MATH20802: STATISTICAL METHODS
LECTURE NOTES

1 Moment Generating Function (MGF)

Definition

Let X be a random variable. We define the mgf of X by

$$M_X(t) = E(\exp(tX)) \tag{1}$$

where t is a dummy variable. Note that $M_X(t)$ always exists at $t = 0$ in which case $M_X(0) = 1$. When X is a discrete rv with pmf $p_X(x)$ then

$$M_X(t) = \sum_{j=1}^{\infty} \exp(tx_j) p_X(x_j) \tag{2}$$

while if X is a continuous rv with pdf $f_X(x)$ then

$$M_X(t) = \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx. \tag{3}$$

The mgf does not have any obvious meaning by itself but it is very useful for distribution theory. Its most basic property is that it can be used to generate moments of a distribution.

Properties

1. $E(X^k) = M_X^{(k)}(0)$ where $M_X^{(k)}(t) = d^k M_X(t)/dt^k$.

2. $Var(X) = M_X^{(2)}(0) - [M_X^{(1)}(0)]^2$.

3. If X has mgf $M_X(t)$ then the mgf of $Y = aX + b$ is $\exp(bt) M_X(at)$.

4. If X and Y are random variables with identical mgfs then they must have identical probability distributions.

5. Let $X_1, X_2, \ldots, X_n$ be independent random variables with mgf $M_{X_j}(t)$, $j = 1, 2, \ldots, n$. Then the mgf of $T = X_1 + X_2 + \ldots + X_n$ is

$$M_T(t) = \prod_{j=1}^{n} M_{X_j}(t). \tag{4}$$

If X_j are independent and identically distributed then

$$M_T(t) = M_X^n(t). \tag{5}$$

2 Some Discrete Distributions

Bernoulli Distribution

A random variable X taking the values $X = 1$ (success) and $X = 0$ (failure) with probabilities p and $1 - p$, respectively, is said to have the Bernoulli distribution, written as $X \sim Bernoulli(p)$.

Uniform Distribution

A random variable X taking the n different values $\{x_1, x_2, \ldots, x_n\}$ with equal probability is said to have the discrete uniform distribution. Its pmf is $p(x) = 1/n$ for $x = x_1, x_2, \ldots, x_n$, where $x_i \neq x_j$ for $i \neq j$. If $\{x_1, x_2, \ldots, x_n\} = \{1, 2, \ldots, n\}$ then $E(X) = (n + 1)/2$ and $Var(X) = (n^2 - 1)/12$.

Binomial Distribution

Consider an experiment of m repeated trials where the following are valid:

1. all trials are statistically independent (in the sense that knowing the outcome of any particular one of them does not change one's assessment of chance related to any others);
2. each trial results in only one of two possible outcomes, labeled as "success" and "failure";
3. and, the probability of success on each trial, denoted by p, remains constant.

Then the random variable X that equals the number of trials that result in a success is said to have the binomial distribution with parameters m and p, written as $X \sim Bin(m, p)$. The pmf of X is:

$$p(x) = \binom{m}{x} p^x (1 - p)^{m-x} \tag{6}$$

for $x = 0, 1, \ldots, m$, where

$$\binom{m}{x} = \frac{m!}{x!(m - x)!}. \tag{7}$$

The cdf is:

$$F(x) = \sum_{i=0}^{x} \binom{m}{i} p^i (1 - p)^{m-i}. \tag{8}$$

The expected value is $E(X) = mp$ and the variance is $Var(X) = mp(1 - p)$.

Negative Binomial Distribution

Consider again a sequence of trials where the following are valid:

1. all trials are statistically independent (in the sense that knowing the outcome of any particular one of them does not change one's assessment of chance related to any others);
2. each trial results in only one of two possible outcomes, labeled as "success" and "failure";
3. and, the probability of success on each trial, denoted by p, remains constant.

Then the random variable X that equals the number of trials up to including the rth success is said to have the negative binomial distribution with parameters r and p, written as $X \sim NB(r, p)$. The pmf of X is:

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \tag{9}$$

for $x = r, r+1, \ldots$. The expected value is $E(X) = r/p$ and the variance is $Var(X) = r(1-p)/p^2$.

Geometric Distribution

The geometric distribution is the special case of the negative binomial for $r = 1$. If a random variable X has this distribution then we write $X \sim Geom(p)$. The pmf of X is:

$$p(x) = p(1-p)^{x-1} \tag{10}$$

for $x = 1, 2, \ldots$. The expected value is $E(X) = 1/p$ and the variance is $Var(X) = (1-p)/p^2$.

Poisson Distribution

Given a continuous interval (in time, length, etc), assume discrete events occur randomly throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

1. the probability of more than one occurrence in a subinterval is zero;
2. the probability of one occurrence in a subinterval is the same for all subintervals and proportional to the length of the subinterval;
3. and, the occurrence of an event in one subinterval has no effect on the occurrence or non-occurrence in another non-overlapping subinterval,

If the mean number of occurrences in the interval is λ, the random variable X that equals the number of occurrences in the interval is said to have the Poisson distribution with parameter λ, written as $X \sim Po(\lambda)$. The pmf of X is:

$$p(x) = \frac{\lambda^x \exp(-\lambda)}{x!} \tag{11}$$

for $x = 0, 1, 2, \ldots$. The cdf is:

$$F(x) = \sum_{i=0}^{x} \frac{\lambda^i \exp(-\lambda)}{i!}. \tag{12}$$

The expected value is $E(X) = \lambda$ and the variance is $Var(X) = \lambda$. The Poisson distribution can be derived as the limiting case of the binomial under the conditions that $m \to \infty$ and $p \to 0$ in such a manner that mp remains constant, say $mp = \lambda$.

3 Some Continuous Distributions

Normal Distribution

If a random variable X has the pdf

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < x < \infty \tag{13}$$

then it is said to have the normal distribution with parameters μ ($-\infty < \mu < \infty$) and σ ($\sigma > 0$), written as $X \sim N(\mu, \sigma^2)$. A normal distribution with $\mu = 0$ and $\sigma = 1$ is called the standard normal distribution. A random variable having the standard normal distribution is denoted by Z. The normal random variable X has the following properties:

1. if X has the normal distribution with parameters μ and σ then $Y = \alpha X + \beta$ is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha\sigma$. In particular, $Z = (X - \mu)/\sigma$ has the standard normal distribution.
2. the normal pdf is a bell-shaped curve that is symmetric about μ and that attains its maximum value of:

   $$\frac{1}{\sqrt{2\pi}\sigma} = \frac{0.399}{\sigma} \tag{14}$$

   at $x = \mu$.
3. 68.26% of the total area bounded by the curve lies between $\mu - \sigma$ and $\mu + \sigma$.
4. 95.44% is between $\mu - 2\sigma$ and $\mu + 2\sigma$.
5. 99.74% is between $\mu - 3\sigma$ and $\mu + 3\sigma$.
6. the expected value is $E(X) = \mu$.

