Monotonicity of a Power Function: An Elementary Probabilistic Proof
Author(s): David Gilat
Source: *The American Statistician*, Vol. 31, No. 2 (May, 1977), pp. 91-93
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2683050
Accessed: 12-03-2015 09:24 UTC

# Monotonicity of a Power Function: an Elementary Probabilistic Proof

DAVID GILAT*

*Abstract*

It is pointed out that in many one-sided testing situations for a real-valued parameter $\theta$, the monotonicity of the power function hinges on the stochastic order of the underlying family of distributions $[F_\theta]$ rather than on the stronger property of monotone likelihood ratio of the family. An elementary proof, accessible to students of introductory probability and statistics, is presented.

KEY WORDS: Power function; Sums of independent random variables; Stochastic order; Distributions with given marginals.

## 1. Introduction

It has been customary in introductory probability and statistics texts, to interrupt the discussion of probability by a chapter on statistical inference based on the binomial distribution (e.g., [1] and [2]). This practice serves the desirable purpose of introducing the basic ideas of statistical methodology at the earliest possible stage for the student. However, when it comes to one-sided tests for proportions or later for general means as well as other parameters, the monotonicity of the power function is taken for granted or at best argued on intuitive grounds and illustrated numerically and graphically for a particular test.

Consider, e.g., the binomial case, and let $S_n$ be the number of successes in $n$ independent binomial trials with probability $\theta$ $(0 < \theta < 1)$ of success on each trial. The good (nonrandomized) tests for

$$H_0\colon \theta \le \theta_0 \text{ vs } H_1\colon \theta > \theta_0$$

are, of course, those which reject $H_0$ in favor of $H_1$ when $S_n \ge c$, where $c$ is some fixed integer between 0 and $n$. The power function $\pi = \pi_c$ of such a test is

$$\pi(\theta) = P_\theta[S_n \ge c]. \tag{1.1}$$

*Obviously, the larger the probability, $\theta$, of success in a single trial, the more likely it is to obtain a large number of successes in any given finite number of trials.* A similar statement is typically presented in introductory textbooks as an argument for the monotonicity of $\pi$, while more advanced texts, like [3] and [4], deduce it from monotone likelihood ratio via the Neyman-Pearson lemma. It is the purpose of this note to provide an elementary proof of this "obviousity" and thus establish the monotonicity of the power-function (1.1) without recourse to either Neyman-Pearson or monotone likelihood ratio. The method of proof applies to one-sided situations, whenever the test statistic has a representation as the sum, $S_n$, of independent random variables with a common distribution $F_\theta$, and the family $[F_\theta]$ satisfies

$$1 - F_{\theta_1}(t) \le 1 - F_{\theta_2}(t) \tag{1.2}$$

for all $t$, whenever $\theta_1 < \theta_2$. A one-parameter family $[F_\theta]$, of distributions which satisfies (1.2) is said to be *stochastically increasing*. It is thus argued that monotonicity of $\pi(\theta) = P_\theta[S_n \ge c]$ does not require the full force of monotone likelihood ratio. The strictly weaker ([4], p. 75) stochastic order, (1.2), of the underlying family of distributions of a single (possibly transformed) observation is sufficient.

## 2. The Binomial Case

Observe that in this case

$$\pi(\theta) = P_\theta[S_n \ge c] = \sum_{k=c}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \tag{2.1}$$

is simply a polynomial in $\theta$. When $\pi(\theta)$ is expressed in ascending powers of $\theta$, some of the coefficients are positive while some are negative, so that monotonicity of $\pi(\theta)$ for $0 \le \theta \le 1$ is not in evidence at all.

