, $a < b$, $c < d$, $[a, b] \subset [c, d]$ then

$$H_{c,d} \subset H_{a,b}. \tag{4.44}$$

If $x \in H_{c,d}$ then $r_{c,d}x \in (r_{c,d}F)^{-}$ so there exist $y_n \in F$ such that $d_{c,d}(r_{c,d}x, r_{c,d}y_n) \to 0$. Two applications of Lemma 4.16(b) ($d_{c,d} \to 0$ implies $d_{c,b} \to 0$ implies $d_{a,b} \to 0$) results in $d_{a,b}(r_{a,b}x, r_{a,b}y_n) \to 0$. Hence $r_{a,b}x \in (r_{a,b}F)^{-}$ whence $x \in r_{a,b}^{-1}((r_{a,b}F)^{-}) = H_{a,b}$.

The last fact needed is that if $s_k$ and $t_k \in T_{X_0}$, $1 \leq t_k \uparrow \infty$, $s_k \downarrow 0$ then

$$F = \lim_{k \to \infty} \downarrow H_{s_k, t_k} = \bigcap_k H_{s_k, t_k}. \tag{4.45}$$

Since we know from (4.43) that $F \subset H_{a,b}$ we must prove $\bigcap_k H_{s_k, t_k} \subset F$. If $x \in \bigcap_k H_{s_k, t_k}$ we will show $x \in F$ by showing $d(x, F) = 0$. Given any $\varepsilon$, there is an integer $p$ such that if $k \geq p$ we have $s_k \vee e^{-t_k} < \varepsilon$. Since $x \in H_{s_p, t_p}$ we have $r_{s_p, t_p}x \in (r_{s_p, t_p}F)^{-}$ and so there exist $y_n \in F$ such that $d_{s_p, t_p}(r_{s_p, t_p}x, r_{s_p, t_p}y_n) \to 0$ as $n \to \infty$. Thus

$$\begin{aligned}
d(x, F) &\leq \limsup_{n \to \infty} d(x, y_n) \\
&= \limsup_{n \to \infty} \int_0^1 ds \int_1^\infty dt\, e^{-t} d_{s,t}(r_{s,t}x, r_{s,t}y_n) \wedge 1) \\
&= \limsup_{n \to \infty} \int_0^{s_p} ds \int_1^\infty dt\, e^{-t}(d_{s,t} \wedge 1) \\
&\quad + \limsup_{n \to \infty} \int_{s_p}^1 ds \int_1^{t_p} dt\, e^{-t}(d_{s,t} \wedge 1) \\
&\quad + \limsup_{n \to \infty} \int_{s_p}^1 ds \int_{t_p}^\infty dt\, e^{-t}(d_{s,t} \wedge 1)
\end{aligned}$$

and because of the choice of $p$ this is bounded by

$$\leq \varepsilon + \limsup_{n \to \infty} \int_{s_p}^1 ds \int_1^{t_p} dt\, e^{-t}(d_{s,t} \wedge 1) + \varepsilon. \tag{4.46}$$

Since $d_{s_p,t_p}(r_{s_p,t_p}x, r_{s_p,t_p}y_n) \to 0$ we get by Lemma 4.16(b) that $d_{s,t}$ in (3.46) converges to zero almost everywhere in $(s,t) \in [s_p, 1] \times [1, t_p]$ and hence by dominated convergence

$$d(x, F) \le 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary we get $d(x, F) = 0$ as desired.

With 4.43, (4.44), and (4.45) checked it is now easy to get to the desired conclusion (4.42). From (4.44) and (4.45) it is evident that there exist $a, b \in T_{X_0}$ such that

$$\mathbf{P}[X_0 \in H_{a,b}] \le \mathbf{P}[X_0 \in F] + \varepsilon.$$

Now

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{P}[X_n \in F] \\
&\quad \le \limsup_{n\to\infty} \mathbf{P}[X_n \in H_{a,b}] \qquad \text{(by 4.43)} \\
&\quad = \limsup_{n\to\infty} \mathbf{P}[r_{a,b}X_n \in (r_{a,b}F)^-].
\end{aligned}$$

Since we assume $r_{a,b}X_n \Rightarrow r_{a,b}X_0$ in $D[a, b]$, if we apply the criterion for weak convergence in terms of closed sets (cf. (3.16) or Billingsley, 1968, Theorem 2.1) we get the foregoing probability bounded above by

$$\begin{aligned}
\mathbf{P}[r_{a,b}X_0 \in (r_{a,b}F)^-] &= \mathbf{P}[X_0 \in r_{a,b}^{-1}((r_{a,b}F)^-)] \\
&= \mathbf{P}[X_0 \in H_{a,b}] \le \mathbf{P}[X_0 \in F] + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that (4.42) is correct. □

## Exercises

4.4.1.1. Analyze the omitted case (2) after (4.38) in the first part of Lemma 4.16. Prove $d_{s,t}$ is left continuous at $t \in \mathscr{C}(x) \cap \mathscr{C}(y)$.

4.4.1.2. Check $d_{a,b}$ is a metric for $0 < a < b$ and then check (4.35) defines a metric on $D(0, \infty)$.

4.4.1.3. Consider the following alternative topology for $D(0, \infty)$, which is constructed by imperfect analogy with the method of Lindvall (1973). For $k \ge 2$ define

$$(D_k, d_k) = (D[k^{-1}, k], d_{k^{-1},k})$$

and set

$$D_\infty = \prod_{k=2}^{\infty} D_k$$

where $D_\infty$ is the combinatorial product. A typical element of $D_\infty$ is $\mathbf{x} = (x_2, x_3, \ldots)$, $x_k \in D_k$, $k \ge 2$ and on $D_\infty \times D_\infty$ define a metric $d_\infty$ by

$$d_\infty(\mathbf{x}, \mathbf{y}) = \sum_{k=2}^{\infty} 2^{-k}(1 \wedge d_k(x_k, y_k)).$$

(a) Show $(D_\infty, d_\infty)$ is a separable metric space. If each $d_k$ is complete, so is $d_\infty$.
(b) Define $r_k = r_{k-1,k}$ and $\varphi$: $D(0, \infty) \to D_\infty$ by

$$\phi x = (r_k x, k \geq 2).$$

Check $\varphi$ is 1–1 and $\varphi(D(0, \infty))$ is closed in $D_\infty$.
(c) Define a metric $d^*$ on $D(0, \infty) \times D(0, \infty)$ by

$$d^*(x, y) = d_\infty(\phi x, \phi y).$$

Compare $(D(0, \infty), d^*)$ with $(D(0, \infty), d)$.
(d) Find a criterion for weak convergence in $(D(0, \infty), d^*)$.

4.4.1.4. Prove $X \stackrel{\text{d}}{=} Y$ in $D(0, \infty)$ iff for all $0 < a < b$, $a$ and $b \in T_X \cap T_Y$,

$$r_{a,b} X \stackrel{\text{d}}{=} r_{a,b} Y$$

in $D[a, b]$.

4.4.1.5. Let $\{X_n, n \geq 0\}$ be random elements of $D(0, \infty)$ and suppose $1 \geq s_k \downarrow 0$, $1 \leq t_k \uparrow \infty$, $s_k$ and $t_k \in T_{X_0}$. Show $X_n \Rightarrow X_0$ in $D(0, \infty)$ iff

$$(r_{s_k, t_k} X_n, k \geq 1) \Rightarrow (r_{s_k, t_k} X_0, k \geq 1)$$

in $\prod_{k=1}^{\infty} D[s_k, t_k]$.

4.4.1.6. Prove addition and multiplication is continuous $D(0, \infty) \times D(0, \infty) \to D(0, \infty)$ at those $(x, y)$ for which $\mathscr{C}(x))^c \cap (\mathscr{C}(y))^c = \varnothing$. In particular, if $X$ and $Y$ are stochastically continuous random elements of $D(0, \infty)$, then these operations are a.s. continuous (Whitt, 1980).

4.4.1.7. The map $T$: $D(0, \infty) \to D(0, \infty)$ defined by

$$Tx(t) = \sup\{(x(s)): 0 < s \leq t\}$$

is continuous and in fact

$$d(Tx, Ty) \leq d(x, y)$$

(Whitt, 1980).

4.4.1.8. (a) What is the continuity set of the map from $D(0, \infty) \to D(0, \infty)$ defined by

$$x \to \left\{ \sup_{0 < s \leq t} ((x(x) - x(s-)) \vee 0), t > 0 \right\}?$$

(b) Check continuity for the map from $D[0, 1] \to \mathbb{R}^\infty$ defined by

$$x \to \{\varphi_i x, i \geq 1\}$$

where $\varphi_i x$ is the $i$th largest positive jump of $x$ in $[0, 1]$.

4.4.1.9. Let $\{\xi_n, n \geq 0\}$ be random elements of $M_p([0, 1])$ such that $\xi_0$ has no multiple points and no atoms at 0 or 1. If

$$\xi_n \Rightarrow \xi_0$$

in $M_p([0, 1])$ then setting $T\xi_n(t) = \xi_n[0, t]$, $n \geq 0$, $0 \leq t \leq 1$ gives

$$T\xi_n \Rightarrow T\xi_0$$

in $D[0, 1]$ (Jagers, 1974).

4.4.1.10. Let

$$\Lambda = \{\lambda: \lambda \text{ is a strictly increasing homeomorphism of } (0,\infty) \text{ onto } (0,\infty), \lambda(0+) = 0, \lambda(\infty) = \infty\}.$$

Show for $x_n$, $n \geq 0$ in $D(0, \infty)$ that

$$d(x_n, x_0) \to 0$$

iff there exist $\lambda_n$, $n \geq 1$, $\lambda_n \in \Lambda$, and

$$x_n \circ \lambda_n \to x_0$$

locally uniformly and

$$\|\lambda_n - e\|_{(0,\infty)} \to 0$$

(Lindvall, 1973).

4.4.1.11. The following maps are not continuous $\mathbb{R}^\infty \to \mathbb{R}^\infty$

$$(x_1, x_2, \ldots) \to (x_1, x_1 + x_2, x_1 + x_2 + x_3, \ldots)$$

$$(x_1, x_2, \ldots) \to (x_1, x_1 \vee x_2, x_1 \vee x_2 \vee x_3, \ldots).$$

## 4.4.2. Weak Convergence of Maximal Processes to Extremal Processes via Weak Convergence of Induced Point Processes

Rather than prove weak convergence of maxima to limiting extremal processes directly we first prove weak convergence of induced point processes to limiting Poisson processes and then apply an a.s. continuous functional. For a rough sketch of the following results, let $\{X_n, n \geq 1\}$ be iid with common distribution $F$ and suppose there exist $a_n > 0$, $b_n$, $n \geq 1$ such that

$$P[M_n \leq a_n x + b_n] = F^n(a_n x + b_n) \to G(x). \tag{4.47}$$

From the following, we will find (almost) that

$$\sum_k \varepsilon_{(k/n,\, a_n^{-1}(X_k - b_n))} \Rightarrow \sum_k \varepsilon_{(t_k, j_k)} \tag{4.48}$$

in $M_p([0,\infty) \times \mathbb{R})$ where the limit point process is PRM. We then apply the map which takes $M_p([0,\infty) \times \mathbb{R}) \to D(0,\infty)$ via $\sum_k \varepsilon_{t_k, j_k} \to \{\bigvee_{0<t_k\leq t} j_k, t > 0\}$.

The point process convergence is a direct application of Proposition 3.21, which we now repeat for convenience.

**Proposition 3.21.** *Suppose $E$ is locally compact with countable base and $\mathscr{E}$ is the Borel $\sigma$-algebra. For each $n$ suppose $\{X_{n,j}, j \geq 1\}$ are* iid *random elements of $(E, \mathscr{E})$ and $\mu$ is a Radon measure on $(E, \mathscr{E})$. Define $\xi_n := \sum_{j=1}^\infty \varepsilon_{(jn^{-1}, X_{n,j})}$ and suppose $\xi$ is* PRM *on$[0,\infty] \times E)$ with mean measure $dt \times d\mu$. Then $\xi_n \Rightarrow \xi$ in $M_p([0,\infty) \times E)$* iff

$$nP[X_{n,1} \in \cdot\,] \xrightarrow{v} \mu \qquad \text{on } E. \tag{3.19}$$

In using Proposition 3.21 to discuss the relation between (4.47) and (4.48) there are two small but potentially annoying problems that must be overcome. The first is to translate the condition $F \in D(G)$, or what is the same (4.47), into an equivalent statement (3.19) about vague convergence of measures. For example, if $F \in D(\Lambda)$ then (4.47) is equivalent to

$$n(1 - F(a_n x + b_n)) = nP[a_n^{-1}(X_1 - b_n) > x] \to e^{-x}, \qquad x \in \mathbb{R}.$$

This will not be equivalent to (3.19), viz.

$$nP[a_n^{-1}(X_1 - b_n) \in \cdot\,] \to \mu$$

where $\mu(x, \infty) = e^{-x}$, $x \in \mathbb{R}$, unless neighborhoods of $+\infty$ are compact. Thus the first difficulty can be overcome by correct choice of topology: $E = [-\infty, \infty] \backslash \{-\infty\} = (-\infty, \infty]$, the homeomorphic image of $(0, 1]$. Closed neighborhoods of $+\infty$ are compact.

The second difficulty occurs when $F \in D(\Phi_\alpha)$. The regular variation in the right tail does not offer us any control over points $(jn^{-1}, X_j/a_n) \in [0, \infty] \times (-\infty, 0)$, and such points will be simply neglected by our point processes.

**Corollary 4.19.** *Let $\{X_n, n \geq 1\}$ be* iid *with distribution $F \in D(G)$ where $G$ is an extreme value distribution. Set $M_n = \bigvee_{i=1}^n X_i$ so there exist $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$ such that*

$$P[a_n^{-1}(M_n - b_n) \leq x] = F^n(a_n x + b_n) \to G(x). \tag{4.47}$$

*Suppose, for convenience, that the norming constants are chosen in the canonical way as described in Propositions* 1.9, 1.11, *and* 1.13.

(i) *If $G = \Lambda$, set $E = (-\infty, \infty]$, $\nu(x, \infty] = e^{-x}$, $x \in \mathbb{R}$, and then* (4.47) *is equivalent to*

$$\xi_n := \sum_{k=1}^{\infty} \varepsilon_{(k/n,\, a_n^{-1}(X_k - b_n))} \Rightarrow \xi = \mathrm{PRM}(dt \times d\nu)$$

in $M_p([0, \infty) \times (-\infty, \infty])$.

(ii) *If $G = \Phi_\alpha$, suppose $F(0) = 0$ (so that $X_i > 0$* a.s.*) and set $E = (0, \infty]$, $\nu(x, \infty] = x^{-\alpha}$, $x > 0$. Then* (4.47) *is equivalent to*

$$\xi_n := \sum_{k=1}^{\infty} \varepsilon_{(k/n,\, X_k/a_n)} \Rightarrow \xi = \mathrm{PRM}(dt \times d\nu)$$

*in $M_p([0, \infty) \times (0, \infty])$.*

(iii) *If $G = \Psi_\alpha$ so that $x_0 = \sup\{x\colon F(x) < 1\} < \infty$, then set*

$$E = (-\infty, 0], \nu(x, 0] = (-x)^\alpha, x < 0,$$

*and* (4.47) *is equivalent to*

$$\xi_n := \sum_{k=1}^{\infty} \varepsilon_{(k/n,\, (X_k - x_0)/(x_0 - \gamma_n))} \Rightarrow \xi = \mathrm{PRM}(dt \times d\nu)$$

*in $M_p([0, \infty) \times (-\infty, 0])$.*

Proof. This is a direct application of Proposition 3.21. Note that (4.47) is equivalent to

$$nP[(X_1 - b_n)/a_n > x] \to -\log G(x)$$

for $x$ such that $G(x) > 0$ and this last convergence statement is equivalent to (3.19) because of the way $E$ is topologized. □

We may now prove the invariance principle first given by Lamperti (1964) with a traditional proof (finite dimensional distributions converge plus tightness).

**Proposition 4.20.** *Let $\{X_n, n \geq 1\}$ be iid rv's with common* df $F(x)$. *Set $M_n = \bigvee_{i=1}^n X_i$ and suppose there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that for a nondegenerate limit* df $G(x)$

$$P[M_n \leq a_n x + b_n] = F^n(a_n x + b_n) \to G(x) \tag{4.47}$$

*weakly. Set*

$$Y_n(t) = \begin{cases} (M_{[nt]} - b_n)/a_n & t \geq n^{-1} \\ (X_1 - b_n)/a_n & 0 < t < n^{-1} \end{cases}$$

*and suppose $(Y(t), t > 0)$ is an extremal process generated by $G$. Then*

$$Y_n \Rightarrow Y$$

*in $D(0, \infty)$ is equivalent to* (4.47).

Proof. Consider first the case $G = \Lambda$. From Corollary 4.19 we have

$$\xi_n := \sum_k \varepsilon_{(k/n, a_n^{-1}(X_k - b_n))} \Rightarrow \xi := \sum_k \varepsilon_{(t_k, j_k)}$$

in $M_p([0, \infty) \times (-\infty, \infty])$ where the limit is PRM$(dt \times dv)$, $v(x, \infty) = e^{-x}$, $x \in \mathbb{R}$. Let $T_1$ be the functional from $M_p([0, \infty) \times (-\infty, \infty]) \to D(0, \infty)$ defined by

$$(T_1 m)(t) = \left(T_1 \sum_k \varepsilon_{(\tau_k, y_k)}\right)(t) = \bigvee_{\tau_k \leq t} y_k$$

provided $m([0, t] \times (-\infty, \infty]) > 0$ for all $t$. Otherwise set $t^* = \sup\{s > 0: m((0, s] \times (-\infty, \infty]) = 0\}$ and

$$(T_1 m)(t) = \bigvee_{\tau_k = t^*} y_k.$$

The functional $T_1$ is defined (except at $m \equiv 0$) and a.s. continuous with respect to $\xi$ (see later discussion), and so by the continuous mapping theorem

$$T_1 \xi_n \Rightarrow T_1 \xi$$

in $D(0, \infty)$ where

$$(T_1\xi)(t) = \bigvee_{t_k \le t} j_k = Y(t)$$

(cf. Section 4.3) and

$$(T_1\xi_n)(t) = Y_n(t).$$

The treatment for $G = \Psi_\alpha$ is similar so consider now the case $G = \Phi_\alpha$. Again we recall the problem of how to handle points $(k/n, X_k/a_n)$ such that $X_k \le 0$, since the regular variation in the right tail offers no control over such points. One method is to neglect these points by using the following device dating back to P. Lévy.

### *The "Découpage de Lévy"*

Suppose $\{X_n, n \ge 1\}$ are iid random elements of a metric space $S$ with Borel sets $\mathscr{S}$. Fix a set $B \in \mathscr{S}$ such that $\mathbf{P}[X_1 \in B] > 0$. Let $K^+(i)$ be those indices $j$ for which $X_j \in B$; i.e., let $K^+(0) = 0$ and $K^+(i) = \inf\{j > K^+(i-1): X_j \in B\}$, $i \ge 1$. Similarly define $\{K^-(i)\}$ by $K^-(0) = 0$ and $K^-(i) = \inf\{j > K^-(i-1): X_j \in B^c\}$. Also define $N(n) = \sup\{i: K^+(i) \le n\}$. Then it follows that $\{X_{K^+(i)}\}$, $\{X_{K^-(i)}\}$, $\{N(i), i \ge 1\}$ are independent and $\{X_{K^\pm(i)}\}$ is iid with

$$P[X_{K^+(i)} \in A] = P[X_1 \in A \mid X_1 \in B], \qquad A \subset B, \qquad A \in \mathscr{S},$$

$$P[X_{K^-(i)} \in A] = P[X_1 \in A \mid X_1 \notin B], \qquad A \subset B^c, \qquad A \in \mathscr{S}.$$

Furthermore $N(n)$, $n \ge 1$ is a renewal counting function and $EN(n) = nP[X_1 \in B]$.

To use this in our problem let $S = (-\infty, \infty]$, $B = (0, \infty]$ so that for $x > 0$

$$\begin{aligned} P[X_{K^+(1)} > x] &= P[X_1 > x]/P[X_1 > 0] \\ &\sim x^{-\alpha}L(x)/(1 - F(0)), \qquad x \to \infty. \end{aligned}$$

From Corollary 4.19(ii)

$$\sum_{i=1}^{\infty} \varepsilon_{(i/n,\, X_{K^+(i)}/a_n)} \Rightarrow \sum_i \varepsilon_{(t_i, j_i)} \tag{4.49}$$

where the limit in (4.49) is PRM. To compute the mean measure we assume that $a_n$ is canonically chosen so that

$$nP[X_1 > a_n x] \to x^{-\alpha}, \qquad x > 0$$

and therefore

$$\begin{aligned} nP[X_{K^+(1)}/a_n > x] &= nP[X_1/a_n > x]/(1 - F(0)) \\ &\to x^{-\alpha}/(1 - F(0)) \end{aligned}$$

so the mean measure, by Proposition 3.21, is $dt \times \alpha x^{-\alpha-1}\, dx/(1 - F(0))$. Applying the analogue of $T_1$ used earlier for the case $G = \Lambda$ we get in $D(0, \infty)$

$$\bigvee_{i=1}^{[n\cdot]} X_{K^+(i)}/a_n \Rightarrow \bigvee_{t_k \le \cdot} j_k =: Y(\cdot) \tag{4.50}$$

where $Y(\cdot)$ has marginals determined by

$$\begin{aligned} P[Y(t) \le x] &= \exp\{-tx^{-\alpha}/(1 - F(0))\} \\ &= (\Phi_\alpha(x))^{t/(1-F(0))}. \end{aligned}$$

If $Y_\alpha$ is the extremal process generated by $\Phi_\alpha$ then we have

$$Y(\cdot) \stackrel{d}{=} Y_\alpha((\cdot)/(1 - F(0))) \tag{4.51}$$

in $D(0, \infty)$.

Because $\{N(n), n \ge 1\}$ is a renewal function, $N(n)/n \to P[X_1 > 0] = (1 - F(0))$ a.s. in $\mathbb{R}$, and it is easy to extend this to (recall $e(t) = t$)

$$N([n\cdot])/n \Rightarrow (1 - F(0))e \tag{4.52}$$

in $D(0, \infty)$. (See the following if you are skeptical.) Since the découpage gives $\{N(n), n \ge 1\}$ independent of $\{X_{K^+(i)}, i \ge 1\}$ one readily combines (4.50) and (4.52) into a joint statement:

$$\left(\bigvee_{i=1}^{[n\cdot]} X_{K^+(i)}/a_n, N([n\cdot])/n\right) \Rightarrow (Y_\alpha((\cdot)/(1 - F(0))), (1 - F(0))e) \tag{4.53}$$

in $D(0, \infty) \times D(0, \infty)$. Composing the two components (an a.s. continuous operation; cf. Whitt, 1980, and Exercise 4.4.2.2) gives

$$\bigvee_{i=1}^{N([n\cdot])} X_{K^+(i)}/a_n \Rightarrow Y_\alpha(\cdot). \tag{4.54}$$

Note

$$\bigvee_{i=1}^{N([n\cdot])} X_{K^+(i)}/a_n = T_1\left(\sum_{k=1}^{\infty} 1_{[X_k > 0]}\varepsilon_{(k/n, X_k/a_n)}\right)$$

so we have managed to neglect the points in $[0, \infty] \times (-\infty, 0]$. However, (4.54) is not quite the desired $Y_n \Rightarrow Y$, but this will be achieved if

$$d\left(\bigvee_{i=1}^{N([n\cdot])} X_{K^+(i)}/a_n, Y_n\right) \xrightarrow{P} 0$$

and by Proposition 4.8 it suffices to show the foregoing with $d_{a,b}$ replacing $d$ for $0 < a < b$. Since Skorohod distance is bounded by uniform distance we show

$$\lim_{n\to\infty} P\left[\sup_{a \le t \le b}\left|\bigvee_{i=1}^{N([nt])} X_{K^+(i)}/a_n - \bigvee_{i=1}^{[nt]} X_i/a_n\right| > \varepsilon\right] = 0 \tag{4.55}$$

for any given $\varepsilon > 0$. However, observe that

$$\begin{aligned} \left[\bigvee_{i=1}^{[na]} X_i/a_n > 0\right] &\subset [N([na]) \ge 1] \\ &\subset \left[\sup_{a \le t \le b}\left|\bigvee_{i=1}^{N([nt])} X_{K^+(i)}/a_n - \bigvee_{i=1}^{[nt]} X_i/a_n\right| = 0\right], \end{aligned}$$

and therefore the probability in (4.55) is bounded by

$$P\left[\bigvee_{i=1}^{[na]} X_i/a_n \le 0\right] = F^{[na]}(0) \to 0$$

as required.

Now let us see why $T_1$ is a.s. continuous. Consider as an illustration the case $G = \Lambda$. It suffices to show that $T_1$ is continuous in $D[a, b]$ (cf. Lemma 4.16(iii)) at $m \in M_p([0, \infty] \times (-\infty, \infty])$, where $m$ satisfies the following:

$$m(\{a\} \times (-\infty, \infty]) = m(\{b\} \times (-\infty, \infty]) = m([0, \infty) \times \{\infty\}) = 0,$$

$$m([0, t] \times (x, \infty)) < \infty, \qquad m([s, t] \times (-\infty, x]) = \infty$$

for any $a < s < t < b$, $x \in \mathbb{R}$. Note that PRM $\xi$ lives in the set of $m$ with these properties.

Let $m_n \in M_p([0, \infty) \times (-\infty, \infty])$ and suppose $m_n \xrightarrow{v} m$. Suppose for concreteness that $T_1 m(a) < T_1 m(b)$. Choose $\delta < T_1 m(a)$ such that $m([0, b] \times \{\delta\}) = 0$. For large enough $n$,

$$m_n([0, b] \times (\delta, \infty]) = m([0, b] \times (\delta, \infty]) = p,$$

$1 \le p < \infty$, and there is an enumeration of the points of $m_n$, call it $((t_i^{(n)}, j_i^{(n)}),\ 1 \le i \le p)$ with $0 < t_1^{(n)} < \cdots < t_q^{(n)} < a < t_{q+1}^{(n)} < \cdots < t_p^{(n)} < b$, $q < p$, such that (Proposition 3.13)

$$\lim_{n\to\infty} ((t_i^{(n)}, j_i^{(n)}), 1 \le i \le p) = ((t_i, j_i), 1 \le i \le p)$$

where $((t_i, j_i), 1 \le i \le p)$ is the analogous enumeration of points of $m$ in $[0, b] \times (\delta, \infty]$. Pick $\delta < \frac{1}{2}\min(t_i - t_{i-1})$ small enough that $\delta$-spheres about the distinct points of the set $\{(t_i, j_i)\}$ are disjoint and in $[0, b] \times [\delta, \infty]$; for $q + 1 \le i \le p$, the $\delta$-spheres should be in $[a, b] \times [\delta, \infty]$. Pick $n$ so large that each $\varepsilon$-sphere contains the same number of points of $m_n$ as of $m$. Define $\lambda_n$ as a homeomorphism of $[a, b]$ onto $[a, b]$ by

$$\lambda_n(a) = a, \qquad \lambda_n(b) = b \qquad \text{and for } q + 1 < i < p$$

$\lambda_n(t_i) = \inf\{t_k^{(n)}: (t_k^{(n)}, j_k^{(n)}) \in \text{sphere of radius } \delta \text{ about } (t_i, j_i)\}$ and $\lambda_n$ is linearly interpolated elsewhere on $[a, b]$. Then

$$\sup_{a \le t \le b} |T_1 m_n(t) - T_1 m(\lambda_n(t))| < \delta$$

$$\sup_{a \le t \le b} |\lambda_n(t) - t| < \delta$$

showing $T_1 m_n$ and $T_1 m$ are at a Skorohod distance $< \delta$ in $D[a, b]$.

We now comment on assertion (4.52). Since $N(n) = \sum_{i=1}^{n} 1_{[X_i > 0]}$ we get for any $t > 0$ by the strong law of large numbers

$$n^{-1} N([nt]) \to (1 - F(0))t \qquad \text{a.s.}$$

and hence for any $\{t_i)$ dense in $(0, \infty)$ we get in $\mathbb{R}^\infty$

$$\{n^{-1}N([nt_i]), i \geq 1\} \to \{(1 - F(0))e(t_i), i \geq 1\}$$

so that for almost all $\omega$ the monotone functions $n^{-1}N([n\cdot])$ converge on a dense set, and hence weakly, to the continuous limit $(1 - F(0))e$. Local uniform convergence ensues from (0.1). This suffices for the Skorohod convergence in (4.52).

Finally we return to the découpage and give an indication of how this is proved. Since $N(n) = \sum_{i=1}^{n} 1_{[X_i \in B]}$ it is enough to show

$$\{X_{K^+(i)}\}, \qquad \{X_{K^-(i)}\}, \qquad \{1_{[X_j \in B]}\}$$

are independent with the given distributions. For integers $k$, $l$, and $m$ we look at the joint probability

$$\mathbf{P}\left\{\bigcap_{i=1}^{k} [X_{K^+(i)} \in B_i] \bigcap_{j=1}^{l} [X_{K^-(j)} \in A_j] \bigcap_{\alpha=1}^{m} [1_{[X_\alpha \in B]} = \delta_\alpha]\right\}$$

where $B_i \subset B$, $B_i \in \mathscr{S}$, $A_i \subset B^c$, $A_i \in \mathscr{S}$, and $\{\delta_\alpha, 1 \leq \alpha \leq m\} \in \{0, 1\}^m$. For concreteness we suppose the number of ones $\sum_{\alpha=1}^{m} \delta_\alpha \geq k$ and the number of zeros $\sum_{\alpha=1}^{m} (1 - \delta_\alpha) \geq m$. Suppose ones occur in the sequence $\delta_1, \ldots, \delta_m$ at indices $i(1), \ldots, i(\sum_{\alpha=1}^{m} \delta_\alpha)$ and zeros occur at indices $j(1), \ldots, j(m - \sum_{\alpha=1}^{m} \delta_\alpha)$. The preceding joint probability is then

$$\mathbf{P}\left\{\bigcap_{p=1}^{k} [X_{i(p)} \in B_p] \bigcap_{q=1}^{l} [X_{j(q)} \in A_q] \bigcap_{p=k+1}^{\Sigma\delta_\alpha} [X_{i(p)} \in B] \bigcap_{q=l+1}^{m-\Sigma\delta_\alpha} [X_{j(q)} \in B^c]\right\}$$

which by independence is

$$\begin{aligned}
&\prod_{p=1}^{k} P[X_p \in B_p] \prod_{q=1}^{l} P[X_q \in A_q] P[X_1 \in B]^{\Sigma\delta_\alpha - k} P[X_1 \in B^c]^{m - \Sigma\delta_\alpha - l} \\
&= \prod_{p=1}^{k} \left(\frac{P[X_p \in B_p]}{P[X_p \in B]}\right) \prod_{q=1}^{l} \left(\frac{P[X_q \in A_q)}{P[X_q \in B^c]}\right) P[X_1 \in B]^{\Sigma\delta_\alpha} P[X_1 \in B^c]^{m - \Sigma\delta_\alpha} \\
&= \prod_{p=1}^{k} P[X_{K^+(p)} \in B_p] \prod_{q=1}^{l} P[X_{K^-(q)} \in A_q] \prod_{\alpha=1}^{m} P[1_{[X_\alpha \in B]} = \delta_\alpha]
\end{aligned}$$

as required. □

We now look at some weak convergence applications of Proposition 4.20 which follow by the continuous mapping theorem. For each of these, we suppose that the assumptions of Proposition 4.20 hold, i.e., that (4.47) is valid.

If $Y$ is extremal-$G$ where $G$ is continuous with left and right endpoints $x_l$ and $x_0$, respectively, we may consider along with $Y$ its path inverse $Y^{\leftarrow} = \{Y^{\leftarrow}(x), x_l < x < x_0\}$ defined by

$$Y^{\leftarrow}(x) = \inf\{t \colon Y(t) > x\}.$$

$Y^{\leftarrow}$ is a process with independent increments (Proposition 4.8). (Note here we take right continuous inverses to keep all paths in $D(x_l, x_0)$.)

**Corollary 4.21.** *If* (4.47) *holds then*

$$Y_n^{\leftarrow} \Rightarrow Y^{\leftarrow}$$

*in*

$$\begin{array}{ll} D(0,\infty) & \textit{if } G = \Phi_\alpha \\ D(-\infty,\infty) & \textit{if } G = \Lambda \\ D(-\infty,0) & \textit{if } G = \Psi_\alpha. \end{array}$$

*Remark.* $D(-\infty,\infty)$ and $D(-\infty,0)$ are defined analogously with $D(0,\infty)$.

PROOF. In case $G = \Lambda$ or $\Psi_\alpha$, we return to Corollary 4.19 and apply the functional $T_2$: $M_p((0,\infty) \times (-\infty,\infty]) \to D(-\infty,\infty)$ (in the case of $\Lambda$, say) defined on $m = \sum \varepsilon_{(\tau_k, y_k)}$ by

$$(T_2 m)(t) = \inf\{\tau_k: y_k > t\}.$$

Then $T_2(\sum \varepsilon_{(t_k, j_k)}) = Y^{\leftarrow}$ and $T_2$ is a.s. continuous by an argument similar to the one used to show $T_1$ a.s. continuous. Hence

$$T_2\xi_n = Y_n^{\leftarrow} \Rightarrow T_2\xi = Y^{\leftarrow}.$$

In case $G = \Phi_\alpha$ we again use the découpage: For $t > 0$ let

$$M^{\leftarrow}(t) = \inf\{i: X_{K^+(i)} > t\}$$

and note

$$Y_n^{\leftarrow}(t) = \inf\{i/n: X_i/a_n > t\}$$

and because $t > 0$ this

$$\begin{aligned} &= \inf\{K^+(i): X_{K^+(i)}/a_n > t\}/n \\ &= K^+(\inf\{i: X_{K^+(i)}/a_n > t\})/n \\ &= K^+(M^{\leftarrow}(a_n t))/n \end{aligned}$$

and so

$$Y_n^{\leftarrow}(t) = K^+(n(n^{-1}M^{\leftarrow}(a_n t))/n. \tag{4.56}$$

Applying $T_2$ to (4.49) we get in $D(0,\infty)$

$$M^{\leftarrow}(a_n t)/n \Rightarrow Y^{\leftarrow} \tag{4.57}$$

where $Y$ appears in (4.50). From (4.51) we have in $D(0,\infty)$

$$Y(\cdot) \stackrel{d}{=} Y_\alpha((\cdot)/(1 - F(0)))$$

and therefore we have

$$Y^{\leftarrow}(\cdot) \stackrel{d}{=} (1 - F(0))Y_\alpha^{\leftarrow}(\cdot)$$

in $D(0,\infty)$ as well. Since by renewal theory we have

$$K^+(n)/n \to (1 - F(0))^{-1} \qquad \text{a.s.}$$

we get by the argument leading to (4.52)

$$K^+([n\cdot])/n \to (1 - F(0))^{-1}e \qquad \text{a.s.} \tag{4.58}$$

in $D(0, \infty)$. Combine (4.57) and (4.58) into a joint statement (cf. Billingsley, 1968, page 27, and Exercise 4.4.2.1)

$$(M^{\leftarrow}(a_n\cdot)/n, K^+([n\cdot])/n) \Rightarrow ((1 - F(0))Y_\alpha^{\leftarrow}, (1 - F(0))^{-1}e)$$

in $D(0, \infty) \times D(0, \infty)$ and composing components (an a.s. continuous operation; cf. Exercise 4.4.2.2 and Whitt, 1980) we get from (4.56)

$$Y_n^{\leftarrow} = n^{-1}K^+([n\cdot n^{-1}M^{\leftarrow}(a_n\cdot)]) \Rightarrow (1 - F(0))^{-1}(1 - F(0))Y_\alpha^{\leftarrow} = Y_\alpha^{\leftarrow}$$

in $D(0, \infty)$. □

*Remark.* One might be tempted to proceed from Proposition 4.20 via the map $T_3$ defined on nondecreasing functions by $T_3 x = x^{\leftarrow}$. However, this is not continuous on $D(0, \infty)$. Consider $x_n(t)$, $t > 0$ defined by

$$x_n(t) = \begin{cases} \frac{1}{2} - n^{-1}, & 0 < t < \frac{1}{2} \\ \frac{1}{2} + n^{-1}, & \frac{1}{2} \le t < 1 \\ t & t \ge 1 \end{cases}$$

$$x(t) = \begin{cases} \frac{1}{2} & 0 < t < 1 \\ t & t \ge 1. \end{cases}$$

For any $0 < a < b$, $\sup_{t \in [a,b]} |x_n(t) - x(t)| \le 1/n \to 0$ and so in $D(0, \infty)$, $d(x_n, x) \to 0$. However,

$$x_n^{\leftarrow}(t) = \begin{cases} 0 & 0 < t < \frac{1}{2} - n^{-1} \\ \frac{1}{2} & \frac{1}{2} - n^{-1} \le t < \frac{1}{2} + n^{-1} \\ 1 & \frac{1}{2} + n^{-1} \le t < 1 \\ t & t > 1 \end{cases}$$

$$x^{\leftarrow}(t) = \begin{cases} 0 & t < \frac{1}{2} \\ 1 & \frac{1}{2} \le t < 1 \\ t & t > 1 \end{cases}$$

and so matter how time is dilated, the distance between the graphs is at least 1/2.

Next consider $D^{\uparrow}(0, \infty)$, the subset of $D(0, \infty)$ consisting of nondecreasing jump functions; these are nondecreasing functions which are constant between isolated jumps. In a compact subset of $(0, \infty)$, such functions have only finitely many jumps. Define

$$T_4: D^{\uparrow}(0, \infty) \to M_p(0, \infty)$$

at $x \in D^{\uparrow}(0, \infty)$ by

$$(T_4 x) = \sum \varepsilon_{t_i}$$

where $\{t_i\}$ are the points of discontinuity of $x$. This is relevant because extremal processes live in $D^{\uparrow}(0, \infty)$ and if $Y$ is extremal

$$T_4 Y = \sum_n \varepsilon_{\tau_n}$$

where $\{\tau_n\}$ are the jump times of $Y$. The map $T_4$ is continuous when restricted to $D^{\uparrow}(0, \infty)$. For suppose $x_n$, $n \geq 0$ are functions in $D^{\uparrow}(0, \infty)$ and $d(x_n, x_0) \to 0$ in the Skorohod topology. If $f \in C_K^+(0, \infty)$, the support of $f$ is a compact set contained in $[a, b]$ for some $0 < a < b$ with $a$ and $b \in \mathscr{C}(x_0)$. There exist $\lambda_n \in \Lambda_{a,b}$ such that

$$\sup_{t \in [a,b]} |x(\lambda_n(t)) - x_0(t)| \to 0 \tag{4.59}$$

$$\sup_{t \in [a,b]} |\lambda_n(t) - t| \to 0. \tag{4.60}$$

If $T_4 x_n = \sum_i \varepsilon_{t_i^{(n)}}$ then we must check

$$\sum_i f(t_i^{(0)}) 1_{[a,b]}(t_i^{(0)}) = \lim_n \sum_i f(t_i^{(n)}) 1_{[a,b]}(t_i^{(n)}) \tag{4.61}$$

and since $x_n \in D^{\uparrow}(0, \infty)$ for $n \geq 0$, the sums in each case involve only a finite number of nonzero terms. From (4.59) and (4.60) we see that the jump points of $x_n$ on $[a, b]$ must be close to those of $x_0$ and (4.61) follows.

This means that $T_4$ is a.s. continuous with respect to the distribution of $Y$ and from the continuous mapping theorem we get the following result.

**Corollary 4.22.** *If* (4.47) *holds then*

$$T_4 Y_n \Rightarrow T_4 Y,$$

*i.e.,*

$$\sum_{k=1}^{\infty} \varepsilon_{n^{-1}L(k)} \Rightarrow \mu_\infty = \sum \varepsilon_{\tau_n}$$

*in* $M_p(0, \infty)$ *where* $\mu_\infty$ *is* PRM$(t^{-1}dt)$.

The last statement follows from Proposition 4.9. Note

$$T_4 Y_n = T_4 M_{[n \cdot]} = \sum_k \varepsilon_{n^{-1}L(k)}$$

since the function $\{M_{[nt]}, t > 0\}$ jumps at $\{L(k)/n, k \geq 1\}$.

Compare this result with Corollary 4.5 where $F$ was required to be continuous. Here we require $F \in D(G)$.

Now apply $T_4$ to the convergence in Corollary 4.21 and use again the continuous mapping theorem.

**Corollary 4.23.** *If* (4.47) *holds then*

$$T_4 Y_n^{\leftarrow} \Rightarrow T_4 Y^{\leftarrow}$$

*i.e.*,

$$\sum_{k=1}^{\infty} \varepsilon_{(X_{L(k)}-b_n)/a_n} \Rightarrow \sum_k \varepsilon_{Y(\tau_k)}$$

*in*

$$M_p(0,\infty) \qquad \text{if } G = \Phi_\alpha$$

$$M_p(-\infty,\infty) \qquad \text{if } G = \Lambda$$

$$M_p(-\infty,0) \qquad \text{if } G = \Psi_\alpha.$$

*Recall that* $\sum_k \varepsilon_{Y(\tau_k)}$ *is* PRM *with mean measure determined by* $S(x) = -\log(-\log G(x))$.

Again compare this result with Proposition 4.1(iii).

Remark on the magic of the invariance principle: Suppose $\{X_n, n \geq 1\}$ is any sequence (for example, stationary) which is not necessarily iid but for which the point process convergence conclusion of Corollary 4.19 is valid. Then for such a sequence Proposition 4.20 and Corollaries 4.21, 4.22, and 4.23 all hold. This remark will be illustrated in the next section, where we study extreme values in the important example of moving average processes.

The preceding results detail the basic convergences. We now give some further illustrations of the power of the invariance principle.

For $m = \sum \varepsilon_{t_i} \in M_p(0,\infty)$, such that $m(1,\infty) = \infty$, the map

$$T_5 m \to (t_1, t_2, \ldots)$$

where $1 < t_1 < t_2 \cdots$ is a.s. continuous from $M_p(0,\infty) \to (0,\infty]^\infty = \prod_1^\infty (0,\infty)$ (cf. Proposition 3.13 and Exercise 4.4.2.11). Apply $T_5$ to the convergence in Corollary 4.22 so that for the terms to the right of 1 we get in $\mathbb{R}^\infty$

$$\{L(k)/n \colon L(k) > n\} \Rightarrow \{\tau_i, i \geq 1\} \stackrel{d}{=} \{e^{\Gamma_i}, i \geq 1\}. \tag{4.62}$$

Letting $\mu(n) = \mu[1,n] =$ number of records among $X_1, \ldots, X_n$, we rephrase (4.62) as

$$n^{-1}(L(\mu(n)+1), L(\mu(n)+2), \ldots) \Rightarrow \{e^{\Gamma_i}; i \geq 1\}.$$

In particular we get for the index of the first record past $n$ that for $x > 0$ as $n \to \infty$

$$P[(L(\mu(n)+1) - n)/n \leq x] \to P[e^{\Gamma_1} - 1 \leq x] = P[\Gamma_1 \leq \log(1+x)]$$
$$= 1 - e^{-\log(1+x)} = x/(1+x).$$

Likewise for the index of the last record at or before $n$ we get for $0 < x < 1$ as $n \to \infty$

$$P[(n - L(\mu(n)))/n \leq x] \to P[1 - \tau_{-1} \leq x] = P\left[\sum_{k=-\infty}^{\infty} \varepsilon_{\tau_k}(1-x, 1) \geq 1\right]$$

and using the fact that $\sum \varepsilon_{\tau_k}$ is PRM$(t^{-1}dt)$ this probability is

$$1 - \exp\{-\log(1/(1 - x))\} = 1 - (1 - x) = x$$

giving an asymptotic uniform distribution.