7. the variance is $Var(X) = \sigma^2$.

8. the mgf is

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right). \tag{15}$$

9. the cdf for the normal random variable X is

$$F(x) = \Pr(X \le x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy. \tag{16}$$

   It can be rewritten as

$$F(x) = \Pr\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = \Pr\left(Z < \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right), \tag{17}$$

   where $\Phi(\cdot)$ is the cdf for the standard normal random variable:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left\{-\frac{y^2}{2}\right\} dy. \tag{18}$$

10. the $100(1-\alpha)\%$ percentile of the normal variable X is given by the simple formula:

$$x_\alpha = \mu + \sigma z_\alpha, \tag{19}$$

   where z_α is the $100(1-\alpha)\%$ percentile of the standard normal random variable Z.

11. $z_\alpha = -z_{1-\alpha}$.

12. if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent then $aX_1 + bX_2 + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$.

13. if $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ are iid then $\overline{X} \sim N(\mu, \sigma^2/n)$, where

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{20}$$

   is the sample mean.

14. if $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ are iid then $\overline{X}$ and S^2 are independently distributed, where

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{21}$$

   is the sample variance.

Log–normal Distribution

If a random variable X has the pdf

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ -\frac{(\log x - \mu)^2}{2\sigma^2} \right\}, \qquad x > 0 \tag{22}$$

then it is said to have the lognormal distribution with parameters μ $(-\infty < \mu < \infty)$ and σ $(\sigma > 0)$, written as $X \sim LN(\mu, \sigma^2)$. The lognormal random variable X has the following properties:

1. $\log(X)$ has the normal distribution with parameters μ and σ.
2. the cdf is
$$F(x) = \Pr(X \le x) = \Pr(\log X \le \log x) = \Phi \left(\frac{\log x - \mu}{\sigma} \right), \tag{23}$$
where Φ is the cdf of the standard normal distribution.
3. the $100(1-\alpha)\%$ percentile is
$$x_\alpha = \exp\left(\mu + \sigma z_\alpha\right), \tag{24}$$
where z_α is the $100(1-\alpha)\%$ percentile of the standard normal distribution.
4. the expected value is:
$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right). \tag{25}$$
5. the variance is:
$$Var(X) = \left\{\exp\left(\sigma^2\right) - 1\right\} \exp\left(2\mu + \sigma^2\right). \tag{26}$$

Uniform Distribution

If a random variable X has the pdf

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b, \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

then it is said to have the uniform distribution over the interval (a, b), written as $X \sim Uni(a, b)$. The uniform random variable X has the following properties:

1. the cdf is
$$F(x) = \begin{cases} 0, & \text{if } x \le a, \\ \frac{x-a}{b-a}, & \text{if } a < x < b, \\ 1, & x \ge b. \end{cases} \tag{28}$$
2. the expected value is $E(X) = (a+b)/2$.
3. the variance is $Var(X) = (b-a)^2/12$.
4. the mgf is $M_X(t) = \{\exp(bt) - \exp(at)\}/((b-a)t)$.

Exponential Distribution

If a random variable X has the pdf

$$f(x) = \lambda \exp\left(-\lambda x\right), \qquad x \geq 0, \quad \lambda > 0 \tag{29}$$

then it is said to have the exponential distribution with parameter λ, written as $X \sim Exp(\lambda)$. This parameter λ represents the mean number of events per unit time, e.g. the rate of arrivals or the rate of failure. The exponential random variable X has the following properties:

1. closely related to the Poisson distribution – if X describes say the time between two failures then the number of failures per unit time has the Poisson distribution with parameter λ.
2. the cdf is
$$F(x) = \lambda \int_0^x \exp\left(-\lambda y\right) dy = 1 - \exp\left(-\lambda x\right). \tag{30}$$
3. the $100(1-\alpha)\%$ percentile is
$$x_\alpha = -\frac{1}{\lambda}\log(\alpha). \tag{31}$$
4. the expected value is:
$$E(X) = \frac{1}{\lambda}. \tag{32}$$
5. the variance is:
$$Var(X) = \frac{1}{\lambda^2}. \tag{33}$$
6. the mgf is:
$$M_X(t) = \frac{\lambda}{\lambda - t}. \tag{34}$$

Gamma Distribution

If a random variable X has the pdf

$$f(x) = \frac{\lambda^a x^{a-1}\exp(-\lambda x)}{\Gamma(a)}, \qquad x \geq 0, \quad a > 0, \quad \lambda > 0 \tag{35}$$

then it is said to have the gamma distribution with parameters a and λ, written as $X \sim Ga(a, \lambda)$. Here,

$$\Gamma(a) = \int_0^\infty x^{a-1}\exp(-x)dx \tag{36}$$

is the gamma function. It satisfies the recurrence relation

$$\Gamma(a+1) = a\Gamma(a). \tag{37}$$

If a is a positive integer $\Gamma(a) = (a-1)!$. The gamma random variable X has the following properties:

1. if $a = 1$ then X is an exponential random variable with $\lambda = 1$.
2. the cdf is
$$F(x) = \frac{\lambda^a}{\Gamma(a)} \int_0^x y^{a-1} \exp(-\lambda y) dy. \tag{38}$$
3. the expected value is $E(X) = a/\lambda$.
4. the variance is $Var(X) = a/\lambda^2$.
5. the mgf is:
$$M_X(t) = \left(\frac{\lambda}{\lambda - t} \right)^a. \tag{39}$$

Beta Distribution

If a random variable X has the pdf
$$f(x) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} x^{a_1 - 1}(1 - x)^{a_2 - 1}, \qquad 0 \le x \le 1, \quad a_1 > 0, \quad a_2 > 0 \tag{40}$$
then it is said to have the beta distribution with parameters a_1 and a_2, written as $X \sim B(a_1, a_2)$. The beta random variable X has the following properties:

1. If $a_1 = 1$ and $a_2 = 1$ then X has the uniform distribution on $(0, 1)$.
2. the cdf is
$$F(x) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \int_0^x y^{a_1 - 1}(1 - y)^{a_2 - 1} dy. \tag{41}$$
3. the expected value is:
$$E(X) = \frac{a_1}{a_1 + a_2}. \tag{42}$$
4. the variance is:
$$Var(X) = \frac{a_1 a_2}{(a_1 + a_2)^2 (a_1 + a_2 + 1)}. \tag{43}$$

Gumbel Distribution

If a random variable X has the pdf
$$f(x) = \frac{1}{\sigma} \exp\left[-\frac{x - \mu}{\sigma} - \exp\left\{ -\frac{x - \mu}{\sigma} \right\} \right], \qquad -\infty < x < \infty, \tag{44}$$
then it is said to have the Gumbel distribution with parameters μ $(-\infty < \mu < \infty)$ and σ $(\sigma > 0)$, written as $X \sim Gum(\mu, \sigma)$. The Gumbel random variable X has the following properties:

1. the cdf is

$$F(x) = \exp\left[-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right]. \tag{45}$$

2. the $100(1-\alpha)\%$ percentile is

$$x_\alpha = \mu - \sigma \log\log\left(\frac{1}{1-\alpha}\right). \tag{46}$$

3. the expected value is:

$$E(X) = \mu + 0.57722\sigma. \tag{47}$$

4. the variance is:

$$Var(X) = 1.64493\sigma^2. \tag{48}$$

Fréchet Distribution

If a random variable X has the pdf

$$f(x) = \frac{\lambda}{\sigma}\left(\frac{\sigma}{x}\right)^{\lambda+1} \exp\left\{-\left(\frac{\sigma}{x}\right)^{\lambda}\right\}, \qquad x \geq 0, \tag{49}$$

then it is said to have the Fréchet distribution with parameters σ ($\sigma > 0$) and λ ($\lambda > 0$), written as $X \sim Frechet(\lambda, \sigma)$. The Fréchet random variable X has the following properties:

1. $\log(X)$ has the Gumbel distribution.
2. the cdf is

$$F(x) = \exp\left\{-\left(\frac{\sigma}{x}\right)^{\lambda}\right\}. \tag{50}$$

3. the $100(1-\alpha)\%$ percentile is

$$x_\alpha = \sigma\left\{\log\left(\frac{1}{1-\alpha}\right)\right\}^{-1/\lambda}. \tag{51}$$

4. the expected value is:

$$E(X) = \sigma\Gamma\left(1-\frac{1}{\lambda}\right), \qquad \lambda > 1. \tag{52}$$

5. the variance is:

$$Var(X) = \sigma^2\left\{\Gamma\left(1-\frac{2}{\lambda}\right) - \Gamma^2\left(1-\frac{1}{\lambda}\right)\right\}, \qquad \lambda > 2. \tag{53}$$

Weibull Distribution

If a random variable X has the pdf

$$f(x) = \frac{\lambda}{\sigma}\left(\frac{x}{\sigma}\right)^{\lambda-1} \exp\left\{-\left(\frac{x}{\sigma}\right)^{\lambda}\right\}, \qquad x \geq 0, \tag{54}$$

then it is said to have the Weibull distribution with parameters σ ($\sigma > 0$) and λ ($\lambda > 0$), written as $X \sim We(\lambda, \sigma)$. The weibull random variable X has the following properties:

1. the special case for $\sigma = 1$ and $\lambda = 1$ is the exponential distribution.
2. $-\log(X)$ has the Gumbel distribution.
3. the cdf is
$$F(x) = 1 - \exp\left\{-\left(\frac{x}{\sigma}\right)^{\lambda}\right\}. \tag{55}$$
4. the $100(1-\alpha)\%$ percentile is
$$x_\alpha = \sigma\left\{\log\left(\frac{1}{\alpha}\right)\right\}^{1/\lambda}. \tag{56}$$
5. the expected value is:
$$E(X) = \sigma\Gamma\left(1 + \frac{1}{\lambda}\right). \tag{57}$$
6. the variance is:
$$Var(X) = \sigma^2\left\{\Gamma\left(1 + \frac{2}{\lambda}\right) - \Gamma^2\left(1 + \frac{1}{\lambda}\right)\right\}. \tag{58}$$

4 Parameter Estimation

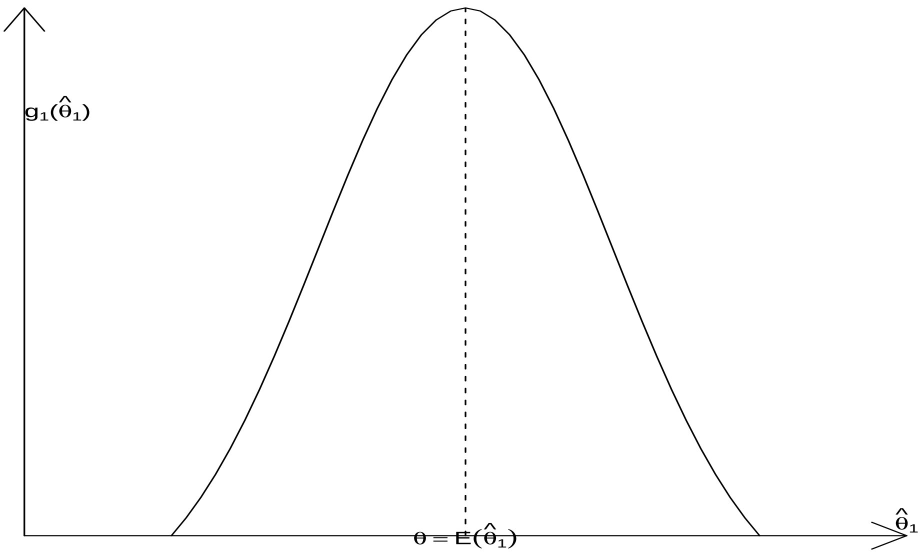
The objective of statistics is to make inferences about a population based on the information contained in a sample. The usual procedure is to firstly hypothesize a probability model to describe the variation observed in the data. Such a model will contain parameters whose values need to be estimated using the sample data. Sometimes these parameters correspond to those which are of direct interest, such as the mean and variance of a normal distribution or the probability p in a binomial distribution. In other circumstances we require estimates of the model parameters in order that we can fit the model to the data and then use it to make inferences about other population parameters, such as the minimum or maximum value in a fixed size sample.