Instead of viewing $\pi(\theta)$ as a polynomial, consider a sequence, $(U_1, \ldots, U_n)$ of $n$ independent random variables with a uniform distribution on the unit interval [0, 1]. For each fixed $\theta$, $0 \le \theta \le 1$, observe which of the $U_i$'s do not exceed $\theta$, and let $S_n(\theta)$ be the number of them. Clearly, $S_n(\theta)$ has the binomial $(n, \theta)$ distribution. Furthermore, for $0 \le \alpha < \beta \le 1$, the joint distribution of $S_n \equiv S_n(\alpha)$ and $T_n \equiv S_n(\beta)$ is such that

$$\Pr[S_n \le T_n] = 1$$

and

$$\begin{aligned} \Pr[S_n < T_n] &> \Pr[\alpha < U_1 \le \beta, \ldots, \alpha < U_n \le \beta] \\ &= \prod_{i=1}^{n} \Pr[\alpha < U_i \le \beta] = (\beta - \alpha)^n > 0. \end{aligned} \tag{2.2}$$

Thus for every $0 < c \le n$, the event $[S_n \ge c]$ strictly (with positive probability to spare) implies the event $[T_n \ge c]$. Consequently,

$$\pi(\alpha) = \Pr[S_n \ge c] < \Pr[T_n \ge c] = \pi(\beta),$$

which demonstrates the strict monotonicity of the power function $\pi$ on [0, 1].

This elegant argument, which I learned from Erich Lehmann, does not seem to generalize to situations

* David Gilat is currently with the University of Minnesota, Minneapolis, MN 55455 on leave from Tel-Aviv University, Israel. This article was revised while the author was at the University of California, Berkeley, July 1976.

other than the binomial. An alternative approach to the binomial case, later to be generalized, therefore, is presented next. For $0 \le \alpha < \beta \le 1$, consider a pair $(X, Y)$ of random variables whose joint distribution given by $\Pr[X = 0, Y = 0] = 1 - \beta$, $\Pr[X = 0, Y = 1] = \beta - \alpha$, and $\Pr[X = 1, Y = 1] = \alpha$, is summarized in the following table.

| $X \backslash Y$ | 0 | 1 | $X$-marginal |
|---|---|---|---|
| 0 | $1 - \beta$ | $\beta - \alpha$ | $1 - \alpha$ |
| 1 | 0 | $\alpha$ | $\alpha$ |
| $Y$-marginal | $1 - \beta$ | $\beta$ | |

As is evident from this table, $X$ and $Y$ have the binomial $(1, \alpha)$ and $(1, \beta)$ as their respective marginal distributions. Furthermore, the joint distribution of $(X, Y)$ is such that

$$\Pr[X \le Y] = 1 \text{ and } \Pr[X < Y] = \beta - \alpha > 0. \quad (2.3)$$

To conclude the argument, take $n$ independent observations, $(X_1, Y_1), \ldots, (X_n, Y_n)$, on the pair $(X, Y)$, let $S_n = X_1 + \cdots + X_n$, and let $T_n = Y_1 + \cdots + Y_n$. Then marginally $S_n$ and $T_n$ have, respectively, the binomial $(n, \alpha)$ and $(n, \beta)$ distributions, while jointly, the pair $(S_n, T_n)$ satisfies

$$\Pr[S_n \le T_n] = 1$$

and

$$\Pr[S_n < T_n] \ge \Pr[X_1 < Y_1, \ldots, X_n < Y_n] \quad (2.4)$$
$$= \prod_1^n \Pr[X_i < Y_i] = (\beta - \alpha)^n > 0.$$

Finally, from (2.4) one proceeds exactly as in the previous proof from (2.2).

## 3. The General Case

The two different proofs presented above for the binomial case have an important common ingredient. The underlying idea in both is the construction of a probabilistic experiment on which the given random variables, $S_n$ and $T_n$, can be jointly realized in such a way that $S_n$ never exceeds $T_n$. In the first proof the experiment consists of $n$ independent observations on a uniform random variable, whereas in the second proof, the probabilistic setup is formed by $n$ independent observations on the special pair $(X, Y)$. It will now be shown that a suitable joint distribution can be constructed not only for the binomial case but, in fact, for any two random variables whose distribution functions satisfy the relation (1.2). To state the result, let $F$ be the distribution function of the random variable $X$, and let $G$ be the distribution function of $Y$.