In a similar way we find for $x > 0$ as $n \to \infty$

$$P[n^{-1}(L(\mu(n) + 1) - L(\mu(n))) \le x] \to P[\tau_1 - \tau_{-1} \le x]$$
$$= \begin{cases} x - \log(1 + x), & x \le 1 \\ 1 - \log(x^{-1}(1 + x)), & x > 1. \end{cases}$$

The last distribution may be computed by using the fact that $(\tau_1 - 1)$ and $1 - \tau_{-1}$ are independent random variables (since PRM's on $(0, \infty)$ have independent increments) so that the distribution of $\tau_1 - \tau_{-1} = (\tau_1 - 1) + (1 - \tau_{-1})$ is a convolution of the two previous limit distributions.

Some cheap variants of (4.62) are

$$n^{-1}(L(k + 1) - L(k): L(k) > n) \Rightarrow \{\tau_{i+1} - \tau_i, i \ge 1\} \stackrel{d}{=} \{e^{\Gamma_{i+1}} - e^{\Gamma_i}, 1 \ge 1\}$$
$$(L(k + 1)/L(k): L(k) > n) \Rightarrow \{e^{E_i}, i \ge 1\}$$
$$(\log(L(k)/n): L(k) > n) \Rightarrow \{\Gamma_i, i \ge 1\}.$$

Now apply $T_5$ in Corollary 4.23. For points to the right of 1 we get when $G = \Lambda$

$$((X_{L(k)} - b_n)/a_n: X_{L(k)} > a_n + b_n) \Rightarrow \{1 + \Gamma_i, i \ge 1\} \tag{4.63}$$

so that

$$((X_{L(k+1)} - X_{L(k)})/a_n: X_{L(k)} > a_n + b_n) \Rightarrow (E_1, E_2, \ldots). \tag{4.64}$$

These results can be made marginally neater by employing a change of variable: Recall that possible choices of $a_n$ and $b_n$ are

$$b_n = F^{\leftarrow}(1 - n^{-1}), \qquad a_n = F^{\leftarrow}(1 - (ne)^{-1}) - b_n$$

so that

$$a_n + b_n = F^{\leftarrow}(1 - (ne)^{-1}).$$

Setting $T = a_n + b_n$ we invert and recalling that $a_n$ is the retraction to the integers of a slowly varying function $a(\cdot)$ we get, for example, in (4.64) that

$$(X_{L(k+1)} - X_{L(k)})/a\left(\frac{1}{1 - F(t)}\right): X_{L(k)} > T) \Rightarrow (E_1, E_2, \ldots)$$

as $T \to \infty$.

When $G = \Phi_\alpha$ similar results hold:

$$(X_{L(k)}/a_n: X_{L(k)} > a_n) \Rightarrow (Y(\tau_i): Y(\tau_i) \ge 1) \stackrel{d}{=} (\exp\{\alpha^{-1}\Gamma_i\}, i \ge 1).$$

Changing variables $T = a_n$ gives a somewhat neater limit theorem as $T \to \infty$. A variant is

$$(X_{L(k+1)}/X_{L(k)}: X_{L(k)} > T) \Rightarrow (e^{\alpha^{-1}E_i}, i \geq 1) \stackrel{d}{=} (U_i^{1/\alpha}, i \geq 1)$$

where $U_i$, $i \geq 1$, are iid uniform (0, 1) random variables.

By now, the idea will have become clear.

EXERCISES

4.4.2.1. Suppose $X_n$, $Y_n$, and $n \geq 0$ are random elements of $D(0, \infty)$ and all are defined on the same probability space. If

$$X_n \Rightarrow X_0$$

in $D(0, \infty)$ and

$$Y_n \Rightarrow Y_0$$

in $D(0, \infty)$ where $Y_0$ is a.s. constant then show

$$(X_n, Y_n) \Rightarrow (X_0, Y_0)$$

in $D(0, \infty) \times D(0, \infty)$ (Billingsley, 1968, page 27).

4.4.2.2. (a) Suppose $x_n$, $n \geq 0$ are functions in $D(0, \infty)$ and $\tau_n \in D(0, \infty)$ is nondecreasing for $n \geq 0$, $\tau_n$: $(0, \infty) \to (0, \infty)$, $d(x_n, x_0) \to 0$, $d(\tau_n, \tau_0) \to 0$. If $\tau_0$ is continuous, show

$$d(x_n \circ \tau_n, x_0 \circ \tau_0) \to 0.$$

(b) Assume in addition $x_n$ is nondecreasing, $x_n \in D(0, \infty)$ and $x_n$: $(0, \infty) \to (0, \infty)$. Show

$$d(\tau_n \circ x_n, \tau_0 \circ x_0).$$

Hint: Try using Exercise 4.4.1.10. See Whitt (1980) for details and refinements.

4.4.2.3. Combine Proposition 4.20 and Corollary 4.22 to show

$$(Y_n, Y_n^{\leftarrow}) \Rightarrow (Y, Y^{\leftarrow})$$

$$\begin{aligned} \text{in } D(0, \infty) \times D(-\infty, \infty) &\qquad \text{if } G = \Lambda \\ D(0, \infty) \times D(0, \infty) &\qquad \text{if } G = \Phi_\alpha \\ D(0, \infty) \times D(-\infty, 0) &\qquad \text{if } G = \Psi_\alpha. \end{aligned}$$

4.4.2.4. Let $\{X_{n,k}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of dependent random variables defined on the same space and suppose $\{\mathscr{F}_{n,k}, 0 \leq k \leq n, n \geq 1\}$ is an array of $\sigma$-algebras such that $X_{n,k}$ is $\mathscr{F}_{n,k}$ measurable and for each $n$, $\mathscr{F}_{n,k-1} \subset \mathscr{F}_{n,k}$. If for a measure $\nu$

(a) $\sum_{k=1}^{n} P[X_{n,k} > x | \mathscr{F}_{n,k-1}] \xrightarrow{P} \nu(x, \infty)$

$\sum_{k=1}^{n} P[X_{n,k} \leq y | \mathscr{F}_{n,k-1}] \xrightarrow{P} \nu(-\infty, y]$

for $y < 0 < x$, $x$ and $y$ not atoms of $\nu$ and

(b) $\max_{1\le j\le n} P[|X_{n,j}| > x|\mathscr{F}_{n,j-1}] \xrightarrow{P} 0 \qquad \text{for } x > 0$

then

$$\sum_1^n \varepsilon_{X_{n,k}} \Rightarrow \sum_1^\infty \varepsilon_{j_k}$$

in $M_p([-\infty,\infty]\backslash\{0\})$ where $\sum \varepsilon_{j_k}$ is PRM($\nu$).
Hint: Review Proposition 4.8. Define the random measure

$$\mu_n(\omega,\cdot) = \sum_1^n P[X_{n,k} \in \cdot\,|\mathscr{F}_{n,k-1}]$$

so that (a) is equivalent to

$$\mu_n \xrightarrow{P} \nu$$

in $M_+([-\infty,\infty]\backslash\{0\})$ (Durrett and Resnick, 1978).

4.4.2.5. Let $F_n$, $n \ge 0$ be probability distributions such that $F_n \to F_0$ weakly. If $Y^{(n)}$ is extremal-$F_n$ show

$$Y^{(n)} \Rightarrow Y^{(0)}$$

in $D(0,\infty)$.

4.4.2.6. Let $\tau(v) = \inf\{n\colon M_n \ge v\}$. If $F \in D(\Phi_\alpha)$, find a limit law for $M_{\tau(v)}$ as $v \to \infty$.

4.4.2.7. If $Y_\alpha$ is extremal-$\Phi_\alpha$ show for any $c > 0$

$$Y_\alpha(c\cdot) \stackrel{d}{=} c^{1/\alpha} Y_\alpha \qquad \text{in } D(0,\infty).$$

If $Y$ is extremal-$\Lambda$ show for $c > 0$

$$Y(c\cdot) \stackrel{d}{=} Y(\cdot) + \log c \qquad \text{in } D(0,\infty).$$

4.4.2.8. Suppose $\{X_n, n \ge 1\}$ are iid random variables satisfying for $0 < p$, $q < 1$, $p + q = 1$, $\alpha > 0$:

$$P[|X_1| > x] \sim x^{-\alpha}L(x), \qquad x \to \infty$$

$$\lim_{x\to\infty} P[X_1 > x]/P[|X_1| > x] = p$$

$$\lim_{x\to\infty} P[X_1 \le -x]/P[|X_1| > x] = q.$$

Define for $x > 0$

$$\nu(x,\infty] = px^{-\alpha}$$

$$\nu[-\infty,-x] = qx^{-\alpha}$$

and set $E = [-\infty,\infty]\backslash\{0\}$ so compact sets are those closed sets bounded away from zero.
(a) Show

$$\xi_n := \sum \varepsilon_{(k/n,\,X_k/a_n)} \Rightarrow \xi = \text{PRM}(dt \times d\nu)$$

on $M_p([0,\infty) \times E)$, where $a_n$ is chosen so that

$$a_n = (1/P[|X_1| > \cdot\,)^{\leftarrow}(n).$$

(b) Let $X_n^{(1)}$ be the term of maximum modulus among $X_1, \ldots, X_n$. Find a limit law for $X_n^{(1)}$ by applying a functional to the result in (a).
(c) Let $T_8^{[a,b]}$: $M_p([0,\infty) \times E) \to D[0,\infty)$ be defined by $(0 < a < b)$

$$T_8^{[a,b]}\left(\sum_k \varepsilon_{(\tau_k, y_k)}\right)(t) = \sum_{\tau_k \le t} y_k 1_{[|y_k| \in [a,b]]}.$$

Show that $T_8$ is a.s. continuous with respect to $\xi$ and compute

$$E(T_8^{[a,b]}\xi)(t), \qquad \operatorname{Var}(T_8^{[a,b]}\xi)(t).$$

(d) For $1 = \delta_0 > \delta_1 > \cdots \to 0$, show by using the Kolmogorov convergence criterion (summing the variances) that

$$\sum_{i=0}^{\infty} ((T_8^{[\delta_{i+1},\delta_i]}\xi)(t) - E(T_8^{[\delta_{i+1},\delta_i]}\xi)(t))$$

converges a.s.
(e) Pick $\{\delta_i\}$ at your convenience to guarantee that convergence in (d) is uniform for $t \in [0,1]$ (Kolmogorov inequality).
(f) Show that for any $i$

$$X_n^{\delta_i}(t) = a_n^{-1} \sum_1^{[nt]} X_j 1_{[|X_j| > a_n\delta_i]} - na_n^{-1} tEX_1 1_{[a_n^{-1}|X_1| \in (\delta_i, 1]}$$

$$\Rightarrow \int_{[|u| > \delta_i, s \le t]} u\xi(ds, du) - t \int_{[1 > |u| > \delta_i]} u\nu(du) = X^{\delta_i}(\cdot)$$

in $D[0,1]$ and that almost surely and uniformly on $[0,1]$

$$X^{\delta_i} \to X$$

as $i \to \infty$, where $X$ is a stable process.
(g) Show

$$X_n^0(t) = a_n^{-1} \sum_{i=1}^{[nt]} X_i - na_n^{-1} tEX_1 1_{[a_n^{-1}|X_1| \le 1]} \Rightarrow X(t)$$

in $D[0,\infty)$ by showing, using Kolmogorov's inequality and Karamata's theorem,

$$\lim_{i\to\infty} \limsup_{n\to\infty} P[d(X_n^0, X_n^{\delta_i}) > \varepsilon] = 0$$

where $d$ is the Skorohod metric on $D[0,\infty)$ (Resnick, 1986; Durrett and Resnick, 1978).

4.4.2.9. If $F \in D(\Lambda)$ show for $x > 0$

$$P[a_n^{-1}(M_n - M_{n-1}) > x \mid n \in \{L(j), j \ge 1\}] \to e^{-x}$$

as $n \to \infty$ (McCormick, 1983).

4.4.2.10. Suppose $X_k$, $k \ge 1$ are iid, $X_1 > 0$, and

$$P[X_1 > x] \sim x^{-\alpha}L(x), \qquad x \to \infty, \qquad \alpha > 1.$$

Let $\tau(\beta) = \sup\{k \ge 1: X_k > \beta k\}$.
(a) Check $\tau(\beta) < \infty$.

(b) The convergence in Corollary 4.19 holds with $n$ replaced by a continuous variable, $u$ say.
(c) Let $T_9$: $M_p[0, \infty) \times (0, \infty) \to M_p([0, \infty) \times [0, \infty))$ be the map induced by the transformation of points

$$(t, y) \to (ty^{-1}, t).$$

Why is this map a.s. continuous?
(d) Apply $T_9$ and then $T_1$ and conclude as $u \to \infty$

$$\tau(u^{-1}a(u)/(\cdot))/u \Rightarrow \sup\{t_k\colon t_k j_k^{-1} < (\cdot)\}.$$

Change variables to get in $\mathbb{R}$ as $s \to \infty$

$$\tau(1/s)/s\alpha(s) \Rightarrow Y^{\#}(1)$$

where $P[Y^{\#}(1) \le x] = (\Phi_\alpha(x))^{(\alpha-1)^{-1}}$. What is $\alpha(s)$? (Husler, 1979; Resnick, 1986)

4.4.2.11. Check $T_5$ is continuous when restricted to $\{m \in M(0, \infty)\colon m(1, \infty) = \infty\}$.

4.4.2.12. Suppose $\{X_n, n \ge 1\}$ is iid with common distribution $F$ such that for $a_n > 0$, $b_n \in \mathbb{R}$

$$F^n(a_n x + b_n) \to G(x)$$

nondegenerate. Let $\xi = \sum_k \varepsilon_{(t_k, u_k)}$ be homogeneous PRM on $[0, \infty)^2$ and let the points in $[0, \infty) \times [0, n)$ be $\{t_k^{(n)}, u_k^{(n)}\}$ where $0 \le t_1^{(n)} < t_2^{(n)} < \cdots$.
(a) $\{u_k^{(n)}\}$ is iid, uniform on $[0, n)$.
(b) For each $n$

$$\{a_n^{-1}(X_k - b_n), k \ge 1\} \overset{d}{=} \{a_n^{-1}(F^{\leftarrow}(\{(1 - n^{-1}u_k^{(n)})^n\}^{1/n}) - b_n), k \ge 1\}$$

in $\mathbb{R}^\infty$.
(c) We have $(F^{\leftarrow}(y^{1/n}) - b_n)/a_n \to G^{\leftarrow}(y)$.
(d) In $M_p([0, \infty) \times (-\infty, \infty])$

$$\sum \varepsilon_{(kn^{-1}, a_n^{-1}(F^{\leftarrow}(\{(1-n^{-1}u_k^{(n)})^n\}^{1/n}) - b_n))} \to \sum_k \varepsilon_{(t_k, G^{\leftarrow}(e^{-u_k}))}$$

almost surely if $G = \Lambda$ with similar results in the other cases.
(e) Hence

$$\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1}(X_k - b_n))} \Rightarrow \text{PRM}$$

and the corollary about weak convergence to extremal processes follows (Pickands, 1971; Resnick, 1975; de Haan, 1984a).

## 4.5. Extreme Value Theory for Moving Averages

In this section we analyze stationary moving average processes where the averaged variables have distributions with regularly varying tails. Such processes are worthy of our attention for at least two reasons. First of all, from the didactic perspective, the analysis of such processes offers additional excellent illustrations of the usefulness and power of the probabilistic and analytic tools thus far developed. Secondly, the autoregressive moving average processes of orders $p$ and $q$ (ARMA $(p, q)$) are among the most frequently used models in time series analysis and ARMA's driven by noise sequences with regularly varying tail probabilities will satisfy the hypotheses of the results to be given later.

Suppose $\{Z_k, -\infty < k < \infty\}$ is a sequence of iid random variables and assume

$$P[|Z_k| > x] \in RV_{-\alpha}, \qquad \alpha > 0 \tag{4.65}$$

and

$$\lim_{x\to\infty} \frac{P[Z_k > x]}{P[|Z_k| > x]} = p, \quad \lim_{x\to\infty} \frac{P[Z_k \le -x]}{P[|Z_k| > x]} = q, \quad 0 \le p \le 1, \quad p + q = 1. \tag{4.66}$$

The sequence of real constants $\{c_j, -\infty < j < \infty\}$ satisfies

$$\sum_{j-\infty}^{\infty} |c_j|^\delta < \infty \qquad \text{for some } 0 < \delta < \alpha \wedge 1. \tag{4.67}$$

The strictly stationary sequence of moving averages is given by

$$X_n := \sum_{j-\infty}^{\infty} c_j Z_{n-j}, \qquad -\infty < n < \infty. \tag{4.68}$$

We study the weak limit behavior of various quantities related to the extremes of $\{X_n\}$.

An immediate issue is whether the series in (4.68) converges. Since $0 < \delta < 1 \wedge \alpha$ we have by the triangle inequality

$$E|X_n|^\delta \le \sum_j |c_j|^\delta E|Z_{n-j}|^\delta = E|Z_1|^\delta \sum_j |c_j|^\delta < \infty,$$

using (4.67) and the fact that $E|Z_1|^\delta < \infty$ (Exercise 1.2.2). Thus the series in (4.68) must be almost surely convergent.

Note that (4.65) and (4.66) are conditions defining global regular variation involving both tails and is thus stronger than the right tail regular variation conditions tupically encountered in extreme value theory. Since $X_n$ is defined by an infinite series, it is convenient for analysis of extremes and other functionals to have firm control of left tail behavior.

The one point uncompactification: The global regular variation conditions (4.65) and (4.66) lead to consideration of such state spaces as $[0, \infty]^d \setminus \{\mathbf{0}\}$, $[-\infty, \infty]^d \setminus \{\mathbf{0}\}$ for some $d \ge 1$ where $\{\mathbf{0}\}$ is understood as the origin of $\mathbb{R}^d$. In order that (4.65) and (4.66) be equivalent to appropriate statements about vague convergence of measures, it is essential that these spaces be understood to have topologies obtained by removing the origin from compact sets. Thus $[-\infty, \infty]^d \setminus \{\mathbf{0}\}$ is the compact set $[-\infty, \infty]^d$ with $\mathbf{0}$ removed, and so on. In such punctured spaces, compact sets must be bounded away from $\mathbf{0}$. The spaces may be metrized by interchanging the roles of zero and infinity. For example, in $(0, \infty]$ a suitable metric is

$$d(x_1, x_2) = |x_1^{-1} - x_2^{-1}|$$

for $x_1 > 0$, $x_2 > 0$. In $[-\infty, \infty]^2 \backslash \{\mathbf{0}\}$ each of the following sets is compact:

$$\{\mathbf{x} \in [-\infty, \infty]^2 \backslash \{\mathbf{0}\}: |x_1| + |x_2| \geq 1\}, \qquad \{\mathbf{x} \in [-\infty, \infty]^2 \backslash \{\mathbf{0}\}: \|\mathbf{x}\| > 1\}$$

where $\mathbf{x} = (x_1, x_2)$ and $\|\mathbf{x}\| = (x_1^2 + x_2^2)^{1/2}$.

For further results and other approaches see Rootzen (1978), Finster (1982), Hannan and Kanter (1977), Kanter and Steiger (1974), and Davis and Resnick (1985a and b, 1986).

Here is an outline of our approach: First, notice that (4.65) and (4.66) are equivalent to a convergence of point processes result. To see this, let $E = [-\infty, \infty] \backslash \{0\}$ and

$$a_n = (1/P[|Z_1| > \cdot\,])^{\leftarrow}(n)$$

so that $a_n$ is the inverse function of $1/P[|Z_1| > x]$ evaluated at $n$. Then (4.65) and (4.66) are equivalent to

$$nP[a_n^{-1} Z_1 \in \cdot\,] \xrightarrow{v} \nu \tag{4.69}$$

in $[-\infty, \infty] \backslash \{0\}$ where

$$\nu(dx) = p\alpha x^{-\alpha-1} dx\, 1_{(0,\infty]}(x) + q\alpha(-x)^{-\alpha-1} dx\, 1_{[-\infty,0)}(x),$$

and therefore Proposition 3.21 applies and gives

$$\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1} Z_k)} \Rightarrow \sum_k \varepsilon_{(t_k, j_k)} \tag{4.70}$$

as $n \to \infty$ in $M_p([0, \infty) \times [-\infty, \infty] \backslash \{0\})$ where the limit is $\mathrm{PRM}(dt \times d\nu)$. We want a similar result involving the $X$'s. Define for $m > 1$

$$X_n^{(m)} = \sum_{|j| \leq m} c_j Z_{n-j}, \qquad -\infty < n < \infty$$

and think of $X_n^{(m)}$ as a simple functional of the vector

$$\mathbf{Z}_n^{(m)} = (Z_{n-j}, |j| \leq m)$$

which suggests looking at the point processes

$$\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1} \mathbf{Z}_k^{(m)})}$$

and in fact as $n \to \infty$ we are able to show weak convergence of these processes. A simple continuity argument gets us to convergence of point processes based on $(X_n^{(m)}, -\infty < n < \infty)$ and then a Slutsky-style approximation argument allows us to remove $m$.

To carry out this program in detail we need some preliminaries. The first result is a special case of a theorem of Cline (1983a and b).

**Lemma 4.24.** *If* $\{Z_n\}$ *satisfies* (4.65) *and* (4.66) *and* $\{c_n\}$ *satisfies* (4.67) *then*

$$\lim_{x\to\infty}\frac{P[\sum_j|c_j||Z_j|>x]}{P[|Z_1|>x]}=\sum_j|c_j|^\alpha. \tag{4.71}$$

Proof. We begin by showing a tamer result, namely

$$\lim_{n\to\infty}\frac{P[|c_1|Z_1|+|c_2||Z_2|>x]}{P[|Z_1|>x]}=|c_1|^\alpha+|c_2|^\alpha. \tag{4.72}$$

This is a standard analytical exercise (Feller, 1971), but the following alternative approach (Resnick, 1986) is more fun and some of the mechanics will be needed later. The regular variation conditions (4.65) and (4.66) imply

$$nP[a_n^{-1}(|Z_1|,|Z_2|)\in\cdot\,]\xrightarrow{v}\mu \tag{4.73}$$

on $[0,\infty]^2\backslash\{\mathbf{0}\}$ where $\mu$ concentrates on the axes $\{(y,0),y>0\}\cup\{(0,y),y>0\}$ and for $x>0$

$$\mu\{(y,0)\colon y>x\}=\mu\{(0,y)\colon y>x\}=x^{-\alpha}.$$

To check this note for $x_1>0,x_2>0$

$$\begin{aligned}&nP[a_n^{-1}(|Z_1|,|Z_2|)\in(x_1,\infty]\times(x_2,\infty]]\\&=nP[a_n^{-1}|Z_1|>x_1]P[a_n^{-1}|Z_2|>x_2]\\&\to x_1^{-\alpha}0=0\end{aligned}$$

so $\mu$ has no mass in the interior of $[0,\infty]^2$. However,

$$\begin{aligned}\mu\{(y,0)\colon y>x\}\}&=\lim_{n\to\infty}nP[a_n^{-1}|Z_1|>x,a_n^{-1}|Z_2|\geq 0]\\&=\lim_{n\to\infty}nP[a_n^{-1}|Z_1|>x]=x^{-\alpha}.\end{aligned}$$

Now let $\{Z_n',Z_n'',n\geq 1\}$ be iid with $(Z_n',Z_n'')\stackrel{d}{=}(Z_1,Z_2)$, $n\geq 1$, and applying Proposition (3.21) to (4.73) gives

$$\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}(|Z_k'|,|Z_k''|))}\Rightarrow\sum_k(\varepsilon_{(t_k',(j_k',0))}+\varepsilon_{(t_k'',(0,j_k''))}) \tag{4.74}$$

on $M_p([0,\infty)\times([0,\infty]^2\backslash\{\mathbf{0}\}))$ where the limit is PRM($\mu$). Now define $T\colon[0,\infty]^2\backslash\{\mathbf{0}\}\to(0,\infty]$ by

$$T(x_1,x_2)=|c_1|x_1+|c_2|x_2.$$

If for any compact set $K\subset(0,\infty]$, $T^{-1}(K)$ is compact in $[0,\infty]^2\backslash\{\mathbf{0}\}$ then one readily checks that Proposition 3.18 may be applied to (4.74) to obtain

$$\begin{aligned}&\sum_{k=1}^{n}\varepsilon_{(kn^{-1},a_n^{-1}(|c_1||Z_k'|+|c_2||Z_k''|))}\\&\Rightarrow\sum_k\varepsilon_{(t_k',|c_1|j_k')}+\sum_k\varepsilon_{(t_k'',|c_2|j_k'')}\end{aligned} \tag{4.75}$$

where the limit is the sum of two independent Poisson processes and is hence Poisson. The mean measure of $\sum_k \varepsilon_{(t'_k, |c_1|j_k)}$ is easily calculated to be $dt \times |c_1|^\alpha \alpha x^{-\alpha-1}\,dx$, and hence the limit in (4.75) is $\mathrm{PRM}(dt \times (|c_1|^\alpha + |c_2|^\alpha)\alpha x^{-\alpha-1}\,dx)$. Applying Proposition 3.21 in reverse order gives that (4.75) implies

$$nP[a_n^{-1}(|c_1||Z'_k| + |c_2||Z''_k| \in \cdot\,] \xrightarrow{v} (|c_1|^\alpha + |c_2|^\alpha)\alpha x^{-\alpha-1}\,dx$$

on $(0, \infty]$, and this is equivalent to

$$\lim_{x\to\infty} \frac{P[|c_1||Z'_1| + |c_2||Z''_1| > a_n x]}{P[|Z_1| > a_n]} = (|c_1|^\alpha + |c_2|^\alpha)x^{-\alpha}$$

and this is readily seen to be the same as (4.72).

Thus the last detail in the verification of (4.72) is to check the compactness condition of Proposition 3.18, viz

$$T^{-1}(K) \qquad \text{is compact in } [0,\infty]^2\backslash\{0\} \text{ if } K \text{ is compact in } (0,\infty]. \tag{4.76}$$

Note first $T$ is continuous so if $K$ is compact $T^{-1}(K)$ is closed. If $0 < \delta_k \downarrow 0$, we have $\{(\delta_k, \infty], k \geq 1\}$ is an open cover of $(0, \infty]$ and hence an open cover of $K$, and therefore for some $\delta$, $(\delta, \infty] \supset K$. If $T^{-1}(\delta, \infty]$ is compact in $[0, \infty]^2\backslash\{\mathbf{0}\}$, then $T^{-1}(K)$, being a closed subset of $T^{-1}(\delta, \infty]$, must also be compact. Thus it remains to prove $T^{-1}(\delta, \infty]$ is compact, but since

$$T^{-1}(\delta, \infty] = \{(x_1, x_2) \in [0, \infty]^2\backslash\{\mathbf{0}\}: |c_1|x_1 + |c_2|x_2 \geq \delta\}$$

is obviously bounded away from $\mathbf{0}$, the result is clear.

We now must leap from (4.72) to (4.71). For $x > 0$, write

$$\begin{aligned}
&P\Big[\sum_j |c_j||Z_j| > x\Big] \\
&= P\Big[\sum_j |c_j||Z_j| > x, \bigvee_j |c_j||Z_j| > x\Big] + P\Big[\sum_j |c_j||Z_j| > x, \bigvee_j |c_j||Z_j| \leq x\Big] \\
&\leq P\Big\{\bigcup_j \Big[|c_j||Z_j| > x\Big]\Big\} + P\Big[\sum_j |c_j||Z_j| 1_{[|c_j||Z_j| \leq x]} > x, \bigvee_j |c_j||Z_j| \leq x\Big] \\
&\leq \sum_j P\Big[|Z_j| > x|c_j|^{-1}\Big] + P\Big[\sum_j |c_j||Z_j| 1_{[|c_j||Z_j| \leq x]} > x\Big]
\end{aligned}$$

and therefore applying Markov's inequality

$$\begin{aligned}
&P\Big[\sum_j |c_j||Z_j| > x\Big]\Big/P\Big[|Z_1| > x\Big] \\
&\leq \sum_j P[|Z_1| > x|c_j|^{-1}]/P[|Z_1| > x] \\
&\quad + x^{-1}\sum_j |c_j| E|Z_1| 1_{[|Z_1| \leq x|c_j|^{-1}]}/P[|Z_1| > x] \\
&= I + II.
\end{aligned}$$

For $I$ we have by Proposition 0.8(ii) that for all $j$ such that $|c_j| < 1$ (i.e., all but a finite number of $j$) there exists $x_0$ such that $x > x_0$ implies

$$\begin{aligned} &P[|Z_1| > x|c_j|^{-1}]/P[|Z_1| > x] \\ &\leq (1 + \delta)|c_j|^{\delta}. \end{aligned}$$

This bound is summable because of (4.67) and hence by dominated convergence

$$\lim_{x \to \infty} I = \sum |c_j|^{\alpha}.$$

In considering $II$, suppose temporarily that $0 < \alpha < 1$. From an integration by parts

$$\frac{E|Z_1| 1_{[|Z_1| \leq x]}}{xP[|Z_1| > x]} = \frac{\int_0^x P[|Z_1| > u]du}{xP[|Z_1| > x]} - 1$$

and applying Karamata's theorem 0.6 this converges, as $x \to \infty$, to

$$(1 - \alpha)^{-1} - 1 = \alpha(1 - \alpha)^{-1}.$$

Thus $E|Z_1| 1_{[|Z_1| \leq x]} \in RV_{1-\alpha}$ and hence applying again Proposition 0.8(ii) we have, for all but a finite number of $j$, that for $x$ sufficiently large and some constant $k > 0$

$$\begin{aligned} &\frac{|c_j| E|Z_1| 1_{[|Z_1| \leq x|c_j|^{-1}]}}{xP[|Z_1| > x]} \\ &= |c_j| \left( \frac{E|Z_1| 1_{[|Z_1| \leq x|c_j|^{-1}]}}{E|Z_1| 1_{[|Z_1| \leq x]}} \right) \frac{E|Z_1| 1_{[|Z_1| \leq x]}}{xP[|Z_1| > x]} \\ &\leq k|c_j|(|c_j|^{-1})^{1-\alpha+\alpha-\delta} = k|c_j|^{\delta} \end{aligned}$$

which is summable. So we conclude

$$\limsup_{x \to \infty} II \leq k \sum_j |c_j||c_j|^{\alpha-1} = k \sum_j |c_j|^{\alpha}$$

and hence when $0 < \alpha < 1$ for some $k' > 0$

$$\limsup_{x \to \infty} P\left[\sum_j |c_j||Z_j| > x\right] \Big/ P\left[|Z_1| > x\right] \leq k' \sum_j |c_j|^{\alpha}. \tag{4.77}$$

If $\alpha \geq 1$ we get a similar inequality by reduction to the case $0 < \alpha < 1$ as follows: Pick $\gamma \in (\alpha, \alpha\delta^{-1})$ and by Jensen's inequality (e.g., Feller, 1971, page 153) if we set $c = \sum_j |c_j|$, $p_j = |c_j|/c$ we get

$$\begin{aligned} \left(\sum_j |c_j||Z_j|\right)^{\gamma} &= c^{\gamma}\left(\sum_j p_j|Z_j|\right)^{\gamma} \leq c^{\gamma} \sum_j p_j|Z_j|^{\gamma} \\ &= c^{\gamma-1} \sum_j |c_j||Z_j|^{\gamma}. \end{aligned}$$

Then

$$\frac{P[\sum_j |c_j||Z_j| > x]}{P[|Z_1| > x]} \leq \frac{P[\sum_j |c_j||Z_j|^\gamma > c^{1-\gamma}x^\gamma]}{P[|Z_1|^\gamma > x^\gamma]}.$$

Now use the fact that $P[|Z_1|^\gamma > x] \in RV_{-\alpha\gamma^{-1}}$, $\delta < \alpha\gamma^{-1} < 1$, and the preceding result to obtain

$$\limsup_{x\to\infty} \frac{P[\sum_j |c_j||Z_j| > x]}{P[|Z_1| > x]} \leq k' \sum_j |c_j|^{\alpha\gamma^{-1}} c^{\alpha(1-\gamma^{-1})}) < \infty \tag{4.77'}$$

which is similar to (4.77).

Now we are ready to prove (4.71): For any integer $m > 0$

$$\frac{P[\sum_j |c_j||Z_j| > x]}{P[|Z_1| > x]} \geq \frac{P[\sum_{|j|\leq m} |c_j||Z_j| > x]}{P[|Z_1| > x]} \to \sum_{|j|\leq m} |c_j|^\alpha$$

by the obvious extension of (4.72) and since $m$ is arbitrary

$$\liminf_{x\to\infty} P\left[\sum_j |c_j||Z_j| > x\right] \Big/ P\left[|Z_1| > x\right] \geq \sum_j |c_j|.$$

On the other hand for any $\varepsilon > 0$

$$\frac{P[\sum_j |c_j||Z_j| > x]}{P[|Z_1| > x]} \leq \frac{P[\sum_{|j|\leq m} |c_j||Z_j| > (1-\varepsilon)x]}{P[|Z_1| > x]} + \frac{P[\sum_{|j|>m} |c_j||Z_j| > \varepsilon x]}{P[|Z_1| > x]}$$

and so from (4.72) and (4.77) for some $k' > 0$

$$\limsup_{x\to\infty} \frac{P[\sum_j |c_j||Z_j| > x]}{P[|Z_1| > x]} \leq (1-\varepsilon)^{-\alpha} \sum_{|j|\leq m} |c_j|^\alpha + k'\varepsilon^{-\alpha} \sum_{|j|>m} |c_j|^\alpha$$

for the case $0 < \alpha < 1$, with a similar bound for the second piece provided by (4.77') when $\alpha \geq 1$. Let $m \to \infty$ and then send $\varepsilon \to 0$ to obtain

$$\limsup_{x\to\infty} P\left[\sum_j |c_j||Z_j| > x\right] \Big/ P\left[|Z_1| > x\right] \leq \sum_j |c_j|^\alpha$$

and this combined with the liminf statement proves (4.71). □

The following exceedingly useful variant of Slutsky's lemma is Billingsley's (1968) Theorem 4.2.

**Lemma 4.25.** *Let $X_{un}$, $X_u$, $Y_n$, and $X$ be random elements of a complete, separable metric space $S$ with metric $\rho$, such that for each $n$, $Y_n$, $X_{un}$, $u \geq 1$ are defined on a common domain. Suppose for each $u$, as $n \to \infty$,*

$$X_{un} \Rightarrow X_u$$

*and as $u \to \infty$*

$$X_u \Rightarrow X.$$

*Suppose further that for all* $\varepsilon > 0$

$$\lim_{u\to\infty} \limsup_{n\to\infty} \mathbf{P}[\rho(X_{un}, Y_n) > \varepsilon] = 0.$$

*Then we have*

$$Y_n \Rightarrow X$$

*as* $n \to \infty$.

PROOF. See Billingsley, 1968, page 25 or consider the following: We must show $\lim_{n\to\infty} \mathbf{E}f(Y_n) = \mathbf{E}f(X)$ for bounded continuous $f$ on $S$, and in fact it suffices to suppose $f$ is uniformly continuous and bounded (Billingsley, 1968, page 12, or examine the statements equivalent to weak convergence at the beginning of Section 3.5). Now write

$$|\mathbf{E}f(Y_n) - \mathbf{E}f(X)| \le \mathbf{E}|f(Y_n) - f(X_{un})| + |\mathbf{E}f(X_{un}) - \mathbf{E}f(X_u)| \\ + |\mathbf{E}f(X_u) - \mathbf{E}f(X)|$$

so that

$$\begin{aligned} \limsup_{n\to\infty} |\mathbf{E}f(Y_n) - \mathbf{E}f(X)| &\le \lim_{u\to\infty} \limsup_{n\to\infty} \mathbf{E}|f(Y_n) - f(X_{un})| \\ &\le \lim_{u\to\infty} \limsup_{n\to\infty} \mathbf{E}(|f(Y_n) - f(X_{un})|; \rho(Y_n, X_{un}) \le \varepsilon) \\ &\quad + \lim_{u\to\infty} \limsup_{n\to\infty} 2 \sup_{x\in S} |f(x)| \mathbf{P}[\rho(Y_n, X_{un}) > \varepsilon] \\ &\le \sup\{|f(x) - f(y)|: \rho(x, y) \le \varepsilon\} + 0 \to 0 \end{aligned}$$

as $\varepsilon \to 0$ since $f$ is uniformly continuous. □

With these preliminaries out of the way we now prove the basic convergence.

**Proposition 4.26.** *Suppose* (4.65), (4.66), *and* (4.67) *hold so that* (4.70) *follows; viz in* $M_p([0,\infty) \times [-\infty,\infty]\backslash\{0\})$

$$\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1} Z_k)} \Rightarrow \sum_k \varepsilon_{(t_k, j_k)} \tag{4.70}$$

*where the limit is* $\mathrm{PRM}(dt \times dv)$, $v(dx) = \alpha p x^{-\alpha-1}\, dx\, 1_{(0,\infty]}(x) + \alpha q |x|^{-\alpha-1} \cdot dx\, 1_{[-\infty,0)}(x)$. *Fix an integer* $m$ *and define vectors* $\mathbf{e}_j$, $-m \le j \le m$ *of length* $2m+1$ *by*

$$\mathbf{e}_{-m} = (1, 0, \ldots, 0), \ldots, \qquad \mathbf{e}_m = (0, \ldots, 1)$$

*and define random vectors*

$$\mathbf{Z}_k = (Z_{k+j}, -m \le j \le m).$$

*Then in* $M_p([0,\infty)\times[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\})$

$$\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}\mathbf{Z}_k)}\Rightarrow\sum_k\sum_{|i|\le m}\varepsilon_{(t_k,j_k\cdot\mathbf{e}_i)}. \tag{4.78}$$

*Remark.* The limit process in (4.78) is obtained by taking the one-dimensional $j_k$'s in (4.70) and laying them down on axis $\mathbf{e}_0$, and then repeating deterministically this pattern on each axis $\mathbf{e}_{\pm 1},\ldots,\mathbf{e}_{\pm m}$. In the proof of Lemma 4.24, points on different axes were independent but the situation is very different here since the pattern of points on each axis is the same.

To see why (4.78) is plausible consider the following: Any limit point process in (4.78) can have no points off the axes. Let us check this. Define

$$\begin{aligned} SLEEVE = SL &= \{\mathbf{x}\in[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}\colon \text{At most one component of } \mathbf{x} \text{ has modulus greater than } \delta\}\\ &= \{\mathbf{x}\colon |x_l|\le\delta,\ -m\le l\le m\}\\ &\quad\bigcup\bigcup_{-m\le i\le m}\{\mathbf{x}\colon |x_i|>\delta, |x_l|\le\delta, l\ne i,\ -m\le l\le m\}\end{aligned}$$

and

$$(SL)^c = \{\mathbf{x}\in[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}\colon \text{At least two components of } \mathbf{x} \text{ have moduli greater than } \delta\}$$

so that *SLEEVE* consists of narrow sleeves about the axes. The significant characteristic is that a limit process can have no points in $[0,t]\times(SL)^c$ for any $t>0$ since

$$\begin{aligned}\mathbf{E}\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}\mathbf{Z}_k)}([0,t]\times(SL)^c) &= [nt]P[a_n^{-1}\mathbf{Z}_k\in(SL)^c]\\ &\le ntP[a_n^{-1}|Z_1|>\delta, a_n^{-1}|Z_2|>\delta]\binom{2m+1}{2}\\ &\to t\binom{2m+1}{2}\delta^{-\alpha}\cdot 0 = 0.\end{aligned}$$

One can guess the form of the limit in (4.78) by supposing $(kn^{-1},a_n^{-1}\mathbf{Z}_k)\in[0,t]\times SLEEVE$; for instance, suppose $|a_n^{-1}Z_{k+m}|>\delta$, $a_n^{-1}|Z_{k+i}|\le\delta$, and $-m\le i<m$. This supposition sets up a reproducing pattern on the other axes since points eventually leave $[0,t]\times(SL)^c$ and we are likely to have $\mathbf{Z}_{k+1}$ contributing $a_n^{-1}|Z_{k+m}|>\delta$, $a_n^{-1}|Z_{k+1+i}|\le\delta$, $-m\le i\le m$, $i\ne m-1$, and $\mathbf{Z}_{k+2}$ contributing $a_n^{-1}|Z_{k+m}|>\delta$, $a_n^{-1}|Z_{k+2+i}|\le\delta$, $-m\le i\le m$, $i\ne m-2$, and so on. This also suggests comparing $I_n=\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}\mathbf{Z}_k)}$ with

$$I_n^* = \sum_{k=1}^{\infty}\sum_{|i|\le m}\varepsilon_{(kn^{-1},a_n^{-1}Z_k\cdot\mathbf{e}_i)}.$$

PROOF OF PROPOSITION 4.26. We first prove

$$d\left(\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1}\mathbf{Z}_k)}, \sum_{k=1}^{\infty} \sum_{|i| \le m} \varepsilon_{(kn^{-1}, a_n^{-1}Z_k \cdot \mathbf{e}_i)}\right) \xrightarrow{P} 0 \tag{4.79}$$

where $d$ is the vague metric on $M_p([0, \infty) \times [-\infty, \infty]^{2m+1} \backslash \{\mathbf{0}\})$. Because of the definition of $d$ (cf. Proposition 3.17), it suffices to show for any $f \in C_K^+([0, \infty) \times [-\infty, \infty]^{2m+1} \backslash \{\mathbf{0}\})$

$$I_n(f) - I_n^*(f) \xrightarrow{P} 0. \tag{4.80}$$

The support of $f$ must be contained in $[0, t] \times \{\mathbf{x} \in [-\infty, \infty]^{2m+1} \backslash \{\mathbf{0}\}: \bigvee_{i=-m}^{m} |x_i| > \delta\}$ for some $t > 0$, $\delta > 0$ and for notational simplicity take $t = 1$.