General Framework

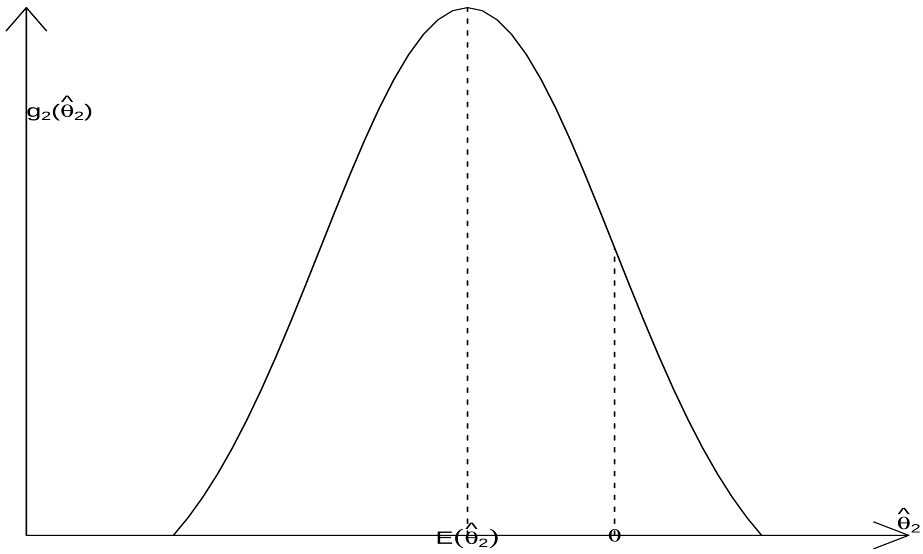
Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution $F(x;\theta)$ where θ is a parameter whose values is unknown. A **point estimator** of θ, denoted by $\widehat{\theta}$ is a real single–valued function of the sample $X_1, X_2, \ldots, X_n$ while the value it assumes for a set of actual data $x_1, x_2, \ldots, x_n$ is a called a **point estimate** of θ, i.e. $\widehat{\theta} = h(X_1, X_2, \ldots, X_n)$ which assumes the numerical values $h(x_1, x_2, \ldots, x_n)$ thus giving a single number or point that estimates the target parameter. Clearly, $\widehat{\theta} = h(\mathbf{X})$ is a random variable with a probability distribution called its **sampling distribution**. Note that θ maybe a vector of t parameters in which case we require t separate estimators. For example, the normal pdf depends on two parameter μ and σ.

We would like an estimator $\widehat{\theta}$ of θ such that:

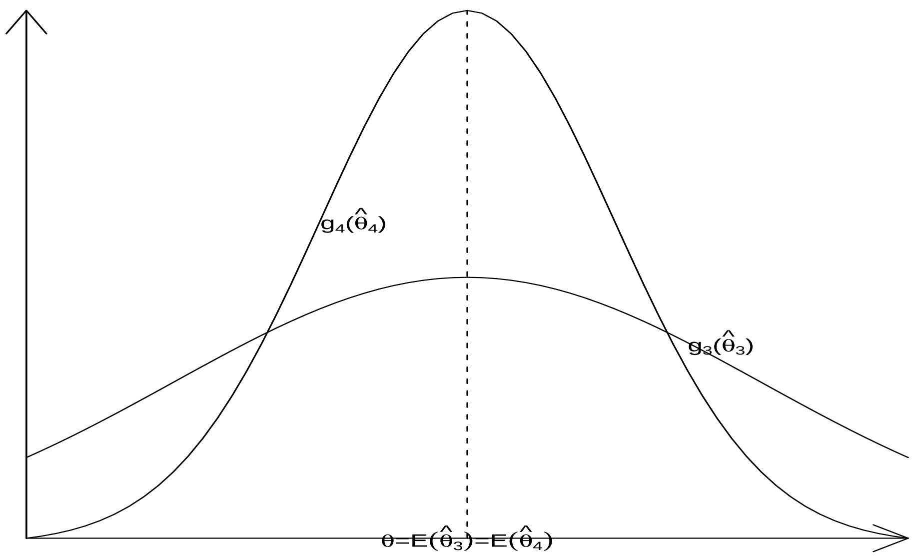
1. the sampling distribution of $\widehat{\theta}$ is centered about the target parameter, θ. In other words, we would like the expected value of $\widehat{\theta}$ to equal θ. So, if $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are two estimators of θ one would like to choose



rather than



2. the spread of the sampling distribution of $\widehat{\theta}$ to be as small as possible. So, if $\widehat{\theta}_3$ and $\widehat{\theta}_4$ are two estimators of θ with densities



we would prefer $\widehat{\theta}_4$ because it has smaller variance.

Definitions

1. $\widehat{\theta}$ is an **unbiased estimator** of θ if $E(\widehat{\theta}) = \theta$. Otherwise, $\widehat{\theta}$ is said to be **biased**.

2. The **bias** of a point estimator $\widehat{\theta}$ is given by $bias(\widehat{\theta}) = E(\widehat{\theta}) - \theta$.

3. The **mean squared error** of a point estimator $\widehat{\theta}$ is given by $E(\widehat{\theta} - \theta)^2$ and is denoted by $MSE(\widehat{\theta})$.

Properties

1. Certain biased estimators can be modified to obtain unbiased estimators, e.g. if $E(\widehat{\theta}) = k\theta$ then $\widehat{\theta}/k$ is unbiased for θ.

2. If $\lim_{n\to\infty} bias(\widehat{\theta}) = 0$ then $\widehat{\theta}$ is said to be **asymptotically unbiased**.

3. $MSE(\widehat{\theta}) = Var(\widehat{\theta}) + [bias(\widehat{\theta})]^2$. Prove this.

4. If $E(\widehat{\theta}) = \theta$ then $MSE(\widehat{\theta}) = Var(\widehat{\theta})$.

5. Given two estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ of estimator $\widehat{\theta}$ we prefer to use the estimator with the smallest MSE, i.e. $\widehat{\theta}_1$ if $MSE(\widehat{\theta}_1) < MSE(\widehat{\theta}_2)$, otherwise choose $\widehat{\theta}_2$.

6. We say that $\widehat{\theta}$ is a **consistent** estimator of θ if $\lim_{n\to\infty} MSE(\widehat{\theta}) = 0$. Since $MSE(\widehat{\theta}) = Var(\widehat{\theta}) + [bias(\widehat{\theta})]^2$, we require both $Var(\widehat{\theta})$ and $bias(\widehat{\theta})$ to approach zero as $n \to \infty$.