(A) *Proposition:* If $1 - F(t) \le 1 - G(t)$ for all $t$ (i.e., if $Y$ is stochastically larger than $X$), then the pair $(X, Y)$ admits a joint distribution $H$, with the prescribed marginals $F$ and $G$, for which the event $[X \le Y]$ has probability 1, and, unless $F = G$, the event $[X < Y]$ has positive probability.

This proposition is well-known. Its standard proof (e.g., [4], p. 73, Lemma 1), however, is inaccessible to students before they reach measure-theoretic probability and statistical theory. An alternative, more elementary, and perhaps new proof of (A) is, therefore suggested here.

As is customary, write $a \wedge b$ for the minimum of the numbers $a$ and $b$. Given the distribution functions $F$ and $G$, define the function $H$ on the plane by

$$H(x, y) = F(x) \wedge G(y). \quad (3.1)$$

It is easily seen that $H$ is a two-dimensional distribution function having $F$ and $G$ for its marginals. Next let $(X, Y)$ be a pair of random variables with the joint distribution $H$. Then for any $x_1 < x_2$ and $y_1 < y_2$,

$$\Pr[x_1 < X \le x_2, y_1 < Y \le y_2] \quad (3.2)$$
$$= H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$
$$= F(x_2) \wedge G(y_2) - F(x_2) \wedge G(y_1)$$
$$- F(x_1) \wedge G(y_2) + F(x_1) \wedge G(y_1).$$

Finally, observe that under the condition $F \ge G$, when the rectangle $(x_1, x_2] \times (y_1, y_2]$ is situated (strictly) below the main diagonal $D = \{(x, y): x = y\}$ of the plane (i.e., when $x_1 > y_2$), then, since $F$ is nondecreasing, the last expression in (3.2) reduces to $G(y_2) - G(y_1) - G(y_2) + G(y_1) = 0$. Consequently, the event $[X \le Y]$ has probability 1, under $H$. Similarly, it can be argued that unless $F = G$, the $H$ probability of the event $[X = Y]$ is strictly less than 1, which completes the proof of (A).

(B) *Corollary:* Let $S_n$ be the sum of $n$ independent observations on the random variable $X$, and let $T_n$ be the sum of $n$ independent observations on the random variable $Y$. If $\Pr[X > t] \le \Pr[Y > t]$ for all $t$, then $\Pr[S_n > t] \le \Pr[T_n > t]$ for all $t$. Furthermore, unless $X$ and $Y$ have the same distribution, the last inequalities are strict for all values of $t$ for which $\Pr[T_n > t] > 0$.

*Proof:* (B) is argued from (A) in exactly the same way as, in the binomial case, (2.4) was argued from (2.3).

## 4. Concluding Remarks

(i) The distribution of $S_n$ in (B) is of course the $n$-fold convolution of the distribution of $X$. In this terminology, (B) merely says that stochastic order is preserved under convolution. The purpose of this note was to provide an elementary proof of this known fact and to relate it to the monotonicity of the power function whenever the test statistic is of a suitable structure.

(ii) For arbitrary distribution functions $F$ and $G$, the two-dimensional distribution function $H$ defined by

(3.1) was first introduced by Hoeffding [8] and later independently rediscovered by Fréchet [7] as the maximal distribution in the plane with given marginals $F$ and $G$. Feller ([6], p. 166) refers to $H$ as the Fréchet maximal distribution. $H$ has many interesting properties. For example, the $L_1$ distance, $E\,|X - Y|$, between the random variables $X$ and $Y$ with respective marginal distributions $F$ and $G$, is minimized when their joint distribution is $H$. This fact is reestablished by Dubins and Meilijson ([5], Lemma 3.2.1) in their study of stability of certain optimization problems. More recently Schaffer [9] has shown that the expected range, $E[X_1 \vee \cdots \vee X_k) - (X_1 \wedge \cdots \wedge X_k)]$, of a finite-dimensional random vector $(X_1, \ldots, X_k)$ with prescribed marginals $F_1, \ldots, F_k$, is minimized, when the joint distribution of $(X_1, \ldots, X_k)$ is given by $H(x_1, \ldots, x_k) = F_1(x_1) \wedge \cdots \wedge F_k(x_k)$. Some new properties of $H$ have most recently been obtained by Tchen [10].