From the remarks subsequent to the statement of the proposition

$$E(I_n([0, 1] \times (SL)^c)) \to 0$$

and this readily implies

$$\begin{aligned} I_n(f) = \int f\, dI_n &= \int_{[0,1] \times SL} + \int_{[0,1] \times (SL)^c} \\ &= \int_{[0,1] \times SL} f\, dI_n + o_p(1). \end{aligned}$$

In addition it is evident that

$$I_n^*(f) = \int_{[0,1] \times SL} f\, dI_n^*,$$

and therefore to show (4.80) one needs to show (recall $f(s, \mathbf{z}) = 0$ if $\bigvee_{i=-m}^{m} |z_i| < \delta$)

$$\begin{aligned} &\sum_{l=-m}^{m} \sum_{k=1}^{n} f(kn^{-1}, a_n^{-1}\mathbf{Z}_k) 1_{[a_n^{-1}|Z_{k+l}| > \delta, \bigvee_{i=-m, i \ne l}^{m} a_n^{-1}|Z_{k+i}| \le \delta]} \\ &\quad - \sum_{l=-m}^{m} \sum_{k=1}^{n} f(kn^{-1}, a_n^{-1} Z_k \cdot \mathbf{e}_l) \xrightarrow{P} 0, \end{aligned}$$

for which it suffices to check for fixed $l$

$$\begin{aligned} &\sum_{k=1}^{n} f(kn^{-1}, a_n^{-1}\mathbf{Z}_k) 1_{[a_n^{-1}|Z_{k+l}| > \delta, \bigvee_{i=-m, i \ne l}^{m} a_n^{-1}|Z_{k+i}| \le \delta]} \\ &\quad - \sum_{k=1}^{n} f(kn^{-1}, a_n^{-1} Z_k \cdot \mathbf{e}_l) 1_{[a_n^{-1}|Z_k| > \delta]} \xrightarrow{P} 0 \end{aligned} \tag{4.81}$$

where the indicator on the last sum was added without harm because of the compact support of $f$. Suppose for concreteness $l > 0$. Change dummy indices $k' = k + l$ and the difference in (4.81) can be bounded by

$$\begin{aligned}
&\sum_{k=1-l}^{0} f((k+l)n^{-1}, a_n^{-1} Z_{k+l}\cdot \mathbf{e}_l)\\
&\quad + \sum_{k=n-l+1}^{n} f(kn^{-1}, a_n^{-1} Z_k) 1_{[a_n^{-1}|Z_{k+l}|>\delta, \bigvee_{i=-m, i\neq l}^{m} a_n^{-1}|Z_{k+i}|\le\delta]}\\
&\quad + \sum_{k=1}^{n-l} |f(kn^{-1}, a_n^{-1}\mathbf{Z}_k)\\
&\quad - f((k+l)n^{-1}, a_n^{-1} Z_{k+l}\cdot \mathbf{e}_l)| 1_{[a_n^{-1}|Z_{k+l}|>\delta, \bigvee_{i=-m, i\neq l}^{m} a_n^{-1}|Z_{k+i}|\le\delta]}\\
&\quad + \sum_{k=1}^{n-l} f((k+l)n^{-1}, a_n^{-1} Z_{k+l}\cdot \mathbf{e}_l) 1_{[a_n^{-1}|Z_{k+l}|>\delta, \bigvee_{i=-m, i\neq l}^{m} a_n^{-1}|Z_{k+i}|>\delta]}\\
&= I + II + III + IV.
\end{aligned}$$

Now $E(IV) \le nP[a_n^{-1}\mathbf{Z}_k \in (SL)^c](\sup f(x)) \to 0$ and

$$P[I > 0] \le P\left[\bigvee_{k=1-l}^{0} a_n^{-1}|Z_{k+l}| > \delta\right]$$

(where the last inequality follows by considering the support of $f$)

$$= P\left[\bigvee_{i=1}^{l} |Z_i| > a_n\delta\right] \to 0$$

and a similar argument shows $II = o_p(1)$. This clears away the rubble resulting from the reindexing and we focus on $III$. The indicator function associated with $III$ is bounded by $(0 < \eta < \delta)$

$$1_{[a_n^{-1}|Z_{k+l}|>\delta, \bigvee_{i=-m, i\neq l}^{m} a_n^{-1}|Z_{k+i}|\le\eta]} + 1_{[a_n^{-1}|Z_{k+l}|>\eta, \bigvee_{i=-m, i\neq l}^{m} a_n^{-1}|Z_{k+i}|>\eta]}$$

so that the expectation of $III$ is bounded by

$$\begin{aligned}
&\sup\Big\{|f(s,\mathbf{x}) - f(t, x_l\mathbf{e}_l|: |x_l| > \delta,\\
&\qquad \bigvee_{\substack{i=-m\\ i\neq l}}^{m} |x_i| \le \eta, |t-s| \le (2m+1)n^{-1}\Big\} nP[a_n^{-1}|Z_1| > \delta]\\
&\quad + (\text{constant})nP[a_n^{-1}|Z_1| > \eta, a_n^{-1}|Z_2| > \eta].
\end{aligned}$$

Since $f$ is uniformly continuous on its compact support, the sup can be made as small as we like, say less than $\varepsilon\delta^\alpha$, by choosing $\eta$ small and $n$ large. So the bound for III converges as $n \to \infty$ to

$$\varepsilon\delta^\alpha \lim_{n\to\infty} nP[a_n^{-1}|Z_1| > \delta] + 0 = \varepsilon\delta^\alpha\delta^{-\alpha} = \varepsilon$$

and hence (4.81), (4.80), and (4.79) follow.

This means that (4.78) will be true if we show that $I_n^*$ converges weakly to the indicated limit. This follows straightaway from (4.70) and the continuous mapping theorem since $T\colon M_p([0,\infty)\times[-\infty,\infty]\backslash\{0\})\to M_p([0,\infty)\times[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\})$ defined by

$$T\sum_k \varepsilon_{(t_k, j_k)} = \sum_k \sum_{|i|\le m} \varepsilon_{(t_k, j_k \mathbf{e}_i)}$$

is continuous. Thus

$$T\left(\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1} Z_k)}\right) = I_n^* \Rightarrow T\left(\sum_k \varepsilon_{(t_k, j_k)}\right)$$
$$= \sum_k \sum_{|i|\le m} \varepsilon_{(t_k, j_k \mathbf{e}_i)}$$

in $M_p([0,\infty)\times[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\})$ as required. □

Proposition 4.26 is now used to get point processes based on $\{X_n\}$ to converge weakly. Recall that $X_n$ is defined in (4.68) and now define

$$X_n^{(m)} = \sum_{|j|\le m} c_j Z_{n-j}$$

as an approximation to $X_n$. We seek to apply Proposition 3.18 by using the map $(t,\mathbf{z})\in[0,\infty)\times([-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\})\to(t,\sum_{|i|\le m} c_i z_i)\in[0,\infty)\times([-\infty,\infty]\backslash\{0\})$, but there is a difficulty since this map is not well defined if, for example, $z_1 = +\infty$, $z_2 = -\infty$. So define $T\colon [0,\infty)\times([-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\})\to[-\infty,\infty]\backslash\{0\}$ by

$$T(t,z) = \begin{cases} \left(t, \sum_{|i|\le m} c_i z_i\right) & \text{if } \bigvee_{i=-m}^{m} |z_i| < \infty \quad \text{and} \quad \sum_{i=-m}^{m} c_i z_i \ne 0; \\ (t,\infty) & \text{otherwise.} \end{cases}$$

and define $\hat{T}\colon M_p([0,\infty)\times([-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}))\to M_p([0,\infty)\times([-\infty,\infty]\backslash\{0\}))$ by $\hat{T}m = m\circ T^{-1}$. If we can show that $\hat{T}$ is almost surely continuous with respect to the distribution of $\sum_k\sum_{|i|\le m}\varepsilon_{(t_k, j_k\cdot \mathbf{e}_i)}$, then applying the continuous mapping theorem and (4.78) gives the desired

$$\hat{T}\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1}\mathbf{Z}_k)} = \sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1} X_k^{(m)})}$$
$$\Rightarrow \hat{T}\sum_k \sum_{|i|\le m} \varepsilon_{(t_k, j_k\cdot \mathbf{e}_i)} = \sum_k \sum_{|i|\le m} \varepsilon_{(t_k, c_i j_k)}. \tag{4.82}$$

If (4.82) seems obvious and you are eager to get on with the story then skip this paragraph. For those with more patience for details, note that since

$$P\left[\sum_k \sum_{|i|\le m} \varepsilon_{(t_k, j_k\cdot \mathbf{e}_i)}\left([0,\infty)\times\left\{\mathbf{z}\in[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}: \bigvee_{i=-m}^{m} |z_i| = \infty\right\}\right) > 0\right]$$
$$= 0,$$

it suffices to show that $\hat{T}$ is continuous at

$$m_0 \in M_p([0,\infty)\times([-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}))$$

when

$$m_0\left([0,\infty)\times\left\{\mathbf{z}: \bigvee_{i=-m}^{m}|z_i|=\infty\right\}\right)=0.$$

Suppose $f\in C_K^+([0,\infty)\times([-\infty,\infty]\backslash\{0\}))$ and $m_n\xrightarrow{v} m_0$. We want $\hat{T}m_n\xrightarrow{v}\hat{T}m_0$ or what is equivalent $\hat{T}m_n(f)=m_n\circ T^{-1}(f)=m_n(f\circ T)\to m_0(f\circ T)=\hat{T}m_0(f)$. The support of $f$ is contained in a compact set of the form $[0,t]\times\{x\in[-\infty,\infty]\backslash\{0\}: |x|>\delta\}$ and hence the support of $f\circ T=T^{-1}$ (support of $f$) $\subset[0,\infty)\times(\{\mathbf{z}\in[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}: |\sum_{|i|\le m}c_iz_i|>\delta\}\cup\{\mathbf{z}\in[-\infty,\infty]^{2m+1}\backslash\{\mathbf{0}\}: \bigvee_{i=-m}^{m}|z_i|=\infty\})$ so that having already flexed our muscles on a similar problem in (4.76) we conclude that $f\circ T$ has compact support. Further, $f\circ T$ is certainly continuous on $[0,\infty)\times\mathbb{R}^{2m+1}\backslash\{\mathbf{0}\}$. Pick compact $K$ containing the support of $f\circ T$ and such that $m_n(K)=m_0(K)$ for $n$ large and set $P_n=m_n(\cdot)/m_n(K)$, $n\ge 0$. Then for large $n$

$$\hat{T}m_n(f)=m_0(K)\int f\circ T\,dP_n\to m_0(K)\int f\circ T\,dP_0=\hat{T}m_0(f)$$

by the continuous mapping theorem since $f\circ T$ is a.s. continuous with respect to the probability measure $P_0$.

We can now convert (4.82) into the desired point process convergence.

**Proposition 4.27.** *If* (4.65), (4.66), and (4.67) *hold then*

$$\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}X_k)}\Rightarrow\sum_k\sum_{i=-\infty}^{\infty}\varepsilon_{(t_k,c_ij_k)} \tag{4.83}$$

*in* $M_p([0,\infty)\times([-\infty,\infty]\backslash\{0\}))$.

PROOF. Observe that as $m\to\infty$

$$\sum_k\sum_{|i|\le m}\varepsilon_{(t_k,c_ij_k)}\to\sum_k\sum_{i=-\infty}^{\infty}\varepsilon_{(t_k,c_ij_k)}$$

pointwise in the vague metric. This, (4.82), and Lemma 4.25 show that it is enough to prove for any $\xi>0$

$$\lim_{m\to\infty}\limsup_{n\to\infty}P\left[d\left(\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}X_k^{(m)})},\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}X_k)}\right)>\xi\right]=0 \tag{4.84}$$

where $d$ is the vague metric. As in the proof of (4.79), if we take account of the definition of $d$ in Proposition 3.17, we see that it suffices to check for any $f\in C_K^+([0,\infty)\times([-\infty,\infty]\backslash\{0\}))$

$$\lim_{m\to\infty}\limsup_{n\to\infty}P\left[\left|\sum_{k=1}^{\infty}f(kn^{-1},a_n^{-1}X_k^{(m)})-\sum_{k=1}^{\infty}f(kn^{-1},a_n^{-1}X_k)\right|>\xi\right]=0. \tag{4.85}$$

Without loss of generality suppose the compact support of $f$ is contained in $[0,t]\times\{x\in[-\infty,\infty]\backslash\{0\}: |x|>\delta\}$ and for typographical simplicity suppose $t = 1$. Set $\omega(\theta) = \sup\{|f(t,x) - f(t,y)|: x, y \in (0,\infty)$ or $x, y \in (-\infty, 0)$ and $|x - y| \le \theta, 0 \le t \le 1\}$, and since $f$ has compact support it is uniformly continuous and therefore $\omega(\theta) \to 0$ as $\theta \to 0$. Because of the compact support each of the infinite sums $\sum_{k=1}^{\infty}$ in (4.85) can be replaced by $\sum_{k=1}^{n}$. Decompose the probability in (4.85) according to whether

$$\text{LESS} := \left[a_n^{-1} \bigvee_{k=1}^{n} |X_k^{(m)} - X_k| \le \theta\right]$$

or its complement occurs. Note by stationarity

$$\begin{aligned} P[(\text{LESS})^c] &\le nP[|X_k^{(m)} - X_k| > a_n\theta] \\ &= nP\left[\left|\sum_{|j|>m} c_j Z_{k-j}\right| > a_n\theta\right] \\ &\le nP\left[\left|\sum_{|j|>m} |c_j||Z_j|\right| > a_n\theta\right] \end{aligned}$$

and applying Lemma 4.24 this is asymptotic to

$$\sim \theta^{-\alpha} \sum_{|j|>m} |c_j|^{\alpha}.$$

Therefore

$$\begin{aligned} &\lim_{m\to\infty} \limsup_{n\to\infty} P\left[\left|\sum_{k=1}^{n} f(kn^{-1}, a_n^{-1}X_k^{(m)}) - \sum_{k=1}^{n} f(kn^{-1}, a_n^{-1}X_n)\right| > \xi; (\text{LESS})^c\right] \\ &\quad \le \lim_{m\to\infty} \theta^{-\alpha} \sum_{|j|>m} |c_j|^{\alpha} = 0. \end{aligned}$$

On the other hand, on LESS, assuming $\theta < \delta/2$, if $a_n^{-1}|X_k^{(m)}| \le \delta/2$, $f(kn^{-1}, a_n^{-1}|X_k^{(m)}|) = f(kn^{-1}, a_n^{-1}|X_k|) = 0$ and if $a_n^{-1}|X_k^{(m)}| > \delta/2$ we have

$$|f(kn^{-1}, a_n^{-1}|X_k^{(m)}|) - f(kn^{-1}, a_n^{-1}|X_k|)| \le \omega(\theta)$$

and so

$$\begin{aligned} &P\left[\left|\sum_{k=1}^{n} f(kn^{-1}, a_n^{-1}X_k^{(m)}) - \sum_{k=1}^{n} f(kn^{-1}, a_n^{-1}X_k)\right| > \xi; \text{LESS}\right] \\ &\quad \le P\left[\omega(\theta) \sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1}X_k^{(m)})}([0,1] \times \{y: |y| > \delta/2\}) > \xi\right], \end{aligned}$$

and via (4.82) this converges as $n \to \infty$ to

$$P\left[\omega(\theta) \sum_{k=1}^{\infty} \sum_{|i|\le m} \varepsilon_{(t_k, c_i j_k)}([0,1] \times \{y: |y| > \delta/2\}) > \xi\right],$$

and as $m \to \infty$ this converges to

$$P\left[\omega(\theta) \sum_{k=1}^{\infty} \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, c_i j_k)}([0,1] \times \{y: |y| > \delta/2\}) > \xi\right].$$

Since $\sum_{k=1}^{\infty}\sum_{i=-\infty}^{\infty}\varepsilon_{(t_k,c_ij_k)}([0,1]\times\{y:|y|>\delta/2\})<\infty$ a.s. the preceding probability goes to zero as $\theta\to 0$, proving the result. □

We now consider various applications of Proposition 4.27. Set

$$c_+ = \max\{c_j \vee 0: -\infty < j < \infty\}, \qquad c_- = \max\{-c_j \vee 0: -\infty < j < \infty\}.$$

**Proposition 4.28.** *Suppose* (4.65), (4.66), and (4.67) *hold and set*

$$Y_n(t) = \begin{cases} a_n^{-1} \bigvee_{i=1}^{[nt]} X_i & \text{if } t \geq n^{-1} \\ a_n^{-1} X_1 & \text{if } 0 < t < n^{-1}. \end{cases}$$

*If either* $c_+p > 0$ *or* $c_-q > 0$ *and* $Y$ *is an extremal process generated by the extreme value distribution* $\exp\{-(c_+^\alpha p + c_-^\alpha q)x^{-\alpha}\}$ *for* $x > 0$, *then*

$$Y_n \Rightarrow Y$$

*in* $D(0,\infty)$.

Proof. Define $T_1: M_p([0,\infty)\times([-\infty,\infty]\backslash\{0\}))\to D(0,\infty)$ in a manner identical to the $T_1$ of Proposition 4.20 so that if $m=\sum_k \varepsilon_{(u_k,v_k)}$ satisfies

$$m([0,t]\times([-\infty,\infty]\backslash\{0\}))>0 \qquad \text{for all } t>0$$

set

$$T_1 m = \bigvee\{v_k: u_k \leq t\}.$$

As in Proposition 4.20, $T_1$ is almost surely continuous so by the continuous mapping theorem (Section 3.5 or Billingsley, 1968) applied to (4.83)

$$T_1\left(\sum_{k=1}^{\infty}\varepsilon_{(kn^{-1},a_n^{-1}X_k)}\right) = Y_n(\cdot) \Rightarrow T_1\left(\sum_{i=-\infty}^{\infty}\sum_k \varepsilon_{(t_k,c_ij_k)}\right)$$

$$= T_1\left(\sum_{i=-\infty}^{\infty}\sum_k \left(1_{(0,\infty]}(j_k)\varepsilon_{(t_k,c_ij_k)} + 1_{[-\infty,0)}(j_k)\varepsilon_{(t_k,c_ij_k)}\right)\right)$$

and because of the definition of $T_1$ this is the same as

$$= T_1\left(\sum_k \varepsilon_{(t_k,c_+j_k)}(\cdot\cap[0,\infty)\times(0,\infty]) + \sum_k \varepsilon_{(t_k,-c_-j_k)}(\cdot\cap[0,\infty)\times(0,\infty])\right)$$

$$= Y.$$

Note that in the preceding line, $T_1$ operates on the sum of two independent Poisson processes; the first comes from those points $(t_k, j_k)$ with $j_k > 0$. The second process results from considering the influence on the maximum of negative $j$'s multiplied by negative $c$'s. It should be clear that the sum of the two processes is Poisson with mean measure of $(x,\infty]$ equal to $(pc_+^\alpha + qc_-^\alpha)x^{-\alpha}$, $x > 0$. Thus $Y$ is extremal as described. □

Next let $M_n^{(r)}$ be the $r$th largest order statistic from the sample $\{X_1, \ldots, X_n\}$ so that in this notation $M_n^{(1)} = \bigvee_{i=1}^n X_i$. It is important to realize that the methods just discussed can be extended to get joint convergence of the $k$ processes based on $(M_{[nt]}^{(i)}, i \le k)$. One merely needs to note that

$$[a_n^{-1} M_{[nt]}^{(r)} \le x] = \sum_{k=1}^{\infty} \varepsilon_{(k/n, X_k/a_n)}([0, t] \times (x, \infty]) \le r - 1].$$

The joint limiting distribution for any collection of upper extremes can thus be determined through the limiting point process. For example, letting

$$N(\cdot) = \sum_k \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, j_k c_i)}([0, 1] \times \cdot)$$

we have for $0 < y < x$,

$$P(a_n^{-1} M_n \le x, a_n^{-1} M_n^{(2)} \le y) \to P(N((x, \infty)) = 0, N((y, x]) \le 1).$$

For convenience, suppose that $c_- = 0$, define $c_{+2}$ = second largest of $(c_i \vee 0)$ and for $y > 0$ set $G(y) = \exp\{-pc_+^\alpha y^{-\alpha}\}$. Then the preceding limit (neglect $t$'s so that $\sum_k \varepsilon_{j_k}$ is $\mathrm{PRM}(px^{-\alpha-1}dx)$ on $(0, \infty]$) becomes

$$\begin{aligned}
&P\left(\sum_k \varepsilon_{j_k}(y/c_+, x/c_+ \wedge y/c_{+2}) \le 1, \sum_k \varepsilon_{j_k}((x/c_+ \wedge y/c_{+2}, \infty]) = 0\right) \\
&\quad = G(x \wedge (c_+ y/c_{+2}))G(y)/G(x \wedge (c_+ y/c_{+2}))(1 - \log(G(y)/G(x \wedge (c_+ y/c_{+2})))) \\
&\quad = G(y)(1 - \log(G(y)/G(x \wedge (c_+ y/c_{+2})))).
\end{aligned}$$

By choosing $\rho(s) = 1 - s(1 \vee (s^{-1}(c_{+2}/c_+)^\alpha))$ the limit distribution of $a_n^{-1}(M_n^{(1)}, M_n^{(2)})$ may be rewritten as

$$\begin{cases} G(x) & \text{if } x \le y \\ G(y)(1 - \rho(\log G(x)/\log G(y))\log G(y)) & \text{if } x > y \end{cases}$$

(cf. Mori 1976, 1977; Welsch, 1972).

As a result of conditions (4.65) and (4.66) giving control over behavior governed by both tails of the distribution, it is possible to determine joint limiting behavior for any collection of upper and lower extremes by using Proposition (4.27). We shall concentrate on the specific case of the maximum $M_n = \bigvee_{j=1}^n X_j$ and the minimum $W_n = \bigwedge_{j=1}^n X_j$.

**Proposition 4.29.** *Suppose* (4.65), (4.66), *and* (4.67) *all hold. Then we have*

$$P(a_n^{-1} M_n \le x, a_n^{-1} W_n \le y) \to G^p(x, \infty)G^q(\infty, x) - G^p(x, -y)G^q(-y, x)$$

*where*

$$G(x, y) = \begin{cases} \exp\{-c_+^\alpha x^{-\alpha}\} \wedge \exp\{-c_-^\alpha y^{-\alpha}\} & \textit{for } x > 0, y > 0 \\ 0 & \textit{otherwise.} \end{cases}$$

**Proof.** We have for $x > 0$, $y < 0$

$$
\begin{aligned}
&P(a_n^{-1}M_n \le x, a_n^{-1}W_n > y)\\
&= P\left(\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1}X_k)}([0,1] \times ([-\infty, y) \cup (x, \infty])) = 0\right)\\
&\to P\left[\sum_k \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, j_k c_i)}([0,1] \times ([-\infty, y) \cup (x, \infty])) = 0\right]\\
&= P\Bigg[\sum_k \varepsilon_{(t_k, j_k)}([0,1] \times ([-\infty, -x/c_-) \cup (x/c_+, \infty] \cup [-\infty, y/c_+)\\
&\qquad \cup (-y/c_-, \infty])) = 0\Bigg]\\
&= P\Bigg[\sum_k \varepsilon_{(t_k, j_k)}([0,1] \times ([-\infty, (-x/c_-) \vee (y/c_+))\\
&\qquad \cup ((x/c_+) \wedge (-y/c_-), \infty])) = 0\Bigg]\\
&= \exp\{-[p(c_+^\alpha x^{-\alpha} \vee c_-^\alpha(-y)^{-\alpha}) + q(c_-^\alpha x^{-\alpha} \vee c_+^\alpha(-y)^{-\alpha})]\}\\
&= G^p(x, -y)G^q(-y, x).
\end{aligned}
$$

Thus,

$$P(a_n^{-1}M_n \le x, a_n^{-1}W_n \le y) = P(a_n^{-1}M_n \le x) - P(a_n^{-1}M_n \le x, a_n^{-1}W_n > y)$$

has the desired limit. □

Now consider inverses, overshoots, and ranges. It is convenient to modify Proposition 4.27 trivially by substituting the continuous variable $s$ for the discrete variable $n$. Combining the method of Proposition 4.20 and Corollary 4.21 we obtain

$$(Y_s, Y_s^{\leftarrow}) \Rightarrow (Y, Y^{\leftarrow})$$

in $D(0, \infty) \times D(0, \infty)$ where

$$a(s) = (1/P[|Z_1| > \cdot])^{\leftarrow}(s)$$

$$Y_s^{\leftarrow}(x) = \inf\{u\colon Y_s(u) > x\} = \inf\left\{k\colon \bigvee_{i=1}^k X_i > a(s)x\right\}\Big/ s$$

and a similar definition holds for $Y^{\leftarrow}$. So setting $\tau(x) = \inf\{k\colon \bigvee_{i=1}^k X_i > x\}$ we have $\tau(a(s)\cdot)/s \Rightarrow Y^{\leftarrow}$ and changing variables we get

$$(1 - F(s))\tau(s\cdot) \Rightarrow Y^{\leftarrow}(\cdot) \tag{4.86}$$

as $s \to \infty$ in $D(0, \infty)$. Recall for $x > 0$, $t > 0$

$$P[Y(t) \le x] = \exp(-t(pc_+^\alpha + qc_-^\alpha)x^{-\alpha}$$

and therefore

$$P[Y^{\leftarrow}(x) \le t] = 1 - \exp(-(pc_+^\alpha + qc_-^\alpha)x^{-\alpha}t).$$

Now define $L(a(s), 1) = \inf\{k: X_k > a(s)\}$, $L(a(s), 2) = \inf\{k > L(a(s), 1): X_k > X_{L(a(s), 1)}\}$, and so on. Then $X_{L(a(s),k)}/a(s), k \ge 1\}$ are those record values of $(X_k/a(s)\}$ bigger than 1. As in Corollary 4.23 this sequence converges weakly in $\mathbb{R}^\infty$ to the range of $Y$ above 1, which is a Poisson process with mean measure of $(a, b]$ equal to $\alpha \log b/a$ (Proposition 4.8). In particular for $x > 0$

$$\lim_{s\to\infty} P[(X_{L(a(s), 1)}/a(s)) - 1 \le x] = 1 - (1 + x)^{-\alpha}. \tag{4.87}$$

(Of course we may change variables $t = a(s)$ to get the limit distribution for the overshoot past $t$.)

Consider jointly $(\{X_{L(a(s),k)}/a(s), k \ge 1\}, Y_s^{\leftarrow}(1))$ on $\mathbb{R}^\infty \times \mathbb{R}$. By the continuous mapping theorem this converges as $s \to \infty$ weakly to

$$\begin{aligned}&(\{\text{points hit by } Y \text{ above } 1\};\ Y^{\leftarrow}(1))\\ &\quad = (\{\text{times of jumps of } Y^{\leftarrow}(x), x > 1\};\ Y^{\leftarrow}(1)).\end{aligned}$$

Since $Y^{\leftarrow}$ has independent increments (Proposition 4.8)

$$\begin{aligned}&\{\text{times of jumps of } Y^{\leftarrow}(x), x > 1\}\\ &\quad = \{\text{times of jumps of } Y^{\leftarrow}(x) - Y^{\leftarrow}(1), x > 1\}\end{aligned}$$

is independent of $Y^{\leftarrow}(1)$. So, for instance, if we combine (4.86) and (4.87) jointly we get as $s \to \infty$

$$P[(1 - F(s))\tau(s) \le x, (X_{L(s,1)} - s)/s > y] \to P[Y^{\leftarrow}(1) \le x](1 + y)^{-\alpha}$$

for $x > 0, y > 0$.

As a last example we discuss excedances as in Rootzen (1978) and Leadbetter, Lindgren, and Rootzen (1983). Consider the indices when observations $X_k/a_n$ exceed a given level $x > 0$. Suppose as a convenience for this discussion that $|c_j| \le 1$ for all $j$. The sequence of point processes of points with ordinates bigger than $x > 0$ converges as $n \to \infty$; that is,

$$\begin{aligned}&\sum_{k=1}^{\infty} \varepsilon_{(k/n, X_k/a_n)}(\cdot \cap ([0, \infty) \times (x, \infty]))\\ &\quad \Rightarrow \sum_k \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, c_i j_k)}(\cdot \cap ([0, \infty) \times (x, \infty]))\end{aligned}$$

in $M_p([0, \infty) \times (x, \infty])$ from Proposition 4.27 and the fact that the map $m \to m(\cdot \cap ([0, \infty) \times (x, \infty]))$ from $M_p([0, \infty) \times [-\infty, \infty]\backslash\{0\})$ to $M_p([0, \infty) \times (x, \infty])$ is a.s. continuous. To evaluate the structure of the limit consider the following: Let $\{\Gamma_n, n \ge 1\}$ be the points of a homogeneous PRM on $[0, \infty)$ with rate $x^{-\alpha}$. Suppose $(J_k, k \ge 1\}$ are iid on $(x, \infty] \cup [-\infty, -x)$ independent of $\{\Gamma_n\}$ and with common density

$$f(y) = (p\alpha y^{-\alpha-1} 1_{(x,\infty)}(y) + q\alpha(-y)^{-\alpha-1} 1_{(-\infty,-x)}(y))x^\alpha.$$

Then from Proposition 3.8

$$\sum_k \varepsilon_{(t_k, j_k)}(\cdot \cap [0,\infty) \times ((x,\infty] \cup [-\infty,-x))) \stackrel{d}{=} \sum_{k=1}^{\infty} \varepsilon_{(\Gamma_k, J_k)}$$

on $M_p([0,\infty) \times ((x,\infty] \cup [-\infty,-x)))$. Therefore, the weak limit of the preceding point processes is

$$\sum_{i=-\infty}^{\infty} \sum_k \varepsilon_{(t_k, c_i j_k)}(\cdot \cap ([0,\infty) \times (x,\infty]))$$

$$\stackrel{d}{=} \sum_{i=-\infty}^{\infty} \sum_{k=1}^{\infty} \varepsilon_{(\Gamma_k, c_i J_k)}(\cdot \cap ([0,\infty) \times (x,\infty])).$$

Finally define $\xi_k = \#\{c_i: c_i J_k > x\}$ so that $\{\xi_k, k \geq 1\}$ is iid. In the limit the point process of times of excedances is the compound Poisson point process $\sum_{k=1}^{\infty} \xi_k \varepsilon_{\Gamma_k}$ where $\{\xi_k\}$ and $\{\Gamma_k\}$ are independent.

Exercises

4.5.1. In Proposition 4.29 show that the maximum and minimum are asymptotically independent iff all of the $c_j$'s have the same sign (Davis and Resnick, 1985a).

4.5.2. Suppose $\{Z_n, n \geq 1\}$ is iid with common distribution $F \in D(G)$, where $G$ is an extreme value distribution. Set $M_n^{(i)} = i$th largest order statistic from $\{Z_1, \ldots, Z_n\}$. Show $\{M_n^{(i)}, i \leq r\}$ has a limit distribution for every fixed $r$ and find it. Conversely, if for some fixed $r$, $M_n^{(r)}$ has a limit distribution, then $F \in D(G)$ (Smirnov, 1952).

4.5.3. As an extension of Proposition 4.27 show for any positive integer $l$

$$\sum_{k=1}^{\infty} \varepsilon_{(kn^{-1}, a_n^{-1}(X_k, X_{k-1}, \ldots, X_{k-l}))}$$

$$\Rightarrow \sum_k \sum_{i=-\infty}^{\infty} \varepsilon_{(t_k, j_k(c_i, c_{i-1}, \ldots, c_{i-l}))}$$

in $M_p([0,\infty) \times ([-\infty,\infty]^{l+1} \backslash \{\mathbf{0}\}))$.

4.5.4. If (4.65), (4.66), and (4.67) hold and $0 < \alpha < 2$ then in $\mathbb{R}$

$$a_n^{-1}\left(\sum_{i=1}^{n} X_i - n\left(\sum_{j=-\infty}^{\infty} c_j\right) b_n\right) \Rightarrow \left(\sum_{j=-\infty}^{\infty} c_j\right) S$$

as $n \to \infty$ where $S$ has a stable distribution with index $\alpha$. Hint: First establish a corresponding result for $\sum_{i=1}^{n} X_i^{(m)}$ (Davis and Resnick, 1985a).

## 4.6. Independence of $k$-Record Processes

In the discussion of record values presented thus far there are two potential sources of dissatisfaction. First, the most interesting and precise results are restricted to the case where the underlying distribution is continuous. Second, from the statistical point of view, studying only records is inefficient as much information about the distribution tail is thereby wasted. In this section, we do not assume that $F$ is necessarily continuous, and we study the point process of $k$-records, i.e., those observations which have relative rank $k$ upon being observed. Surprisingly, for different values of $k$, these point processes turn out to be iid.

This topic has been an area of recent fruitful research. See Deheuvels (1983), Goldie (1983), Goldie and Rogers (1984), Ignatov (1977, 1978), and Stam (1985). The present discussion is based on a nice simplification due to Wim Vervaat (1986) which uses discretization and the découpage de Lévy discussed in the proof of Proposition 4.20.

We now present more precisely the processes under study. Suppose $\{X_n, n \geq 1\}$ are iid random variables with common distribution function $F$. We do not require that $F$ is continuous but only suppose continuity at $x_0 = \sup\{x: F(x) < 1\}$ since otherwise if $F(x_0-) < 1$ there would be a finite number of records. As in Proposition 4.3 we have need of rank variables $R_n$, $n \geq 1$. Without continuity of $F$ as an assumption, ties among the $X$'s are possible and care must be taken in defining $R_n$. We set for $n = 1, 2, \ldots$

$$R_n = \sum_{k=1}^{n} 1_{[X_k \geq X_n]}$$

so $R_n$ is the (relative) rank of $X_n$ among $X_1, \ldots, X_n$ and equals 1 iff $X_n > \max\{X_1, \ldots, X_{n-1}\}$. Similarly $R_n = 2$ iff there is exactly one value among $X_1, \ldots, X_{n-1}$ at least as large as $X_n$. For each fixed $k \geq 1$ define

$$L_k(0) = 0, \qquad L_k(n+1) = \inf\{i > L_k(n): R_i = k\}$$

for $n \geq 0$ and the times $L_k(1), L_k(2), \ldots$ are indices at which observations with relative rank $k$ occur. The values of these observations, namely

$$\{X_{L_k(n)}, n \geq 1\},$$

are the $k$-record values. Note that with $k = 1$, everything reduces to the definitions given prior to Proposition 4.1.

**Proposition 4.30.** *Suppose $F$ has left endpoint $x_l$, right endpoint $x_0$, $-\infty \leq x_l < x_0 \leq \infty$, and atoms $\mathscr{D} = \{b_1, b_2, \ldots\}$. The point processes of k-records*

$$\{N_k = \sum_{n=1}^{\infty} \varepsilon_{X_{L_k(n)}}, k \geq 1\}$$

*are* iid *random elements of $M_p[x_l, x_0]$. The structure of $N_1$ (and hence of any $N_k$, $k \geq 1$) is as follows: Define the monotone function*

$$A(t) = \int_{[x_l, t]} \frac{F(ds)}{F[s, \infty)}, \qquad x_l \leq t \leq x_0$$

*which has the decomposition*

$$A(t) = A^{(c)}(t) + A^{(d)}(t) = \int_{[x_l, t] \cap \mathscr{D}^c} \frac{F(ds)}{F[s, \infty)} + \sum_{\substack{b \in \mathscr{D} \\ b \leq t}} \frac{F\{b\}}{F[b, \infty)}$$

*so that* $A^{(c)}$ *is the function obtained from* $A$ *by removing the jumps. Then* $N_1$ *is an independent sum of two point processes*

$$N_1 \stackrel{d}{=} N_1^{(c)} + N_1^{(d)}$$

*where* $N_1^{(c)}$ *is* $\mathrm{PRM}(A^{(c)})$ *and*

$$N_1^{(d)} = \sum_{b \in \mathscr{D}} \delta_b \varepsilon_b$$

*and* $\{\delta_b, b \in \mathscr{D}\}$ *are independent Bernoulli random variables with*

$$P[\delta_b = 1] = F\{b\}/F[b, \infty) = 1 - P[\delta_b = 0].$$

Proof. The proof starts with the case that $F$ is purely discrete with atoms $\{b_1, b_2, \ldots\} = \mathscr{D}$ and thus $A^{(c)} \equiv 0$. Define the event

$$\begin{aligned}[b \text{ is a } k\text{-record}] &= \bigcup_{n=1}^{\infty} [X_n = b, R_n = k] \\ &= [b \in \{X_{L_k(n)}, n \geq 1\}]\end{aligned}$$

for each $b \in \mathscr{D}$, $k \geq 1$. Then

$$N_k = \sum_{b \in \mathscr{D}} 1_{[b \text{ is a } k\text{-record}]} \varepsilon_b$$

and we must show that

$$\{1_{[b \text{ is a } k\text{-record}]}, b \in \mathscr{D}\}$$

are independent Bernoulli random variables with $P[b \text{ is a } k\text{-record}] = F\{b\}/F[b, \infty)$ and also that for $k \neq l$

$$\{1_{[b \text{ is a } k\text{-record}]}, b \in \mathscr{D}\} \quad \text{and} \quad \{1_{[b \text{ is an } l\text{-record}]}, b \in \mathscr{D}\}$$

are independent sequences. This is achieved by the découpage: Fix $b', b \in \mathscr{D}$, $b' < b$, and let

$$K^+(0) = 0, \qquad K^+(i) = \inf\{j > K^+(i-1) \colon X_j \geq b\}$$

$$K^-(0) = 0, \qquad K^-(i) = \inf\{j > K^-(i-1) \colon X_j < b\}$$

so that $\{X_{K^+(i)}\}$, $\{X_{K^-(i)}\}$, $\{1_{[X_n \geq b]}, n \geq 1\}$ are independent and

$$P[X_{K^+(i)} = x] = P[X_1 = x \mid X_1 \geq b] = F\{x\}/F[b, \infty)$$

if $x \geq b$. (See the proof of the découpage on page 215.) Observe that

$$P[b \text{ is a } k\text{-record}] = P[X_{K^+(k)} = b] = F\{b\}/F[b, \infty).$$

Furthermore for any $l \geq 1$

$$[b' \text{ is a } l\text{-record}] = \bigcup_{n=1}^{\infty} \left[ X_{K^-(n)} = b', \sum_{i=1}^{n} 1_{[b', b)}(X_{K^-(i)}) + \sum_{i=1}^{K^-(n)} 1_{[X_i \geq b]} = l \right]$$

since the right side expresses the requirement the $b'$ must be exceeded $l$ times, and these excedances are apportioned between $X$'s in $[b', b)$ and those in $[b, \infty)$. So for $b' < b$ the event [$b$ is a $k$-record] is in the $\sigma$-field generated by $\{X_{K^+(i)}\}$, and the event [$b'$ is an $l$-record] is in the $\sigma$-field generated by $\{X_{K^-(i)}\}$, and $\{1_{[X_n \geq b]}, n \geq 1\}$, and thus [$b'$ is an $l$-record] and [$b$ is a $k$-record] are independent. If $b' = b$ and $l \neq k$ then

$$[b \text{ is a } k\text{-record}] = [X_{K^+(k)} = b]$$

$$[b \text{ is an } l\text{-record}] = [X_{K^+(l)} = b]$$

and so independence again is verified.

This proves Proposition 4.30 in the case that $F$ is purely discrete. For the general case where $F$ has a continuous part we set

$$X_l^{(n)} = \sum_{-\infty < i < \infty} 2^{-n} i \, 1_{[2^{-n}i, 2^{-n}(i+1))}(X_l)$$

for $l \geq 1$ so that $\lim_{n\to\infty} \uparrow X_l^{(n)} = X_l$. Define also

$$N_k^{(n)} = \sum_i 1_{[2^{-n}i \text{ is a } k\text{-record}]} \varepsilon_{2^{-n}i} = \sum_i \varepsilon_{X^{(n)}_{L_k^{(n)}(i)}}$$

where $\{L_k^{(n)}(i), i \geq 1\}$ are the indices of $k$-records for $\{X_i^{(n)}, i \geq 1\}$. We seek to prove

$$\lim_{n\to\infty} N_k^{(n)} = N_k. \tag{4.88}$$

By the first part of the proof, $\{N_k^{(n)}, k \geq 1\}$ are iid point processes for each $n$ so that if (4.88) holds then it follows that $\{N_k, k \geq 1\}$ are iid point processes as this property is preserved by taking limits. So it remains to check (4.88) and that $\lim_{n\to\infty} N_1^{(n)}$ has the correct distribution.

To check (4.88) observe that since $X_l^{(n)} \uparrow X_l$ we have as $n \to \infty$

$$1_{[X_k^{(n)} \geq X_i^{(n)}]} \to 1_{[X_k \geq X_i]}$$

and therefore

$$R_i^{(n)} \to R_i;$$

i.e.,

$$\{R_i^{(n)}, i \geq 1\} \to \{R_i, i \geq 1\}$$

in $\mathbb{R}^\infty$. This implies for each $k$

$$\{L_k^{(n)}(i), i \geq 1\} \to \{L_k(i), i \geq 1\}$$

and consequently

$$\{X^{(n)}_{L_k^{(n)}(i)}, i \geq 1\} \to \{X_{L_k(i)}, i \geq 1\}, \tag{4.89}$$

whence by the definition of vague convergence (cf. also Proposition 3.13) we have (4.88). (If (4.89) appears puzzling, consider that $\{L_k^{(n)}(i)\}$ are integer valued so that for any integer $m$ and $n$ sufficiently large we have

$$\{L_k^{(n)}(i), i \le m\} = \{L_k(i), i \le m\}$$

and therefore

$$\{X^{(n)}_{L_k^{(n)}(i)}, i \le m\} = \{X^{(n)}_{L_k(i)}, i \le m\} \to \{X_{L_k(i)}, i \le m\}$$

since $\{X_l^{(n)}, l \ge 1\} \to \{X_l, l \ge 1\}$.)

The last step is to show $N_1^{(n)} \Rightarrow N_1^{(c)} + N_1^{(d)}$ where $N_1^{(c)}$ and $N_1^{(d)}$ are independent. Since $N_1^{(c)} + N_1^{(d)}$ is a simple point process, it is convenient to use the powerful Proposition 3.22, which assures us that it is enough to prove for $c, d \in \mathscr{D}^c$

$$\lim_{n\to\infty} P[N_1^{(n)}(c,d] = 0] = P[N_1^{(c)}(c,d] + N_1^{(d)}(c,d] = 0] \tag{4.90}$$

and

$$\begin{aligned}\lim_{n\to\infty} EN_1^{(n)}(c,d] &= E[N_1^{(c)}(c,d] + EN_1^{(d)}(c,d] \\ &= A^{(c)}(c,d] + \sum_{\substack{c<b\le d \\ b\in\mathscr{D}}} (F\{b\}/F[b,\infty)).\end{aligned} \tag{4.91}$$

Provided that $\mathscr{D} \neq \varnothing$ we may reformulate (4.90) as follows. Define a continuous distribution $F^{(c)}$ and a discrete distribution $F^{(d)}$ to satisfy

$$1 - F(x) = (1 - F^{(d)}(x))(1 - F^{(c)}(x))$$

so that for $b \in \mathscr{D}$

$$F\{b\} = (1 - F^{(c)}(b))F^{(d)}\{b\}$$

and thus

$$F\{b\}/F[b,\infty) = F^{(d)}\{b\}/F^{(d)}[b,\infty). \tag{4.92}$$

Furthermore, one readily checks

$$1_{\mathscr{D}^c}(x)F(dx) = 1_{\mathscr{D}^c}(x)F^{(d)}[x,\infty)F^{(c)}(dx) \tag{4.93}$$

(cf. Exercise 4.6.1) so that

$$\begin{aligned}A^{(c)}(t) &= \int_{[x_l,t]\cap\mathscr{D}^c} F(dx)/F[x,\infty) \\ &= \int_{[x_l,t]\cap\mathscr{D}^c} F^{(d)}[x,\infty)F^{(c)}(dx)/F^{(d)}[x,\infty)F^{(c)}[x,\infty) \\ &= \int_{[x_l,t]\cap\mathscr{D}^c} F^{(c)}(dx)/F^{(c)}[x,\infty) = \int_{[x_l,t]} F^{(c)}(dx)/F^{(c)}[x,\infty)\end{aligned}$$

(since $F^{(c)}$ is continuous) whence

$$A^{(c)}(t) = -\log F^{(c)}(x,\infty). \tag{4.94}$$

Thus

$$P[N_1^{(c)}(c,d] + N_1^{(d)}(c,d] = 0] = P[N_1^{(c)}(c,d] = 0]P[N_1^{(d)}(c,d] = 0]$$
$$= \exp\{-A^c(c,d]\} \prod_{\substack{c<b\le d\\ b\in\mathscr{D}}} (1 - F\{b\}/F[b,\infty))$$

and applying first (4.94) and then (4.92) we get

$$= (F^{(c)}(d,\infty)/F^{(c)}(c,\infty)) \prod_{\substack{c<b\le d\\ b\in\mathscr{D}}} (1 - F^{(d)}\{b\}/F^{(d)}[b,\infty))$$
$$= (F^{(c)}(d,\infty)/F^{(c)}(c,\infty)) \prod_{\substack{c<b\le d\\ b\in\mathscr{D}}} F^{(d)}(b,\infty)/F^{(d)}[b,\infty)$$
$$= (F^{(c)}(d,\infty)/F^{(c)}(c,\infty))(F^{(d)}(d,\infty)/F^{(d)}(c,\infty))$$
$$= F(d,\infty)/F(c,\infty)$$

and the desired reformulation of (4.90) becomes

$$\lim_{n\to\infty} P[N_1^{(n)}(c,d] = 0] = F(d,\infty)/F(c,\infty). \tag{4.90'}$$

To prove (4.90′) we write

$$P[N_1^{(n)}(c,d] = 0] = P\left[\sum_i 1_{[2^{-n}i \text{ is a record of } \{X_i^{(n)}, i\ge 1\}]}\varepsilon_{2^{-n}i}(c,d] = 0\right]$$
$$= \prod_{c<2^{-n}i\le d} P[2^{-n}i \text{ is not a record of } \{X_i^{(n)}, i \ge 1\}]$$
$$= \prod_{c<2^{-n}i\le d} (1 - P[X_1^{(n)} = 2^{-n}i]/P[X_1^{(n)} \ge 2^{-n}i])$$
$$= \prod_{c<2^{-n}i\le d} (1 - F[2^{-n}i, 2^{-n}(i+1))/F[2^{-n}i,\infty))$$
$$= \prod_{2^nc<i\le 2^nd} (F[2^{-n}(i+1),\infty)/F[2^{-n}i,\infty))$$
$$= F[([d2^n]+1)2^{-n},\infty)/F[([c2^n]+1)2^{-n},\infty)$$
$$\to F(d,\infty)/F(c,\infty)$$

as $n\to\infty$, since $c$ and $d\in\mathscr{D}^c$. This proves (4.90′). To verify (4.91) write

$$EN_1^{(n)}(c,d] = \sum_{c<i2^{-n}\le d} P[i2^{-n} \text{ is a record of } \{X_i^{(n)}, i\ge 1\}]$$
$$= \sum_{c<i2^{-n}\le d} P[X_1^{(n)} = i2^{-n}]/PX_i^{(n)} \ge i2^{-n}]$$
$$= \sum_{2^nc<i\le 2^nd} F[i2^{-n},(i+1)2^{-n})/F[i2^{-n},\infty)$$
$$= \sum_{2^nc<i\le 2^nd} F[i2^{-n},(i+1)2^{-n})\cap\mathscr{D})/F[i2^{-n},\infty)$$
$$+ \sum_{2^nc<i\le 2^nd} F([i2^{-n},(i+1)2^{-n})\cap\mathscr{D}^c)/F[i2^{-n},\infty)$$
$$= I + II.$$

Now for $I$ we have

$$\sum_{2^n c < i \le 2^n d} \sum_{\substack{b \in \mathscr{D} \\ i2^{-n} \le b < (i+1)2^{-n}}} F\{b\}/F[i2^{-n}, \infty)$$
$$= \sum_{\substack{([2^n c]+1)2^{-n} \le b < ([2^n d]+1)2^{-n} \\ b \in \mathscr{D}}} F\{b\}/F[[b2^n]2^{-n}, \infty).$$

Since $b \ge [b2^n]2^{-n} \to b$ we have

$$F[b, \infty) \le F[[b2^n]2^{-n}, \infty) \to F[b, \infty)$$

and so remembering that $c$ and $d \in \mathscr{D}^c$ we have

$$I \to \sum_{\substack{c < b \le d \\ b \in \mathscr{D}}} F\{b\}/F[b, \infty)$$

as desired. To deal with $II$ note

$$\sum_{c < i2^{-n} \le d} F([i2^{-n}, (i+1)2^{-n}) \cap \mathscr{D}^c)/F[i2^{-n}, \infty) \to \int_{(c,d] \cap \mathscr{D}^c} F(ds)/F[s, \infty)$$
$$= A^{(c)}(c, d]$$

since

$$\begin{aligned}
0 &\le \sum_{c < i2^{-n} \le d} \int_{(i2^{-n}, (i+1)2^{-n}] \cap \mathscr{D}^c} F(ds)/F[s, \infty) \\
&\quad - \sum_{c < i2^{-n} \le d} F([i2^{-n}, (i+1)2^{-n}) \cap \mathscr{D}^c)/F[i2^{-n}, \infty) \\
&\le \sum_{c < i2^{-n} \le d} F([i2^{-n}, (i+1)2^{-n}) \cap \mathscr{D}^c) |(F[(i+1)2^{-n}, \infty))^{-1} \\
&\quad - (F[i2^{-n}, \infty))^{-1}| \\
&\le \sum_{c < i2^{-n} \le d} F([i2^{-n}, (i+1)2^{-n}) \cap \mathscr{D}^c) F[i2^{-n}, (i+1)2^{-n})/F^2[(i+1)2^{-n}, \infty) \\
&\le \frac{F[([2^n c]+1)2^{-n}, ([2^n d]+1)2^{-n})}{F^2[([2^n d]+1)2^{-n}, \infty)} \sup_{c < i2^{-n} \le d} F([i2^{-n}, (i+1)2^{-n}) \cap \mathscr{D}^c) \\
&\to (\text{constant})\, 0 = 0 \qquad \text{as } n \to \infty.
\end{aligned}$$

The verification of (4.92) is now straightforward and the proof is complete.