5 Likelihood Function

Let $x_1, x_2, \ldots, x_n$ be a random sample of observations taken on corresponding iid random variables $X_1, X_2, \ldots, X_n$. If the X_i's are all discrete with the pmf $p(x;\theta)$ then we define the likelihood function as

$$L(\theta; x_1, x_2, \ldots, x_n) = p(x_1;\theta)p(x_2;\theta)\cdots p(x_n;\theta). \tag{59}$$

If the X_i's are all continuous with the pdf $f(x;\theta)$ then we define the likelihood function as

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta). \tag{60}$$

6 Maximum Likelihood Estimator

Definition

Let $L(\theta)$ be the likelihood function for a given random sample $x_1, x_2, \ldots, x_n$. The **maximum likelihood estimator** (MLE) of θ is the value that maximizes $L(\theta)$. It can be found by standard techniques in calculus. The usual approach is to take the log–likelihood:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log p(x_i; \theta) \tag{61}$$

in the discrete case,

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta) \tag{62}$$

in the continuous case, and solving the equation

$$\frac{\partial l(\theta)}{\partial \theta} = 0 \tag{63}$$

for $\theta = \widehat{\theta}$, making sure that

$$\left. \frac{\partial^2 l(\theta)}{\partial \theta^2} \right|_{\theta = \widehat{\theta}} < 0. \tag{64}$$

Then $\widehat{\theta}$ is said to be the MLE of θ. If $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ is a vector then one will need to solve the equations

$$\frac{\partial l(\theta)}{\partial \theta_1} = 0, \tag{65}$$

$$\frac{\partial l(\theta)}{\partial \theta_2} = 0, \tag{66}$$

$$\vdots \tag{67}$$

$$\frac{\partial l(\theta)}{\partial \theta_p} = 0 \tag{68}$$

simultaneously for $\theta_1 = \widehat{\theta}_1$, $\theta_2 = \widehat{\theta}_2$, …, $\theta_p = \widehat{\theta}_p$.

Properties

1. If $\widehat{\theta}$ is the mle of θ and $g(\cdot)$ is a one–to–one function $g(\widehat{\theta})$ is the mle of $g(\theta)$. This is known as the invariance principle.

2. An mle is a consistent estimator, i.e. if $\widehat{\theta}$ is the mle of θ then $\lim_{n\to\infty} MSE(\widehat{\theta}) = 0$.

3. $\widehat{\theta}$ is approximately normally distributed for large n.

7 Central Limit Theorem (CLT)

The central limit theorem says that if $X_1, X_2, \ldots, X_n$ are random observations from a distribution with expected value μ and variance σ^2, then the the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{69}$$

approaches the standard normal as $n \to \infty$, where

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{70}$$

is the sample mean. If X_is have the normal distribution then Z will be exactly standard normal.

8 Sampling Distributions

There are three important distributions (continuous distributions) used to model the random behavior of various statistics. They are the Chi-squared, the t, and the F distributions.

Chi-squared Distribution

If a random variable X has the pdf

$$f(x) = \frac{1}{2\Gamma(\nu/2)}\left(\frac{x}{2}\right)^{(\nu/2)-1} \exp\left\{-\frac{x}{2}\right\}, \qquad x \geq 0, \quad \nu > 0 \tag{71}$$

then it is said to have the chi-squared distribution with parameter ν. The parameter is termed the number of degrees of freedom (df). The chi-squared random variable X has the following properties:

1. if $X_1, X_2, \ldots, X_n$ are random observations from a normal distribution with parameters μ and σ, then the statistic

   $$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}, \tag{72}$$

   where

   $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{73}$$

   is the sample mean.

2. the $100(1-\alpha)\%$ percentile of a chi-squared distribution is usually denoted by $\chi^2_{\nu,\alpha}$.

3. the expected value is $E(X) = \nu$.

4. the variance is $Var(X) = 2\nu$.
5. the mgf is $M_X(t) = (1-2t)^{-\nu/2}$.
6. if $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$.
7. if $X_1 \sim \chi^2_a$ and $X_2 \sim \chi^2_b$ then $X_1 + X_2 \sim \chi^2_{a+b}$.
8. if $Z_i \sim N(0,1)$, $i = 1, 2, \ldots, n$ are independent then

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n. \tag{74}$$

If $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ are independent then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2_n. \tag{75}$$

Student's t Distribution

If a random variable X has the pdf

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \qquad -\infty < x < \infty, \quad \nu > 0 \tag{76}$$

then it is said to have the Student's t distribution with parameter ν. The parameter is again termed the number of degrees of freedom (df). The Student's t random variable X has the following properties:

1. if $Z \sim N(0,1)$ and $U \sim \chi^2_\nu$ are independent

$$\frac{Z}{\sqrt{U/\nu}} \sim t_\nu. \tag{77}$$

2. if $X_1, X_2, \ldots, X_n$ are random observations from a normal distribution with parameters μ and σ, then the statistic

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}, \tag{78}$$

where

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{79}$$

is the sample mean and

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} \tag{80}$$

is the sample variance.

3. the $100(1-\alpha)\%$ percentile of a t distribution is usually denoted by $t_{\nu,\alpha}$.
4. $t_{\nu,\alpha} = -t_{\nu,1-\alpha}$.
5. the expected value is $E(X) = 0$.
6. the variance is $Var(X) = \nu/(\nu-1)$.
7. as $\nu \to \infty$, $t_\nu \to N(0,1)$.

F Distribution

If a random variable X has the pdf

$$f(x) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}x^{(\nu_1/2)-1}(\nu_2+\nu_1 x)^{-(\nu_1+\nu_2)/2}, \qquad x \geq 0 \tag{81}$$

then it is said to have the F distribution with parameters ν_1 ($\nu_1 > 0$) and ν_2 ($\nu_2 > 0$). Both parameters are termed the number of degrees of freedom. The F random variable X has the following properties:

1. if X_1 and X_2 are independent chi-squared random variables with ν_1 and ν_2 degrees of freedom, respectively, then the statistic

$$\frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\nu_2}. \tag{82}$$

2. Consider the two independent random samples: X_{11}, X_{12}, ..., $X_{1n_1} \sim N(\mu_1, \sigma_1^2)$ and X_{21}, X_{22}, ..., $X_{2n_2} \sim N(\mu_2, \sigma_2^2)$. Then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}, \tag{83}$$

where

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \tag{84}$$

$$\overline{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}, \tag{85}$$

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2 \tag{86}$$

and

$$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\overline{X}_2)^2. \tag{87}$$

3. if $X \sim F_{\nu_1,\nu_2}$ then $1/X \sim F_{\nu_2,\nu_1}$.
4. the $100(1-\alpha)\%$ percentile of an F distribution is usually denoted by $F_{\nu_1,\nu_2,\alpha}$.
5. $F_{\nu_1,\nu_2,1-\alpha} = 1/F_{\nu_2,\nu_1,\alpha}$.
6. the expected value is:

$$E(X) = \frac{\nu_2}{\nu_2 - 2}, \qquad \nu_2 > 2. \tag{88}$$

7. the variance is:

$$Var(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \qquad \nu_2 > 4. \tag{89}$$

8. if $X \sim t_\nu$ then $X^2 \sim F_{1,\nu}$.