[*Received January 13, 1976. Revised November 8, 1976.*]

## References


[1] Lindgren, B. W. and McElrath, G. W., *Introduction to Probability and Statistics*, 3rd ed., New York: The Macmillan Co., 1969.

[2] Robbins, H. and Van Ryzin, J., *Introduction to Statistics*, Chicago: SRA Inc., 1975.

[3] Ferguson, T. S., *Mathematical Statistics*, New York: Academic Press, 1967.

[4] Lehmann, E. L., *Testing Statistical Hypotheses*, New York: John Wiley & Sons, 1959.

[5] Dubins, L. E. and Meilijson, I., "On Stability for Optimization Problems," *Annals of Probability*, 2, (1974) 243–55.

[6] Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed. New York: John Wiley & Sons, 1971.

[7] Fréchet, M., "Sur les Tableaux de Corrélation Dont les Marges Sont Données," *Ann. Univ. de Lyon*, 14, (1951) 53–77.

[8] Hoeffding, W., "Masstabinvariante Korrelationstheorie," *Schriften Math. Inst. Univ. Berlin*, 5, (1940) 181–233.

[9] Schaffer, M., "Note on the $k$-Dimensional Jensen Inequality," *Annals of Probability*, 4, (1976) 502–4.

[10] Tchen, A. H., "Inequalities for Distributions with Given Marginals," *Technical Report No. 19, Department of Statistics, Stanford University*, June 1976.


# Construction of a Markov Chain with Given Stationary Distribution

FRED J. CONNELL*


*Abstract*

Given a rational, finite probability vector, a Markov chain is constructed having the given vector as its stationary distribution.


KEY WORDS: Stationary distribution; Markov chain.

Given finite probability vector $P = [p_1, p_2, \ldots, p_k]$, with each $p_i$ rational, we construct a Markov chain whose stationary distribution is $P$. Let $N$ be a common denominator of the $p_i$. Write $p_i = n_i/N$, $i = 1, 2, \ldots, k$, and partition the set $\{0, 1, \ldots, N-1\}$ into $k$ subsets $S_1, S_2, \ldots, S_k$ of cardinality $n_1, n_2, \ldots, n_k$, respectively. The sets $S_i$ will be considered to be the states of a Markov chain. Fix an integer $m$, $1 \leq m < N$. If the system is in state $S_i$ at a given moment, choose at random one of the integers $n \in S_i$. Then transition to state $S_j$ occurs if $n + m \bmod N$ is an element of $S_j$.

* Fred J. Connell is with Department of Mathematics, Millikin University, Decatur, IL 62522.

If $m$ is relatively prime to $N$, the Markov chain so constructed is irreducible and has the given vector $P$ as its stationary distribution. The proof is an interesting exercise.

For example, given the probability vector $P = [\frac{1}{5}\ \frac{3}{10}\ \frac{1}{2}]$, we take $N = 10$. We might partition the set $[0, 1, \ldots, 9]$ into subsets $S_1 = [0, 5]$, $S_2 = [3, 6, 9]$, and $S_3 = [1, 2, 4, 7, 8]$, and choose $m = 1$. Focusing attention on state $S_2$, we see $3 + 1 = 4 \in S_3$, $6 + 1 = 7 \in S_3$, and $9 + 1 = 0 \in S_1$. Thus $p_{21} = \frac{1}{3}$, $p_{22} = 0$, and $p_{23} = \frac{2}{3}$, where the $p_{ij}$ are the transition probabilities. We find

$$[p_{ij}] = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{bmatrix},$$

and $P$ is the probability eigenvector for this matrix.

[*Received August 30, 1976. Revised December 16, 1976.*]