## Exercises

4.6.1. Verify (4.93) and (4.94) (cf. Shorrock, 1972; Resnick, 1974; Goldie and Rogers, 1984).

4.6.2. Verify directly

$$P[\text{no record values in } (c, d]] = F(d, \infty)/F(c, \infty).$$

Do also for 2-records.

4.6.3. Verify directly that

$$N_1^{(n)} \Rightarrow N_1^{(c)} + N_1^{(d)}$$

using Laplace functionals. (This will increase your appreciation of Proposition 3.22.)

4.6.4. (a) Let $\{Y(t), t > 0\}$ be extremal-$F$. Formulate and prove the analogue of Proposition 4.30 for the range of $Y$. Prove $Y^{\leftarrow}$ has independent increments.

(b) Let

$$\xi = \sum_k \varepsilon_{(t_k, j_k)}$$

be PRM$(dt \times dv)$ and for each point $(t_k, j_k)$ define a relative rank

$$r(t_k, j_k) = \xi((0, t_k] \times [j_k, \infty)).$$

Define a point process on $(0, \infty)$ by

$$N_l = \sum_k 1_{[r(t_k, j_k) = l]} \varepsilon_{(t_k, j_k)}$$

for $l = 1, 2, \ldots$. Prove $(N_l, l \geq 1)$ are iid and give the distribution (Goldie and Rogers, 1984).

CHAPTER 5

# Multivariate Extremes

We now consider extremes of multivariate data. Let $\{\mathbf{X}_n, n \geq 1\}$ be iid random vectors in $\mathbb{R}^d$. When $d = 1$, concepts such as extreme values, order statistics and record values have natural definitions but when $d > 1$ this is no longer the case as several different concepts of ordering are possible.

Consider the following: In hydrological settings, data may be collected at several sites and there is interest in the maximum flow at each site. Similarly in meteorology, wind speeds impacting different sides of a skyscraper may be recorded and interest lies in modeling the behavior of maximum windspeeds on each side of the building. Bearing these and similar circumstances in mind, it seems that for a random sample from a multivariate distribution, the following definition has some usefulness: The maxima of the sample is the vector of componentwise maxima. Thus if $\mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(d)})$, then for $n \geq 1$ set

$$\mathbf{M}_n = (M_n^{(1)}, \ldots, M_n^{(d)}) = \left( \bigvee_{j=1}^{n} X_j^{(1)}, \ldots, \bigvee_{j=1}^{n} X_j^{(d)} \right).$$

In this chapter we consider various problems related to the asymptotic distribution theory of $\{\mathbf{M}_n\}$.

Some notation: Vectors in $\mathbb{R}^d$ will be denoted by $\mathbf{x} = (x^{(1)}, \ldots, x^{(d)})$ and relations and operations are taken componentwise. Thus for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

$$\begin{aligned} \mathbf{x} < \mathbf{y} \qquad & \text{means } x^{(i)} < y^{(i)}, \qquad 1 \leq i \leq d, \\ \mathbf{x} \leq \mathbf{y} \qquad & \text{means } x^{(i)} \leq y^{(i)}, \qquad 1 \leq i \leq d, \\ \mathbf{x} \vee \mathbf{y} &= (x^{(1)} \vee y^{(1)}, \ldots, x^{(d)} \vee y^{(d)}), \end{aligned}$$

and so on. Rectangles will be denoted by

$$\begin{aligned} (\mathbf{a}, \mathbf{b}] &= \{\mathbf{x} \in \mathbb{R}^d \colon \mathbf{a} < \mathbf{x} \leq \mathbf{b}\} \\ &= \{(x^{(1)}, \ldots, x^{(d)}) \colon a^{(i)} < x^{(i)} \leq b^{(i)}, 1 \leq i \leq d\} \end{aligned}$$

for $\mathbf{a}$ and $\mathbf{b} \in \mathbb{R}^d$ and similarly

$$(-\infty, \mathbf{a}] = \{\mathbf{x} \in \mathbb{R}^d \colon \mathbf{x} \leq \mathbf{a}\}.$$

Later, we will see that there is need also to work in $\bar{\mathbb{R}}^d = [-\infty, \infty]^d$ and so we will require sets of the form

$$\begin{aligned}
[-\infty, \mathbf{a}] &= \{\mathbf{x} \in \bar{\mathbb{R}}^d \colon -\infty \le x^{(i)} \le a^{(i)}, 1 \le i \le d\} \\
(\mathbf{a}, \infty] &= \{\mathbf{x} \in \mathbb{R}^d \colon a^{(i)} < x^{(i)} \le \infty, 1 \le i \le d\} \\
[-\infty, \mathbf{a}]^c &= \bar{\mathbb{R}}^d \backslash [-\infty, \mathbf{a}] \\
&= \{\mathbf{x} \in \bar{\mathbb{R}}^d \colon x^{(i)} > a^{(i)} \text{ for some } i = 1, \dots, d\}
\end{aligned}$$

As in the case of weak limit theory for partial sums, the best understanding of the asymptotic distribution theory for $\{\mathbf{M}_n\}$ comes in the context of an infinite divisibility concept, and this is the topic of the initial sections of Chapter 5. The class of limit distributions for multivariate extremes is characterized in Section 5.4.1, and in 5.4.2 we characterize domains of attraction of these limit distributions. Domain of attraction criteria are based on change of variable techniques and a theory of multivariate regularly varying functions, and the presentation partially parallels the developments in Section 1.2. Section 5.5 considers when limit distributions for maxima are product measures, and Section 5.6 shows that multivariate limit distributions for extremes possess a positive dependence property called *association*.

## 5.1. Max-Infinite Divisibility

The discussion of Chapter 4 has shown that an effective way to study extremes is through extremal processes. What if we attempt to carry out a program in $d > 1$ dimensions paralleling the successful program in one dimension? To construct a multivariate extremal process we could take a multivariate distribution function $F(\mathbf{x})$ and use it to construct finite dimensional distributions as in (4.19). Such a construction requires that $F^t(\mathbf{x})$ be a distribution function for every $t > 0$. Although this is obviously true in case $d = 1$, it is not necessarily valid in $\mathbb{R}^2$ or higher dimensions.

Notational warning: Usually we follow the convention that $F$ stands for the distribution function as well as the measure, and thus we need to emphasize that

$$F^t(\mathbf{x}) \qquad \text{stands for } (F(\mathbf{x}))^t$$

and that for $\mathbf{x} < \mathbf{y}$

$$F^t((\mathbf{x}, \mathbf{y}]) \neq (F(\mathbf{x}, \mathbf{y}])^t.$$

For instance, in $\mathbb{R}^2$ we have

$$F^t((\mathbf{x}, \mathbf{y}]) = F^t(\mathbf{y}) - F^t(x^{(2)}, y^{(1)}) - F^t(x^{(1)}, y^{(2)}) + F^t(\mathbf{x}). \tag{5.1}$$

EXAMPLE. Suppose $\{X_n, n \ge 1\}$ is an iid sequence of $\mathbb{R}$ valued random variables and we wish to study the range

$$R_n = \bigvee_{i=1}^{n} X_i - \bigwedge_{i=1}^{n} X_i.$$

Instead of $R_n$ one could study the joint behavior of

$$\left\{\left(\bigvee_{i=1}^{n} X_i, \bigwedge_{i=1}^{n} X_i\right), n \geq 1\right\}$$

or what is theoretically the same one could study

$$\left\{\left(\bigvee_{i=1}^{n} X_i, -\bigwedge_{i=1}^{n} X_i\right), n \geq 1\right\} = \left\{\left(\bigvee_{i=1}^{n} X_i, \bigvee_{i=1}^{n} (-X_i)\right), n \geq 1\right\}$$
$$= \bigvee_{i=1}^{n} (X_i, -X_i).$$

The joint distribution $F(x^{(1)}, x^{(2)})$ of $(X_1, -X_1)$ concentrates on $\{(x^{(1)}, x^{(2)})$: $x^{(1)} + x^{(2)} = 0\}$ and we show that it is *not* the case that $F^t$ is a distribution for every $t > 0$. If it were then the expression in (5.1) would be non-negative for all $t$. However, take $\mathbf{x} = \mathbf{0}$, $\mathbf{y} = (1, 1)$ and observe that

$$F(0, 1) = F\{(u^{(1)}, u^{(2)}): u^{(1)} + u^{(2)} = 0, 0 < u^{(2)} \leq 1\} = p_1$$

say and

$$F(1, 0) = F\{(u^{(1)}, u^{(2)}): u^{(1)} + u^{(2)} = 0, 0 < u^{(1)} \leq 1\} = p_2$$

say, so that if $\mathbf{1} = (1, 1)$

$$F(\mathbf{1}) = p_1 + p_2.$$

Then $F^t((\mathbf{0}, \mathbf{1}]) \geq 0$ requires from (5.1)

$$(p_1 + p_2)^t \geq p_1^t + p_2^t$$

which need not be the case for $t < 1$.

Thus not every distribution function $F$ on $\mathbb{R}^d$ for $d > 1$ has the property that $F^t$ is a distribution. Those which do are called *max-infinitely divisible*. In accordance with tradition the definition is formulated as follows.

**Definition.** The distribution function $F$ on $\mathbb{R}^d$ is max-infinitely divisible (max-id) if for every $n$ there exists a distribution $F_n$ on $\mathbb{R}^d$ such that

$$F = F_n^n;$$

i.e., $F^{1/n}$ is a distribution. For the sake of brevity, a random vector with max-id distribution will be called max-id.

**Proposition 5.1.** *Suppose that for $n \geq 0$ $F_n$ are probability distribution functions on $\mathbb{R}^d$. If*

$$F_n^n \to F_0$$

*weakly (pointwise convergence at continuity points of $F_0$) then $F_0$ is max-id. Consequently*
(a) *$F$ is max-id* iff *$F^t$ is a distribution function for every $t > 0$;*
(b) *The class of max-id distributions is closed in $\mathbb{R}^d$ with respect to weak convergence: If $G_n$ are max-id distributions converging weakly to a distribution $G_0$, then $G_0$ is max-id.*

PROOF. Suppose $F_n^n \to F_0$ weakly and that $\mathbf{x}$ is a continuity point of $F_0$. We show for every $t > 0$ that $F_n^{[nt]}(\mathbf{x}) \to F_0^t(\mathbf{x})$. If $F_0(\mathbf{x}) = 0$ then

$$F_n^{[nt]}(\mathbf{x}) = (F_n^n(\mathbf{x}))^{[nt]/n} \to 0 = F_0^t(\mathbf{x}).$$

If $F_0(\mathbf{x}) > 0$ then $F_n(\mathbf{x}) \to 1$ and

$$\begin{aligned} -\log F_n^{[nt]}(\mathbf{x}) &= [nt](-\log F_n(\mathbf{x})) \sim nt(-\log F_n(\mathbf{x})) \\ &= t(-\log F_n^n(\mathbf{x})) \to t(-\log F_0(\mathbf{x})) = -\log F_0^t(\mathbf{x}). \end{aligned}$$

Thus $F_n^{[nt]}(\mathbf{x}) \to F_0^t(\mathbf{x})$ whence $F_0^t$ is a distribution function. Hence $F_0^{1/n}$ is a distribution for any $n$ and $F_0$ is max-id.

The proof of (a) is now clear from the foregoing paragraph. As for (b) merely observe that $G_0 = \lim G_n = \lim(G_n^{1/n})^n$ in order to conclude that $G_0$ is max-id. $\square$

So starting with a multivariate distribution $F$, we may generate an extremal process whose finite dimensional distributions are given as in (4.19), provided $F$ is max-id.

The following proposition collects some further simple properties.

**Proposition 5.2.** (i) *If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent max-id random vectors in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively, then $(\mathbf{X}_1, \mathbf{X}_2)$ is max-id in $\mathbb{R}^{d_1+d_2}$. In particular if $X_1, \ldots, X_d$ are independent random variables, then $(X_1, \ldots, X_d)$ is max-id in $\mathbb{R}^d$.*
(ii) *Products of max-id distributions on $\mathbb{R}^d$ are max-id on $\mathbb{R}^d$. If $\mathbf{X}$, $\mathbf{Y}$ are independent max-id random vectors in $\mathbb{R}^d$, then $\mathbf{X} \vee \mathbf{Y}$ is max-id.*
(iii) *If $\mathbf{X} = \{X^{(1)}, \ldots, X^{(d)})$ is max-id and $f_i\colon \mathbb{R} \to \mathbb{R}$ is nondecreasing for $1 \le i \le d$, then $(f_1(X^{(1)}), \ldots, f_d(X^d))$ is max-id in $\mathbb{R}^d$.*

The proof of this proposition is Exercise 5.1.3. For (iii) one can save work by considering the extremal process $(\mathbf{Y}(t), t > 0)$ such that $\mathbf{Y}(1) \stackrel{d}{=} \mathbf{X}$.

The analogue of (iii) where we assume $f_i\colon \mathbb{R}^d \to \mathbb{R}$ is much harder and it is not clear what would be sensible conditions which would guarantee $(f_1(\mathbf{X}), \ldots, f_d(\mathbf{X}))$ be max-id when we start with the vector $\mathbf{X}$ max-id. Such an analogue would be helpful in deciding, for example, when the multivariate normal (see Section 5.2) is max-id; we could take $\mathbf{X}$ to be a $d$-dimensional vector of iid normals.

A general criterion for a distribution $F$ to be max-id is presented in Section 5.3. When $F$ is absolutely continuous on $\mathbb{R}^2$, there is the following criterion. Although the given criterion is not simple to apply as it involves more than the density, it is sufficient for resolving in Section 5.2 whether the bivariate normal is max-id. The generalization of Proposition 5.3 to dimensions $d > 2$ is given in Exercise 5.1.4.

In what follows we represent partial derivatives by subscripts; viz

$$F_x = \frac{\partial}{\partial x}F, \qquad F_y = \frac{\partial}{\partial y}F, \qquad F_{x,y} = \frac{\partial^2}{\partial y\,\partial x}F,$$

and so on. The notation $[F > 0]$ means the set $\{\mathbf{x}: F(\mathbf{x}) > 0\}$.

**Proposition 5.3.** *Let $F$ be a distribution on $\mathbb{R}^2$ with continuous density $F_{x,y}$. Then $F$ is max id* iff *$Q := -\log F$ satisfies*

$$Q_{x,y} \le 0 \qquad \text{on } [F > 0]$$

*or equivalently* iff

$$F_x F_y \le F_{xy} F \qquad \text{a.s. on } \mathbb{R}^2.$$

Proof. Since $F^t = e^{-tQ}$ we have on (the open) set $[F > 0]$

$$\begin{aligned}\frac{\partial}{\partial y}\frac{\partial}{\partial x}F^t &= \frac{\partial}{\partial y}(-te^{-tQ}Q_x) = \frac{-\partial}{\partial y}(tF^tQ_x)\\ &= -t(Q_{x,y}F^t - tF^tQ_yQ_x)\\ &= tF^t(tQ_xQ_y - Q_{x,y})\end{aligned}$$

and $F$ is max id iff this latter expression is non-negative for all $t$, as occurs iff

$$tQ_xQ_y - Q_{x,y} \ge 0 \tag{5.2}$$

for all $t$. Since $Q_x = -F_x/F \le 0$ and $Q_y \le 0$ we have (5.2) holds for all $t$ iff $Q_{x,y} \le 0$ as asserted. The rest follows by differentiation:

$$\begin{aligned}0 \ge Q_{x,y} &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}(-\log F) = \frac{-\partial}{\partial x}(F_y/F)\\ &= -(FF_{x,y} - F_yF_x)/F^2.\end{aligned}$$

□

Note that the condition $Q_{x,y} \le 0$ suggests that $Q$ is the tail of a measure, and this is discussed in connection with the representation theorem in Section 5.3.

## Exercises

5.1.1. Is the following distribution on $\mathbb{R}^2$ max-id:

$$F((0,0)) = F((0,1)) = F((1,0)) = 1/3?$$

What about the distribution $F$ on $\mathbb{R}^2$ defined by

$$F((0,0)) = F((0,1)) = F((1,0)) = F((1,1)) = 1/4?$$

5.1.2. Check that the uniform distribution on $[0,1]^d$ is max-id on $\mathbb{R}^d$.

5.1.3. Prove the assertions of Proposition 5.2.

5.1.4. Suppose $F$ is an absolutely continuous distribution on $\mathbb{R}^d$ with continuous density. Prove that $F$ is max-id iff for each $1 \le l \le d$

$$Q_{x^{(i_1)}, x^{(i_2)}, \ldots, x^{(i_l)}} \le 0$$

a.e. on $[F > 0]$ for all $1 \le i_1 < \cdots < i_l \le d$.

5.1.5. If $F$ is max-id on $\mathbb{R}^d$ and $l \in \mathbb{R}^d$ show

$$F(\mathbf{x}) 1_{[l, \infty]}(\mathbf{x})$$

is still max-id.

## 5.2. An Example: The Bivariate Normal

In this section we exhibit an example worked out by Dr. A.A. Balkema which shows that the bivariate normal distribution is max-id iff the correlation is non-negative.

The verification requires two lemmas. Let $N$ and $n$ be the univariate standard normal distribution and density, respectively.

**Lemma 5.4.** *The function $n/N$ is strictly decreasing on $\mathbb{R}$; i.e.,* $\log N$ *is strictly concave.*

Proof. Observe that

$$(n/N)' = (Nn' - n^2)/N^2$$

so it suffices to check

$$Nn' < n^2;$$

since $n'(x) = -xn(x)$ we need

$$-xN(x) < n(x).$$

When $x$ is positive this is clear. When $x$ is negative we need

$$N(-|x|) = 1 - N(|x|) < n(-x)/|x| = n(|x|)/|x|$$

and so we want for $y > 0$

$$1 - N(y) < n(y)/y. \tag{5.3}$$

As in the proof of Mills' ratio (Feller, 1968) note that

$$\begin{aligned}(x^{-1}n(x))' &= -x^{-2}n(x) + x^{-1}n'(x) = -x^{-2}n(x) - n(x) \\ &= -n(x)(x^{-2} + 1) < -n(x).\end{aligned}$$

Therefore for $y > 0$

$$1 - N(y) = \int_y^\infty n(u)du < \int_y^\infty -(u^{-1}n(u))'du$$
$$= y^{-1}n(y). \qquad \square$$

**Lemma 5.5.** *Suppose that $f$ and $g$ are strictly positive functions on* $[0, \infty)$ *and that $\mu$ is a measure on* $[0, \infty)$.

(i) *If $f/g$ is decreasing and $\int f\,d\mu < \infty$ then*

$$\int f\,d\mu \Big/ \int g\,d\mu \leq f(0)/g(0).$$

(ii) *If $f/g$ is increasing and $\int g\,d\mu < \infty$ then*

$$\int f\,d\mu \Big/ \int g\,d\mu \geq f(0)/g(0).$$

*The inequalities in* (i) *and* (ii) *are strict unless*

$$f/g = f(0)/g(0) \qquad \mu - \text{a.e.}$$

PROOF. (i) Suppose $\int g\,d\mu = k < \infty$ (otherwise there is nothing to prove). Define

$$d\nu := k^{-1}g\,d\mu$$

so that $\nu$ is a probability measure on $[0, \infty)$. Then

$$\int f\,d\mu \Big/ \int g\,d\mu = \int fk^{-1}\,d\mu = \int (f/g)k^{-1}g\,d\mu$$
$$= \int (f/g)d\nu \leq (f(0)/g(0))\nu[0, \infty) = f(0)/g(0).$$

(ii) Take reciprocals. $\square$

Now we wish to check that any bivariate normal distribution is max-id iff the correlation is non-negative. In view of Proposition 5.2(iii) it suffices to let $(U, V)$ be iid $N(0, 1)$ random variables and then prove the pair

$$(U + cV, V)$$

is max-id iff $c \geq 0$. We have the following formulas

$$F(x, y) = P[U + cV \leq x, V \leq y]$$
$$= \int_{-\infty}^{y} N(x - cv)n(v)dv$$
$$F_x(x, y) = \int_{-\infty}^{y} n(x - cv)n(v)dv$$
$$F_y(x, y) = N(x - cy)n(y)$$
$$F_{x,y} = n(x - cy)n(y).$$

From Proposition 5.3, $F$ is max-id iff

$$F_x F_y \le F_{x,y} F;$$

i.e.,

$$F_x/F \le F_{x,y}/F_y,$$

and in the present example this is

$$\frac{\int_{-\infty}^{y} n(x-cv)n(v)dv}{\int_{-\infty}^{y} N(x-cv)n(v)dv} \le \frac{n(x-cy)n(y)}{N(x-cy)n(y)} = \frac{n(x-cy)}{N(x-cy)}. \tag{5.4}$$

In the integrals on the left set $s = -v + y$ and (5.4) is equivalent to

$$\frac{\int_0^\infty n(x-cy+cs)\mu(ds)}{\int_0^\infty N(x-cy+cs)\mu(ds)} \le \frac{n(x-cy)}{N(x-cy)} \tag{5.5}$$

where we consider $x$ and $y$ fixed and $\mu(ds) = n(y-s)ds$. We now apply the Lemmas 5.4 and 5.5. If $c > 0$ then for fixed $x$, $y$

$$n(x-cy+cs)/N(x-cy+cs) \tag{5.6}$$

is strictly decreasing and 5.5 is true by Lemma 5.5(i). If $c < 0$ the ratio in 5.6 is strictly increasing and so (5.5) fails. If $c = 0$, max-infinite divisibility follows because $F$ is a product measure.

## 5.3. Characterizing Max-id Distributions

We now present a discussion which culminates in a characterization of max-id distributions. The presentation incorporates improvements by Gerritse (1986) into the approach given in Balkema and Resnick (1977). We begin with two examples.

EXAMPLE 5.6 (Compound Poisson Analogue). Let $\{\mathbf{U}_n, n \ge 1\}$ be iid $\mathbb{R}^d$ valued random vectors bounded below; i.e., suppose there exists $\boldsymbol{l} \in \mathbb{R}^d$ and $\mathbf{U}_i \ge \boldsymbol{l}$ a.s. Let $\{N(t), t \ge 0\}$ be homogeneous PRM on $[0,\infty)$ so that $EN(t) = t$ and suppose $\{N(t), t \ge 0\}$ is independent of $\{\mathbf{U}_n, n \ge 1\}$. Define

$$\mathbf{Y}(t) = \bigvee_{i=0}^{N(t)} \mathbf{U}_i$$

where by definition $\mathbf{U}_0 \equiv \boldsymbol{l}$. Then $\mathbf{Y}(t) \ge \boldsymbol{l}$ and for $\mathbf{x} \ge \boldsymbol{l}$ we have

$$\begin{aligned} P[\mathbf{Y}(t) \le \mathbf{x}] &= e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} P\left[\bigvee_{i=0}^{n} \mathbf{U}_i \le \mathbf{x}\right] \\ &= e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} P^n[\mathbf{U}_1 \le \mathbf{x}] = \exp\{-t(1 - P[\mathbf{U}_1 \le \mathbf{x}])\} \\ &= \exp\{-tP([\mathbf{U}_1 \le \mathbf{x}]^c)\}. \end{aligned}$$

Since $\mathbf{Y}(t)$ is a random vector for each $t > 0$, we get $\exp\{-tP([\mathbf{U}_1 \le \mathbf{x}]^c)\}$ is a distribution for each $t > 0$ and hence

$$F(\mathbf{x}) := \exp\{-P([\mathbf{U}_1 \leq \mathbf{x}]^c)\}$$

is max-id. Note that

$$F(\mathbf{l}) = \exp\{-(1 - P[\mathbf{U}_1 = \mathbf{l}])\} > 0.$$

EXAMPLE 5.7. Suppose $\mathbf{l} \in [-\infty, \infty)^d$ and consider $E = [\mathbf{l}, \mathbf{\infty}] \backslash \{\mathbf{l}\}$. As in Section 4.5, this is the compact set $[\mathbf{l}, \mathbf{\infty}]$ punctured by the removal of one point. Compact subsets of $E$ are those closed sets bounded away from $\mathbf{l}$. Suppose $\mu$ is a Radon measure on $E$ and let

$$\xi = \sum_k \varepsilon_{(t_k, \mathbf{j}_k)}$$

be PRM on $[0, \infty) \times E$ with mean measure $dt \times d\mu$. Define for $t > 0$

$$\mathbf{Y}(t) = \left(\bigvee_{t_k \leq t} \mathbf{j}_k\right) \bigvee \mathbf{l}.$$

Then for $\mathbf{y} \in \mathbb{R}^d$, $\mathbf{y} \geq \mathbf{l}$, and $[-\mathbf{\infty}, \mathbf{y}]^c := E \backslash [-\mathbf{\infty}, \mathbf{y}]$

$$\begin{aligned} P[\mathbf{Y}(t) \leq \mathbf{y}] &= P[N([0, t] \times [-\mathbf{\infty}, \mathbf{y}]^c) = 0] \\ &= \exp\{-t\mu([-\mathbf{\infty}, \mathbf{y}]^c)\}. \end{aligned}$$

Provided that $\mathbf{Y}(1)$ is $\mathbb{R}^d$ valued we have shown that

$$F(\mathbf{y}) = \begin{cases} \exp\{-\mu([-\mathbf{\infty}, \mathbf{y}]^c)\} & \mathbf{y} \geq \mathbf{l} \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

is max-id. To check $\mathbf{Y}(1)$ is $\mathbb{R}^d$ valued we need

$$F(\mathbf{\infty}) = 1 \tag{5.8}$$

$$\begin{aligned} F(-\infty, x^{(1)}, \ldots, x^{(d-1)}) &= F(x^{(1)}, -\infty, \ldots, x^{(d-1)}) \\ &= \cdots = F(x^{(1)}, \ldots, x^{(d-1)}, -\infty) = 0 \end{aligned} \tag{5.9}$$

for all $(x^{(1)}, \ldots, x^{(d-1)}) \in \mathbb{R}^{d-1}$. If $\mathbf{y} > \mathbf{l}$, $\mu$ Radon entails $\mu([-\mathbf{\infty}, \mathbf{y}]^c) < \infty$ and (5.8) is equivalent to $\lim_{\mathbf{y} \to \mathbf{\infty}} \mu([-\mathbf{\infty}, \mathbf{y}]^c) = 0$, and this means that $\mu$ places no mass on lines through $\mathbf{\infty}$; i.e., (5.8) is the same as

$$\mu(E \backslash [-\infty, \infty)^d) = \mu\left(\bigcup_{i=1}^{d} \{\mathbf{y} \in E \colon y^{(i)}) = \infty\}\right) = 0. \tag{5.8$'$}$$

For (5.9) we observe that if $\mathbf{l} > -\mathbf{\infty}$ then (5.9) is obvious by the definition of $F$ in (5.7). Example 5.6 exemplifies this situation. In the contrary case, where $l^{(i)} = -\infty$ for some $i = 1, \ldots, d$, if $\mathbf{x} \geq \mathbf{l}$ and $x^{(i)} = -\infty$ we need $\mu([-\mathbf{\infty}, \mathbf{x}]^c) = \infty$. So (5.9) is equivalent to

$$\begin{gathered} \text{Either } \mathbf{l} > -\mathbf{\infty} \quad \text{or} \quad \mathbf{x} \geq \mathbf{l} \quad \text{and} \quad x^{(i)} = -\infty \\ \text{for some } i \leq d \text{ implies } \mu([-\mathbf{\infty}, \mathbf{x}]^c) = \infty. \end{gathered} \tag{5.9$'$}$$

For example, if $d = 2$ and $\boldsymbol{l} = -\infty$, (5.9′) becomes

$$\mu([-\infty, \infty] \times (-\infty, \infty]) = \mu((-\infty, \infty] \times [-\infty, \infty]) = \infty. \quad (5.9'')$$

Observe that for distributions constructed according to (5.7), $F(\mathbf{x}) > 0$ if $\mathbf{x} > \boldsymbol{l}$; this follows because $\mu$ is Radon and $[-\infty, \mathbf{x}]^c$ is relatively compact in $E$. The measure $\mu$ in (5.7) satisfying (5.8′) and (5.9′) is called the *exponent measure*. The construction leading to (5.7) has been designed to allow $\mu$ to vanish on the interior of $[\boldsymbol{l}, \infty]$. Such a possibility must be allowed in order to be able to construct max-id distributions which are product measures. For instance, suppose that

$$F(\mathbf{x}) = \prod_{i=1}^{d} \Lambda(x^{(i)})$$

where as usual $\Lambda(x) = \exp\{-e^{-x}\}$, $x \in \mathbb{R}$. Define $\mu$ by

$$\mu\left(\bigcup_{1 \le i \le j \le d} \{\mathbf{y} \in [-\infty, \infty]^d \setminus \{-\infty\} : y^{(i)} > -\infty, y^{(j)} > -\infty\}\right) = 0$$

$$\begin{aligned} \mu((x, \infty] \times \{-\infty\} \times \cdots \times \{-\infty\}) &= \cdots \\ &= \mu(\{-\infty\} \times \cdots \times \{-\infty\} \times (x, \infty]) \\ &= e^{-x}, \qquad x \in \mathbb{R} \end{aligned}$$

so that $\mu$ concentrates on the lines through $-\infty$. Thus for $\mathbf{x} \in \mathbb{R}^d$

$$\begin{aligned} \mu([-\infty, \mathbf{x}]^c) &= \mu\left(\bigcup_{i=1}^{d} \{\mathbf{y}: y^{(i)} > x^{(i)}\}\right) \\ &= \mu\left(\bigcup_{i=1}^{d} \{(-\infty, \ldots, -\infty, y^{(i)}, -\infty, \ldots, -\infty): y^{(i)} > x^{(i)}\}\right) \\ &= \sum_{i=1}^{d} \mu\{-\infty, \ldots, -\infty, y^{(i)}, -\infty, \ldots, -\infty): y^{(i)} > x^{(i)}\} \\ &= \sum_{i=1}^{d} e^{-x^{(i)}}. \end{aligned}$$

Therefore,

$$\begin{aligned} F(\mathbf{x}) = \exp\{-\mu([-\infty, \mathbf{x}]^c)\} &= \exp\left\{-\sum_{i=1}^{d} e^{-x^{(i)}}\right\} \\ &= \prod_{i=1}^{d} \Lambda(x^{(i)}). \end{aligned}$$

In this case the points of the PRM live on the lines through $-\infty$.

The situation in Example 5.7 is one-half the characterization of max-id distributions.

**Proposition 5.8.** *The following are equivalent*:
(i) *$F$ is max-id.*
(ii) *For some $\boldsymbol{l} \in [-\infty, \infty)^d$, there exists an exponent measure $\mu$ on $E := [\boldsymbol{l}, \boldsymbol{\infty}] \backslash \{\boldsymbol{l}\}$ ($\mu$ is Radon and satisfies* (5.8′) *and* (5.9′)) *such that* (5.7) *holds.*
(iii) *For some $\boldsymbol{l} \in [-\infty, \infty)^d$, there exists an exponent measure $\mu$ on $E := [\boldsymbol{l}, \boldsymbol{\infty}] \backslash \{\boldsymbol{l}\}$ and* PRM*($dt \times d\mu$)*

$$\xi = \sum_k \varepsilon_{(t_k, \mathbf{j}_k)}$$

*on $[0, \infty) \times E$ such that*

$$F(\mathbf{y}) = P\left[\bigvee_{t_k \leq 1} \mathbf{j}_k \vee \boldsymbol{l} \leq \mathbf{y}\right].$$

*Remark.* We will show that if $F$ is max-id then $[F > 0] \subset \mathbb{R}^d$ is a rectangle of the form $A_1 \times \cdots \times A_d$ where $A_i = [l^{(i)}, \infty)$ or $(l^{(i)}, \infty)$ and $\boldsymbol{l} = (l^{(1)}, \ldots, l^{(d)}) = \inf[F > 0]$. Note that $F$ is a probability distribution on $\mathbb{R}^d$ and $\mu$ is defined on $[\boldsymbol{l}, \boldsymbol{\infty}] \backslash \{\boldsymbol{l}\} = E \subset [-\infty, \infty]^d$.

PROOF. It suffices to show (i) implies (ii). We start by showing that if $F$ is max-id, then $[F > 0]$ is a rectangle. To do this we need to verify two properties of $[F > 0]$:

$$\mathbf{x} \in [F > 0] \quad \text{and} \quad \mathbf{x} \leq \mathbf{y} \qquad \text{implies } \mathbf{y} \in [F > 0] \tag{5.10}$$

$$\mathbf{x}, \mathbf{y} \in [F > 0] \qquad \text{implies } \mathbf{x} \wedge \mathbf{y} \in [F > 0]. \tag{5.11}$$

The first is obvious, and for the second it suffices to show that

$$F(\mathbf{x} \wedge \mathbf{y}) \geq F(\mathbf{x})F(\mathbf{y})$$

or equivalently

$$Q(\mathbf{x} \wedge \mathbf{y}) \leq Q(\mathbf{x}) + Q(\mathbf{y}) \tag{5.12}$$

(recall $Q = -\log F$). However, suppose that $\{\mathbf{Y}(t), t > 0\}$ is extremal-$F$. Then

$$\begin{aligned}
F^{n^{-1}}(\mathbf{x}) &= P[\mathbf{Y}(n^{-1}) \leq \mathbf{x}] \\
&= P[\mathbf{Y}(n^{-1}) \leq \mathbf{x}, \mathbf{Y}(n^{-1}) \leq \mathbf{y}] + P([\mathbf{Y}(n^{-1}) \leq \mathbf{x}] \cap [\mathbf{Y}(n^{-1}) \leq \mathbf{y}]^c) \\
&\leq P[\mathbf{Y}(n^{-1}) \leq \mathbf{x} \wedge \mathbf{y}] + P([\mathbf{Y}(n^{-1}) \leq \mathbf{y}]^c) \\
&= F^{n^{-1}}(\mathbf{x} \wedge \mathbf{y}) + 1 - F^{n^{-1}}(\mathbf{y})
\end{aligned}$$

and therefore

$$n(1 - F^{n^{-1}}(\mathbf{x} \wedge \mathbf{y})) \leq n(1 - F^{n^{-1}}(\mathbf{x})) + n(1 - F^{n^{-1}}(\mathbf{y})). \tag{5.13}$$

For fixed $\mathbf{x} \in [F > 0]$ we have as $n \to \infty$

$$n(1 - F^{n^{-1}}(\mathbf{x})) \sim -n \log F^{n^{-1}}(\mathbf{x}) = Q(\mathbf{x})$$

and letting $n \to \infty$ in (5.13) gives the desired (5.12).

Based on (5.10) and (5.11) the verification that $[F > 0]$ is a rectangle can proceed: Define projection maps as usual by

$$\pi_i \mathbf{x} = x^{(i)}, \qquad i = 1, \dots, d$$

for $x \in \mathbb{R}^d$. We assert

$$[F > 0] = \pi_1[F > 0] \times \cdots \times \pi_d[F > 0] =: \bigtimes_{i=1}^{d} \pi_i[F > 0] \tag{5.14}$$

and from this the result directly follows since $\pi_i[F > 0]$ is an interval of the form $(l^{(i)}, \infty)$ or $[l^{(i)}, \infty)$ by (5.10). Half the proof of (5.14) is easy, for if $\mathbf{x} \in [F > 0]$ then of course $x^{(i)} \in \pi_i[F > 0]$ implying $\mathbf{x} \in \bigtimes_{i=1}^{d} \pi_i[F > 0]$. Conversely suppose that $x \in \bigtimes_{i=1}^{d} \pi_i[F > 0]$ so that for $i = 1, \dots, d$, $x^{(i)} \in \pi_i[F > 0]$ and thus there exists $\mathbf{y}_i \in [F > 0]$ with $\pi_i \mathbf{y}_i = x^{(i)}$. Since $\mathbf{y}_i \in [F > 0]$ we get from (5.11) that $\mathbf{y} := \bigwedge_{i=1}^{d} \mathbf{y}_i \in [F > 0]$. However, $\pi_i \mathbf{y} \le \pi_i \mathbf{y}_i = x^{(i)}$, and thus $\mathbf{y} \le \mathbf{x}$ and therefore by (5.10) we get $\mathbf{x} \in [F > 0]$. Thus (5.14) is verified.

With $\boldsymbol{l} = \inf[F > 0]$ consider $E = [\boldsymbol{l}, \boldsymbol{\infty}] \setminus \{\boldsymbol{l}\}$ and define on $E$ measures

$$\mu_n := nF^{n^{-1}}.$$

Since $F^{n^{-1}}$ is only defined on $\mathbb{R}^d$ one must extend the definition of $F^{n^{-1}}$ in the obvious way to $[-\infty, \infty]^d$ in order to get $\mu_n$ defined on $E$. Sets of the form $[-\boldsymbol{\infty}, \mathbf{x}]^c = E \setminus [-\boldsymbol{\infty}, \mathbf{x}]$ for $\mathbf{x} > \boldsymbol{l}$ are relatively compact subsets of $E$ and as $n \to \infty$

$$\mu_n([-\boldsymbol{\infty}, \mathbf{x}]^c) = n(1 - F^{n^{-1}}(\mathbf{x})) \to Q(\mathbf{x}) < \infty,$$

so that for such $\mathbf{x} > \boldsymbol{l}$

$$\sup_n \mu_n([-\boldsymbol{\infty}, \mathbf{x}])^c < \infty.$$

Since $E = \lim_{\mathbf{x} \downarrow \boldsymbol{l}, \mathbf{x} > \boldsymbol{l}} [-\infty, \mathbf{x}]^c$ it follows that for any relatively compact subset $B$ of $E$

$$\sup_n \mu_n(B) < \infty$$

and hence $\{\mu_n\}$ is vaguely relatively compact by Proposition 3.16. Let $\mu_{\mathrm{I}}$ and $\mu_{\mathrm{II}}$ be two vague limit points of $\{\mu_n\}$. Then for any $\mathbf{x} > \boldsymbol{l}$

$$\mu_{\mathrm{I}}([-\boldsymbol{\infty}, \mathbf{x}]^c) = \mu_{\mathrm{II}}([-\boldsymbol{\infty}, \mathbf{x}]^c) = Q(\mathbf{x}) = -\log F(\mathbf{x})$$

and thus

$$\mu_{\mathrm{I}} = \mu_{\mathrm{II}}.$$

So all limits points of $\{\mu_n\}$ are equal and hence there is a limit measure $\mu$ on $E$ with

$$\mu_n \xrightarrow{v} \mu.$$

Thus for $\mathbf{x} > \boldsymbol{l}$

$$\mu([-\boldsymbol{\infty}, \mathbf{x}]^c) = -\log F(\mathbf{x})$$

and so

$$F(\mathbf{x}) = \exp\{-\mu([-\infty, \mathbf{x}]^c)\}$$

as desired. That $\mu$ is an exponent measure follows from the equivalence of (5.8) and (5.9) with (5.8′) and (5.9′). □

EXERCISES

5.3.1. (i) Suppose $F$ is max-id on $\mathbb{R}^d$ with exponent measure $\mu$. Show $F(\mathbf{x}) = \prod_{i=1}^d F_i(x^{(i)})$ where $F_i$ is a distribution on $\mathbb{R}$, iff $\mu$ vanishes on $\bigcup_{1 \le i < j \le d} \{\mathbf{y} \in [\mathbf{l}, \infty]: y^{(i)} > l^{(i)}, y^{(j)} > l^{(j)}\}$.

(ii) If $\mathbf{X}$ is max-id in $\mathbb{R}^4$ give conditions on the exponent measure $\mu$ necessary and sufficient for $(X^{(1)}, X^{(2)})$ to be independent of $(X^{(3)}, X^{(4)})$.

(iii) Give an example of a max-id random vector in $\mathbb{R}^3$ such that the exponent measure vanishes on $(\mathbf{l}, \infty]$ but no subset of $X^{(1)}$, $X^{(2)}$, $X^{(3)}$ is independent of the complementary subset of variables.

5.3.2. Give examples to show that if $F$ is max-id and $\mathbf{x}$ is on the boundary of $[\mathbf{l}, \infty]$, both $F(\mathbf{x}) > 0$ and $F(\mathbf{x}) = 0$ are possible.