9 Hypotheses Testing

As we have discussed earlier, a principal objective of statistics is to make inferences about the unknown values of population parameters based on a sample of data from the population. In this section, we will explore how to test hypotheses about the parameter values.

Definition

A statistical hypothesis is a conjecture or proposition regarding the distribution of one or more random variables. We need to specify the functional form of the underlying distribution as well as the values of any parameters. A **simple** hypothesis completely specifies the distribution whereas a **composite** hypothesis does not. For example, "the data come from a normal distribution with $\mu = 5$ and $\sigma = 1$" is a simple hypothesis while "the data come from a normal distribution with $\mu > 5$ and $\sigma = 1$" is a composite hypothesis.

Elements of a Statistical Test

Null hypothesis (H_0), the hypothesis to be tested.

Alternative hypothesis (H_1), a research hypothesis about the population parameters which we will accept if the data do not provide sufficient support for H_0.

The null hypothesis we will be considering are simple whereas the alternative maybe simple or composite. For example, while testing hypotheses about the mean μ of a normal distribution with known variance σ^2, we may take
$H_0 : \mu = \mu_0$ (where μ_0 is a specified value) and $H_1 : \mu \neq \mu_0$;

$H_0 : \mu = \mu_0$ (where μ_0 is a specified value) and $H_1 : \mu > \mu_0$; or
$H_0 : \mu = \mu_0$ (where μ_0 is a specified value) and $H_1 : \mu < \mu_0$.

Test statistic, a function of the sample data whose value we will use to decide between which of H_0 or H_1 to accept.

Acceptance and Rejection Regions: the set of all possible values a test statistic can take is divided into two non–overlapping subsets called the acceptance and rejection regions. If the value of the test statistic falls in the acceptance region then we accept the claim made under H_0. However, if the value falls into the rejection region then we reject H_0 in favor of the claim made under H_1.

Type I error occurs when we reject H_0 when it is in fact true. The probability of this error is denoted by α and is called the significance level or size of the test. Usually, the value of α is decided in advance, e.g. $\alpha = 0.05$.

Type II error occurs when we accept H_0 when it is in fact false. The probability of this error is denoted by β.

Inferences about Population Mean

Suppose $x_1, x_2, \ldots, x_n$ is a random sample from a normal population with mean μ and variance σ^2 (assumed known). For

$$H_0 : \mu = \mu_0 \tag{90}$$

versus

$$H_1 : \mu \neq \mu_0 \tag{91}$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}}{\sigma} |\overline{x} - \mu_0| > z_{\alpha/2}, \tag{92}$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)\%$ percentile of the standard normal distribution. For

$$H_0 : \mu = \mu_0 \tag{93}$$

versus

$$H_1 : \mu > \mu_0 \tag{94}$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}}{\sigma} (\overline{x} - \mu_0) \geq z_\alpha. \tag{95}$$

For

$$H_0 : \mu = \mu_0 \tag{96}$$

versus

$$H_1 : \mu < \mu_0 \tag{97}$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}}{\sigma} (\overline{x} - \mu_0) \leq -z_\alpha. \tag{98}$$

Suppose $x_1, x_2, \ldots, x_n$ is a random sample from a normal population with mean μ and variance σ^2 (assumed unknown). Denote the sample variance by:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 . \tag{99}$$

For

$$H_0 : \mu = \mu_0 \tag{100}$$

versus

$$H_1 : \mu \neq \mu_0 \tag{101}$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}}{s} |\overline{x} - \mu_0| > t_{n-1,\alpha/2}, \tag{102}$$

where $t_{n-1,\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the t distribution with $n-1$ degrees of freedom. For

$$H_0 : \mu = \mu_0 \tag{103}$$

versus

$$H_1 : \mu > \mu_0 \tag{104}$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}}{s} (\overline{x} - \mu_0) \geq t_{n-1,\alpha}. \tag{105}$$

For

$$H_0 : \mu = \mu_0 \tag{106}$$

versus

$$H_1 : \mu < \mu_0 \tag{107}$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}}{s} (\overline{x} - \mu_0) \leq -t_{n-1,\alpha}. \tag{108}$$

Inferences about Population Variance

Suppose $x_1, x_2, \ldots, x_n$ is a random sample from a normal population with mean μ and variance σ^2. Denote the sample variance by:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 . \tag{109}$$

For

$$H_0 : \sigma = \sigma_0 \tag{110}$$

versus

$$H_1 : \sigma \neq \sigma_0 \tag{111}$$

reject H_0 at significance level α if

$$\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha/2} \tag{112}$$

or

$$\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha/2}, \tag{113}$$

where $\chi^2_{n-1,\alpha/2}$ and $\chi^2_{n-1,1-\alpha/2}$ are the $100(1-\alpha/2)\%$ and $100\alpha/2\%$ percentiles of the chi-square distribution with $n-1$ degrees of freedom. For

$$H_0 : \sigma = \sigma_0 \tag{114}$$

versus

$$H_1 : \sigma > \sigma_0 \tag{115}$$

reject H_0 at significance level α if

$$\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha}. \tag{116}$$

For

$$H_0 : \sigma = \sigma_0 \tag{117}$$

versus

$$H_1 : \sigma < \sigma_0 \tag{118}$$

reject H_0 at significance level α if

$$\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha}. \tag{119}$$

Inferences about Population Proportion

Suppose $x_1, x_2, \ldots, x_n$ is a random sample from the Bernoulli distribution with parameter p and assume $n \geq 25$. For

$$H_0 : p = p_0 \tag{120}$$

versus

$$H_1 : p \neq p_0 \tag{121}$$

reject H_0 at significance level α if

$$\sqrt{\frac{n}{\overline{x}(1-\overline{x})}} \left| \overline{x} - p_0 \right| \geq z_{\alpha/2}, \tag{122}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. For

$$H_0 : p = p_0 \tag{123}$$

versus

$$H_1 : p > p_0 \tag{124}$$

reject H_0 at significance level α if

$$\sqrt{\frac{n}{\overline{x}(1-\overline{x})}} \left(\overline{x} - p_0 \right) \geq z_{\alpha}. \tag{125}$$