5.3.3. Prove Proposition 5.1(b) by appealing to Proposition 5.8(iii).

5.3.4. Let $F$ be a distribution on $\mathbb{R}^2$. Show $F$ is max-id iff for $\mathbf{a} < \mathbf{b}$ in $\mathbb{R}^2$

$$F(a^{(1)}, a^{(2)})F(b^{(1)}, b^{(2)}) \ge F(a^{(1)}, b^{(2)})F(b^{(1)}, a^{(2)})$$

(Balkema and Resnick, 1977).

5.3.5. Let $\mathbf{X}$ be a max-id random element of $\mathbb{R}^d$ with distribution $F$. Suppose $F(\mathbf{l}) > 0$ where $\mathbf{l} = \inf[F > 0]$, $\mathbf{l} \in \mathbb{R}^d$. Then $F$ has the form given in Example 5.6 (Balkema and Resnick, 1977).

5.3.6. Any max-id distribution is the weak limit of a sequence $\{F_n\}$ where each $F_n$ has the form given in Example 5.6. Hint: Review Exercise 5.1.5 (Balkema and Resnick, 1977).

5.3.7. Suppose $\{\mathbf{X}_{n,k}, 1 \le k \le n\}$ are independent random vectors in $\mathbb{R}^d$ for each $n \ge 1$. Suppose $\bigvee_{k=1}^n \mathbf{X}_{n,k}$ converges weakly to a random vector $\mathbf{X}$ in $\mathbb{R}^d$ and that

$$\lim_{n\to\infty} \bigvee_{k=1}^n P([\mathbf{X}_{n,k} \le \mathbf{x}]^c) = 0$$

for each continuity point $\mathbf{x}$ of the distribution of $\mathbf{X}$ such that $P[\mathbf{X} \le \mathbf{x}] > 0$. Show $\mathbf{X}$ is max-id (Balkema and Resnick, 1977).

5.3.8. (a) Let $F$ be max-id on $\mathbb{R}^d$ with support $S$. If $\mu$ is an exponent measure with support $S_1$ show

$$S = \overline{[F > 0]} \cap (S_1 \vee \cdots \vee S_1)$$

($d$ times) (Balkema and Resnick, 1977).

(b) Let $\mathbf{Y}$ be extremal-$F$ and define

$$R(\mathbf{Y}) = \{\mathbf{x}\text{: for all open sets } O \text{ containing } \mathbf{x}\text{: } P[\mathbf{Y}(t) \in O \text{ for some } t > 0] > 0\}.$$

Show $R(\mathbf{Y}) = S$ (de Haan and Resnick, 1977).

(c) Define a measure $\mu$ on $[0, \infty]^2 \setminus \{\mathbf{0}\}$ as follows: If $G$ is a probability measure on $[0, \pi/2]$ and if $(r, \theta)$ are the polar coordinates of $x$:

$$\mu\{\mathbf{x}: r > t, \theta_1 < \theta \le \theta_2\} = t^{-\alpha} G(\theta_1, \theta_2]$$

for $t > 0, 0 \le \theta_1 < \theta_2 \le \pi/2$. Check that $\mu$ is an exponent measure. Find $S_1$ and $S$.

(d) A distribution on $\mathbb{R}^2$ which concentrates on $\{\mathbf{x}: x^{(1)} + x^{(2)} = 0\}$ cannot be max-id.

(e) If $F$ is max-id on $\mathbb{R}^d$, then all distributions $F^t$, $t > 0$ have the same support.

5.3.9. Suppose $F$ is max-id on $\mathbb{R}^d$ with exponent measure $\mu$ and let $\mathbf{Y}$ be extremal-$F$. Show that $\mathbf{Y}$ is Markov with stationary transition probabilities and exhibit the transition probabilities. Check that the holding time in state $\mathbf{x}$ is exponential with parameter $\mu([-\infty, \mathbf{x}]^c)$. Given that a sojourn in state $\mathbf{x}$ is ending, the next state visited is in $[-\infty, \mathbf{y}]^c$ with probability

$$\mu([-\infty, \mathbf{y}]^c)/\mu([-\infty, \mathbf{x}]^c) \qquad \text{for } \mathbf{y} \ge \mathbf{x}.$$

5.3.10. Let $\mu$ be an exponent measure on $[\boldsymbol{l}, \infty] \setminus \{\boldsymbol{l}\} =: E$ for $\boldsymbol{l} \in [-\infty, \infty)^d$ and suppose

$$\xi = \sum_k \varepsilon_{(t_k, \mathbf{j}_k)}$$

is PRM$(dt \times d\mu)$ on $[0, \infty) \times E$. Define

$$\mathbf{X} := \boldsymbol{l} \vee \bigvee_k \{\mathbf{j}_k - t_k \mathbf{1}\}$$

where $\mathbf{1} = (1, \dots, 1)$ is the vector in $\mathbb{R}^d$ all of whose components equal 1. Show that $\mathbf{X}$ is max-self-decomposable; i.e., for each $t$ there is a random vector $\mathbf{X}_t$ independent of $\mathbf{X}$ such that

$$\mathbf{X} \stackrel{d}{=} (\mathbf{X} - t\mathbf{1}) \vee \mathbf{X}_t$$

(Gerritse, 1986).

## 5.4. Limit Distributions for Multivariate Extremes

Suppose as before that $\{\mathbf{X}_n = (X_n^{(1)}, \dots, X_n^{(d)}), n \ge 1\}$ are iid random $d$-dimensional vectors with common distribution $F(\mathbf{x})$. Let the marginal distributions of $F(\mathbf{x})$ be $F_1, \dots, F_d$ so that $F_1(x) = F(x, \infty, \dots, \infty)$, and so on. Assume that there exist normalizing constants $a_n^{(i)} > 0$, $b_n^{(i)} \in \mathbb{R}$, $1 \le i \le d$, $n \ge 1$ such that as $n \to \infty$

$$\begin{aligned} P[(M_n^{(i)} - b_n^{(i)})/a_n^{(i)} &\le x^{(i)}, 1 \le i \le d] \\ &= F^n(a_n^{(1)} x^{(1)} + b_n^{(1)}, \dots, a_n^{(d)} x^{(d)} + b_n^{(d)}) \to G(x) \end{aligned} \tag{5.15}$$

for the limit distribution $G$ such that each marginal $G_i$, $i = 1, \dots, d$ is non-degenerate. The class of limits $G$, called *extreme value distributions*, must be characterized, and we seek necessary and sufficient conditions on $F$ for (5.15) to hold. Any $F$ giving rise to (5.15) will be said to be in the *domain of attraction* of $G$ and as in 0.4 this will be written $F \in D(G)$. We emphasize that the marginal distributions of a multivariate extreme value distribution are assumed to be nondegenerate.

Looking at the $i$th marginal convergence in (5.15) we get

$$F_i^n(a_n^{(i)}x + b_n^{(i)}) \to G_i(x), \qquad \text{nondegenerate, } 1 \le i \le d. \tag{5.16}$$

From Theorem 0.3 it follows that each $G_i$ is a one-dimensional extreme value distribution. We assume here and throughout that the normalizing constants $a_n^{(i)}$ and $b_n^{(i)}$ have been chosen so that $G_i$ is either $\Phi_\alpha$, $\Psi_\alpha$, or $\Lambda$ as given in Theorem 0.3.

Bearing in mind the procedure used to derive the form of the one-dimensional extreme value distributions in Theorem 0.3, we say that a distribution $G(x)$ is *max-stable* if for $i = 1, \ldots, d$ and every $t > 0$ there exist functions $\alpha^{(i)}(t) > 0$, $\beta^{(i)}(t)$ such that

$$G^t(\mathbf{x}) = G(\alpha^{(1)}(t)x^{(1)} + \beta^{(1)}(t), \ldots, \alpha^{(d)}(t)x^{(d)} + \beta^{(d)}(t)). \tag{5.17}$$

It is clear from (5.17) that for every $t > 0$, $G^t$ is a distribution function and hence every max-stable distribution is max-id. The relevance of max-stable distributions is obvious from the next result.

**Proposition 5.9.** *The class of multivariate extreme value distributions is precisely the class of max-stable distribution functions with nondegenerate marginals.*

Proof. It is clear that if $G$ has nondegenerate marginals and is max-stable, then (5.15) holds; take $F = G$. Conversely suppose (5.15) holds. From marginal convergence (5.16) and (0.18) there exist functions $\alpha^{(i)}(t) > 0$, $\beta^{(i)}(t) \in \mathbb{R}$ such that for $t > 0$, $1 \le i \le d$

$$\lim_{n\to\infty} a_n^{(i)}/a_{[nt]}^{(i)} = \alpha^{(i)}(t), \qquad \lim_{n\to\infty} (b_n^{(i)} - b_{[nt]}^{(i)})/a_{[nt]}^{(i)} = \beta^{(i)}(t). \tag{5.18}$$

Suppose $\mathbf{Y}(t)$ is a vector with distribution $G^t(\mathbf{x})$. (It is clear from (5.20) later that $G^t$ is a distribution.) Then for $t > 0$ we have on the one hand

$$((M_{[nt]}^{(i)} - b_{[nt]}^{(i)})/a_{[nt]}^{(i)}, 1 \le i \le d) \Rightarrow \mathbf{Y}(1) \tag{5.19}$$

and on the other

$$((M_{[nt]}^{(i)} - b_n^{(i)})/a_n^{(i)}, 1 \le i \le d) \Rightarrow \mathbf{Y}(t) \tag{5.20}$$

since $P[\mathbf{M}_{[nt]} \le \mathbf{x}] = F^{[nt]}(\mathbf{x}) = (F^n(\mathbf{x}))^{[nt]/n}$. Using (5.18), (5.19), and (5.20) we have

$$\begin{aligned}
&((M_{[nt]}^{(i)} - b_{[nt]}^{(i)})/a_{[nt]}^{(i)}, 1 \le i \le d)\\
&= \left(\left(\frac{M_{[nt]}^{(i)} - b_n^{(i)}}{a_n^{(i)}}\right)\frac{a_n^{(i)}}{a_{[nt]}^{(i)}} + \frac{b_n^{(i)} - b_{[nt]}^{(i)}}{a_{[nt]}^{(i)}}, 1 \le i \le d\right)\\
&\Rightarrow (\alpha^{(i)}(t)Y^{(i)}(t) + \beta^{(i)}(t), 1 \le i \le d) \overset{\mathrm{d}}{=} \mathbf{Y}(1)
\end{aligned}$$

which is the same as (5.17). □

To characterize max-stable distributions with nondegenerate marginals, it is an enormous help to standardize the problem so that $G$ has specified marginals. How one standardizes is somewhat arbitrary and depends on taste. Different specifications have led to (superficially) different representations in the literature. For the purposes of making connections later with the theory of multivariate regularly varying functions we standardize so that the marginals of $G$ are $\Phi_1(x) = \exp\{-x^{-1}\}$, $x > 0$. The standardization does not introduce difficulties, as shown next.

**Proposition 5.10.** (a) *Suppose $G$ is a multivariate distribution function with continuous marginals. Define for $i = 1, \dots, d$*

$$\psi_i(x) = (1/(-\log G_i))^{\leftarrow}(x), \qquad x > 0$$

*and*

$$G_*(\mathbf{x}) = G(\psi_1(x^{(1)}), \dots, \psi_d(x^{(d)})), \qquad \mathbf{x} \geq \mathbf{0}.$$

*Then $G_*$ has marginal distributions $G_{*i}(x) = \Phi_1(x)$ and $G$ is a multivariate extreme value distribution* iff *$G_*$ is also.*
(b) *Define $U_i = 1/(1 - F_i)$, $1 \leq i \leq d$ and let $F_*$ be the distribution of $(U_1(X_1^{(1)}), \dots, U_d(X_1^{(d)}))$ so that*

$$F_*(\mathbf{x}) = F(U_1^{\leftarrow}(x^{(1)}), \dots, U_d^{\leftarrow}(x^{(d)})).$$

*If* (5.15) *holds, so that $F \in D(G)$, then $F_* \in D(G_*)$ and*

$$P\left[\bigvee_{j=1}^{n} U_i(X_j^{(i)})/n \leq x^{(i)}, 1 \leq i \leq d\right] = F_*^n(n\mathbf{x}) \to G_*(\mathbf{x}). \tag{5.21}$$

*Conversely if* (5.21) *holds as well as* (5.16) *and $G_*$ has nondegenerate marginals then $F \in D(G)$ and* (5.15) *is true.*

Proof. Suppose (5.15) holds so that (5.16) is true as well. From (5.16) we get

$$n(1 - F_i(a_n^{(i)}x + b_n^{(i)})) \to -\log G_i(x)$$

so that

$$U_i(a_n^{(i)}x + b_n^{(i)})/n \to 1/(-\log G_i(x)) \tag{5.22}$$

and an inversion yields

$$(U_i^{\leftarrow}(ny) - b_n^{(i)})/a_n^{(i)} \to \psi_i(y), \qquad y > 0. \tag{5.23}$$

Thus for $\mathbf{x} > \mathbf{0}$

$$\begin{aligned} F_*^n(n\mathbf{x}) &= P\left[\bigvee_{j=1}^{n} X_j^{(i)} \leq U_i^{\leftarrow}(nx^{(i)}), 1 \leq i \leq d\right] \\ &= P[(M_n^{(i)} - b_n^{(i)})/a_n^{(i)} \leq (U_i^{\leftarrow}(nx^{(i)}) - b_n^{(i)})/a_n^{(i)}, 1 \leq i \leq d] \end{aligned}$$

and applying (5.23) and (5.15) we get

$$F_*^n(n\mathbf{x}) \to G_*(\mathbf{x}).$$

To check that $G_*$ has the correct marginals merely observe

$$\begin{aligned} G_i \circ \psi_i(x) &= \exp\{-(-\log G_i((1/-\log G_i)^{\leftarrow}(x))\} = \exp\{-x^{-1}\} \\ &= \Phi_1(x). \end{aligned}$$

Conversely given (5.21) and (5.16), (5.22) also holds so that

$$\begin{aligned} &\lim_{n\to\infty} F^n(a_n^{(1)}x^{(1)} + b_n^{(1)}, \dots, a_n^{(d)}x^{(d)} + b_n^{(d)}) \\ &\quad = \lim_{n\to\infty} F_*^n(n(U_1(a_n^{(1)}x^{(1)} + b_n^{(1)})/n), \dots, n(U_d(a_n^{(d)}x^{(d)} + b_n^{(d)})/n)) \\ &\quad = G_*(\psi_1^{\leftarrow}(x^{(1)}), \dots, \psi_d^{\leftarrow}(x^{(d)})) = G(x). \end{aligned}$$

□

For those wishing to learn about limit distributions for multivariate extremes while standing on one leg (i.e. quickly), Proposition 5.10 tells pretty much the whole story. The rest is commentary.

### 5.4.1. Characterizing Max-Stable Distributions

Suppose (5.15) and hence (5.21) hold. For $G_*$, (5.17) becomes

$$G_*^t(t\mathbf{x}) = G_*(\mathbf{x}) \tag{5.24}$$

which follows from (5.18) upon taking $a_n^{(i)} = n$, $b_n^{(i)} = 0$. Since $G_*$ is max-id there is an exponent measure $\mu_*$. Each marginal of $G_*$ is $\Phi_1$, which concentrates on $[0, \infty)$, and therefore it is appropriate to take $\boldsymbol{l}$ from Proposition 5.8 equal to $\mathbf{0}$ so that $\mu_*$ concentrates on $E := [0, \infty]^d \setminus \{\mathbf{0}\}$ and $\mu_*$ has no mass on the lines through $+\infty$; i.e., $\mu_*(\bigcup_{i=1}^d \{\mathbf{y} \in E: y^{(i)} = \infty\}) = 0$. We may translate (5.24) into a homogeneity property for $\mu_*$:

$$\mu_*([\mathbf{0}, \mathbf{x}]^c) = t\mu_*([\mathbf{0}, t\mathbf{x}]^c) \tag{5.25}$$

where $t > 0$, $\mathbf{x} \geq \mathbf{0}$, $\mathbf{x} \neq \mathbf{0}$, and $[\mathbf{0}, \mathbf{x}]^c = E \setminus [\mathbf{0}, \mathbf{x}]$. Note that

$$\begin{aligned} [\mathbf{0}, t\mathbf{x}]^c &= \bigcup_{i=1}^d \{\mathbf{y} \in E: y^{(i)} > tx^{(i)}\} \\ &= \bigcup_{i=1}^d \{t\mathbf{y} \in E: y^{(i)} > x^{(i)}\} = t[\mathbf{0}, \mathbf{x}]^c \end{aligned}$$

where for a Borel set $B \subset E$ we write $tB = \{tb: b \in B\}$. Thus (5.25) can be rewritten as

$$\mu_*([\mathbf{0}, \mathbf{x}]^c) = t\mu_*(t[\mathbf{0}, \mathbf{x}]^c). \tag{5.25'}$$

For a fixed $t > 0$, the equation (5.25′) can readily be extended to hold for all rectangles contained in $E$, and so the equality

$$\mu_*(B) = t\mu_*(tB) \tag{5.26}$$

holds on a generating class closed under intersections and is hence true for all Borel subsets of $E$ (cf. Exercise 3.1.3.). Thus (5.24) and (5.26) are equivalent.

Now pick your favorite norm on $\mathbb{R}^d$ so that $\|\mathbf{x}\|$ is the distance induced by this norm of the point $\mathbf{x}$ from $\mathbf{0}$. We will pretend $\|\cdot\|$ is defined on all of $E$; the definition of $\|\cdot\|$ on $E\backslash[0,\infty)^d = \bigcup_{i=1}^d \{\mathbf{y}\in E: y^{(i)} = \infty\}$ will be immaterial for our purposes since by (5.8′) any exponent measure places zero mass on $\bigcup_{i=1}^d \{\mathbf{y}\in E: y^{(i)} = \infty\}$. Let $\aleph = \{\mathbf{y}\in E: \|\mathbf{y}\| = 1\}$ be the unit sphere in $E$. Because all norms on $\mathbb{R}^d$ are equivalent (this means for two norms $\|\cdot\|$ and $\|\cdot\|^*$ there exist $0 < c_1 < c_2$ such that $c_1\|\mathbf{x}\| \le \|\mathbf{x}\|^* \le c_2\|\mathbf{x}\|$ for all $\mathbf{x}\in\mathbb{R}^d$—see Simmons, 1963, p. 223) $\aleph$ is bounded away from $\mathbf{0}$ and is hence compact. For a Borel subset $A \subset \aleph$ write

$$S(A) = \mu_*\{\mathbf{x}: \|\mathbf{x}\| > 1, \|\mathbf{x}\|^{-1}\mathbf{x}\in A\},$$

which is a measure concentrating on $\aleph$. Since $\mu_*$ is finite on compact sets, $S$ is a finite measure on $\aleph$. The transformation $\mathbf{x} \to (\|\mathbf{x}\|, \|\mathbf{x}\|^{-1}\mathbf{x})$, a kind of polar coordinate transformation, allows us to capitalize on the homogeneity property (5.26) since for $r > 0$ and Borel $A \subset \aleph$ we have

$$\begin{aligned}
&\mu_*\{\mathbf{y}\in E: \|\mathbf{y}\| > r, \|\mathbf{y}\|^{-1}\mathbf{y}\in A\}\\
&= r^{-1}\mu_*\{r^{-1}\mathbf{y}: \|\mathbf{y}\| > r, \|\mathbf{y}\|^{-1}\mathbf{y}\in A\}\\
&= r^{-1}\mu_*\{r^{-1}\mathbf{y}: \|r^{-1}\mathbf{y}\| > 1, \|r^{-1}\mathbf{y}\|^{-1}(r^{-1}\mathbf{y})\in A\}\\
&= r^{-1}\mu_*\{\mathbf{x}\in E: \|\mathbf{x}\| > 1, \|\mathbf{x}\|^{-1}\mathbf{x}\in A\} = r^{-1}S(A).
\end{aligned}$$

Thus with respect to the new coordinates $(\|\mathbf{y}\|, \|\mathbf{y}\|^{-1}\mathbf{y})$ we have that $\mu_*$ is a product measure. Put another way, if $T: E \to ((0,\infty] \times \aleph)$ via $T\mathbf{y} = (\|\mathbf{y}\|, \|\mathbf{y}\|^{-1}\mathbf{y})$ then

$$\mu_* \circ T^{-1}(dr, d\mathbf{a}) = r^{-2}\,dr S(d\mathbf{a}).$$

Therefore for $\mathbf{x}\in E$

$$\mu_*([0,\mathbf{x}])^c = \mu_* \circ T^{-1} \circ T([\mathbf{0},\mathbf{x}]^c)$$

and since

$$\begin{aligned}
T([\mathbf{0},\mathbf{x}]^c) &= T\{\mathbf{y}\in E: y^{(i)} > x^{(i)} \text{ for some } i = 1,\dots,d\}\\
&= \{(r,\mathbf{a})\in((0,\infty] \times \aleph): (r\mathbf{a})^{(i)} > x^{(i)}, \text{for some } i = 1,\dots,d\}\\
&= \left\{(r,\mathbf{a}): r > \frac{x^{(i)}}{a^{(i)}}, \text{ for some } i = 1,\dots,d\right\}\\
&= \left\{(r,\mathbf{a}): r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{a^{(i)}}\right\}
\end{aligned}$$

we have

$$\begin{aligned}\mu_*([0,\mathbf{x}]^c) &= \int\int_{T([0,\mathbf{x}]^c)} r^{-2}\,dr S(d\mathbf{a})\\ &= \int_{\aleph} S(d\mathbf{a})\left(\int_{\left[r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{a^{(i)}}\right]} r^{-2}\,dr\right)\\ &= \int_{\aleph} S(d\mathbf{a})\left(\bigwedge_{i=1}^{d} \frac{x^{(i)}}{a^{(i)}}\right)^{-1} = \int_{\aleph} \bigvee_{i=1}^{d}\left(\frac{a^{(i)}}{x^{(i)}}\right) S(d\mathbf{a}).\end{aligned}$$

We now summarize this discussion. Recall that $E = [0,\infty]^d \backslash \{\mathbf{0}\}$.

**Proposition 5.11.** *The following statements are equivalent.*

(i) $G_*(\mathbf{x})$ *is a multivariate extreme value distribution with* $\Phi_1$ *marginals.*

(ii) *There exists a finite measure S on*

$$\aleph = \{\mathbf{y} \in E : \|\mathbf{y}\| = 1\}$$

*satisfying*

$$\int_{\aleph} a^{(i)} S(d\mathbf{a}) = 1, \qquad 1 \le i \le d \tag{5.27}$$

*such that for* $\mathbf{x} \in \mathbb{R}^d$

$$G_*(\mathbf{x}) = \exp\left\{-\int_{\aleph} \bigvee_{i=1}^{d}\left(\frac{a^{(i)}}{x^{(i)}}\right) S(d\mathbf{a})\right\}. \tag{5.28}$$

(iii) *There exist non-negative Lebesgue integrable functions* $f_i(s)$, $1 \le i \le d$, *and* $0 \le s \le 1$ *on* $[0,1]$ *satisfying*

$$\int_{[0,1]} f_i(s)\,ds = 1, \qquad 1 \le i \le d \tag{5.29}$$

*such that for* $x \ge 0$

$$G_*(\mathbf{x}) = \exp\left\{-\int_{[0,1]} \bigvee_{i=1}^{d}\left(\frac{f_i(s)}{x^{(i)}}\right) ds\right\}. \tag{5.30}$$

(iv) *There exists* $\sum_k \varepsilon_{(t_k,\mathbf{j}_k)}$, PRM$(dt \times d\mu_*)$ *on* $[0,\infty) \times E$ *with*

$$\mu_*\{\mathbf{y} : \|\mathbf{y}\| > r, \|\mathbf{y}\|^{-1}\mathbf{y} \in A\} = r^{-1}S(A)$$

*and S a finite measure satisfying* (5.27) *such that for* $\mathbf{x} \ge \mathbf{0}$

$$G_*(\mathbf{x}) = P\left[\bigvee_{t_k \le 1} \mathbf{j}_k \le \mathbf{x}\right].$$

(v) *There exists* $\sum_k \varepsilon_{(t_k,r_k,u_k)}$, PRM$(dt \times d\mu)$ *on* $[0,\infty) \times (0,\infty] \times [0,1]$ *with*

$$\mu(dr, du) = S(\aleph) r^{-2}\,dr\,du$$

*and non-negative Lebesgue integrable functions* $f_i(s)$, $1 \le i \le d$, *and* $0 \le s \le 1$

*on* $[0, 1]$ *satisfying* (5.29) *such that for* $\mathbf{x} \geq \mathbf{0}$

$$G_*(\mathbf{x}) = P\left[\bigvee_{t_k \leq 1} (f_i(u_k)r_k, 1 \leq i \leq d) \leq \mathbf{x}\right].$$

*Remark.* Representations (iii) and (v) appear in de Haan (1984b), where they are used to obtain a spectral representation for max-stable processes.

PROOFS. It is readily verified that each representation in (ii)–(v) is of a max-stable distribution. The side conditions (5.27) and (5.29) appear because of the requirement that each marginal of $G_*$ be $\Phi_1$. The representation in (5.28) was derived before the statement of Proposition 5.11, and (iv) is just a restatement of Proposition 5.8. We may readily obtain (iii) from (ii) by noting that since $S(d\mathbf{a})/S(\aleph)$ is a probability measure on the complete, separable metric space $\aleph$, there exists a random element $\mathbf{f} = (f_1, \ldots, f_d)$ from the Lebesgue interval $[0, 1]$ considered as a probability space into $\aleph$ such that $\mathbf{f}$ has distribution $S(d\mathbf{a})/S(\aleph)$. This generalization of the probability integral transform in 0.2 is well known; see Billingsley, 1971, pages 6–7 or Skorohod, 1956, for example. This result allows the integration in the integral in (5.28) to be performed on $[0, 1]$:

$$S(\aleph)\int_{\aleph} \bigvee_{i=1}^{d}\left(\frac{a^{(i)}}{x^{(i)}}\right) S(d\mathbf{a})/S(\aleph) = S(\aleph)\int_{[0,1]} \bigvee_{i=1}^{d}\left(\frac{f_i(s)}{x^{(i)}}\right) ds$$

which apart from the factor $S(\aleph)$ is (5.30).

Similarly we may obtain (v) from (iv): If $\{\mathbf{A}_i, i \geq 1\}$ are iid random elements of $\aleph$ with distribution $S(d\mathbf{a})/S(\aleph)$, there exists $\mathbf{f} = (f_1, \ldots, f_d)\colon [0, 1] \to \aleph$ such that if $(U_k, k \geq 1)$ are iid uniformly distributed random variables then

$$(\mathbf{A}_i, i \geq 1) \stackrel{d}{=} (\mathbf{f}(U_i), i \geq 1)$$

in $\aleph^\infty$. If $\sum_k \varepsilon_{(t_k, r_k)}$ is PRM on $[0, \infty) \times (0, \infty]$ with mean measure $dt \times S(\aleph)r^{-2}\,dr$ and if $\sum_k \varepsilon_{(t_k, r_k)}$ is independent of $\{U_k\}$ we have by Proposition 3.8 that

$$\sum \varepsilon_{(t_k, r_k, \mathbf{f}(U_k))}$$

is PRM on $[0, \infty) \times (0, \infty] \times \aleph$ with mean measure $dt \times r^{-2}\,dr \times S(d\mathbf{a})$. It follows from Proposition 3.7 that

$$\sum_k \varepsilon_{(t_k, r_k\mathbf{f}(U_k))}$$

is PRM$(dt \times d\mu_*)$. Therefore

$$\sum_k \varepsilon_{(t_k, r_k\mathbf{f}(U_k))} \stackrel{d}{=} \sum_k \varepsilon_{(t_k, \mathbf{j}_k)}$$

in $M_p([0, \infty) \times E)$, which entails

$$\bigvee_{t_k \leq 1} r_k\mathbf{f}(U_k) \stackrel{d}{=} \bigvee_{t_k \leq 1} \mathbf{j}_k$$

as desired. □

We now consider some examples in $R^2$. When $d = 2$ we have $\aleph$ one-dimensional, and it is frequently most convenient to parameterize $\aleph$ by an interval.

EXAMPLE 5.12. Suppose for $(x, y) > 0$ we have in $[0, \infty)^2$

$$G_*(x, y) = \exp\{-(x^{-2} + y^{-2})^{1/2}\}$$

so that

$$\mu_*([0, (x, y)]^c) = (x^{-2} + y^{-2})^{1/2}$$

and the density of $\mu_*$ is

$$\mu_*(dx, dy) = (x^2 + y^2)^{-3/2}.$$

Defining $\|(x, y)\|^2 = x^2 + y^2$, $\aleph = \{(x, y) \geq \mathbf{0}: x^2 + y^2 = 1\}$ and $H[0, \theta_0] = \mu_*\{(x, y): x^2 + y^2 \geq 1, \arctan(y/x) \leq \theta_0\}$, we have upon switching to polar coordinates

$$\begin{aligned} H[0, \theta_0] &= \iint_{\substack{\|(x,y)\| \geq 1 \\ \arctan(y/x) \leq \theta_0}} \mu_*(dx, dy) \\ &= \int_{\theta=0}^{\theta_0} \int_{r=1}^{\infty} r^{-3} r \, dr \, d\theta = \theta_0 \int_1^{\infty} r^{-2} \, dr = \theta_0 \end{aligned}$$

and so $H$ is Lebesque measure on $[0, \pi/2]$. Now define $h: [0, \pi/2] \to \aleph$ by $h(\theta) = (\cos\theta, \sin\theta)$ and

$$\begin{aligned} S(d\mathbf{a}) &= \mu_*\{(x, y): x^2 + y^2 \geq 1, (x^2 + y^2)^{-1/2}(x, y) \in d\mathbf{a}\} \\ &= \mu_*\{(x, y): r \geq 1, (\cos\theta, \sin\theta) \in d\mathbf{a}\} \\ &= H \circ h^{-1}(d\mathbf{a}). \end{aligned}$$

Therefore for $(x, y) \geq 0$

$$G_*(x, y) = \exp\left\{-\int_{\aleph} \left(\frac{a^{(1)}}{x} \vee \frac{a^{(2)}}{y}\right) S(da)\right\}$$

and by the transformation theorem for integrals, since $S(d\mathbf{a}) = H \circ h^{-1}(d\mathbf{a})$ we have

$$\begin{aligned} G_*(x, y) &= \exp\left\{-\int_{[0, \pi/2]} \left(\frac{\cos\theta}{x} \vee \frac{\sin\theta}{y}\right) d\theta\right\} \\ &= \exp\left\{-\int_{[0,1]} \left(\frac{\cos(2^{-1}\pi s)}{x} \vee \frac{\sin(2^{-1}\pi s)}{y}\right) \frac{\pi}{2} ds\right\} \end{aligned}$$

and thus $\mathbf{f}(\theta)$ in (5.30) may be written as

$$\mathbf{f}(s) = (\pi/2)(\cos(2^{-1}\pi s), \sin(2^{-1}\pi s)).$$

EXAMPLE 5.13. Now suppose in $[0, \infty)^2$ we have

$$G_*(x, y) = \exp\{-(x^{-p} + y^{-p})^{1/p}\}$$

for $p > 1$. Then

$$\mu_*([\mathbf{0}, (x, y)]^c) = (x^{-p} + y^{-p})^{1/p}$$

and taking partial derivatives with respect to $x$ and then $y$, we find

$$\mu_*(dx, dy) = (p - 1)(x^p + y^p)^{p^{-1}-2}/(xy)^{2-p}.$$

Define the norm to be

$$\|(x, y)\| = (x^p + y^p)^{1/p}$$

so that $\aleph = \{(x, y) \geq \mathbf{0}: x^p + y^p = 1\}$. Make the change of variable in $\mu_*(dx, dy)$ by setting

$$r = (x^p + y^p)^{1/p}, \qquad \theta = x/(x^p + y^p)^{1/p}$$

so that

$$x = \theta r, \qquad y = r(1 - \theta^p)^{1/p}$$

and the Jacobean of this inverse transformation is

$$J = r(1 - \theta^p)^{p^{-1}-1}.$$

So the density $\mu_*$ after transforming to new variables $r$ and $\theta$ becomes

$$(p - 1)r^{-2}\, dr\theta^{p-2}(1 - \theta^p)^{-1/p}\, d\theta. \tag{5.31}$$

If we define $H(\theta_0) = \mu_*\{(x, y) \geq \mathbf{0}: r > 1, \theta \leq \theta_0\}$ then the expression in (5.31) yields

$$H(\theta_0) = \int_{[0,\theta_0]} (p - 1)\theta^{p-2}(1 - \theta^p)^{-1/p}\, d\theta.$$

If we define $h: [0, 1] \to \aleph$ by

$$h(\theta) = (\theta, (1 - \theta^p)^{1/p})$$

then $S = H \circ h^{-1}$ and therefore

$$\mu_*[\mathbf{0}, (x, y)]^c = \int_\aleph \left(\frac{a^{(1)}}{x} \vee \frac{a^{(2)}}{y}\right) S(d\mathbf{a})$$

becomes by transforming the integral:

$$\begin{aligned}
&\int_\aleph \left(\frac{a^{(1)}}{x} \vee \frac{a^{(2)}}{y}\right) H \circ h^{-1}(d\mathbf{a}) \\
&= \int_{[0,1]} \left(\frac{s}{x} \vee \frac{(1 - s^p)^{1/p}}{y}\right)(p - 1)s^{p-2}(1 - s^p)^{-1/p}\, ds \\
&= \int_{[0,1]} \left(\frac{s^{p-1}(1 - s^p)^{-1/p}}{x} \vee \frac{s^{p-2}}{y}\right)(p - 1)ds.
\end{aligned}$$

Thus we may take $\mathbf{f}(s)$ in (5.30) as

$$\mathbf{f}(s) = (p-1)(s^{p-1}(1-s^p)^{-1/p}, s^{p-2}).$$

For $p = 2$, the choice of $\mathbf{f}$ is different than in Example 5.12.

For convenience we now restate Proposition 5.11 for the case that the marginal distributions of a multivariate extreme value distribution $G(\mathbf{x})$ are all equal to $\Lambda(x) = \exp\{-e^{-x}\}$, $x \in \mathbb{R}$. As before $E = [0, \infty]^d \backslash \{\mathbf{0}\}$.

**Proposition 5.11′.** *The following are equivalent*:
(i) $G(\mathbf{x})$ *is a multivariate extreme value distribution with* $\Lambda(x)$ *marginals.*
(ii) *There exists a finite measure S on*

$$\aleph = \{\mathbf{y} \in E \colon \|\mathbf{y}\| = 1\}$$

*satisfying*

$$\int_{\aleph} a^{(i)} S(d\mathbf{a}) = 1, \qquad 1 \le i \le d \tag{5.27}$$

*such that for* $\mathbf{x} \in \mathbb{R}^d$

$$G(\mathbf{x}) = \exp\left\{-\int_{\aleph} \bigvee_{i=1}^{d} (a^{(i)} e^{-x^{(i)}}) S(d\mathbf{a})\right\}.$$

(iii) *There exist non-negative Lebesgue integrable functions* $f_i(s)$, $1 \le i \le d$, *and* $0 \le s \le 1$ *on* $[0, 1]$ *satisfying*

$$\int_{[0,1]} f_i(s) ds = 1, \qquad 1 \le i \le d \tag{5.29}$$

*such that for* $\mathbf{x} \in \mathbb{R}^d$

$$G(\mathbf{x}) = \exp\left\{-\int_{[0,1]} \bigvee_{i=1}^{d} (f_i(s) e^{-x^{(i)}}) ds\right\}$$

*or equivalently if we set* $f_i = \exp\{g_i\}$

$$G(\mathbf{x}) = \exp\left\{-\int_{[0,1]} \exp\left\{-\bigwedge_{i=1}^{d} (x^{(i)} - g_i(s)) ds\right\}.\right.$$

(iv) *There exist* $\sum_k \varepsilon_{(t_k, \mathbf{j}_k)}$, $\mathrm{PRM}(dt \times d\mu_*)$ *on* $[0, \infty) \times E$ *with*

$$\mu_*\{\mathbf{y} \in E \colon \|\mathbf{y}\| > r, \|\mathbf{y}\|^{-1}\mathbf{y} \in A\} = r^{-1} S(A)$$

*and S a finite measure satisfying* (5.27) *such that for* $\mathbf{x} \in \mathbb{R}^d$

$$G(\mathbf{x}) = P\left[\bigvee_{t_k \le 1} \log \mathbf{j}_k \le \mathbf{x}\right]$$

($\log \mathbf{j}_k = (\log j_k^{(i)}, 1 \le i \le d)$).
(v) *There exists* $\sum_k \varepsilon_{(t_k, v_k, u_k)}$, $\mathrm{PRM}(dt \times d\mu_\Lambda)$ *on* $[0, \infty) \times (-\infty, \infty] \times [0, 1]$ *with*

$$\mu_\Lambda(dv, du) = S(\aleph)e^{-v}\,dv\,du$$

*and non-negative Lebesgue integrable fuctions* $f_i(s) = \exp\{g_i(s)\}$, $1 \le i \le d$, $0 \le s \le 1$ *on* $[0,1]$ *satisfying* (5.29) *such that for* $\mathbf{x} \in \mathbb{R}^d$

$$G(\mathbf{x}) = P\left[\bigvee_{t_k \le 1} (g_i(u_k) + v_k, 1 \le i \le d) \le \mathbf{x}\right].$$

PROOF. We merely need to observe that from Proposition 5.10

$$G(\mathbf{x}) = G_*(e^{\mathbf{x}})$$

where $\mathbf{x} \in \mathbb{R}^d$ and $e^{\mathbf{x}} = (e^{x^{(i)}}, 1 \le i \le d)$. To get (v), note that if $\sum \varepsilon_{(t_k, \mathbf{j}_k)}$ is PRM$(dt \times d\mu_*)$ then

$$\sum_k \varepsilon_{(t_k, \log \mathbf{j}_k)} \overset{\mathrm{d}}{=} \sum_k \varepsilon_{(t_k, (\log r_k)\mathbf{1} + \log \mathbf{f}(u_k))} \overset{\mathrm{d}}{=} \sum_k \varepsilon_{(t_k, v_k \mathbf{1} + \mathbf{g}(u_k))}$$

where $\mathbf{1} = (1, \ldots, 1)$ and $\log \mathbf{f} = (\log f_i, 1 \le i \le d)$ and $\sum_k \varepsilon_{(t_k, r_k, u_k)}$ is PRM$(dt \times S(\aleph)r^{-2}\,dr \times du)$. Therefore $\sum_k \varepsilon_{(t_k, \log r_k, u_k)} \overset{\mathrm{d}}{=} \sum \varepsilon_{(t_k, v_k, u_k)}$ by Proposition 3.7. □

EXAMPLE 5.14. Consider the multivariate extreme value distribution on $\mathbb{R}^2$

$$G(x, y) = \exp\{-(e^{-x} + e^{-y} - (e^x + e^y)^{-1})\}$$

which is the limit distribution arising from a bivariate exponential distribution (Mardia, 1970; Galambos, 1978; Marshall and Olkin, 1983). We observe first that for $x > 0$, $y > 0$

$$\begin{aligned} G_*(x, y) &= G(\log x, \log y) \\ &= \exp\{-(x^{-1} + y^{-1} - (x + y)^{-1})\} \end{aligned}$$

and thus, since $\mu_*([\mathbf{0}, (x, \infty)]^c) = x^{-1}$ we have for $x > 0$, $y > 0$

$$\mu_*((x, \infty] \times (y, \infty]) = (x + y)^{-1}$$

leading to

$$\mu_*(dx, dy) = 2(x + y)^{-3}, \qquad (x, y) \in E = [0, \infty]^2 \backslash \{\mathbf{0}\}.$$

Thus it is natural to take as norm

$$\|(x, y)\| = x + y$$

so that

$$\aleph = \{(s, (1 - s)) \colon 0 \le s \le 1\}$$

and $\aleph$ is the line through $(0, 1)$ and $(1, 0)$, considered parameterized by $s$, $0 \le s \le 1$. Make the change of variable

$$(x, y) \to \left(x + y, \frac{x}{x + y}\right) = (r, s)$$

and in terms of these new coordinates the density is

$$2r^{-3}r\,dr\,ds = r^{-2}\,dr\,2ds$$

so that $S$ is a multiple of Lebesgue measure on $\aleph$. The representation for $G$ is

$$G(x, y) = \exp\left\{-\int_{[0,1]} (se^{-x} \vee (1-s)e^{-y})2ds\right\}.$$

## Exercises

5.4.1.1. Give an example of a distribution $G$ which is max-id but not max-stable.

5.4.1.2. Give an example of a max-stable distribution which does not have all marginals nondegenerate.

5.4.1.3. Give an example of a max-stable distribution which is not absolutely continuous.

5.4.1.4. Verify directly in Proposition 5.11(v) that $\bigvee_{t_k \le 1} \mathbf{f}(U_k)r_k$ has distribution $G_*$ by computing $P[\bigvee_{t_k \le 1} \mathbf{f}(U_k)r_k \le \mathbf{x}]$ for $\mathbf{x} \ge 0$ (de Haan, 1984, page 1195).

5.4.1.5. On the basis of (5.15) give the appropriate generalizations of Corollary 4.19 and Proposition 4.20.

5.4.1.6. Show that $\mathbf{Y}$ is a random element of $[0, \infty)^d$ with extreme value distribution $G_*$ (with $\Phi_1$ marginals) iff for every $t > 0$ $\bigvee_{i=1}^d t^{(i)} Y^{(i)}$ is a multiple of a random variable with distribution function $\Phi_1$ (de Haan, 1978). State the result when the marginals are $\Lambda(x)$ or $\Psi_1(x)$.

5.4.1.7. State and prove the analogue of Proposition 5.11 when all marginals are equal to $\Psi_1(x) = \exp\{-|x|^\alpha\}$, $x < 0$.

5.4.1.8. Comparing the representation in Example 1 with the one in Example 2 specialized to the case $p = 2$ shows that the choice of $\mathbf{f}$ is not unique. Given $\mathbf{f}$ and $\mathbf{f}^\#$ both satisfying (5.30), what is the relation between $\mathbf{f}$ and $\mathbf{f}^\#$ (de Haan and Pickands, 1986)?

5.4.1.9. Let $\{\mathbf{X}(t), t > 0\}$ be a Levy process in $\mathbb{R}^d$, i.e., a process with stationary, independent increments. Suppose $\{\mathbf{X}(t)\}$ is stable with index 1 so that for all $a > 0$

$$\mathbf{X}(a\cdot) \overset{\mathrm{d}}{=} a\mathbf{X}(\cdot) + \mathbf{c}(a)$$

in $D([0, \infty), [-\infty, \infty]^d \backslash \{\mathbf{0}\})$, where $\mathbf{c}(a)$ is some nonrandom vector. Suppose the Levy measure of $\{X^{(i)}(t), t > 0\}$ is $v^{(i)}(x, \infty) = x^{-1}, x > 0$. Show that

$$G_*(\mathbf{x}) := P\left[\sup_{t \le 1} (\mathbf{X}(t) - \mathbf{X}(t-)) \le \mathbf{x}\right]$$

is a multidimensional extreme value distribution with $\Phi_1$ marginals and all such distributions can be obtained in this manner (de Haan and Resnick, 1977).

5.4.1.10. The following are multivariate extreme value distributions. Pick an appropriate norm, define $\aleph$, and give the representation of Proposition 5.11 or 5.11′.

(i) $G_*(x, y) = \exp\{-(x^{-1} + y^{-1} - (x^2 + y^2)^{-1/2}\}, (x, y) \geq \mathbf{0}$.
(ii) $G_*(x, y, z) = \exp\{-2^{-1}(x^{-1} \vee y^{-1} + y^{-1} \vee z^{-1} + x^{-1} \vee z^{-1})\}, (x, y, z) \geq \mathbf{0}$.
(iii) $G_*(x, y, z) = \exp\{-2^{-1}((x^{-2} + y^{-2})^{1/2} + (x^{-2} + z^{-2})^{1/2} + (y^{-2} + z^{-2})^{1/2})\}$, $(x, y, z) \geq \mathbf{0}$ (de Haan and Resnick, 1977).
(iv) $G(x, y) = \exp\{-(e^{-x} + e^{-y} - e^{-(x \vee y)})\}$, $(x, y) \geq \mathbf{0}$ (Marshall and Olkin, 1983; Galambos, 1978).
(v) $G(x, y) = \exp\{-((e^{-x} + (1 - \theta)e^{-y}) \vee e^{1-y})\}, 0 \leq \theta \leq 1, (x, y) \in \mathbb{R}^2$.

Check that this $G$ contains a singular part and that if $X$ and $Z$ are iid with distribution $\Lambda$ then

$$G(x, y) = P[X \leq x, (X + \log \theta) \vee (Z + \log(1 - \theta)) \leq y]$$

(Tiago de Oliveira, 1980).

5.4.1.11. Find $G_*$ in the following cases when $d = 2$:
(i) $\aleph = \{(x, y) \geq \mathbf{0}\colon\ x^p + y^p = 1\}$, $p \geq 1$ and $S = H \circ h^{-1}$ where $h(x) = (x, (1 - x^p)^{1/p})\colon [0, 1] \to \aleph$ and $H(t) = t^\alpha, 0 \leq t \leq 1, \alpha > 0$.
(ii) $\aleph = \{(x, y) \geq \mathbf{0}\colon x \vee y = 1\}$ and for monotone functions $U_1(t)$, $U_2(t)$, $0 \leq t \leq 1$

$$S([0, t] \times \{1\}) = U_1(t), \qquad S(\{1\} \times [0, t]) = U_2(t).$$

What are the conditions on $U_1$ and $U_2$ in order that (5.27) be satisfied?
(iii) $\aleph = \{(x, y) \geq \mathbf{0}\colon x^2 + y^2 = 1\}$ and $S\{(0, 1)\} = S\{(1, 0)\} = 1$ and $S$ has no mass elsewhere.
(iv) $\mathbf{f}(u)$, $0 \leq u \leq 1$ satisfies (5.24) and $f_1(u) f_2(u) = 0$ a.e. Give concrete examples of $\mathbf{f}$ satisfying these conditions. This provides another graphic illustration of the nonuniqueness of $\underline{\mathbf{f}}$.
(v) $\aleph = \{(x, y) \geq \mathbf{0}\colon x^2 + y^2 = 1\}, S\{(\sqrt{\tfrac{1}{2}}, \sqrt{\tfrac{1}{2}})\} = \sqrt{2}$ and $S$ places no mass elsewhere.

5.4.1.12. If $G_*$ is a multivariate extreme value distribution as in Proposition 5.11, give conditions in order that $G_*$ be the distribution of an exchangeable random vector in $\mathbb{R}^d$.

5.4.1.13. Analogue of the Cramer–Wold device (Billingsley, 1968): Suppose that $\mathbf{X}_n$, $n \geq 0$ are random vectors in $\mathbb{R}_+^d = [0, \infty)^d$. Prove

$$\mathbf{X}_n \Rightarrow \mathbf{X}_0$$

in $\mathbb{R}_+^d$ iff for every $\mathbf{t} \in \mathbb{R}_+^d$

$$\bigvee_{i=1}^{d} t^{(i)} X_n^{(i)} \Rightarrow \bigvee_{i=1}^{d} t^{(i)} X_0^{(i)}$$

in $\mathbb{R}_+$. Hint: Use distribution functions.

5.4.1.14. If $G(x, y)$ is a bivariate extreme value distribution with $\Lambda$-marginals, show that the correlation coefficient is non-negative.

5.4.1.15. (a) Show $G_*(x, y)$ is a bivariate extreme value distribution with $\Phi_1$-marginals iff

$$G_*(x, y) = \exp\{-(x^{-1} + y^{-1} + y^{-1}\chi(yx^{-1}))\}$$

$x, y > 0$ where $\chi(t)$, $t \geq 0$ satisfies

(i) $\chi(t)$ is continuous and convex
(ii) $\max(-t, -1) \le \chi(t) \le 0, \qquad t \ge 0.$

(b) Show $G(x, y)$ is a bivariate extreme value distribution with $\Lambda$-marginals iff

$$G(x, y) = \Lambda(x)\Lambda(y)\}^{k(y-x)}, \qquad (x, y) \in \mathbb{R}^2.$$

What are the analogues of Conditions (i) and (ii) in Part (a) which $k(t)$ must satisfy? Express the correlation coefficient of $G$ in terms of $k$ (cf. 5.4.1.14).

(c) For Examples 1, 2, and 3 and for the distributions of Problem 5.4.1.10(i), (iv), and (v) give $\chi$ or $k$ as appropriate (Sibuya, 1960; Geffroy, 1958, 1959; Finkelshteyn, 1953; Tiago de Oliveira, 1975a, b, 1980; de Haan and Resnick, 1977).

5.4.1.16. Suppose $G_*(\mathbf{x})$ is a multivariate extreme value distribution with $\Phi_1$-marginals. Let $\{V_i, i \ge 1\}$ be iid random variables with common distribution $\Phi_1$. A random vector $\mathbf{Y} \in \mathbb{R}^d$ has distribution $G_*$ iff $\mathbf{Y}$ is the limit in distribution of random vectors of the form

$$\mathbf{Y}_n = \left(\bigvee_{k=1}^{n} a_{nk}^{(i)} p_{nk} V_k, 1 \le i \le d\right)$$

where

$$p_{nk} \ge 0, \qquad \sum_{k=1}^{n} p_{nk} = S(\aleph)$$

$$\sum_{k=1}^{n} a_{nk}^{(i)} p_{nk} = 1, \qquad 1 \le i \le d, \qquad n \ge 1$$

and

$$(a_{nk}^{(i)}, 1 \le i \le d) \in \aleph.$$

Hint: Approximate $S$ in (5.28) by a discrete measure concentrating on $n$ points $a_{nk} \in \aleph$, $k = 1, \ldots, n$ (Pickands, 1981; de Haan, 1985).

5.4.1.17. Show that the limit distribution in Proposition 4.29 for $a_n^{-1}(M_n, W_n)$ is of the form

$$H(x, \infty) - H(x, -y)$$

where $H(x, y)$ is a bivariate extreme value distribution. Give a representation for $H$ (Davis, 1982a; Davis and Resnick, 1985a).

## 5.4.2. Domains of Attraction; Multivariate Regular Variation

Suppose $C \subset \mathbb{R}^d$ is a cone, i.e., $\mathbf{x} \in C$ iff $t\mathbf{x} \in C$ for every $t > 0$. For concreteness suppose $\mathbf{1} \in C$. A function $h: C \to (0, \infty)$ is monotone if it is either nondecreasing in each component or nonincreasing in each component. So for instance, $h$ is monotone nondecreasing on $C$ if for $\mathbf{x}$ and $\mathbf{y} \in C$, $\mathbf{x} \le \mathbf{y}$ we have $h(\mathbf{x}) \le h(\mathbf{y})$. We say a measurable function $h$ is regularly varying on $C$ with limit function $\lambda$ if $\lambda(\mathbf{x}) > 0$, $\mathbf{x} \in C$, and for all $\mathbf{x} \in C$

$$\lim_{t \to \infty} h(t\mathbf{x})/h(t\mathbf{1}) = \lambda(\mathbf{x}) \tag{5.32}$$

so that $\lambda(\mathbf{1}) = 1$. For each fixed $\mathbf{x} \in C$ we have for $s > 0$

$$\lim_{t\to\infty} \frac{h(ts\mathbf{x})}{h(t\mathbf{x})} = \lim_{t\to\infty} \frac{h(ts\mathbf{x})}{h(t\mathbf{1})} \Big/ \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \frac{\lambda(s\mathbf{x})}{\lambda(\mathbf{x})}$$

and applying Proposition 0.4 to $h(t\mathbf{x})$ considered as a function of $t$ we find there exists $\rho \in \mathbb{R}$ such that $h(t\mathbf{x}) \in RV_\rho$ and $\lambda(s\mathbf{x})/\lambda(\mathbf{x}) = s^\rho$. Provided $\rho$ does not depend on $\mathbf{x}$, we find that $\lambda$ is homogeneous:

$$\lambda(s\mathbf{x}) = s^\rho \lambda(\mathbf{x}), \qquad s > 0. \tag{5.33}$$

We must now check that $\rho$ does not depend on $\mathbf{x}$. Temporarily supposing it does and writing $\rho(\mathbf{x})$ to indicate this dependence we have for $\mathbf{x}$ and $\mathbf{y} \in C$

$$\begin{aligned}
\rho(\mathbf{y}) &= \lim_{t\to\infty} (\log h(ts\mathbf{y}) - \log h(t\mathbf{y}))/\log s \\
&= \lim_{t\to\infty} (\log s)^{-1} \left( \log \frac{h(ts\mathbf{y})}{h(ts\mathbf{x})} - \log \frac{h(t\mathbf{y})}{h(t\mathbf{x})} + \{\log h(ts\mathbf{x}) - \log h(t\mathbf{x})\} \right) \\
&= (\log s)^{-1} \left( \log \frac{\lambda(\mathbf{y})}{\lambda(\mathbf{x})} - \log \frac{\lambda(\mathbf{y})}{\lambda(\mathbf{x})} \right) + \rho(\mathbf{x}) = \rho(\mathbf{x}).
\end{aligned}$$

Thus we find that the definition (5.32) can be rephrased: $h$ is regularly varying with limit function $\lambda$ iff there exist $V: (0, \infty) \to (0, \infty)$, $V \in RV_\rho$, and for all $\mathbf{x} \in C$

$$\lim_{t\to\infty} h(t\mathbf{x})/V(t) = \lambda(\mathbf{x}). \tag{5.34}$$

It follows from (5.33) that if $\lambda$ is monotone then $\lambda$ is continuous on $C \cap (\bigcap_{i=1}^d \{\mathbf{x}: x^{(i)} \neq 0\})$: Supposing $\lambda$ nondecreasing and $C \subset (0, \infty)^d$ for concreteness we have for any $\mathbf{y} \in [(1 - \varepsilon)\mathbf{x}, (1 + \varepsilon)\mathbf{x}]$ and $(1 - \varepsilon)\mathbf{x} \in C$ that

$$(1 - \varepsilon)^\rho \lambda(\mathbf{x}) = \lambda((1 - \varepsilon)\mathbf{x}) \leq \lambda(\mathbf{y}) \leq \lambda((1 + \varepsilon)\mathbf{x}) = (1 + \varepsilon)^\rho \lambda(\mathbf{x})$$

and the continuity of $\lambda$ at $\mathbf{x}$ is thus clear.

With these preliminaries digested we may discuss domain of attraction criteria. Before proceeding you may wish to review Proposition 5.10.

**Proposition 5.15.** *As in Proposition 5.10, for a multivariate extreme value distribution $G$, define*

$$G_*(\mathbf{x}) = G(\psi_1(x^{(1)}), \ldots, \psi_d(x^{(d)}))$$

*where*

$$\psi_i(x) = (1/(-\log G_i))^{\leftarrow}(x), \qquad x \geq 0, \qquad 1 \leq i \leq d.$$

*For a distribution $F$, define $U_i = 1/(1 - F_i)$, $1 \leq i \leq d$ so $U_i$ has range $[1, \infty]$ and $U_i^{\leftarrow}$ has domain $[1, \infty]$. Set*

$$F_*(\mathbf{x}) = F(U_1^{\leftarrow}(x^{(1)}), \ldots, U_d^{\leftarrow}(x^{(d)})), \qquad \mathbf{x} \geq \mathbf{1}.$$

(a) $F_* \in D(G_*)$ iff $1 - F_*$ *is regularly varying on the cone* $(0, \infty)^d$ *with limit function* $-\log G_*(\mathbf{x})/(-\log G_*(\mathbf{1}))$; *i.e., for* $\mathbf{x} > \mathbf{0}$

$$\lim_{t\to\infty} (1 - F_*(t\mathbf{x}))/(1 - F_*(t\mathbf{1})) = (-\log G_*(\mathbf{x}))/(-\log G_*(\mathbf{1})). \tag{5.35}$$

(b) $F \in D(G)$ iff *marginal convergences* (5.16) *hold and* $F_* \in D(G_*)$.

PROOF. (a) Given (5.35) we may define $a_n$ to satisfy

$$1 - F_*(a_n\mathbf{1}) \sim n^{-1}(-\log G_*(\mathbf{1}))$$

so that (5.35) gives (with $t$ replaced by $a_n$):

$$n(1 - F_*(a_n\mathbf{x})) \to -\log G_*(\mathbf{x}), \qquad \mathbf{x} > \mathbf{0}.$$

From this it readily follows, as for the $d = 1$ case, that

$$F_*^n(a_n\mathbf{x}) \to G_*(\mathbf{x}), \qquad \mathbf{x} > \mathbf{0}$$

and so $F_* \in D(G_*)$.

Conversely, suppose $F_* \in D(G_*)$. Since the marginals of $G_*$ are $\Phi_1$, we take $b_n^{(i)} = 0$, $1 \le i \le d$, and $F_* \in D(G_*)$ means

$$F_*^n(a_n^{(1)}x^{(1)}, \ldots, a_n^{(d)}x^{(d)}) \to G_*(\mathbf{x}), \qquad \mathbf{x} > \mathbf{0}.$$

We show we may take $a_n^{(i)} = n$ and to check this, marginal considerations are enough; it suffices to show $1 - F_{*i}(x) \sim x^{-1}$, $x \to \infty$, $1 \le i \le d$ (cf. 1.12 in Proposition 1.11). Since $F_* \in D(G_*)$ implies $F_{*i} \in D(\Phi_1)$ which is equivalent to $1/(1 - F_{*i}) = U_i \circ U_i^{\leftarrow} \in RV_1$ we need to check that

$$U_i \circ U_i^{\leftarrow} \in RV_1 \qquad \text{implies } U_i \circ U_i^{\leftarrow}(x) \sim x \tag{5.36}$$

as $x \to \infty$. From (0.6(b))

$$U_i \circ U_i^{\leftarrow}(x) \ge x.$$

Also, for any $\delta > 0$, (0.6(c)) implies

$$x > U(U^{\leftarrow}((1 - \delta)x)) \qquad \text{iff } U^{\leftarrow}(x) > U^{\leftarrow}((1 - \delta)x). \tag{5.37}$$

The second inequality on the right must be true for all large $x$ since otherwise there would be a sequence $x_n \uparrow \infty$ and $U^{\leftarrow}(x_n) = U^{\leftarrow}((1 - \delta)x_n)$ leading to $\lim_{n\to\infty} U(U^{\leftarrow}(x_n))/U(U^{\leftarrow}((1 - \delta)x_n)) = 1$, which contradicts $U \circ U^{\leftarrow} \in RV_1$. Thus the first inequality on the left of (5.37) ultimately holds. We therefore have the string of inequalities

$$\begin{aligned} 1 &\le \liminf_{x\to\infty} \frac{U \circ U^{\leftarrow}(x)}{x} \le \limsup_{x\to\infty} \frac{U \circ U^{\leftarrow}(x)}{x} \\ &\le \limsup_{x\to\infty} \frac{U \circ U^{\leftarrow}(x)}{U(U^{\leftarrow}((1 - \delta)x))} = (1 - \delta)^{-1} \end{aligned}$$

and since $\delta > 0$ is arbitrary, we conclude that (5.36) is true.

Thus $F_* \in D(G_*)$ implies

$$F_*^n(n\mathbf{x}) \to G_*(\mathbf{x}), \qquad \mathbf{x} > \mathbf{0}.$$

Taking logarithms and as usual using $-\log z \sim (1 - z)$, $z \to 1$, we get

$$n(1 - F_*(n\mathbf{x})) \to -\log G_*(\mathbf{x}) \tag{5.38}$$

and thus

$$(1 - F_*(n\mathbf{x}))/(1 - F_*(n\mathbf{1})) \to -\log G_*(\mathbf{x})/(-\log G_*(\mathbf{1}))$$

and a simple monotonicity argument allows $n$ to be replaced by $t$, giving the desired (5.35).
(b) We checked in (a) that $F_* \in D(G_*)$ iff $F_*^n(n\mathbf{x}) \to G_*(\mathbf{x})$, $\mathbf{x} > 0$ so the statement in (b) is equivalent to Proposition 5.10(b). □

To apply Proposition 5.12 one can follow this sequence of steps:
(i) Compute marginals $F_i$, $1 \le i \le d$ and check marginal convergences (5.16).
(ii) Compute $F_*$ and then check the regular variation condition (5.35).
(iii) From (5.35) obtain $-\log G_*(\mathbf{x})$ and hence $G_*(\mathbf{x})$. Compute $G(x) = G_*(\psi_1^{\leftarrow}(x^{(1)}), \ldots, \psi_d^{\leftarrow}(x^{(d)}))$ where $\psi_i$ is obtained from marginal convergence in step (i).

EXAMPLE 5.16 (Galambos, 1978, page 249; Marshall and Olkin, 1983, page 176; Mardia, 1970). Let

$$1 - F(x, y) = e^{-x} + e^{-y} - (e^x + e^y - 1)^{-1} \qquad (x, y) \ge \mathbf{0}.$$

Then

$$1 - F_i(x) = e^{-x}, \qquad x > 0, \qquad i = 1, 2$$

and

$$U_i(x) = 1/(1 - F_i(x)) = e^x, \qquad x > 0, \qquad i = 1, 2$$
$$U_i^{\leftarrow}(y) = \log y, \qquad y > 1, \qquad i = 1, 2$$

so therefore

$$1 - F_*(x, y) = 1 - F(\log x, \log y) = x^{-1} + y^{-1} - (x + y - 1)^{-1},$$
$$(x, y) \ge \mathbf{1}.$$

Thus for $(x, y) > \mathbf{0}$ and $t$ large enough that $t(x, y) > (1, 1)$ we have

$$\frac{1 - F_*(tx, ty)}{1 - F_*(t, t)} = \frac{x^{-1} + y^{-1} - (x + y - t^{-1})^{-1}}{2 - (2 - t^{-1})^{-1}} \to \frac{x^{-1} + y^{-1} - (x + y)^{-1}}{3/2}$$

as $t \to \infty$, verifying (5.35), and we can identify $G_*$ by

$$-\log G_*(x, y) = x^{-1} + y^{-1} - (x + y)^{-1}, \qquad (x, y) > \mathbf{0}.$$

This checks $F_* \in D(G_*)$. To identify $G$ we have from marginal convergences

$$n(1 - F_i(x + \log n)) \to e^{-x}, \qquad x \in \mathbb{R}, \qquad i = 1, 2.$$

Thus

$$\psi_i^{\leftarrow}(x) = 1/(-\log G_i(x)) = e^x, \qquad x \in \mathbb{R}$$

and

$$G(x, y) = G_*(\psi_1^{\leftarrow}(x), \psi_2^{\leftarrow}(y)) = G_*(e^x, e^y)$$
$$= \exp\{-(e^{-x} + e^{-y} - (e^x + e^y)^{-1})\}, \qquad (x, y) \in \mathbb{R}^2$$

which is the $G$ in Example 5.14 of Section 5.4.1.

The next result is a useful alternative criterion which is easy to apply in cases of spherical symmetry. As in the previous section, let $\|\mathbf{x}\|$ be your favorite norm for $\mathbf{x} \in \mathbb{R}^d$.

**Proposition 5.17.** *Suppose $F$, $F_*$, $G$, and $G_*$ are as in Proposition* 5.15, *and $\mu_*$, $S$, and $\aleph$ are as described in Proposition* 5.11. *Let $\mathbf{X}_*$ be a random element of $[1, \infty)^d$ with distribution $F_*$. The following are equivalent*:

(i) $F_* \in D(G_*)$.

(ii) $\displaystyle\lim_{t\to\infty} \frac{1 - F_*(t\mathbf{x})}{1 - F_*(t\mathbf{1})} = \frac{-\log G_*(\mathbf{x})}{-\log G_*(\mathbf{1})} = \frac{\mu_*([\mathbf{0}, \mathbf{x}]^c)}{\mu_*([\mathbf{0}, \mathbf{1}]^c)}$

*for* $\mathbf{x} > \mathbf{0}$.

(iii) $nF_*(n\,\cdot) = nP[n^{-1}\mathbf{X}_* \in \cdot\,] \xrightarrow{v} \mu_*$ *on* $E$.

(iv) $nP[(n^{-1}\|\mathbf{X}_*\|, \|\mathbf{X}_*\|^{-1}\mathbf{X}_*) \in \cdot\,] \xrightarrow{v} r^{-2}\,dr \times S$ *on* $(0, \infty] \times \aleph$.

*Remark.* Since $\mu_*$ puts zero mass on INF $:= \bigcup_{i=1}^d \{\mathbf{x} \in E: x^{(i)} = \infty\}$ it will be immaterial how we define

$$P\mathbf{x} := (\|\mathbf{x}\|, \|\mathbf{x}\|^{-1}\mathbf{x})$$

on INF. In the following discussion, we ignore INF.

Proof. The equivalence of (i) and (ii) is given in Proposition 5.12, where it is also shown that $F_* \in D(G_*)$ iff

$$F_*(n\mathbf{x}) \to G_*(\mathbf{x}), \qquad \mathbf{x} > \mathbf{0}.$$

This last statement is equivalent to

$$n(1 - F_*(n\mathbf{x})) = nP[n^{-1}\mathbf{X}_* \in [\mathbf{0}, \mathbf{x}]^c] \to -\log G_*(\mathbf{x})$$
$$= \mu_*([\mathbf{0}, \mathbf{x}]^c), \qquad \mathbf{x} > \mathbf{0}. \tag{5.39}$$

It is easy to extend (5.39) to convergence on rectangles and thus to vague convergence on $E$, which gives (iii).

Given (iii) we check (iv) as follows: Suppose $g$: $(0, \infty] \times \aleph \to [0, \infty)$ is continuous with compact support; we must show

$$nEg(n^{-1}\|\mathbf{X}_*\|, \|\mathbf{X}_*\|^{-1}\mathbf{X}_*)$$
$$\to \iint_{(0,\infty]\times\aleph} g(r, \mathbf{a})r^{-2}\,dr\,S(d\mathbf{a}). \tag{5.40}$$

Note that $g(n^{-1}\|\mathbf{X}_*\|, \|\mathbf{X}_*\|^{-1}\mathbf{X}_*) = g\circ P(n^{-1}\mathbf{X}_*)$ and since we know, by assuming (iii) is true that

$$nEf(n^{-1}\mathbf{X}_*) \to \int_E f\,d\mu_*$$

for $f\colon E \to [0, \infty)$ continuous with compact support, it suffices to show that $g \circ P$ is continuous with compact support on $E$. It is clear $g \circ P$ is continuous. Let the support of $g$ be $K'$. Then for some $\delta > 0$ we have $K' \subset [\delta, \infty] \times \aleph$ and the support of $g \circ P$ is

$$\begin{aligned}\{\mathbf{x}\colon g\circ P(\mathbf{x}) > 0\} &= P^{-1}(K') \subset P^{-1}([\delta, \infty] \times \aleph)\\ &= \{\mathbf{x}\in E\colon \|\mathbf{x}\| \geq \delta\},\end{aligned}$$

which is compact in $E$. Since $P$ is continuous, $P^{-1}(K')$ is a closed subset of the compact set $\{\mathbf{x}\colon \|\mathbf{x}\| \geq \delta\}$ and hence is compact.

The proof that (iv) implies (iii) is similar. To get (iii) implies (ii) plug $[0, \mathbf{x}]^c$ into the vague convergence statement and use monotonicity to switch to the continuous variable $t$. □

The information in Propositions 5.15, and 5.17 is now recast in a somewhat more easily applied form which sometimes eliminates the necessity of computing $F_*$.

**Corollary 5.18.** *$F$ is a distribution on $R^d$.*
(a) *If $F$ satisfies the regular variation condition*

$$\lim_{t\to\infty}\frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = W(\mathbf{x}) > 0, \qquad \mathbf{x} > \mathbf{0} \tag{5.41}$$

*and $W(c\mathbf{x}) = c^{-\alpha}W(\mathbf{x}), c > 0, \mathbf{x} > 0, \alpha > 0$, then $F \in D(G)$ where $G(\mathbf{x}) = \exp\{-W(\mathbf{x})\}, \mathbf{x} > \mathbf{0}$.*
(b) *Suppose $\mathbf{X}$ has distribution $F$. If there exist $a_n \to \infty$ such that*

$$\begin{aligned}&nP[(a_n^{-1}\|\mathbf{X}\|, \|\mathbf{X}\|^{-1}\mathbf{X}) \in (dr, d\mathbf{a})]\\ &\quad \xrightarrow{v} \alpha r^{-\alpha-1}\,dr\,S(d\mathbf{a})\end{aligned} \tag{5.42}$$

*on $(0, \infty] \times \{\mathbf{x}\in \mathbb{R}^d\colon \|\mathbf{x}\| = 1\}$ for a finite measure $S$, then $F$ is in the domain of attraction of the extreme value distribution $\exp\{-\mu([-\infty, \mathbf{x}]^c)\}, \mathbf{x} > \mathbf{0}$, where*

$$\mu\{\mathbf{x}\in \mathbb{R}^d\backslash\{\mathbf{0}\}\colon \|\mathbf{x}\| > r, \|\mathbf{x}\|^{-1}\mathbf{x}\in A\} = r^{-\alpha}S(A). \tag{5.43}$$

(c) *Suppose $d = 2$ and (5.42) holds on $(0, \infty] \times \{\mathbf{x} \geq \mathbf{0}\colon \|\mathbf{x}\| = 1\}$ or equivalently suppose there exists $V(t)\in RV_{-\alpha}, \alpha > 0$, and*

$$P[t^{-1}\mathbf{X}\in \cdot\,]/V(t) \xrightarrow{v} \mu$$

*on $E = [0, \infty]^2\backslash\{\mathbf{0}\}$. If in addition for the marginal distributions $F_i, i = 1, 2$ we have*

$$1 - F_i(x) \sim c_i V(x), \qquad x \to \infty, \qquad c_i > 0 \tag{5.44}$$

*then* (5.41) *holds and* $F \in D(G)$ *as in* (a) *where*

$$W(x, y) = c_1 x^{-\alpha} + c_2 y^{-\alpha} - \mu((x, y), \infty)$$

*for* $(x, y) > \mathbf{0}$.

*Remark.* Vague convergence on $E$ does not control mass in $E^c$ for a distribution $F$ which does not concentrate on $[0, \infty)^d$. For $d = 2$, marginal regular variation provides the necessary control. If $d > 2$ see Exercise 5.4.2.5.

Proof. For the proof of (a) examine the first part of the proof of Proposition 5.15. For (b), note that (5.42) implies

$$nP[a_n^{-1}\mathbf{X} \in \cdot\,] \xrightarrow{v} \mu \tag{5.45}$$

on $[-\infty, \infty]^d \backslash \{\mathbf{0}\}$, where $\mu$ has the definition given in (5.43). From (5.45) we get for $\mathbf{x} > \mathbf{0}$

$$\begin{aligned} nP[a_n^{-1}\mathbf{X} \in (-\infty, \mathbf{x}]^c] &= n(1 - F(a_n\mathbf{x})) \\ &\to \mu((-\infty, \mathbf{x}]^c) \end{aligned}$$

and this puts us essentially in the situation covered by (a).

For (c) observe that for $d = 2$ we have for $\mathbf{x} > \mathbf{0}$

$$\begin{aligned} &(1 - F(t\mathbf{x}))/V(t) \\ &\quad = (P[X^{(1)} > tx^{(1)}] + P[X^{(2)} > tx^{(2)}] - P[X^{(1)} > tx^{(1)}, X^{(2)} > tx^{(2)}])/V(t) \\ &\quad = (1 - F_1(tx^{(1)}))/V(t) + (1 - F_2(tx^{(2)}))/V(t) - P[t^{-1}\mathbf{X} \in (\mathbf{x}, \infty)]/V(t) \\ &\quad \to c_1(x^{(1)})^{-\alpha} + c_2(x^{(2)})^{-\alpha} - \mu(\mathbf{x}, \infty) =: W(\mathbf{x}) \end{aligned}$$

and we may now apply (a). □

Example 5.19. Suppose $d = 2$ and $\mathbf{X}$ has two-dimensional Cauchy distribution $F$ with density

$$F'(x, y) = (2\pi)^{-1}(1 + x^2 + y^2)^{-3/2}, \qquad (x, y) \in \mathbb{R}^2$$

(Feller, 1971, page 70). The polar coordinates of $\mathbf{X}$ are $(\|\mathbf{X}\|, \theta(\mathbf{X})) = (((X^{(1)})^2 + (X^{(2)})^2)^{1/2}, \arctan(X^{(2)}/X^{(1)}))$ and have density

$$\begin{aligned} P[\|\mathbf{X}\| \in dr, \theta(\mathbf{X}) \in d\theta] &= F'(r\cos\theta, r\sin\theta) r\,dr\,d\theta \\ &= r(1 + r^2)^{-3/2}\,dr(2\pi)^{-1}\,d\theta \end{aligned}$$

so that $\|\mathbf{X}\|$, and $\theta(\mathbf{X})$ are independent with $\theta(\mathbf{X})$ uniformly distributed on $[0, 2\pi)$ and

$$\begin{aligned} P[\|\mathbf{X}\| > r] &= (1 + r^2)^{-1/2}, \qquad r > 0 \\ &\sim r^{-1}, \qquad r \to \infty. \end{aligned}$$

The condition (5.41) of Corollary 5.18 becomes

$$\begin{aligned}
&nP[(n^{-1}\|\mathbf{X}\|, \|\mathbf{X}\|^{-1}\mathbf{X}) \in (r, \infty] \times A] \\
&= nP[\|\mathbf{X}\| > nr]P[\|\mathbf{X}\|^{-1}\mathbf{X} \in A] \\
&\to r^{-1}P[(\cos\theta(\mathbf{X}), \sin\theta(\mathbf{X})) \in A] = r^{-1}S(A).
\end{aligned}$$

Thus for $(x, y) > \mathbf{0}$ and $(r, \theta)$ the polar coordinates of $(u, v)$:

$$\begin{aligned}
\mu(-\infty, (x, y)] &= \mu\{(u, v): u > x \text{ or } v > y\} \\
&= \mu\{(u, v): r\cos\theta > x \text{ or } r\sin\theta > y, 0 < \theta \le \pi/2\} \\
&\quad + \mu\{(u, v): r\sin\theta > y, \pi/2 < \theta \le \pi\} \\
&\quad + \mu\{(u, v): r\cos\theta > x, 3\pi/2 < \theta \le 2\pi\} \\
&= \iint_{\left[\substack{r > x(\cos\theta)^{-1} \wedge y(\sin\theta)^{-1} \\ 0 \le \theta \le \pi/2}\right]} r^{-2}\, dr\, d\theta/2\pi \\
&\quad + \iint_{\left[\substack{r > y/\sin\theta \\ \pi/2 < \theta \le \pi}\right]} r^{-2}\, dr\, d\theta/2\pi \\
&\quad + \iint_{\left[\substack{r > x/\cos\theta \\ 3\pi/2 < \theta \le 2\pi}\right]} r^{-2}\, dr\, d\theta/2\pi \\
&= \int_0^{\pi/2} (x^{-1}\cos\theta) \vee (y^{-1}\sin\theta)\, d\theta/2\pi \\
&\quad + \int_{\pi/2}^{\pi} y^{-1}\sin\theta\, d\theta/2\pi \\
&\quad + \int_{3\pi/2}^{2\pi} x^{-1}\cos\theta\, d\theta/2\pi \\
&= \int_0^{\tan^{-1}(y/x)} x^{-1}\cos\theta\, d\theta/2\pi + \int_{\tan^{-1}(y/x)}^{\pi/2} y^{-1}\sin\theta\, d\theta/2\pi \\
&\quad + (x^{-1} + y^{-1})/2\pi \\
&= (2\pi)^{-1}(x^{-1} + y^{-1} + (x^{-2} + y^{-2})^{1/2}).
\end{aligned}$$

This gives

$$\begin{aligned}
F^n(n(x, y)) &\to \exp\{-(2\pi)^{-1}(x^{-1} + y^{-1} + (x^{-2} + y^{-2})^{1/2})\} \\
&= \exp\left\{-\int_0^{\pi/2} (x^{-1}\cos\theta \vee y^{-1}\sin\theta)(\varepsilon_0(d\theta) + \varepsilon_{\pi/2}(d\theta) + d\theta/2\pi)\right\}.
\end{aligned}$$

Since most multivariate distributions are specified by densities, not by distribution functions, it is important to have good criteria in terms of densities which imply the regular variation of distribution tails. When $d = 1$, regular variation of the density implies regular variation of the distribution tail via

Karamata's theorem 0.6. This, however, fails to be true in higher dimensions without the imposition of some regularity conditions. Cf. Exercise 5.4.2.11.

What distinguishes multivariate regular variation from the univariate case is that as we move from ray to ray the definition of regular variation exerts no control over the function's variation; there is only radial control as we move out along a ray. Imposing a uniformity condition as we move across rays overcomes this difficulty.

As before let $\mathbf{x} \to \|\mathbf{x}\|$ be a norm on $\mathbb{R}^d$ and set $\aleph = \{\mathbf{x} \in [0, \infty)^d : \|\mathbf{x}\| = 1]$. The following is due to de Haan and Resnick (1987). See also de Haan and Resnick (1979b), de Haan and Omey (1983); and de Haan, Omey and Resnick (1984).

**Proposition 5.20.** *Suppose $F$ concentrates on $[0, \infty)^d$ and has density $F'$ which is regularly varying with limit function $\lambda$ on $[0, \infty)^d \backslash \{\mathbf{0}\}$; i.e., for some regularly varying function $V(t)$ of index $\rho < 0$ we have for $\mathbf{x} \geq \mathbf{0}$, $\mathbf{x} \neq \mathbf{0}$*

$$\lim_{t \to \infty} \frac{F'(t\mathbf{x})}{t^{-d} V(t)} = \lambda(\mathbf{x}) > 0. \tag{5.46}$$

*Necessarily $\lambda$ satisfies $\lambda(t\mathbf{x}) = t^{\rho - d} \lambda(\mathbf{x})$ for $\mathbf{x} > 0$. $\mathbf{x} \neq \mathbf{0}$. Suppose further that $\lambda$ is bounded on $\aleph$ and that the following uniformity condition holds:*

$$\lim_{t \to \infty} \sup_{\mathbf{x} \in \aleph} \left| \frac{F'(t\mathbf{x})}{t^{-d} V(t)} - \lambda(\mathbf{x}) \right| = 0. \tag{5.47}$$

*Then for any $\varepsilon > 0$*

$$\lim_{t \to \infty} \sup_{\|\mathbf{x}\| > \varepsilon} \left| \frac{F'(t\mathbf{x})}{t^{-d} V(t)} - \lambda(\mathbf{x}) \right| = 0. \tag{5.48}$$

*Also $\lambda$ is integrable on $[0, \mathbf{x}]^c$, $\mathbf{x} > \mathbf{0}$ and $1 - F$ is regularly varying on $(0, \infty)^d$, i.e.,*

$$\lim_{t \to \infty} \frac{1 - F(t\mathbf{x})}{V(t)} = \int_{[0, \mathbf{x}]^c} \lambda(\mathbf{u}) d\mathbf{u}, \qquad \mathbf{x} > \mathbf{0}. \tag{5.49}$$

*Remarks.*

(1) If (5.47) holds for $\|\mathbf{x}\|$, it also holds for any other norm $\|\mathbf{x}\|^*$. This follows immediately from (5.48).

(2) Since $\aleph$ is compact, continuity of $\lambda$ on $\aleph$ of course implies that $\lambda$ is bounded on $\aleph$. If $\lambda$ is monotone, then $\lambda$ is continuous on $(0, \infty)^d$ and hence on $(0, \infty)^d \cap \aleph$, but it is not necessarily true that $\lambda$ is continuous on all of $\aleph$.

(3) Note in (5.46) that convergence on the boundary $\bigcup_{i=1}^d \{\mathbf{x} \geq 0 : \mathbf{x} \neq \mathbf{0}, x^{(i)} = 0\}$ is required.

(4) The form of (5.46) is suggested by taking partial derivatives in (5.35).

Proof of Proposition 5.20. We begin by showing that (5.48) follows from (5.47). We have with $h(t) = t^{-d} V(t) \in RV_{\rho - d}$

$$\begin{aligned}
&\sup_{\|\mathbf{x}\|>\varepsilon}\left|\frac{F'(t\mathbf{x})}{t^{-d}V(t)}-\lambda(\mathbf{x})\right| \\
&= \sup_{\|\mathbf{x}\|>\varepsilon}\left|\frac{F'(t\|\mathbf{x}\|(\|\mathbf{x}\|^{-1}\mathbf{x}))}{h(t\|\mathbf{x}\|)}\frac{h(t\|\mathbf{x}\|)}{h(t)}-\lambda(\|\mathbf{x}\|\,\|\mathbf{x}\|^{-1}\mathbf{x})\right| \\
&\le \sup_{\|\mathbf{x}\|>\varepsilon}\|\mathbf{x}\|^{-d+\rho}\left|\frac{F'(t\|\mathbf{x}\|\,\|\mathbf{x}\|^{-1}\mathbf{x})}{h(t\|\mathbf{x}\|)}-\lambda(\|\mathbf{x}\|^{-1}\mathbf{x})\right| \\
&+ \sup_{\|\mathbf{x}\|>\varepsilon}\left|\frac{F'(t\|\mathbf{x}\|(\|\mathbf{x}\|^{-1}\mathbf{x}))}{h(t\|\mathbf{x}\|)}\right|\left|\frac{h(t\|\mathbf{x}\|)}{h(t)}-\|\mathbf{x}\|^{-d+\rho}\right| = I + II. \qquad (5.50)
\end{aligned}$$

Given any $\delta > 0$, there exists $t_0$ such that for $t \ge t_0$ the sup in (5.47) is less than $\delta$. Since $\|\mathbf{x}\|^{-1}\mathbf{x} \in \aleph$ we have for $t \ge \varepsilon^{-1}t_0$ that $I \le \delta\varepsilon^{-d+\rho}$ and since $\delta$ is arbitrary, we conclude that $\lim_{t\to\infty} I = 0$.

For $II$ note that since $\|\mathbf{x}\|^{-1}\mathbf{x} \in \aleph$ we have by (5.47) that for given $\eta > 0$ and large $t$

$$\sup_{\|\mathbf{x}\|>\varepsilon}\frac{F'(t\|\mathbf{x}\|\,\|\mathbf{x}\|^{-1}\mathbf{x})}{h(t\|\mathbf{x}\|)} \le \sup_{x\in\aleph}\lambda(\mathbf{x}) + \eta < \infty \qquad (5.51)$$

by assumption that $\lambda$ is bounded on $\aleph$. Since $h$ is regularly varying with exponent $\rho - d < 0$, one-dimensional uniform convergence (Proposition 0.5) gives

$$\lim_{t\to\infty}\sup_{y>\varepsilon}\left|\frac{h(ty)}{h(t)} - y^{-d+\rho}\right| = 0$$

and thus $\lim_{t\to\infty} \mathrm{II} = 0$. This verifies (5.48).

For the rest of the proof we may without loss of generality suppose that

$$\|\mathbf{x}\| = \bigvee_{i=1}^{d} |x^{(i)}|$$

so that for $y > 0$

$$\int_{\{\mathbf{u}\in\mathbb{R}_+^d:\|\mathbf{u}\|\le y\}} d\mathbf{u} = y^d. \qquad (5.52)$$

With this choice of norm (and hence with any norm) one can readily verify $\int_{[\|\mathbf{x}\|>\varepsilon]}\lambda(\mathbf{x})d\mathbf{x} < \infty$ since by Fubini's theorem or a change of variables and (5.52)

$$\begin{aligned}
\int_{[\|\mathbf{x}\|>\varepsilon]}\lambda(\mathbf{x})d\mathbf{x} &= \int_{[\|\mathbf{x}\|>\varepsilon]}\|\mathbf{x}\|^{-d-\rho}\lambda(\|\mathbf{x}\|^{-1}\mathbf{x})d\mathbf{x} \le \sup_{\mathbf{a}\in\aleph}\lambda(\mathbf{a})\int_{(\varepsilon,\infty)} r^{-d+\rho}\,dr^d \\
&= d\left(\sup_{\mathbf{a}\in\aleph}\lambda(\mathbf{a})\right)\int_{\varepsilon}^{\infty} r^{-d+\rho+d-1}\,dr < \infty
\end{aligned}$$

since $\rho < 0$.

It remains to prove (5.49). Let $A$ be any Borel set such that for some $\varepsilon > 0$ we have $A \subset [\|\mathbf{x}\| \ge \varepsilon]$. For $\mathbf{x} \in A$ we construct an integrable bound for

$$\frac{F'(t\mathbf{x})}{h(t)} = \frac{F'(t\|\mathbf{x}\| \, \|\mathbf{x}\|^{-1}\mathbf{x})}{h(t\|\mathbf{x}\|)} \cdot \frac{h(t\|\mathbf{x}\|)}{h(t)} \tag{5.53}$$

as follows: The first factor on the right side of (5.53) is bounded by a constant as in (5.51) and the second factor has upper bound (cf. Proposition 0.8(ii)) $c\|\mathbf{x}\|^{\rho-d-\gamma}$ for sufficiently large $t$ where $\gamma$ is chosen so small that $0 < \gamma < -\rho$. Thus on $A$ we get for all large $t$

$$\frac{F'(t\mathbf{x})}{h(t)} \leq c_1 \|\mathbf{x}\|^{\rho-d+\gamma}$$

and the right side is Lebesgue integrable on $[\|\mathbf{x}\| \geq \varepsilon]$ and hence on $A$. From dominated convergence and (5.46) we get

$$\lim_{t\to\infty} \int_A \frac{F'(t\mathbf{u})}{h(t)} d\mathbf{u} = \int_A \lambda(\mathbf{u})\, d\mathbf{u}$$

and setting $A = [\mathbf{0}, \mathbf{x}]^c$ for $\mathbf{x} > \mathbf{0}$ gives the desired result (5.49). □

Example 5.21. Consider the bivariate $t$-density (Johnson and Kotz, 1972) defined on $[0, \infty)^2$ by $(-1 < \rho < 1)$

$$F'(x, y) = c(1 + x^2 + 2\rho xy + y^2)^{-2}.$$

Set $\|(x, y)\|^2 = x^2 + 2\rho xy + y^2 = (x + \rho y)^2 + (1 - \rho^2)y^2$ and check that this defines a norm. Thus the density can be expressed as

$$F'(\mathbf{x}) = c(1 + \|\mathbf{x}\|^2)^{-2}$$

and it is obvious that the uniformity condition (5.47) is satisfied.