For

$$H_0 : p = p_0 \tag{126}$$

versus

$$H_1 : p < p_0 \tag{127}$$

reject H_0 at significance level α if

$$\sqrt{\frac{n}{\overline{x}(1-\overline{x})}} \left(\overline{x} - p_0 \right) \leq -z_{\alpha}. \tag{128}$$

Confidence Intervals

Some $100(1-\alpha)\%$ confidence intervals for

- population mean μ when σ is known or $n \geq 25$:

$$\left(\overline{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right). \tag{129}$$

- population mean μ when σ is unknown and $n < 25$:

$$\left(\overline{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right). \tag{130}$$

- population proportion p when $n \geq 25$:

$$\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right). \tag{131}$$

- population variance σ^2:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right). \tag{132}$$

Testing Equality of Means – Variances Known

Let $x_1, x_2, \ldots, x_m$ be a random sample from a normal population with mean μ_X and variance σ_X^2 (assumed known). Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population with mean μ_Y and variance σ_Y^2 (assumed known). Assume independence of the two samples. For

$$H_0 : \mu_X = \mu_Y$$

versus

$$H_1 : \mu_X \neq \mu_Y$$

reject H_0 at level of significance α if

$$\frac{|\ \bar{x} - \bar{y}\ |}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \quad \geq \quad z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(|Z| \geq \frac{|\ \bar{x} - \bar{y}\ |}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}\right),$$

where Z denotes a random variable having the standard normal distribution.

For

$$H_0 : \mu_X \leq \mu_Y$$

versus

$$H_1 : \mu_X > \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \quad \geq \quad z_\alpha.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(Z \geq \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}\right).$$

For

$$H_0 : \mu_X \geq \mu_Y$$

versus

$$H_1 : \mu_X < \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \quad \leq \quad -z_\alpha.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(Z \leq \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}\right).$$

Testing Equality of Means – Variances Unknown but Equal

Let $x_1, x_2, \ldots, x_m$ be a random sample from a normal population with mean μ_X and variance σ^2 (assumed unknown). Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population with mean μ_Y and variance σ^2 (assumed unknown). Assume independence of the two samples. Estimate the common variance by the pooled sample variance:

$$s_p^2 \quad = \quad \frac{(m-1)s_X^2 + (n-1)s_Y^2}{m+n-2},$$

where

$$s_X^2 \quad = \quad \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x})^2$$

and

$$s_Y^2 \quad = \quad \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Then for testing

$$H_0 : \mu_X = \mu_Y$$

versus

$$H_1 : \mu_X \neq \mu_Y$$

reject H_0 at level of significance α if

$$\frac{|\ \bar{x} - \bar{y}\ |}{\sqrt{s_p^2 \left(\frac{1}{m} + \frac{1}{n} \right)}} \quad \geq \quad t_{m+n-2,\alpha/2},$$

where $t_{m+n-2,\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the t distribution with $m+n-2$ degrees of freedom. The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr \left(|T_{m+n-2}| \geq \frac{|\ \bar{x} - \bar{y}\ |}{\sqrt{s_p^2 \left(\frac{1}{m} + \frac{1}{n} \right)}} \right),$$

where T_{m+n-2} denotes a random variable having the t distribution with $m+n-2$ degrees of freedom.

For testing

$$H_0 : \mu_X \leq \mu_Y$$

versus

$$H_1 : \mu_X > \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x} - \bar{y}}{\sqrt{s_p^2 \left(\frac{1}{m} + \frac{1}{n} \right)}} \quad \geq \quad t_{m+n-2,\alpha}.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr \left(T_{m+n-2} \geq \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2 \left(\frac{1}{m} + \frac{1}{n} \right)}} \right).$$

For testing

$$H_0 : \mu_X \geq \mu_Y$$

versus

$$H_1 : \mu_X < \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x}-\bar{y}}{\sqrt{s_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}} \quad \leq \quad -t_{m+n-2,\alpha}.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(T_{m+n-2} \leq \frac{\bar{x}-\bar{y}}{\sqrt{s_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}}\right).$$

Testing Equality of Means – Variances Unknown and Unequal

Let $x_1, x_2, \ldots, x_m$ be a random sample from a normal population with mean μ_X and variance σ_X^2 (assumed unknown). Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population with mean μ_Y and variance σ_Y^2 (assumed unknown). Assume independence of the two samples. Estimate σ_X^2 and σ_Y^2 by

$$s_X^2 \quad = \quad \frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2$$

and

$$s_Y^2 \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2,$$

respectively.

Then for testing

$$H_0: \mu_X = \mu_Y$$

versus

$$H_1: \mu_X \neq \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\mid \bar{x}-\bar{y} \mid}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}} \quad \geq \quad t_{\nu,\alpha/2},$$

where

$$\nu \quad = \quad \frac{\left(\frac{s_X^2}{m}+\frac{s_Y^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{s_X^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{s_Y^2}{n}\right)^2}$$

and $t_{\nu,\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the t distribution with ν degrees of freedom (approximate ν to the nearest integer if it is not an integer). The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(|T_\nu| \geq \frac{|\,\bar{x}-\bar{y}\,|}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}}\right),$$

where T_ν denotes a random variable having the t distribution with ν degrees of freedom.

For testing

$$H_0 : \mu_X \leq \mu_Y$$

versus

$$H_1 : \mu_X > \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}} \quad \geq \quad t_{\nu,\alpha},$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(T_\nu \geq \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}}\right).$$

For testing

$$H_0 : \mu_X \geq \mu_Y$$

versus

$$H_1 : \mu_X < \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}} \quad \leq \quad -t_{\nu,\alpha},$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(T_\nu \leq \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}}\right).$$

Testing Equality of Means – Paired Data

Let $x_1, x_2, \ldots, x_n$ be a random sample from a normal population with mean μ_X and let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population with mean μ_Y. Suppose $(x_1, y_1), (x_2, y_2), \ldots$ are "paired" as if they were mesured off the same "experimental unit". Tests for equality of the means are based on $D_i = X_i - Y_i$. Let $\bar{d}$ and S_d^2 denote the sample mean and sample variance of D_i, i.e.

$$\bar{d} \quad = \quad \frac{1}{n} \sum_{i=1}^{n} d_i$$

and

$$S_d^2 \quad = \quad \frac{1}{n-1} \sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2 .$$

Then for testing

$$H_0 : \mu_X = \mu_Y$$

versus

$$H_1 : \mu_X \neq \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n} \mid \bar{d} \mid}{S_d} \quad \geq \quad t_{n-1,\alpha/2},$$

where $t_{n-1,\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the t distribution with $n-1$ degrees of freedom. The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(|T_{n-1}| \geq \frac{\sqrt{n} \mid \bar{d} \mid}{S_d}\right),$$

where T_{n-1} denotes a random variable having the t distribution with $n-1$ degrees of freedom.