Whenever the density is of the form

$$F'(\mathbf{x}) = c(1 + \|\mathbf{x}\|^\alpha)^{-\beta}$$

$\alpha > 0$, $\beta > 0$ for some clever choice of the norm $\mathbf{x} \to \|\mathbf{x}\|$, the condition (5.47) will be clearly satisfied. Many densities are of this form. The choice of the norm is suggested by the following scheme: If $F'$ is regularly varying with limit function $\lambda$ and $\lambda(t\mathbf{x}) = t^{-\alpha}\lambda(\mathbf{x})$, $\mathbf{x} > \mathbf{0}$, we try setting

$$\|\mathbf{x}\| := \lambda^{-1/\alpha}(\mathbf{x})$$

and hope this defines a norm. The limit function $\lambda$ is a function only of $\|\mathbf{x}\|$, and possibly the same will be true for $F'$. For instance, in the case of Example 5.21 earlier

$$\lambda(x, y) = (x^2 + 2\rho xy + y^2)^{-2}, \qquad \lambda(t(x, y)) = t^{-4}\lambda(x, y)$$

and so

$$\|(x, y)\| = \lambda^{-1/4}(x, y) = (x^2 + 2\rho xy + y^2)^{1/2}.$$

Example 5.22. Consider the multivariate $F$-density (Johnson and Kotz, 1972, page 240): Let $\nu_0, \ldots, \nu_d$ be positive integers, $\nu = \sum_{i=0}^d \nu_j$, $\mathbf{x} > \mathbf{0}$, and for a suitable $c > 0$

$$F'(\mathbf{x}) = c\frac{\prod_{j=1}^{d}(x^{(j)})^{(1/2)\nu_j-1}}{(\nu_0 + \sum_{j=1}^{d}\nu_j x^{(j)})^{\nu/2}}.$$

Set $\|\mathbf{x}\| = \sum_{j=1}^{d}\nu_j x^{(j)}$ and for $\mathbf{x} > \mathbf{0}$ and $t \to \infty$

$$F'(t\mathbf{x}) = \frac{ct^{\sum_{j=1}^{d}((1/2)\nu_j-1)}\prod_{j=1}^{d}(x^{(j)})^{(1/2)\nu_j-1}}{(\nu_0 + t\|\mathbf{x}\|)^{\nu/2}}$$

$$\sim ct^{-(1/2)\nu_0-d}\prod_{j=1}^{d}(x^{(j)})^{(1/2)\nu_j-1}/\|\mathbf{x}\|^{\nu/2}$$

so that $\rho = -(1/2)\nu_0$, $h(t) = ct^{-(1/2)\nu_0-d}$, and $\lambda(\mathbf{x}) = \prod_{j=1}^{d}(x^{(j)})^{(1/2)\nu_j-1}/\|\mathbf{x}\|^{\nu/2}$. So (5.46) holds and for (5.47) we have

$$\sup_{\|\mathbf{x}\|=1}\left|\frac{F'(t\mathbf{x})}{h(t)} - \lambda(\mathbf{x})\right| = \sup_{\|\mathbf{x}\|=1}\prod_{j=1}^{d}(x^{(j)})^{(1/2)\nu_j-1}\left|\frac{t^{\nu/2}}{(\nu_0+t)^{\nu/2}} - 1\right|.$$

The region $[\|\mathbf{x}\| = 1]$ is compact and since $\prod_{j=1}^{d}(x^{(j)})^{(1/2)\nu_j-1}$ is continuous it is also bounded, and hence the condition (5.47) holds.

We have assumed for simplicity that $F$ concentrates on $[0, \infty)^d$. When this is not the case we have the following result. It is stated for $d = 2$. For $d > 2$ a similar but more complicated result can be formulated; see Exercise 5.4.2.6 and Corollary 5.18.

**Corollary 5.23.** *Suppose $\mathbf{X}$ is a random vector in $\mathbb{R}^2$ with distribution function $F$ and density $F'$ satisfying* (5.46) *and* (5.47). *If in addition for $x > 0$*

$$\lim_{t\to\infty} P[X^{(i)} > tx]/V(t) = c_i x^{\rho}, \qquad c_i > 0 \tag{5.44}$$

*for $i = 1, 2$, then $1 - F$ is regularly varying on $(0, \infty)^2$: for $\mathbf{x} > \mathbf{0}$*

$$\lim_{t\to\infty}(1 - F(t\mathbf{x}))/V(t) = c_1(x^{(1)})^{\rho} + c_2(x^{(2)})^{\rho} - \int_{[\mathbf{u}>\mathbf{x}]}\lambda(\mathbf{u})d\mathbf{u}. \tag{5.54}$$

Proof. The proof follows simply from

$$\frac{1 - F(t\mathbf{x})}{V(t)} = \frac{P([\mathbf{X} \le t\mathbf{x}]^c)}{V(t)}$$

$$= (P[X^{(1)} > tx^{(1)}] + P[X^{(2)} > tx^{(2)}] - P[\mathbf{X} > t\mathbf{x}])/V(t).$$

The method in Proposition 5.20 shows that the last term converges. □

Example 5.19 (Continued). The bivariate Cauchy

$$F'(x, y) = \frac{1}{2\pi}(1 + x^2 + y^2)^{-3/2}, \qquad (x, y) \in \mathbb{R}^2$$

obviously satisfies (5.44), and if we set $\|\mathbf{x}\| = ((x^{(1)})^2 + (x^{(2)})^2)^{1/2}$ we get

$$F'(\mathbf{x}) = \frac{1}{2\pi}(1 + \|\mathbf{x}\|^2)^{-3/2}$$

and clearly (5.47) is satisfied with $\lambda(\mathbf{x}) = \|\mathbf{x}\|^{-3}/2\pi$. With $V(t) = t^{-1}$ and $c_1 = c_2 = \pi^{-1}$ the limit is (5.54) at the point $(x, y) > 0$ is

$$\pi^{-1}(x^{-1} + y^{-1}) - \int_x^\infty \int_y^\infty \frac{1}{2\pi(u^2 + v^2)^{3/2}}\,du\,dv.$$

For the double integral make the change of variable $v = u \tan\theta$ to get

$$\begin{aligned}\int_x^\infty \int_y^\infty &= \int_x^\infty (2\pi u^2)^{-1}\,du \int_{\tan^{-1}(u^{-1}y)}^{\pi/2} \cos\theta\,d\theta \\ &= (2\pi)^{-1}\left\{x^{-1} - \int_x^\infty yu^{-2}(u^2 + y^2)^{-1/2}\,du\right\}\end{aligned}$$

and setting u = y tan $\theta$ yields

$$\int_x^\infty \int_y^\infty = (2\pi)^{-1}\{x^{-1} + y^{-1} - (x^{-2} + y^{-2})^{1/2}\},$$

and so the limit in (5.44) is

$$\begin{aligned}&\pi^{-1}(x^{-1} + y^{-1}) - (2\pi)^{-1}\{x^{-1} + y^{-1} - (x^{-2} + y^{-2})^{1/2}\} \\ &= (2\pi)^{-1}\{x^{-1} + y^{-1} + (x^{-2} + y^{-2})^{1/2}\}\end{aligned}$$

which is in agreement with the calculation of the exponent of the limit distribution done previously by using polar coordinates.

All of the criteria given so far are easiest to apply when tails are Pareto-like but may be clumsy for other cases. The multivariate normal is handled by asymptotic independence as discussed in the next section.

It is possible to give partial converses of Proposition 5.20. See de Haan and Resnick (1979b); de Haan and Omey (1983); and de Haan, Omey, and Resnick (1984).

## Exercises

5.4.2.1. Give an example of a function $h$ defined on a cone $C \subset \mathbb{R}^d$ such that for each $\mathbf{x} \in C$, $h(t\mathbf{x})$ is a regularly varying function of $t$ with exponent of variation $\rho$ not depending on $\mathbf{x}$, but yet $h$ does not satisfy Definition 5.32 (Stam, 1977).

5.4.2.2. Discuss a proof of Proposition 5.15 based on Exercise 5.4.1.12.

5.4.2.3. Suppose $F$ concentrates on $[0, \infty)^d$ and for any $\mathbf{t} \in (0, \infty)^d$

$$\lim_{x\to\infty} x \frac{\sum_{i=1}^d t^{(i)} \frac{\partial}{\partial x^{(i)}} F(x\mathbf{t})}{1 - F(x\mathbf{t})} = 1.$$

Then $F \in D(G)$ where $G$ has $\Phi_1$ marginals. (Hint: Use Exercises 5.4.1.6, 5.4.1.12, and (1.19)). Use the criterion to verify that

$$F(x, y) = 1 - \{x^{-1} + y^{-1} - (x + y - 1)^{-1}\}$$

is in a domain of attraction.

5.4.2.4. If $0 \leq h_n \leq g_n$ are real functions on some measure space and $h_n \to h$, $g_n \to g$, and $\int g_n \to \int g$ then $\int h_n \to \int h$ provided $\int h < \infty$ (Johns, 1957; Pratt, 1960).

5.4.2.5. Suppose $V(t) \in RV_{-\alpha}$, $\alpha > 0$, and $\mathbf{X}$ is a random vector in $\mathbb{R}^d$. If for $l = 1, \ldots, d$ and all choices $1 \leq n_1 < \cdots < n_l \leq d$

$$\lim_{t \to \infty} P[X^{(n_i)} > tx^{(n_i)}, i = 1, \ldots, l]/V(t)$$

exists positive and finite for $x^{(n_i)} > 0$, $i = 1, \ldots, l$ then $P([\mathbf{X} \leq \mathbf{x}]^c)$ is regularly varying on $(0, \infty)^d$.

5.4.2.6. Formulate a version of Corollary 5.23 for $d > 2$ by using Exercise 5.4.2.5.

5.4.2.7. Discuss which of the following are in a domain of attraction. Specify the domain and the limit distribution where appropriate.

(i) Bivariate lognormal: Let ($N^{(1)}$ and $N^{(2)}$) have a bivariate normal density and set

$$(X^{(1)}, X^{(2)}) = (\exp\{N^{(1)}\}, \exp\{N^{(2)}\}).$$

(ii) Multivariate $t$-density on $\mathbb{R}^d$:

$$F'(\mathbf{x}) = \frac{\Gamma(2^{-1}(v + d))}{(\pi v)^{d/2}\Gamma(v/2)|R|^{1/2}}(1 + v^{-1}\mathbf{x}R^{-1}\mathbf{x}')^{-(v+d)/2}$$

where $R$ is a $d \times d$ covariance matrix, $v$ a non-negative integer. A special case is the multivariate Cauchy density on $\mathbb{R}^d$

$$F'(\mathbf{x}) = \frac{\Gamma(2^{-1}(1 + d))}{(\pi)^{d/2}\Gamma(1/2)}\left(1 + \sum_{i=1}^{d}(x^{(i)})^2\right)^{-(d+1)/2}$$

(Johnson and Kotz, 1972, page 134).

(iii) A bivariate gamma density; Suppose $x > 0$, $y > 0$, and $w = x \wedge y$, and let $p$ be a positive integer. Then

$$F'(x, y) = e^{-(x+y)}(-1)^p\left\{1 - e^w\left[1 - w + \frac{w^2}{2!} - \cdots + (-1)^{p-1}\frac{w^{p-1}}{(p-1)!}\right]\right\}$$

(Johnson and Kotz, 1972, page 218).

(iv) Multivariate $F$: Let $v_0, \ldots, v_d$ be positive integers, $v = \sum_{j=0}^{d} v_j$, $\mathbf{x} > \mathbf{0}$, and

$$F'(x) = \frac{\Gamma(v/2)\prod_{j=0}^{d} v_j^{v_j/2}}{\prod_{j=0}^{d}\Gamma(v_j/2)} \frac{\prod_{j=1}^{m}(x^{(j)})^{2^{-1}v_j - 1}}{(v_0 + \sum_{j=1}^{d} v_j x^{(j)})^{v/2}}.$$

(Johnson and Kotz, 1972, page 240).

(v) Marshall–Olkin bivariate exponential: Let $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_{12} > 0$, and for $x > 0$, $y > 0$

$$P[X^{(1)} > x, X^{(2)} > y] = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_{12}(x \vee y)\}$$

(Johnson and Kotz, 1972, page 266).

(vi) Multivariate logistic: For $x \in \mathbb{R}^d$

$$F(\mathbf{x}) = \left(1 + \sum_{i=1}^{d} \exp\{-x^{(i)}\}\right)^{-1}$$

(Johnson and Kotz, 1972, page 291).

5.4.2.8. Consider the density

$$F'(x, y) = \begin{cases} (xy)^{-2} & \text{if } x \geq 1, y \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

so that for $x > 0$, $y > 0$, and $t$ large enough for $t(x, y) \geq (1, 1)$

$$F'(tx, ty)/F'(t, t) = (xy)^{-2},$$

whence

$$\lambda(x, y) = (xy)^{-2} \qquad x > 0, \qquad y > 0.$$

Proposition 5.20 fails. Why (de Haan and Omey, 1983)?

5.4.2.9. Let $U$: $(0, \infty)^d \to (0, \infty)$ be monotone and suppose $a_i(t) \in RV_{\alpha_i}$, $\alpha_i > 0$, $i = 1, \ldots, d$. If there exists a function $V(t) > 0$ such that for $\mathbf{x} > \mathbf{0}$

$$\lim_{t \to \infty} U(a_1(t)x^{(1)}, \ldots, a_d(t)x^{(d)})/V(t) = \lambda(\mathbf{x}) > 0$$

then

$$\lim_{t \to \infty} U(a_1(tx^{(1)}), \ldots, a_d(tx^{(d)}))/V(t) = \lambda((x^{(1)})^{\alpha_1}, \ldots, (x^{(d)})^{\alpha_d})$$

and so the function $U(a_1(x^{(1)}), \ldots, a_d(x^{(d)}))$ is regularly varying (de Haan, Omey, and Resnick, 1984).

5.4.2.10. Give an example of $\lambda$ on $[0, \infty)^d$ such that $\lambda$ is monotone, $\lambda(s\mathbf{x}) = s^\rho \lambda(\mathbf{x})$. $\mathbf{x} \geq \mathbf{0}$, $\rho \in \mathbb{R}$, $s > 0$, and $\lambda$ is continuous on $(0, \infty)^d$ but not continuous at points of $[0, \infty)^d \backslash (0, \infty)^d$.

5.4.2.11. Give an example of a regularly varying density on $[0, \infty)^2$ such that the distribution tail fails to be regularly varying (de Haan and Resnick, 1987).

5.4.2.12. Show in Proposition 5.20 that if $F'$ is continuous on $[\|\mathbf{x}\| > 1]$ then $\lambda$ is continuous on $[0, \infty)^d \backslash \{\mathbf{0}\}$. Give an example to show that without (5.47), continuity of $F'$ does not imply $\lambda$ continuous (de Haan and Resnick, 1987).

## 5.5. Independence and Dependence

Suppose $\mathbf{X}$ is a random $d$-dimensional vector with distribution $F$. If $F$ is max-id or a multivariate extreme value distribution, when is $\mathbf{X}$ a vector of independent components so that $F(\mathbf{x}) = \prod_{i=1}^{d} F_i(x^{(i)})$? If we know $F \in D(G)$, what conditions on $F$ guarantee $G$ is a product measure? Similar questions can be posed for full dependence. For example if $F$ is max-id, when is it the case that $P[X^{(i)} = X^{(j)}] = 1$, for all $1 \leq i < j \leq d$?

Initially we discuss independence. (Cf. Exercise 5.3.1.)

**Proposition 5.24.** *Suppose* $\mathbf{X}$ *has a max-id distribution with exponent measure* $\mu$ *concentrating on* $E := [\mathbf{l}, \boldsymbol{\infty}]\backslash\{\mathbf{l}\}$, $\mathbf{l} \in [-\infty, \infty)$. *The following are equivalent.*
(i) *The components of* $\mathbf{X}$, *namely* $X^{(1)}, \ldots, X^{(d)}$, *are independent random variables.*
(ii) *The components of* $\mathbf{X}$ *are pairwise independent: For every* $1 \le i < j \le d$.

$$X^{(i)} \quad \textit{and} \quad X^{(j)}$$

*are independent random variables.*
(iii) *The exponent measure* $\mu$ *concentrates on*

$$\bigcup_{i=1}^{d} \{l^{(1)}\} \times \cdots \times (l^{(i)}, \infty) \times \cdots \times \{l^{(d)}\} \tag{5.55}$$

*so that for* $\mathbf{y} > \mathbf{l}$

$$\mu\left( \bigcup_{1 \le i < j \le d} \{\mathbf{x} \in E : x^{(i)} > y^{(i)}, x^{(j)} > y^{(j)}\} \right) = 0. \tag{5.56}$$

*Remark.* When $d = 2$ and $\mathbf{l} = 0$, (iii) says that $\mu$ concentrates on the positive coordinate axes and has no mass in the interior of the first quadrant.

Proof. Obviously (i) implies (ii) and it is easy to check that (iii) implies (i) as follows: Suppose $\mu$ concentrates on the set in (5.55). Then for $\mathbf{x} > \mathbf{l}$

$$\begin{aligned}
-\log F(\mathbf{x}) &= \mu(E\backslash[\mathbf{l}, \mathbf{x}]) \\
&= \mu\left( \bigcup_{i=1}^{d} \{\mathbf{u} \in E : u^{(i)} > x^{(i)}\} \right) \\
&= \sum_{i=1}^{d} \mu\{\mathbf{u} \in E : u^{(i)} > x^{(i)}\} \\
&\quad - \sum_{1 \le i < j \le d} \mu\{\mathbf{u} \in E : u^{(i)} > x^{(i)}, u^{(j)} > x^{(j)}\} \\
&\quad + \cdots + (-1)^{d+1} \mu\{\mathbf{u} \in E : u^{(i)} > x^{(i)}, i = 1, \ldots, d\}
\end{aligned}$$

so that because of (5.56) we have

$$-\log F(\mathbf{x}) = \sum_{i=1}^{d} \mu\{\mathbf{u} \in E : u^{(i)} > x^{(i)}\}.$$

Set

$$\begin{aligned}
Q_i(x^{(i)}) &= \mu\{\mathbf{u} \in E : u^{(i)} > x^{(i)}\} \\
&= \mu(\{l^{(1)}\} \times \cdots \times (x^{(i)}, \infty) \times \cdots \times \{l^{(d)}\})
\end{aligned}$$

and we have

$$-\log F(\mathbf{x}) = \sum_{i=1}^{d} Q_i(x^{(i)})$$

so that

$$F(\mathbf{x}) = \prod_{i=1}^{d} \exp\{-Q_i(x^{(i)})\}$$

and thus $F$ is a product measure as desired.

It remains to show that (ii) implies (iii). Set $Q_i(y) = -\log P[X^{(i)} \le y]$. We have for $\mathbf{y} > \boldsymbol{l}$ that pairwise independence implies

$$Q_i(y^{(i)}) + Q_j(y^{(j)}) = -\log F(\infty, \ldots, \infty, y^{(i)}, \infty, \ldots, \infty, y^{(j)}, \infty, \ldots, \infty).$$

Since $F(\mathbf{x}) = \exp\{-\mu(E\backslash[\boldsymbol{l}, \mathbf{x}])\}$ for $\mathbf{x} > \boldsymbol{l}$ we have

$$\begin{aligned} Q_i(y^{(i)}) + Q_j(y^{(j)}) &= \mu(\{\mathbf{x}: x^{(i)} > y^{(i)}\} \cup \{\mathbf{x}: x^{(j)} > y^{(j)}\}) \\ &= \mu\{\mathbf{x}: x^{(i)} > y^{(i)}\} + \mu\{\mathbf{x}: x^{(j)} > y^{(j)}\} \\ &\quad - \mu\{\mathbf{x}: x^{(i)} > y^{(i)}, x^{(j)} > y^{(j)}\} \\ &= Q_i(y^{(i)}) + Q_j(y^{(i)}) - \mu\{\mathbf{x}: x^{(i)} > y^{(i)}, x^{(j)} > y^{(j)}\} \end{aligned}$$

and thus

$$\mu\{\mathbf{x} \in E: x^{(i)} > y^{(i)}, x^{(j)} > y^{(j)}\} = 0$$

so that (5.56) holds; this is equivalent to $\mu$ concentrating on the set in (5.55). □

The equivalence of (i) and (ii) in Proposition 5.24 has been observed by various authors from various perspectives. See, for example, Berman (1961) and Marshall and Olkin (1983). An analogous result is also well known in the context of infinite divisibility for sums of independent random variables.

We now restate Proposition 5.24 for the case that $F$ is not only max-id but also a multivariate extreme value distribution. We consider the case where $\boldsymbol{l} = 0$. Set

$$(\mathbf{e}_1, \ldots, \mathbf{e}_d) = ((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1))$$

as the basis vectors in $\mathbb{R}^d$. Suppose the norm $\|\mathbf{x}\|$ on $\mathbb{R}^d$ has been defined in such a way that $\|\mathbf{e}_i\| = 1$, $i = 1, \ldots, d$.

**Corollary 5.25.** *Suppose* $\mathbf{X}$ *is a random vector in* $\mathbb{R}^d$ *with the multivariate extreme value distribution* $G$ *with* $\Phi_1$ *marginals as described in Proposition* 5.11. *The following are equivalent*:

(i) *The components of* $\mathbf{X}$, *namely* $X^{(1)}, \ldots, X^{(d)}$, *are independent random variables.*

(ii) *The components of* $\mathbf{X}$ *are pairwise independent.*

(iii) *The measure* $S$ *defined on* $\aleph = \{\mathbf{y} \in E: \|\mathbf{y}\| = 1\}$, *described in Proposition* 5.11(ii), *concentrates on* $\{\mathbf{e}_i, 1 \le i \le d\}$.

(iv) *The functions* $f_i$, $1 \le i \le d$ *on* $[0, 1]$, *defined in Proposition* 5.11(iii), *satisfy*

$$f_i(s) f_j(s) = 0 \qquad \text{a.e. } i \ne j.$$

Proof. The equivalence of (i) and (ii) is covered in Proposition 5.24. If $S$ concentrates on $\{\mathbf{e}_i, 1 \le i \le d\}$ then from the representation in Proposition 5.11(ii), for $\mathbf{x} > \mathbf{0}$,

$$-\log G(\mathbf{x}) = \int_{\aleph} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}}\right) S(d\mathbf{a})$$
$$= \sum_{j=1}^{d} \int_{\{\mathbf{e}_j\}} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}}\right) S(d\mathbf{a}) = \sum_{j=1}^{d} (1/x^{(j)}) S(\mathbf{e}_j),$$

and thus $G$ is a product of its marginals. Conversely, if $G$ is a product then the exponent measure $\mu$ of $G$ satisfies (5.55); i.e., $\mu$ concentrates on $\bigcup_{i=1}^{d} \{t\mathbf{e}_i: t > 0\}$. Recalling that for a Borel set $A \subset \aleph$

$$S(A) = \mu\{\mathbf{x} \in E: \|\mathbf{x}\| > 1, \|\mathbf{x}\|^{-1}\mathbf{x} \in A\}$$

we have

$$S(A) = \sum_{i=1}^{d} \mu(\{\mathbf{x} \in E: \|\mathbf{x}\| > 1, \|\mathbf{x}\|^{-1}\mathbf{x} \in A\} \cap \{t\mathbf{e}_i: t > 0\})$$
$$= \sum_{i=1}^{d} \mu\{t\mathbf{e}_i: t > 1, \mathbf{e}_i \in A\}.$$

For the special case $A = \{\mathbf{e}_i\}$ this gives

$$S(\{\mathbf{e}_i\}) = \mu\{t\mathbf{e}_i: t > 1\}$$

and thus for general $A$

$$S(A) = \sum_{\mathbf{e}_i \in A} S(\{\mathbf{e}_i\}),$$

showing that $S$ concentrates on $\{\mathbf{e}_i, 1 \le i \le d\}$.

To understand (iv) suppose that (iii) holds and recall from the proof of Proposition 5.11(iii) that $(f_i, 1 \le i \le d)$ can be chosen so that if $U$ is a uniform random variable then

$$(f_i(U), 1 \le i \le d)$$

has distribution $S(\cdot)/S(\aleph)$ and thus

$$1 = P[(f_i(U), 1 \le i \le d) \in \{\mathbf{e}_i, 1 \le i \le d\}]$$

whence

$$P[f_i(U) f_j(U) = 0] = 1$$

for $1 \le i < d$. Conversely if we are given that

$$V := \{s \in [0,1]: f_i(s) f_j(s) = 0, 1 \le i < j \le d\}$$

has Lebesgue measure 1 then for $\mathbf{x} > \mathbf{0}$

$$-\log G(x) = \int_{[0,1]} \bigvee_{i=1}^{d} \left(\frac{f_i(s)}{x^{(i)}}\right) ds = \int_{V} \bigvee_{i=1}^{d} \left(\frac{f_i(s)}{x^{(i)}}\right) ds$$
$$= \sum_{i=1}^{d} \int_{V \cap [f_i > 0]} f_i(s) ds / x^{(i)}$$

and again $G$ is a product. □

This covers the case where the components of $\mathbf{X}$ are independent. At the other extreme the components of $\mathbf{X}$ are fully dependent; that is,

$$P[X^{(1)} = X^{(2)} = \cdots = X^{(d)}] = 1. \tag{5.57}$$

This can be rephrased that there exists a random variable $Y$ such that in $\mathbb{R}^d$

$$\mathbf{X} \stackrel{\mathrm{d}}{=} Y\mathbf{1} \tag{5.58}$$

where recall the notation $\mathbf{1} = (1, \ldots, 1)$. This situation can also be recognized in terms of distribution functions. If $\mathbf{X}$ has distribution $F$ and $Y$ has distribution $H$ then

$$F(\mathbf{x}) = H\left(\bigwedge_{i=1}^{d} x^{(i)}\right). \tag{5.59}$$

**Proposition 5.26** (Conditions for Full Dependence).
(i) *Suppose that $\mathbf{X} > \mathbf{0}$ has max-id distribution $F$ with exponent $\mu$. Then any of* (5.57)–(5.59) *holds* iff *$\mu$ concentrates on the line*

$$\{t\mathbf{1}: t > 0\}.$$

(ii) *Suppose that $\mathbf{X} \geq 0$ has a multivariate extreme value distribution $G$ as described in Proposition* 5.11(ii). *Then any of* (5.57)–(5.59) *holds* iff *the measure $S$ concentrates on* $\{\|\mathbf{1}\|^{-1}\mathbf{1}\}$. *In terms of the representation* (5.30) *of Proposition* 5.11(iv), *full dependence holds* iff

$$f_1 = f_2 = \cdots = f_d$$

a.e. *on* $[0, 1]$.

Proof. (i) If $\mu$ concentrates on the line, we have for $\mathbf{x} > \mathbf{0}$, ($E = [0, \infty]^d \backslash \{\mathbf{0}\}$)

$$\begin{aligned}
-\log F(\mathbf{x}) &= \mu\left(\bigcup_{i=1}^{d} \{\mathbf{y} \in E: y^{(i)} > x^{(i)}\}\right)\\
&= \mu\left(\bigcup_{i=1}^{d} \{\mathbf{y} \in E: y^{(i)} > x^{(i)}\} \cap \{t\mathbf{1}: t > 0\}\right)\\
&= \mu\left(\bigcup_{i=1}^{d} \{t\mathbf{1}: t > x^{(i)}\}\right) = \mu\left\{t\mathbf{1}: t > \bigwedge_{i=1}^{d} x^{(i)}\right\}.
\end{aligned}$$

Set

$$H(u) = \exp\{-\mu\{t\mathbf{1}: t > u\}$$

and we get

$$F(\mathbf{x}) = H\left(\bigwedge_{i=1}^{d} x^{(i)}\right)$$

so that (5.59) holds.

Conversely, if (5.59) holds, let $\{Y(t), t > 0\}$ be the (one-dimensional) extremal-

$H$ process. We know, for instance, from the construction in Proposition 5.8, that

$$nF^{n^{-1}}(\cdot) \xrightarrow{v} \mu(\cdot)$$

on $E$ and this can be rephrased

$$nP[Y(n^{-1})\mathbf{1} \in \cdot] \xrightarrow{v} \mu(\cdot).$$

The measures on the left side of the preceding relation put zero mass off the line $\{t\mathbf{1}: t > 0\}$ and hence the same is true of $\mu$.
(ii) If full dependence holds, then by (i) we know that $\mu$ concentrates on the line $\{t\mathbf{1}: t > 0\}$ and hence for a Borel set $A \subset \aleph$

$$\begin{aligned} S(A) &= \mu\{\mathbf{y} \in E: \|\mathbf{y}\| > 1, \|\mathbf{y}\|^{-1}\mathbf{y} \in A\} \\ &= \mu(\{\mathbf{y} \in E: \|\mathbf{y}\| > 1, \|\mathbf{y}\|^{-1}\mathbf{y} \in A\} \cap \{t\mathbf{1}: t > 0\}) \\ &= \mu\{t\mathbf{1}: t\|\mathbf{1}\| > 1, \|\mathbf{1}\|^{-1}\mathbf{1} \in A\} \\ &= \mu\{t\mathbf{1}: t\|\mathbf{1}\| > 1\}\varepsilon_{\|\mathbf{1}\|^{-1}\mathbf{1}}(A) \end{aligned}$$

so that $S$ concentrates mass $\mu\{t\mathbf{1}: t\|\mathbf{1}\| > 1\}$ on $\|\mathbf{1}\|^{-1}\mathbf{1}$. Conversely if $S$ concentrates on $\{\|\mathbf{1}\|^{-1}\mathbf{1}\}$ then for $\mathbf{x} > \mathbf{0}$

$$\begin{aligned} -\log F(\mathbf{x}) &= \int_{\aleph} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}}\right) S(d\mathbf{a}) \\ &= S\{\|\mathbf{1}\|^{-1}\mathbf{1}\}\,\|\mathbf{1}\|^{-1} \bigvee_{i=1}^{d} (1/x^{(i)}). \end{aligned}$$

If (5.27) is assumed to hold then $S\{\|\mathbf{1}\|^{-1}\mathbf{1}\}\,\|\mathbf{1}\|^{-1} = 1$ and we get

$$F(\mathbf{x}) = \exp\left\{-\left(\bigwedge_{i=1}^{d} x^{(i)}\right)^{-1}\right\} = \Phi_1\left(\bigwedge_{i=1}^{d} x^{(i)}\right)$$

and so (5.59) holds.

If $f_1 = \cdots = f_d$ in (5.30) then for $\mathbf{x} > \mathbf{0}$

$$-\log G(\mathbf{x}) = \int_{[0,1]} \bigvee_{i=1}^{d} \left(\frac{f_i(s)}{x^{(i)}}\right) ds = \left(\int_{[0,1]} f_1(s)ds\right) \bigvee_{i=1}^{d} (1/x^{(i)})$$

and because of (5.29) this is $(\bigwedge_{i=1}^{d} x^{(i)})^{-1}$ and (5.59) again holds. Conversely suppose full dependence holds. If $U$ is a uniform random variable then recall

$$(f_1(U), \dots, f_d(U))$$

has distribution $S(\cdot)/S(\aleph)$ and since $S$ concentrates on $\{\|\mathbf{1}\|^{-1}\mathbf{1}\}$ we easily get

$$P[f_1(U) = \cdots = f_d(U)] = 1. \qquad \square$$

We now make some remarks about asymptotic independence by which we mean that (5.15) holds with the limit $G$ a product of its marginals. Proposition 5.24(ii) makes us hopeful that asymptotic independence in $d > 2$ dimensions can be reduced to the two-dimensional case, and Proposition 5.24(iii) makes us hopeful that bivariate asymptotic independence can be recognized by the behavior of $P[X^{(i)} > t, X^{(j)} > t]$. This is discussed next.

**Proposition 5.27.** *Suppose* $\{\mathbf{X}_n, n \geq 1\}$ *are* iid *random vectors in* $\mathbb{R}^d$, $d \geq 2$, *with common distribution* $F$. *Suppose for simplicity all marginal distributions of* $F$ *are the same and equal* $F_1(x)$, *which we suppose is in the domain of attraction of some univariate extreme value distribution* $G_1(x)$; *i.e., there exist* $a_n > 0$, $b_n$ *such that*

$$F_1^n(a_n x + b_n) \to G_1(x). \tag{5.60}$$

*The following are equivalent.*

(i) *$F$ is in the domain of attraction of a product measure*:

$$F^n(a_n\mathbf{x} + b_n\mathbf{1}) = P\left[\bigvee_{j=1}^{n} \mathbf{X}_j \leq a_n\mathbf{x} + b_n\mathbf{1}\right] \to \prod_{i=1}^{d} G_1(x^{(i)}). \tag{5.61}$$

(ii) *For all* $1 \leq i < j \leq d$

$$P\left[\bigvee_{l=1}^{n} X_l^{(p)} \leq a_n x^{(p)} + b_n, p = i,j\right] \to G_1(x^{(i)})G_1(x^{(j)}). \tag{5.62}$$

(iii) *For* $x^{(p)}$ *such that* $G_1(x^{(p)}) > 0$, $1 \leq p \leq d$

$$\lim_{n\to\infty} nP[X_1^{(p)} > a_n x^{(p)} + b_n, p = i,j] = 0 \tag{5.63}$$

*for* $1 \leq i < j \leq d$.

(iv) *With* $x_0 = \sup\{u: F_1(u) < 1\}$ *and any* $1 \leq i < j \leq d$

$$\lim_{t\to x_0} P[X^{(i)} > t, X^{(j)} > t]/\bar{F}_1(t) = 0. \tag{5.64}$$

Proof. Suppose initially that $d = 2$. Then (i) and (ii) are the same. If (5.62) holds, then in the usual way, by taking logarithms, we get

$$\begin{aligned} &nP([X_1^{(1)} \leq a_n x^{(1)} + b_n, X_1^{(2)} \leq a_n x^{(2)} + b_n]^c) \\ &\to -\log G(x^{(1)}) - \log G(x^{(2)}). \end{aligned} \tag{5.65}$$

The left side of (5.65) is

$$\begin{aligned} &nP[X_1^{(1)} > a_n x^{(1)} + b_n] + nP[X_1^{(2)} > a_n x^{(2)} + b_n] \\ &- nP[X_1^{(1)} > a_n x^{(1)} + b_n, X_1^{(2)} > a_n x^{(2)} + b_n] \end{aligned}$$

so that if we assume (5.60) then it is clear that (5.62) and (5.63) are equivalent.

The equivalence of (5.63) and (5.64) results from the inequalities

$$\frac{n\bar{F}_1(a_n(x^{(1)} \vee x^{(2)}) + b_n)P[X_1^{(1)} > a_n(x^{(1)} \vee x^{(2)}) + b_n, X_1^{(2)} > a_n(x^{(1)} \vee x^{(2)}) + b_n]}{\bar{F}_1(a_n(x^{(1)} \vee x^{(2)}) + b_n)}$$

$$\leq nP[X_1^{(1)} > a_n x^{(1)} + b_n, X_1^{(2)} > a_n x^{(2)} + b_n]$$

$$\leq \frac{n\bar{F}_1(a_n(x^{(1)} \wedge x^{(2)}) + b_n)P[X_1^{(1)} > a_n(x^{(1)} \wedge x^{(2)}) + b_n, X_1^{(2)} > a_n(x^{(1)} \wedge x^{(2)}) + b_n]}{\bar{F}_1(a_n(x^{(1)} \wedge x^{(2)}) + b_n)}$$

since

$$n\bar{F}_1(a_n x + b_n) \to -\log G_1(x) \in (0, \infty)$$

if $0 < G_1(x) < 1$.

This proves the proposition in case $d = 2$ and also checks the equivalence of (ii), (iii), and (iv) in case $d > 2$. So if $d > 2$, the general result will be proved if we verify that (iii) implies (i). However, (5.61) is equivalent to

$$nP([X_1^{(p)} \le a_n x^{(p)} + b_n, 1 \le p \le d]^c) \to \sum_{i=1}^{d} (-\log G(x^{(i)})) \tag{5.66}$$

and the left side of this relation is

$$n \sum_{p=1}^{d} P[X_1^{(p)} > a_n x^{(p)} + b_n] - n \sum_{1 \le p < q \le d} P[X_1^{(p)} > a_n x^{(p)} + b_n, X_1^{(q)} > a_n x^{(q)} + b_n]$$
$$+ \cdots (-1)^{d+1} nP[X_1^{(p)} > a_n x^{(p)} + b_n, p = 1, \ldots, d],$$

which, assuming (5.60) and (5.63) hold, is

$$\sum_{p=1}^{d} (-\log G(x^{(p)})) + o(1)$$

so that (5.66) is true. □

Versions of this proposition have been discussed by Geffroy (1958, 1959), Sibuya (1960), de Haan and Resnick (1977), Galambos (1978), and Marshall and Olkin (1983).

A very interesting corollary, given in Sibuya (1960), applies to the $d$-dimensional multivariate normal distribution.

**Corollary 5.28.** *Let $F$ be the $d$-dimensional multivariate normal with all univariate marginals supposed equal to $N(0, 1)$ for simplicity. If all correlations are less than* 1 *(i.e., $EX_1^{(i)}X_1^{(j)} = \rho_{i,j} < 1$) then asymptotic independence* (5.61) *holds with $G_1(x) = \Lambda(x) = \exp\{-e^{-x}\}$.*

PROOF. According to Proposition 5.27 it suffices to prove that if $\mathbf{X} = (X^{(1)}, X^{(2)})$ is bivariate normal, $EX^{(i)} = 0$, $E(X^{(i)})^2 = 1, i = 1, 2, EX^{(1)}X^{(2)} = \rho < 1$ then

$$\lim_{t \to \infty} P[X^{(1)} > t, X^{(2)} > t]/\bar{N}(t) = 0 \tag{5.67}$$

where, of course, $\bar{N}(t) = 1 - N(t)$ and $N(t) = N(0, 1, t)$. If $\rho = -1$ then we may take $X^{(2)} = -X^{(1)}$ and (5.67) obviously holds, so suppose $|\rho| < 1$. If $U$ and $V$ are iid, $N(0, 1)$, then in $\mathbb{R}^2$

$$(U, \rho U + (1 - \rho^2)^{1/2} V) \stackrel{d}{=} \mathbf{X}$$

and the probability in the numerator of the left side of (5.67) is

$$P[U > t, \rho U + (1 - \rho^2)^{1/2} V > t] \le P[U + \rho U + (1 - \rho^2)^{1/2} V > 2t]$$
$$= P[(1 + \rho)U + (1 - \rho^2)^{1/2} V > 2t]$$

and because $(1+\rho)U + (1-\rho^2)^{1/2}V$ is $N(0, 2(1+\rho))$ this probability is $\bar{N}(2(2(1+\rho))^{-1/2}t)$. Since $N \in D(\Lambda)$ we have for all $A > 1$ that $\bar{N}(At)/\bar{N}(t) \to 0$ (Exercise 1.1.9) and the desired result follows if

$$2(2(1+\rho))^{-1/2} > 1.$$

This is the same as $2 > (2(1+\rho))^{1/2}$ or $2^{1/2} > (1+\rho)^{1/2}$ which is obviously true for $|\rho| < 1$. □

EXERCISES

5.5.1. Suppose $I_1, \ldots, I_k$ is a disjoint partition of $\{1, \ldots, d\}$. If $\mathbf{X}$ is a random vector in $\mathbb{R}^d$ with max-id distribution $F$ with exponent measure $\mu$, give necessary and sufficient conditions for $(\{X^{(i)}, i \in I_j\}, j = 1, \ldots, k)$ to be independent vectors.

5.5.2. Suppose $\{X_n, n \geq 1\}$ are iid one-dimensional random variables with common distribution $F \in D(G)$ where $G$ is an extreme value distribution. Assume $P[X_1 > x] \sim P[-X_1 > x]$ as $x \to x_0$. Then $\bigvee_{i=1}^n X_i$ and $\bigwedge_{i=1}^n X_i$ are asymptotically independent; i.e., suitably normalized $(\bigvee_{i=1}^n X_i, \bigwedge_{i=1}^n X_i)$ has a limit distribution which is a product of its marginals. Consequently the range $\bigvee_{i=1}^n X_i - \bigwedge_{i=1}^n X_i$ has a limit distribution. Hint: Use (5.64) (de Haan and Resnick, 1977; de Haan, 1974b).

5.5.3. Suppose the hypotheses of Proposition 5.27 are satisfied and $d = 2$. Show that the bivariate maxima have a limit distribution concentrating on the line $y = x$ iff

$$\lim_{t \to x_0} P[X_1^{(1)} > t, X^{(2)} > t]/\bar{F}_1(t) = 1.$$

Generalize to $d > 2$ (Sibuya, 1960).

5.5.4. Consider
(i) $F(x, y) = F_1(x)F_1(y)[1 + \alpha\bar{F}_1(x)\bar{F}_1(y)]$, $-1 \leq \alpha \leq 1$.
(ii) $F(x, y) = F_1(x)F_1(y)(1 - \alpha\bar{F}_1(x)\bar{F}_1(y))^{-1}$, $-1 \leq \alpha \leq 1$.
(iii) $P[X^{(1)} > x, X^{(2)} > y] = \bar{F}_1(x)\bar{F}_1(y)\bar{F}_1(x \vee y)$.
(iv) $F(x, y) = (1 + e^{-x} + e^{-y})^{-1}$, $(x, y) \in \mathbb{R}^2$.
Discuss when asymptotic independence or dependence holds (Marshall and Olkin, 1983).

## 5.6. Association

We now discuss a result which unifies several results in the literature and explains why various dependence measures between the components of an extreme value vector are always positive. See, for example, Tiago de Oliveira (1962/63) and de Haan (1985). Marshall and Olkin (1983) have proved that if $\mathbf{Y} \in \mathbb{R}^d$ has a multivariate extreme value distribution, then $Y$ is associated.

A random vector $Y \in \mathbb{R}^d$ is associated (Esary, Proschan, and Walkup, 1967) if

$$\mathrm{Cov}(g_1(\mathbf{Y}), g_2(\mathbf{Y})) \geq 0 \tag{5.68}$$

for all nondecreasing functions $g_i: \mathbb{R}^d \to \mathbb{R}$ for which $E|g_i(\mathbf{Y})| < \infty$, $E|g_1(\mathbf{Y})g_2(\mathbf{Y})| < \infty$ $(i = 1, 2)$. Here $g$ nondecreasing means that $\mathbf{x} \le \mathbf{y}$ implies $g(\mathbf{x}) \le g(\mathbf{y})$, which amounts to $g$'s being monotone in each coordinate. In what follows, when discussing covariances, we shall always assume, without explicit mention, that all relevant expectations are finite.

The motivation behind this dependence measure as explained by Esary, Proschan, and Walkup (1967) is as follows: It is natural to say that random variables $S$ and $T$ exhibit a positive dependence if $\text{Cov}(S, T) \ge 0$. Successively stronger notions of positive dependence are

(i) $\text{Cov}(g_1(S), g_2(T)) \ge 0$ for all pairs $g_i: \mathbb{R} \to \mathbb{R}$, $g_i$ nondecreasing, $i = 1, 2$.
(ii) $\text{Cov}(g_1(S, T), g_2(S, T)) \ge 0$ for all pairs $g_i: \mathbb{R}^2 \to \mathbb{R}$, $g_i$ nondecreasing, $i = 1, 2$.

Note that (ii) implies (i) and (i) implies $\text{Cov}(S, T) \ge 0$. It is (ii) which generalizes nicely to higher dimensions (and even to function spaces). Part of the power and beauty of the definition results from the fact that association is preserved under nondecreasing transformations.

The Marshall and Olkin (1983) proof of the association of an extreme value vector is based on the result in Exercise 5.4.1.16. They use the facts that independent random variables are associated, nondecreasing functions of associated variables are associated, and association is preserved under weak convergence.

We follow a different route and start with the fact that PRM is associated (defined later). Since any vector $\mathbf{Y}$ with max-id distribution is a nondecreasing function of PRM we are able to extend the Marshall and Olkin result.

**Proposition 5.29.** *If $\mathbf{Y}$ is an $\mathbb{R}^d$ valued random vector with max-id distribution, then $\mathbf{Y}$ is associated. Thus extreme value random vectors are associated.*

We now discuss the proof, skipping the proofs of certain needed facts until the end. For a space $E$ which is locally compact with a countable base define an order on $M_p(E)$ (or $M_+(E)$) by

$$\mu \le \nu \quad \text{iff} \quad \mu(A) \le \nu(A), \qquad \text{for all } A \in \mathscr{E}$$

for $\mu, \nu \in M_p(E)$. We say a point process $N$ is associated if for any measurable $F_i: M_p(E) \to \mathbb{R}, i = 1, 2, F_i$ nondecreasing (meaning $\mu \le \nu$ implies $F_i(\mu) \le F_i(\nu)$), we have

$$\text{Cov}(F_1(N), F_2(N)) \ge 0. \tag{5.69}$$

This definition is the most convenient for applications but makes verification that a specific point process is associated difficult. Thus we need the following equivalences (cf. Burton and Waymire, 1985).

**Proposition 5.30.** *For the point process $N$, the following are equivalent:*
(i) *$N$ is associated.*
(ii) *For any $k \ge 1$, and any sets $A_1, \ldots, A_k$ in $\mathscr{E}$, we have*

$$(N(A_1),\ldots,N(A_k))$$

*is an associated random vector in* $\mathbb{R}^k$.

(iii) *For any* $k \geq 1$ *and any functions* $f_1,\ldots,f_k$ *in* $C_K^+(E)$ *we have*

$$(N(f_1),\ldots,N(f_k))$$

*is an associated vector in* $\mathbb{R}^k$.

The proof of Proposition 5.30 comes later. These equivalences make it easy to prove (Burton and Waymire, 1985, page 1271) the following.

**Proposition 5.31.** *If $N$ is a Poisson process (or any completely random measure) then $N$ is associated.*

Now suppose $\mathbf{Y} \in \mathbb{R}^d$ has a max-id distribution with exponent measure $\mu$ on $E := [\mathbf{l}, \infty]\backslash\{\mathbf{l}\}$. According to Proposition 5.8 there exists PRM($\mu$) on $E$,

$$N = \sum_k \varepsilon_{\mathbf{j}_k}$$

such that

$$\mathbf{Y} \stackrel{\mathrm{d}}{=} \bigvee_k \mathbf{j}_k.$$

To prove Proposition 5.29 we need to show for any nondecreasing $g_i$: $\mathbb{R}^d \to \mathbb{R}$ that (5.68) holds. For $m = \sum_k \varepsilon_{\mathbf{v}_k} \in M_p(E)$ define $T$: $M_p(E) \to \mathbb{R}^d$ by

$$Tm = \bigvee_k \mathbf{v}_k = \inf\{\mathbf{x}: m([\mathbf{l},\mathbf{x}]^c) = 0\}$$

so that $\mathbf{Y} = TN$. Observe that if $m_1 \leq m_2$ then $m_2([\mathbf{l},\mathbf{x}]^c) = 0$ implies $m_1([\mathbf{l},\mathbf{x}]^c) = 0$ and therefore

$$\{x: m_2([\mathbf{l},\mathbf{x}]^c) = 0\} \subset \{\mathbf{x}: m_1([\mathbf{l},\mathbf{x}]^c) = 0\}.$$

Thus $Tm_1 \leq Tm_2$, and $T$ is nondecreasing. Furthermore, the composition $g_i \circ T$: $M_p(E) \to \mathbb{R}$ is nondecreasing, and thus if $N$ is associated we have by definition (see 5.69) that

$$\mathrm{Cov}(g_1 \circ T(N), g_2 \circ T(N)) = \mathrm{Cov}(g_1(\mathbf{Y}), g_2(\mathbf{Y})) \geq 0$$

which proves Proposition 5.29.

The proofs of Propositions 5.30 and 5.31 require several facts about association collected together as a lemma. (See Esary, Proschan, and Walkup, 1967; Lehmann, 1966.)

**Lemma 5.32.** (i) *For any random variables $X_1$ and $X_2$ such that $E|X_i| < \infty$, $E|X_1X_2| < \infty$ we have*

$$\mathrm{Cov}(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{Cov}(1_{(x,\infty)}(X_1), 1_{(y,\infty)}(X_2))dx\,dy. \quad (5.70)$$

(ii) *It suffices for association to check* (5.68) *or* (5.69) *for nondecreasing indicator functions.*
(iii) *It suffices for association to check* (5.68) *or* (5.69) *for nondecreasing, bounded continuous functions. Thus association is preserved under weak convergence.*
(iv) *A single random variable $X$ is associated.*
(v) *If $\mathbf{Y}_i \in R^{d_i}$ is associated $i = 1, 2$, and $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent, then $(\mathbf{Y}_1, \mathbf{Y}_2)$ is associated in $R^{d_1+d_2}$.*
(vi) *If $X_1, \ldots, X_d$ are independent random variables, then $\mathbf{X} = (X_1, \ldots, X_d)$ is associated in $\mathbb{R}^d$.*

PROOFS. (i) This charming proof of (5.70) is from Lehmann (1966), where it is attributed to Hoeffding. Let $(X_1, X_2)$ and $(X_1^\#, X_2^\#)$ be independent with $(X_1, X_2) \stackrel{d}{=} (X_1^\#, X_2^\#)$. Then

$$\begin{aligned}
&2\,\mathrm{Cov}(X_1, X_2)\\
&\quad = E(X_1 - X_1^\#)(X_2 - X_2^\#)\\
&\quad = E\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(1_{(u,\infty)}(X_1) - 1_{(u,\infty)}(X_1^\#))(1_{(v,\infty)}(X_2) - 1_{(v,\infty)}(X_2^\#))\,du\,dv
\end{aligned}$$

and passing expectation through the integrals yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2\,\mathrm{Cov}(1_{(u,\infty)}(X_1), 1_{(v,\infty)}(X_2))\,du\,dv.$$

(ii) Suppose (5.69) holds for all nondecreasing indicator functions. For any nondecreasing $F_i$ we get from (5.70)

$$\mathrm{Cov}(F_1(N), F_2(N)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{Cov}(1_{(x,\infty)}(F_1(N)), 1_{(y,\infty)}(F_2(N)))\,dx\,dy. \tag{5.71}$$

For $m \in M_p(E)$ the function $1_{(x,\infty)}(F_1(m))\colon M_p(E) \to \mathbb{R}$ is a nondecreasing indicator function so that by assumption the integrand on the right side of (5.71) is non-negative. Thus $\mathrm{Cov}(F_1(N)), F_2(N)) \geq 0$. A similar proof works for the assertion about (5.68).
(iv) Because of (ii) it suffices to prove that if $\gamma_i\colon \mathbb{R} \to \mathbb{R}$. $\gamma_i$ nondecreasing and the range of $\gamma_i$ is $\{0, 1\}$ then

$$\mathrm{Cov}(\gamma_1(X), \gamma_2(X)) \geq 0.$$

However, the assumptions on $\gamma_i$ imply $\gamma_1 \leq \gamma_2$ or $\gamma_2 \leq \gamma_1$. Suppose for concreteness that $\gamma_1 \leq \gamma_2$. Then $\gamma_1\gamma_2 = \gamma_1$ and

$$\begin{aligned}
\mathrm{Cov}(\gamma_1(X), \gamma_2(X)) &= E\gamma_1(X)\gamma_2(X) - E\gamma_1(X)E\gamma_2(X)\\
&= E\gamma_1(X) - E\gamma_1(X)E\gamma_2(X)\\
&= E\gamma_1(X)\{1 - E\gamma_2(X)\} \geq 0.
\end{aligned}$$

(iii) Suppose (5.69) holds for all bounded, continuous, nondecreasing test functions. We show that it holds for all nondecreasing indicator functions and hence by (ii) $N$ is associated.

Suppose then that $\gamma_i$: $M_p(E) \to \mathbb{R}$ are nondecreasing indicator functions. Set

$$A_i = [\gamma_i = 1] = \{m \in M_p(E): \gamma_i(m) = 1\}.$$

Since $M_p(E)$ is a complete separable metric space, probabilities on $M_p(E)$ are tight: i.e., for $A \in \mathcal{M}_p(E)$

$$P[N \in A] = \sup\{P[N \in C]: C \subset A, C \text{ is compact}\}$$

(Billingsley, 1968, page 9). Thus there exists for any $\varepsilon > 0$ a compact set $C_i \in \mathcal{M}_p(E)$ such that $C_i \subset A_i$ and

$$P[N \in C_i] + \varepsilon \geq P[N \in A_i]. \tag{5.72}$$

Let

$$C_i^{\uparrow} = \{m \in M_p(E): \text{there exists } m_1 \in C_i \text{ and } m \geq m_1\}.$$

Then

$$C_i \subset C_i^{\uparrow} \subset A_i \tag{5.73}$$

as a consequence of $\gamma_1$ being nondecreasing. Furthermore $C_i^{\uparrow}$ is closed in $M_p(E)$, as is readily checked as follows: Suppose $m_n \in C_i^{\uparrow}$ and $m_n \xrightarrow{v} m_0$ in $M_p(E)$. For each $n$ there exists $m_n^{(1)} \in C_i$ such that $m_n \geq m_n^{(1)}$ and since $C_i$ is compact $\{m_n^{(1)}\}$ has a limit point in $C_i$, $m_0^{(1)}$ say; i.e., for some subsequence $\{n'\}$ we have $m_{n'}^{(1)} \xrightarrow{v} m_0^{(1)} \in C_i$. Thus $m_n \xrightarrow{v} m_0$, $m_{n'} \geq m_{n'}^{(1)} \xrightarrow{v} m_0^{(1)}$ imply $m_0 \geq m_0^{(1)}$, and thus $m_0 \in C_i^{\uparrow}$, thus showing that $C_i^{\uparrow}$ is closed.

Since $C_i^{\uparrow}$ is closed, there are bounded, continuous $h_i^{(n)}$ such that $h_i^{(n)} \downarrow 1_{C_i^{\uparrow}}$, $n \to \infty$. (Cf. Billingsley, 1968, page 8.) In order to make $h_i^{(n)}$ non-decreasing we may take

$$h_i^{(n)}(m) = 1 - ((n\rho(m, C_i^{\uparrow})) \wedge 1)$$

where $\rho$ is the vague metric. To check $h_i^{(n)}$ is non-decreasing it suffices to verify that if $C$ is compact in $M_p(E)$ and $m_i \notin C^{\uparrow}$, $i = 1, 2$, then

$$m_1 \leq m_2 \text{ implies } \rho(m_1, C^{\uparrow}) \geq \rho(m_2, C^{\uparrow}).$$

Observe that since $C^{\uparrow}$ is closed, for each $m$ there exists $m^* \in C^{\uparrow}$ such that

$$\rho(m, C^{\uparrow}) = \rho(m, m^*).$$

Since $m_1 \leq m_2$ we have $m_2 - m_1 \in M_p(E)$ and thus $m_1^* + m_2 - m_1 \geq m_1^*$ so that $m_1^* + m_2 - m_1 \in C^{\uparrow}$. Consequently

$$\begin{aligned}\rho(m_2, C^{\uparrow}) &\leq \rho(m_2, m_1^* + m_2 - m_1) = \rho(m_1 + m_2 - m_1, m_1^* + m_2 - m_1)\\ &= \rho(m_1, m_1^*) = \rho(m_1, C^{\uparrow})\end{aligned}$$

thus showing the desired monotonicity.

Because we assume (5.69) holds for bounded, continuous functions we have

$$\mathrm{Cov}(h_1^{(n)}(N), h_2^{(n)}(N)) \geq 0$$

and letting $n \to \infty$ gives

$$\mathrm{Cov}(1_{C_1^\uparrow}(N), 1_{C_2^\uparrow}(N)) \geq 0. \tag{5.74}$$

From (5.73)

$$1_{C_i} \leq 1_{C_i^\uparrow} \leq 1_{A_i} = \gamma_i.$$

Thus

$$E\gamma_1(N)\gamma_2(N) \geq E1_{C_1^\uparrow}(N)1_{C_2^\uparrow}(N)$$

and from (5.72)

$$\begin{aligned} E\gamma_i(N) = P[N \in A_i] &\leq \varepsilon + P[N \in C_i] \\ &\leq \varepsilon + P[N \in C_i^\uparrow] = \varepsilon + E1_{C_i^\uparrow}(N). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathrm{Cov}(\gamma_1(N), \gamma_2(N)) &= E\gamma_1(N)\gamma_2(N) - E\gamma_1(N)E\gamma_2(N) \\ &\geq E1_{C_1^\uparrow}(N)1_{C_2^\uparrow}(N) - E1_{C_1^\uparrow}(N)E1_{C_2^\uparrow}(N) \\ &\quad - \varepsilon^2 - \varepsilon(E\{1_{C_1^\uparrow}(N) + 1_{C_2^\uparrow}(N)\}) \\ &\geq \mathrm{Cov}(1_{C_1^\uparrow}(N), 1_{C_2^\uparrow}(N)) - \varepsilon^2 - 2\varepsilon. \end{aligned}$$

From (5.74) and the arbitrariness of $\varepsilon$ we conclude that $\mathrm{Cov}(\gamma_1(N), \gamma_2(N)) \geq 0$ as desired.

The proof of the result for (5.68) is the same.

(v) Suppose $g_i$: $\mathbb{R}^{d_1+d_2} \to \mathbb{R}$, $g_i$ nondecreasing. Then

$$\begin{aligned} &\mathrm{Cov}(g_1(\mathbf{Y}_1, \mathbf{Y}_2), g_2(\mathbf{Y}_1, \mathbf{Y}_2)) \\ &\quad = Eg_1g_2(\mathbf{Y}_1, \mathbf{Y}_2) - Eg_1(\mathbf{Y}_1, \mathbf{Y}_2)Eg_2(\mathbf{Y}_1, \mathbf{Y}_2) \end{aligned}$$

and letting $\mu_i$ be the distribution of $\mathbf{Y}_i$ this is

$$\begin{aligned} &\int_{\mathbf{y}_1} \mu_1(d\mathbf{y}_1)\left\{\int_{\mathbf{y}_2} \mu_2(d\mathbf{y}_2)g_1g_2 - \int_{\mathbf{y}_2} g_1\,d\mu_2 \int_{\mathbf{y}_2} g_2\,d\mu_2\right\} \\ &\quad + \int_{\mathbf{y}_1} \mu_1(d\mathbf{y}_1)\left\{\int_{\mathbf{y}_2} g_1\,d\mu_2 \int_{\mathbf{y}_2} g_2\,d\mu_2\right\} \\ &\quad - \int_{\mathbf{y}_1} \mu_1(d\mathbf{y}_1) \int_{\mathbf{y}_2} g_1\,d\mu_2 \int_{\mathbf{y}_1} \mu_1(d\mathbf{y}_1) \int_{\mathbf{y}_2} g_2\,d\mu_2 \\ &= \int_{\mathbf{y}_1} \mu_1(d\mathbf{y}_1)\mathrm{Cov}(g_1(\mathbf{y}_1, \mathbf{Y}_2), g_2(\mathbf{y}_1, \mathbf{Y}_2)) \\ &\quad + \mathrm{Cov}\left(\int_{\mathbf{y}_2} g_1(\mathbf{Y}_1, \mathbf{y}_2)\mu_2(d\mathbf{y}_2), \int_{\mathbf{y}_2} g_2(\mathbf{Y}_1, \mathbf{y}_2)\mu_2(d\mathbf{y}_2)\right). \end{aligned}$$

For fixed $\mathbf{y}_1$ we have $g_i(\mathbf{y}_1, \cdot)$ is nondecreasing and also we have $\int g_i(\cdot, \mathbf{y}_2)\mu_2(d\mathbf{y}_2)$ is nondecreasing. Since $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are each associated we get

$$\operatorname{Cov}(g_1, g_2) \geq 0.$$

(vi) Combine (iv) and (v). □

Now consider the proof of Proposition 5.30. If $N$ is associated it is easy to check that $(N(A_i),\ 1 \leq i \leq k)$ is associated in $\mathbb{R}^k$: Suppose $g_i$: $\mathbb{R}^k \to \mathbb{R}$ are monotone and we need to check

$$\operatorname{Cov}(g_1(N(A_1), \ldots, N(A_k)), g_2(N(A_1), \ldots, N(A_k))) \geq 0.$$

Define $F_i$: $M_p(E) \to \mathbb{R}$ by

$$F_i(m) = g_i(m(A_1), \ldots, m(A_k))$$

so that $F_i$ is monotone and therefore since $N$ is assumed associated

$$\begin{aligned} 0 &\leq \operatorname{Cov}(F_1(N), F_2(N)) \\ &= \operatorname{Cov}(g_1(N(A_1), \ldots, N(A_k)), g_2(N(A_1), \ldots, N(A_k))) \end{aligned}$$

as required for showing that (i) implies (ii).

Now given (ii) we prove (iii). For $f_l \in C_K^+(E)$ we show $(N(f_l),\ l \leq k)$ is associated in $\mathbb{R}^k$. Write for $1 \leq l \leq k$

$$\begin{aligned} f_l &= \lim_{n\to\infty} \uparrow f_l^{(n)} := \lim_{n\to\infty} \uparrow \sum_{j=1}^{2^n} j2^{-n} 1_{f_l^{-1}[j2^{-n}, (j+1)2^{-n})} \\ &=: \lim_{n\to\infty} \uparrow \sum_{j=1}^{2^n} j2^{-n} 1_{A_{lj}^{(n)}} \end{aligned}$$

where $A_{lj}^{(n)} \in \mathscr{E}$. Since $f_l^{(n)} \uparrow f_l$ we have $(N(f_l^{(n)}), l \leq k) \to (N(f_l), l \leq k)$ and by Lemma 5.32 (iii) it suffices to verify that $(N(f_l^{(n)}), l \leq k)$ is associated in $\mathbb{R}^k$. For $g_i$: $\mathbb{R}^k \to \mathbb{R}$ nondecreasing

$$g_i(N(f_l^{(n)}), l \leq k) = \Phi_i(N(A_{lj}^{(n)}),\ 1 \leq l \leq k,\ 1 \leq j \leq 2^n)$$

for some nondecreasing $\Phi_i$: $\mathbb{R}^{k2^n} \to \mathbb{R}$, and so from the assumption of (ii)

$$\begin{aligned} &\operatorname{Cov}(g_1(N(f_l^{(n)}), l \leq k), g_2(N(f_l^{(n)}), l \leq k)) \\ &\quad = \operatorname{Cov}(\Phi_1(N(A_{lj}^{(n)}), 1 \leq l \leq k, 1 \leq j \leq 2^n), \Phi_2(N(A_{lj}^{(n)}), 1 \leq l \leq k, 1 \leq j \leq 2^n)) \\ &\quad \geq 0 \end{aligned}$$

as desired.

We now show that (iii) implies (i). Recall from the definition of the vague metric in Proposition 3.17 that there exists a countable family $h_l \in C_K^+(E), l \geq 1$ such that $m \in M_p(E)$ is determined by its values $(m(h_l), l \geq 1)$. In fact $m \to (m(h_l), l \geq 1)$ is an isometry between $M_p(E)$ and the following subset of $\mathbb{R}^\infty$: $\{(m(h_l), l \geq 1): m \in M_p(E)\}$. If $F_i$: $M_p(E) \to \mathbb{R}$ is bounded, continuous, and nondecreasing, there is $\phi_i$: $\mathbb{R}^\infty \to R$ with the same properties and

$$F_i(m) = \Phi_i(m(h_l), l \geq 1).$$

In $\mathbb{R}^\infty$

$$(m(h_l), l \geq 1) = \lim_{n \to \infty} (m(h_l), 1 \leq l \leq n, m(h_n), m(h_n), \ldots)$$

and thus there is a bounded, continuous, nondecreasing $\psi_i^{(n)}\colon \mathbb{R}^n \to \mathbb{R}$ and

$$F_i(m) = \lim_{n \to \infty} \psi_i^{(n)}(m(h_l), 1 \leq l \leq n).$$

Because we assume (iii) we have

$$\begin{aligned} &\operatorname{Cov}(F_1(N), F_2(N)) \\ &= \lim_{n \to \infty} \operatorname{Cov}(\psi_1^{(n)}(N(h_l), 1 \leq l \leq n), \psi_2^{(n)}(N(h_l), 1 \leq l \leq n)) \geq 0 \end{aligned}$$

and this gives (i).

Finally we need to verify Proposition 5.31, that a Poisson process $N$ is associated. We use the criterion of Proposition 5.23(ii) and show for any $k$, $A_1, \ldots, A_k \in \mathscr{E}$

$$(N(A_l), 1 \leq l \leq k)$$

is associated in $\mathbb{R}^k$. If $A_l$, $1 \leq l \leq k$ are disjoint the result follows from Lemma 5.32(vi) since then $N(A_l)$, $1 \leq l \leq k$ are independent. If $A_l$, $1 \leq l \leq k$ are not disjoint, there exist disjoint sets $B_l$, $1 \leq l \leq p$ in $\mathscr{E}$ and for $1 \leq q \leq k$

$$A_q = \bigcup_{j \in I_q} B_\alpha$$

where $I_q \subset \{1, \ldots, p\}$. Thus

$$(N(A_l), 1 \leq l \leq k) = \left(\sum_{j \in I_l} N(B_j), 1 \leq l \leq k\right).$$

The variables $(N(B_j),\ 1 \leq j \leq p)$ being independent are associated and $(N(A_l), 1 \leq l \leq k)$ being sums and hence a nondecreasing function of an associated vector is itself associated. □

## Exercises

5.6.1. If $\psi\colon \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ is nondecreasing and $\mathbf{Y} \in \mathbb{R}^{d_1}$ is associated, then $\psi(\mathbf{Y})$ is associated.

5.6.2. (a) If $N_i$, $1 \leq i \leq k$ are independent, associated point processes, then $\sum_1^k N_i$ is associated (Burton and Waymire, 1985).

(b) Suppose $N_i$, $i \geq 1$ are independent point processes and each $N_i$ is associated. If the positive integer valued random variable $N$ is independent of $\{N_i, i \geq 1\}$ prove $\sum_1^N N_i$ is associated.

(c) Use the construction of PRM in Section 3.3 and Part (b) earlier to give an alternate proof that PRM is associated.

5.6.3. (i) If $X$ and $Y$ are associated random variables, then

$$H(x, y) := P[X > x, Y > y] - P[X > x]P[Y > y] \geq 0.$$

If $X$ and $Y$ also have finite variance

$$|\mathrm{Cov}(e^{irX}, e^{isY})| \le \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ir)(is) \exp\{irx + isy\} H(x, y) dx\, dy \right|$$

$$\le |rs| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) dx\, dy = |r||s| \mathrm{Cov}(X, Y).$$

(ii) If $X_1, \ldots, X_m$ are associated finite variance random variables with joint and marginal characteristic functions $\Phi(r_1, \ldots, r_m)$ and $\Phi_j(r)$ then

$$|\Phi(r_1, \ldots, r_m) - \prod_{j=1}^{m} \Phi_j(r_j)| \le 2^{-1} \sum_{1 \le j \ne k \le m} |r_j||r_k| \mathrm{Cov}(X_j, X_k).$$

(iii) Associated random variables which are uncorrelated are jointly independent (Newman and Wright, 1981).

# References


Ailam, G. (1968). On probability properties of random sets and the asymptotic behavior of empirical distribution functions. *J. Appl. Probability*, *5*, 196–202.

Aljancic, S. and Karamata, J. (1956). Fonctions à comportement regulier et l'integrale de Frullani: *Recueil des Travaux de l'Academic Serbe des Sciences L*, Institut Mathematique No. 5, 239–248.

Anderson, C.W. (1971). *Contributions to the Asymptotic Theory of Extreme Values.* Ph.D. Thesis, University of London.

Anderson, C.W. (1976). Extreme value theory and its approximations. *Proc. Symp. Reliability Technology.* Bradford, U.K. Atomic Energy Authority.

Anderson, C.W. (1978). Super-slowly varying functions in extreme value theory. *J. Roy. Statist. Soc. B*, *40*, 197–202.

Anderson, C.W. (1984). Large deviations of extremes. In: *Statistical Extremes and Applications*, 325–340; ed. J. Tiago de Oliveira, D. Reidel, Dordrecht, Holland.

Balkema, A.A. (1973). *Monotone Transformations and Limit Laws.* Mathematical Centre Tract, Amsterdam, Holland.

Balkema, A.A. and Haan, L. de (1972). On R. Von-Mises' condition for the domain of attraction of $\exp\{-e^{-x}\}$. *Ann. Math. Statist.*, *43*, 1352–1354.

Balkema, A. and Resnick, S. (1977). Max-infinite divisibility. *J. Appl. Probability*, *14*, 309–319.

Ballerini, R. and Resnick, S. (1987). Embedding sequences of successive maxima in extremal processes, with applications. *J. Appl. Probability.*

Barndorff-Nielsen, O. (1963). On the limit behavior of extreme order statistics. *Ann. Math. Statist.* *34*, 992–1002.

Berman, S.M. (1961). Convergence to bivariate limiting extreme value distributions. *Ann. Inst. Statist. Math.*, *13*, 217–223.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Billingsley, P. (1971). *Weak Convergence of Measures: Applications in Probability.* SIAM Monograph No. 5, Philadelphia.

Billingsley, P. (1979). *Probability and Measure.* Wiley, New York.

Bingham, N.H. and Goldie, C.M. (1982). Extensions of regular variation. I: Uniformity and quantifiers. *Proc. London Math., Soc. 44*, 473–496.

Bingham, N., Goldie, C. and Teugels, J. (1987). *Regular Variation.* Encyclopedia of Mathematics and Its Applications, vol. 27. Cambridge University Press, Cambridge, U.K.

Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

Burton, R. and Waymire, E. (1985). Scaling limits and associated random measures. *Ann. Probability, 13*, 1267–1278.

Chung, K.L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

Cinlar, E. (1976). Random measures and dynamic point processes. II: Poisson random measures. Discussion paper II, Center for Statistics and Probability, Northwestern University, Evanston, Ill.

Cline, D. (1983a). Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. Thesis, Dept. of Statistics, Colorado State University, Fort Collins.

Cline, D. (1983b). Infinite series of random variables with regularly varying tails. Tech. Report 83-24, Institute of Applied Mathematics and Statistics, University of British Columbia, Vancouver, B.C.

Cohen, D.L. (1980). *Measure Theory*. Birkhauser, Boston.

Cohn, D.P. (1982a). The penultimate form of approximation to normal extremes. *Advances Appl. Probability, 14*, 324–339.

Cohn, D.P. (1982b). Convergence rates for the ultimate and penultimate approximations in extreme value theory. *Advances Appl. Probability, 14*, 833–854.

Davis, R.A. (1979). Maxima and minima of stationary sequences. *Ann. Probability, 7*, 453–460.

Davis, R.A. (1982a). Limit laws for the maximum of stationary sequences. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 61*, 31–42.

Davis, R.A. (1982b). The rate of convergence in distribution of the maxima. *Statistica Neerlandica, 36*, 31–35.

Davis R.A. (1983). Limit laws for upper and lower extremes from stationary mixing sequences. *J. Multivariate Anal., 13*, 273–286.

Davis, R.A. (1984). On upper and lower extremes in stationary sequences. In: *Statistical Extremes and Applications*, 443–460; ed. J. Tiago de Oliveira, D. Reidel, Dordrecht, Holland.

Davis, R. and Resnick, S. (1985a). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probability, 13*, 179–195.

Davis, R. and Resnick, S. (1985b). More limit theory for the sample correlation function of moving averages. *J. Stochastic Processes Appl., 20*, 257–279.

Davis, R. and Resnick, S. (1986). Limit theory for sample covariance and correlation functions. *Ann. Statist., 14*, 533–558.

Deheuvels, P. (1974). Valeur extremales d'echantillons croissants d'une variable aleatoire reelle. *Ann. Inst. H. Poincare, 10*, 89–114.

Deheuvels, P. (1978). Caracterisation complete des lois extremes multivariees et de la convergence aux types extremes. *Publ. Inst. Statist. Univ. Paris, 23.*

Deheuvels, P. (1980). Some applications of the dependence function to statistical inference: Nonparametric estimates of extreme values distributions, and a Kiefer type universal bound for the uniform test of independence. *Coll. Math. Soc. Janos Bolyai, 32, Nonparametric Statistical Inference*, Budapest, 1980.

Deheuvels, P. (1981). The strong approximation of extremal processes. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 58*, 1–6.

Deheuvels, P. (1982a). Strong approximation in extreme value theory and applications. *Colloquia Math. Soc. Janos Bolya, 36. Limit Theorems in Probability and Statistics*, Veszprém, Hungary.

Deheuvels, P. (1982b). On record times associated with $k$-th extremes. *Proceedings of the 3rd Pannonian Symposium on Mathematical Statistics.* Visegrad, Hungary, ed. J. Mogyorodi, I. Vincze, W. Wertz, 43–51.

Deheuvels, P. (1983). The strong approximation of extremal processes II. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 62,* 7–15.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In: *Statistical Extremes and Applications,* 117–130; ed. J. Tiago de Oliveria, D. Reidel. Dordrecht, Holland.

Durrett, R. and Resnick, S.I. (1978). Functional limit theorems for dependent variables. *Ann. Probability, 6,* 829–846.

Dwass, M. (1960). Some $k$ sample rank order tests. In: *Contributions to Probability and Statistics, Essays in Honour of H. Hotelling,* 198–202; ed. I. Olkin et al. *Stanford Studies in Math. and Stat., 2,* Stanford University Press, Stanford, Calif.

Dwass, M. (1964). Extremal processes. *Ann. Math. Statist., 35,* 1718–1725.

Dwass, M. (1966). Extremal processes II. *Illinois J. Math., 10,* 381–391.

Dwass, M. (1974). Extremal processes III. *Bull. Inst. Math. Acad. Sin., 2,* 255–265.

Esary, J., Proschan, F. and Walkup, D. (1967). Association of random variables, with applications. *Ann. Math. Statist., 38,* 1466–1474.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications,* volume I, 3rd ed. Wiley, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* volume II, 2nd ed. Wiley, New York.

Finkelshteyn, B.V. (1953). Limiting distribution of extremes of a variational series of a two-dimensional random variable. *Dokl. Ak. Nauk. SSSR, 91,* 209–214.

Finster, M. (1982). The maximum term and first passage times for autoregressions. *Ann. Probability, 10,* 737–744.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distributions of largest or smallest member of a sample. *Proc. Cambridge Philos. Soc., 24,* 180–190.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York.

Galambos, J. and Seneta, E. (1975). Record times. *Proc. Amer. Math. Soc., 50,* 383–387.

Geffroy, J. (1958, 1959). Contributions à la theorie des valeurs extremes. *Publ. Inst. Statist. Univ. Paris, 7, 8,* 37–185.

Gerritse, G. (1986). Supremum self-decomposable random vectors. *Probability Theory and Related Fields, 72,* 17–34.

Gnedenko, B.V. (1943). Sur la distribution limité du terme d' une série aléatoire. *Ann. Math., 44,* 423–453.

Goldie, C.M. (1983). On records and related topics in probability theory. Thesis, University of Sussex, Mathematics Department, Brighton, U.K.

Goldie, C.M. and Rogers, L.C.G. (1984). The $k$-record processes are i.i.d. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 67,* 197–211.

Goldie, C. and Smith, R. (1987). Slow variation with remainder: Theory and applications. *Quart. J. Math. Oxford (2), 38,* 45–71.

Gumbel, E.J. (1958). *Statistics of Extremes.* Columbia University Press, New York.

Haan L. de (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tract 32, Mathematics Centre, Amsterdam, Holland.

Haan, L. de (1974a). Equivalence classes of regularly varying functions. *J. Stochastic Processes Appl., 2,* 243–259.

Haan, L. de (1974b). Weak limits of the sample range. *J. Applied Probability, 11*, 836–841.

Haan, L. de (1976a). Sample extremes: An elementary introduction. *Statist. Neerlandica, 30*, 161–172.

Haan, L. de (1976b). An Abel-Tauber theorem for Laplace transforms. *J. London Math. Soc., 13*, 537–542.

Haan, L. de (1978). A characterization of multidimensional extreme value distributions. *Sankhya Ser. A, 40*, 85–88.

Haan, L. de (1984a). Extremal processes. In: *Statistical Extremes and Applications*, 297–309; ed. J. Tiago de Oliveira. Reidel, Dordrecht, Holland.

Haan, L. de (1984b). A spectral representation for max-stable processes. *Ann. Probability, 12*, 1194–1204.

Haan, L. de (1985). Extremes in higher dimensions: The model and some statistics. *Proceedings of the 47th Session of the International Statistical Institute*, 1–16, Amsterdam, Holland.

Haan, L. de and Hordijk, A. (1972). The rate of growth of sample maxima. *Ann. Math. Statist., 43*, 1185–1196.

Haan, L. de and Omey, E. (1983). Integrals and derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions II. *J. Stochastic Processes Appl., 16*, 157–170.

Haan, L. de, Omey, E. and Resnick, S. (1984). Domains of attraction and regular variation in $\mathbb{R}^d$. *J. Multivariate Anal., 14*, 17–33.

Haan, L. de and Pickands, J. (1986). Stationary min-stable stochastic processes. *Probability Theory and Related Fields, 72*, 477–492.

Haan, L. de and Resnick, S.I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete, 40*, 317–337.

Haan, L. de and Resnick, S.I. (1979a). Conjugate Π-variation and process inversion. *Ann. Probability, 7*, 1028–1035.

Haan, L. de and Resnick, S.I. (1979b). Derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. *J. Stochastic Processes Appl., 8*, 349–355.

Haan, L. de and Resnick, S.I. (1982). Local limit theorems for sample extremes. *Ann. Probability, 10*, 396–413.

Haan, L. de and Resnick, S. (1984). Asymptotically balanced functions and stochastic compactness of sample extremes. *Ann. Probability, 12*, 588–608.

Haan, L. de and Resnick, S. (1987). On regular variation of probability densities. *J. Stochastic Processes Appl.*

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probability, 16*, 433–439.

Hall, W.J. and Wellner, J.A. (1979). The rate of convergence in law of the maximum of an exponential sample. *Statistica Neerlandica, 33*, 151–154.

Hannan, E.J. and Kanter, M. (1977). Autoregressive processes with infinite variance. *J. Appl. Probability, 14*, 411–415.

Holmes, P. and Strawderman, B. (1969). A note on the waiting time between record observations. *J. Appl. Probability, 6*, 711–714.

Hüsler, J. (1979). The limiting behavior of the last exit time for sequences of independent identically distributed random variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete, 50*, 159–164.

Ignatov, Z. (1977). Ein von der Variationsreihe erzeugter Poissonscher Punktprozess. *Ann. Univ. Sofia.*

Ignatov, Z. (1978). Point processes generated by order statistics and their applications. In: *Point Processes and Queueing Problems*, 109–116; ed. P. Bartfai and J. Tomko, Keszthlely, Hungary. *Coll. Mat. Soc. James Bolyai* 24, North Holland.

Jagers, P. (1974). Aspects of random measures and point processes. *Advances in Probability*, vol. 3. ed. P. Ney and S. Port. Marcel Dekker, New York.

Johns, M.V. (1957). Non-parametric empirical Bayes procedures. *Ann. Math. Statist.*, *28*, 649–669.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

Kallenberg, O. (1983). *Random Measures*, 3rd ed. Akademie-Verlag, Berlin.

Kanter, M. and Steiger, W.L. (1974). Regression and autoregression with infinite variance. *Advances Appl. Probability*, *6*, 768–783.

Karamata, J. (1962). Some theorems concerning slowly varying functions. Mathematics Research Center, Tech. Rep. No. 369, University of Wisconsin, Madison.

Kuratowski, K. (1966). *Topology I.* Academic Press, New York.

Lamperti, J. (1964). On extreme order statistics. *Ann. Math. Statist.* *35*, 1726–1737.

Landau, E. (1916). *Darstellung und Begrundung Einiger Neueren Ergebnisse der Funktionentheorie.* Springer, Berlin.

Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.

Lehmann, E. (1966). Some concepts of dependence. *Ann. Math. Statist.*, *37*, 1137–1153.

Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0, \infty)$. *J. Appl. Probability*, *10*, 109–121.

Loève, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, N.J.

Maller, R. and Resnick, S. (1984). Limiting behaviour of sums and the term of maximum modulus. *Proc. London Math. Soc.* (*3*), *49*, 385–422.

Mardia, K.V. (1970). *Families of Bivariate Distributions.* Griffin's Statistical Monographs, Griffin, London.

Marshall, A. and Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Probability*, *11*, 168–177.

McCormick, W.P. (1983). A conditioned limit law result for jumps in the sequence of partial maxima of a stationary Gaussian sequence. *Sankhya Ser. A*, *45*, 191–200.

Mejzler, D. (1949). On a theorem of B.V. Gnedenko. *Sb. Trudov Inst. Mat. Akad. Nauk Ukrain. R.S.R.*, *12*, 31–35.

Mises, R. Von (1936). La distribution de la plus grande de *n* valeurs. *Selected Papers II, American Math. Soc.*, 271–294.

Mori, T. (1976). Limit laws for maxima and second maxima from strong mixing processes. *Ann. Probability*, *4*, 122–126.

Mori, T. (1977). Limit distributions of two-dimensional point processes generated by strong-mixing sequences. *Yokohama Math. Journal*, *25*, 155–168.

Mori, T. and Oodaira, H. (1976). A functional law of the iterated logarithm for sample sequences. *Yokohama Math. Journal*, *24*, 35–49.

Neveu, J. (1976). Processus ponctuels. Ecole d'Eté de Probabilitês de Saint-Flour VI. Lecture Notes in Mathematics, 598, Springer, Berlin.

Newman, C.M. and Wright, A.L. (1981). An invariance principle for certain dependent sequences. *Ann. Probability*, *9*, 671–675.

Pickands, J. (1967). Sample sequences of maxima. *Ann. Math. Statist. 38*, 1570–1574.
Pickands, J. (1968). Moment convergence of sample extremes. *Ann. Math. Statist., 39*, 881–889.
Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *J. Appl. Probability, 8*, 745–756.
Pickands, J. (1975). Statistical inference using extr^me order statistics. *Ann. Statist., 3*, 119–131.
Pickands, J. (1981). Multivariate extreme value distributions. *Proceedings, 43rd Session of the International Statistical Institute*. Buenos Aires, Argentina. Book 2, 859–878.
Pratt, J. (1960). On interchanging limits and integrals. *Ann. Math. Statist., 31*, 74–77.
Renyi, A. (1962). Theorie des elements sailants d'une suite d'observations. Colloquium on Combinational Methods in Probability Theory. August 1–10. Mathematisk Institute, Aarhus Universitet, Denmark.
Resnick, S.I. (1971). Tail equivalence and its applications. *J. Appl. Probability, 8*, 136–156.
Resnick, S.I. (1973a). Limit laws for record values. *J. Stochastic Processes Appl., 1*, 67–82.
Resnick, S.I. (1973b). Record values and maxima. *Ann. Probability, 1*, 650–662.
Resnick S.I. (1973c). Extremal processes and record value times. *J. Appl. Probability, 10*, 864–868.
Resnick, S.I. (1974). Inverses of extremal processes. *Advances Appl. Probability, 6*, 392–406.
Resnick, S.I. (1975). Weak convergence to extremal processes. *Ann. Probability, 3*, 951–960.
Resnick, S.I. (1983). Extremal processes. *Encyclopedia of Statistical Science*: ed. S. Johnson and S. Kotz. Wiley, New York.
Resnick, S.I. (1986). Point processes, regular variation and weak convergence. *Advances Appl. Probability, 18*, 66–138.
Resnick, S.I. and Rubinovitch, M. (1973). The structure of extremal processes. *Advances Appl. Probability, 5*, 287–307.
Resnick, S.I. and Tomkins, R. (1973). Almost sure stability of maxima. *J. Appl. Probability, 10*, 387–401.
Robbins, H. and Siegmund, D. (1971). On the law of the iterated logarithm for maxima and minima. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, Calif.
Rootzen, H. (1978). Extremes of moving averages of stable processes. *Ann. Probability, 6*, 847–869.
Rootzen, H. (1984). Attainable rates of convergence of maxima. *Statistics and Probability Letters, 2*, 219–221.
Seneta, E. (1973). A Tauberian theorem of E. Landau and W. Feller. *Ann. Probability, 1*, 1057–1058.
Seneta, E. (1976). *Regularly Varying Functions*. Lecture Notes in Mathematics, 508, Springer, Berlin.
Serfozo, R. (1982). Functional limit theorems for extreme values of arrays of independent random variables. *Ann. Probability, 6*, 295–315.
Shorrock, R. (1970). Caravans in traffic flow. Technical Report No. 20, Dept of Statistics, Stanford University, Stanford, Calif.
Shorrock, R. (1972). On record values and record times. *J. Appl. Probability, 9*, 316–326.

Shorrock, R. (1973). Record values and inter-record times. *J. Appl. Probability, 10*, 543–555.

Shorrock, R. (1974). On discrete time extremal processes. *Advances Appl. Probability, 6*, 580–592.

Shorrock, R. (1975). Extremal processes and random measures. *J. Appl. Probability, 12*, 316–323.

Sibuya, M. (1960). Bivariate extreme statistics. *Ann. Inst. Stat. Math., 11*, 195–210.

Simmons, G. (1963). *Topology and Modern Analysis.* McGraw-Hill, New York.

Shorohod, A.V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl., 1*, 261–290.

Smirnov, N.V. (1952). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Transl. Ser. I, 11*, 82–143.

Smith, R. (1982). Uniform rates of convergence in extreme value theory. *Advances Appl. Probability, 14*, 543–565.

Smith, R. and Miller, J. (1984). Predicting records. Pre-print, Imperial College, London.

Stam, A.J. (1977). Regular variation in $\mathbb{R}_+^d$ and the Abel-Tauber theorem, Pre-print, Rijksuniversiteit Groningen, Hoogbouw WSN, Postbus 800, Holland.

Stam, A.J. (1985). Independent Poisson processes generated by record values and inter-record times. *J. Stochastic Processes Appl., 19*, 315–325.

Stone, C. (1963). Weak convergence of stochastic processes defined on a semi-infinite time interval. *Proc. Amer. Math. Soc. 14*, 694–696.

Straf, M. (1971). Weak convergence of stochastic processes with several parameters. *Proceedings of the Sixth Berkely Symposium Mathematical Statistics and Probability 2*, 187–221. Univ of California Press, Berkeley.

Sweeting, T. (1985). On domains of uniform local attraction in extreme value theory. *Ann. Probability, 13*, 196–205.

Tata, M.N. (1969). On outstanding values in a sequence of random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 12*, 9–20.

Tiago de Oliveira, J. (1958). Extremal distributions. *Rev. Fac. Sci. Lisboa, Ser. A*, 7, 215–227.

Tiago de Oliveira, J. (1961). The asymptotical independence of the sample mean and the extremes. *Rev. Fac. Sci. Lisboa, Ser. A, 8*, 229–310.

Tiago de Oliveira, J. (1962/63). Structure theory of bivariate extremes: Extensions. *Estudos Math. Estat. Econom.*, 7, 165–195.

Tiago de Oliveira, J. (1968). Extremal processes: Definition and properties. *Publ. Inst. Statist. Univ. Paris, 17*, 25–36.

Tiago de Oliveira, J. (1970). Biextremal distributions: Statistical decision. *Trab. Estat. Invest. Oper., 21*, 107–117.

Tiago de Oliveira, J. (1975a). Bivariate extremes: Extensions. *Bull. Int. Stat. Inst., 46*, Book 2, 241–252.

Tiago de Oliveira, J. (1975b). Bivariate and multivariate extremal distributions. In: *Statistical Distributions in Scientific Work*, vol. 1, 355–361. Ed. G.P. Patil et al., Reidel, Dordrecht, Holland.

Tiago de Oliveira, J. (1976). Asymptotic behaviour of maxima with periodic disturbances. *Ann. Inst. Statist. Math., 28*, 19–23.

Tiago de Oliveira, J. (1980). Bivariate extremes: Foundations and statistics. In: *Proceedings of the Fifth International Symposium on Multivariate Analysis.* Ed. P.R. Krishnaia, 349–366. North Holland, Amsterdam, Holland.

Tucker, H.G. (1968). Convolutions of distributions attracted to stable laws. *Ann. Math.*

*Statist.*, *39*, 1381–1390.

Vervaat, W. (1973a). Limit theorems for partial maxima and records. Technical Report, Dept. of Math., University of Washington, Seattle.

Vervaat, W. (1973b). Limit theorems for records from discrete distributions. *Stochastic Process. Appl.*, *1*, 317–334.

Vervaat, W. (1986). Ignatov's theorem: A new and short proof. *IMS Bulletin*, *15*, *2*, Abstract 194–36.

Weissman, I. (1975a). Multivariate extremal processes generated by independent non-identically distributed random variables. *J. Appl. Probability*, *12*, 477–487.

Weissman, I. (1975b). Extremal processes generated by independent nonidentically distributed random variables. *Ann. Probability*, *3*, 172–178.

Weissman, I. (1975c). On location and scale functions of a class of limiting processes with applications to extreme value theory. *Ann. Probability*, *3*, 178–181.

Weissman, I. (1976). On weak convergence of extremal processes. *Ann. Probability*, *4*, 470–473.

Welsch, R.E. (1972). Limit laws for extreme order statistics from strong-mixing processes. *Ann. Math. Statist.*, *43*, 439–446.

Westcott, M. (1977). A note on record times. *J. Appl. Probability*, *14*, 637–639.

Whitt, W. (1980). Some useful functions for functional limit theorems. *Math. Operat. Res.*, *5*, 67–85.

Williams, D. (1973). On Renyi's record problem and Engel's series, *Bull. London Math. Soc.*, *5*, 235–237.

Williams, D. (1979). *Diffusions, Markov Processes and Martingales.* Wiley, New York.

# Index