For testing

$$H_0 : \mu_X \leq \mu_Y$$

versus

$$H_1 : \mu_X > \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}\bar{d}}{S_d} \quad \geq \quad t_{n-1,\alpha}.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(T_{n-1} \geq \frac{\sqrt{n}\bar{d}}{S_d}\right).$$

For testing

$$H_0 : \mu_X \geq \mu_Y$$

versus

$$H_1 : \mu_X < \mu_Y$$

reject H_0 at level of significance α if

$$\frac{\sqrt{n}\bar{d}}{S_d} \quad \leq \quad t_{n-1,\alpha}.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(T_{n-1} \leq \frac{\sqrt{n}\bar{d}}{S_d}\right).$$

Testing Equality of Proportions

Let $x_1, x_2, \ldots, x_m$ be a random sample from a Bernoulli population with parameter p_X and let $y_1, y_2, \ldots, y_n$ be a random sample from a Bernoulli population with parameter p_Y. Assume $m \geq 25$, $n \geq 25$ and the independence of the two samples. Then for testing

$$H_0 : p_X = p_Y$$

versus

$$H_1 : p_X \neq p_Y$$

reject H_0 at level of significance α if

$$\frac{\mid \bar{x} - \bar{y} \mid}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}}} \quad \geq \quad z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(|Z| \geq \frac{\mid \bar{x} - \bar{y} \mid}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}}}\right),$$

where Z denotes a random variable having the standard normal distribution.

For testing

$$H_0 : p_X \leq p_Y$$

versus

$$H_1 : p_X > p_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m}+\frac{\bar{y}(1-\bar{y})}{n}}} \quad \geq \quad z_\alpha.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(Z \geq \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m}+\frac{\bar{y}(1-\bar{y})}{n}}}\right).$$

For testing

$$H_0 : p_X \geq p_Y$$

versus

$$H_1 : p_X < p_Y$$

reject H_0 at level of significance α if

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m}+\frac{\bar{y}(1-\bar{y})}{n}}} \quad \leq \quad -z_\alpha.$$

The corresponding p-value is:

$$p\text{-value} \quad = \quad Pr\left(Z \leq \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m}+\frac{\bar{y}(1-\bar{y})}{n}}}\right).$$

Testing Equality of Variances

Let $x_1, x_2, \ldots, x_m$ be a random sample from a normal population with mean μ_X and variance σ_X^2. Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population with mean μ_Y and variance σ_Y^2. Assume independence of the two samples. Estimate σ_X^2 and σ_Y^2 by

$$s_X^2 \quad = \quad \frac{1}{m-1}\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2$$

and

$$s_Y^2 \quad = \quad \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2,$$

respectively.

Then for testing

$$H_0 : \sigma_X = \sigma_Y$$

versus

$$H_1 : \sigma_X \neq \sigma_Y$$

reject H_0 at level of significance α if

$$\frac{s_X^2}{s_Y^2} \quad > \quad F_{m-1,n-1,\alpha/2}$$

or

$$\frac{s_X^2}{s_Y^2} \quad < \quad F_{m-1,n-1,1-\alpha/2},$$

where $F_{m-1,n-1,\alpha/2}$ and $F^2_{m-1,n-1,1-\alpha/2}$ are the $100(1-\alpha/2)\%$ and $100\alpha/2\%$ percentiles of the F distribution with and $m-1$ and $n-1$ degrees of freedom.

For testing

$$H_0 : \sigma_X \leq \sigma_Y$$

versus

$$H_1 : \sigma_X > \sigma_Y$$

reject H_0 at level of significance α if

$$\frac{s_X^2}{s_Y^2} \quad > \quad F_{m-1,n-1,\alpha}.$$

For testing

$$H_0 : \sigma_X \geq \sigma_Y$$

versus

$$H_1 : \sigma_X < \sigma_Y$$

reject H_0 at level of significance α if

$$\frac{s_X^2}{s_Y^2} \quad < \quad F_{m-1,n-1,1-\alpha}.$$

It is useful to know that $F_{n-1,m-1,\alpha/2} = 1/F_{m-1,n-1,1-\alpha/2}$.

Confidence Intervals

Some $100(1-\alpha)\%$ confidence intervals

- for $\mu_1 - \mu_2$ when σ_X^2 and σ_Y^2 are known:

$$\left(\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}\right).$$

- for $\mu_1 - \mu_2$ when $\sigma_X^2 = \sigma_Y^2$ and the common variance is unknown:

$$\left(\bar{x} - \bar{y} \pm t_{m+n-2,\alpha/2} s_p \sqrt{\frac{1}{m} + \frac{1}{n}}\right).$$

- for $\mu_1 - \mu_2$ when $\sigma_X^2 \neq \sigma_Y^2$ and the variances are unknown:

$$\left(\bar{x} - \bar{y} \pm t_{\nu,\alpha/2}\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}\right).$$

- for $\mu_1 - \mu_2$ for paired samples:

$$\left(\bar{x} - \bar{y} \pm t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}\right).$$

- for $p_1 - p_2$ when $m \geq 25$ and $n \geq 25$:

$$\left(\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}}\right).$$

- for σ_X^2/σ_Y^2:

$$\left(\frac{1}{F_{m-1,n-1,\alpha/2}}\frac{s_X^2}{s_Y^2}, F_{n-1,m-1,\alpha/2}\frac{s_X^2}{s_Y^2}\right).$$

Power Function

We wish to make inferences about a population parameter θ based on a random sample of size n. Suppose Ω is the set of all possible values of θ and let θ_0 be a fixed value in Ω. We wish to test

$$H_0 : \theta = \theta_0$$

versus

$$H_1 : \theta \neq \theta_0$$

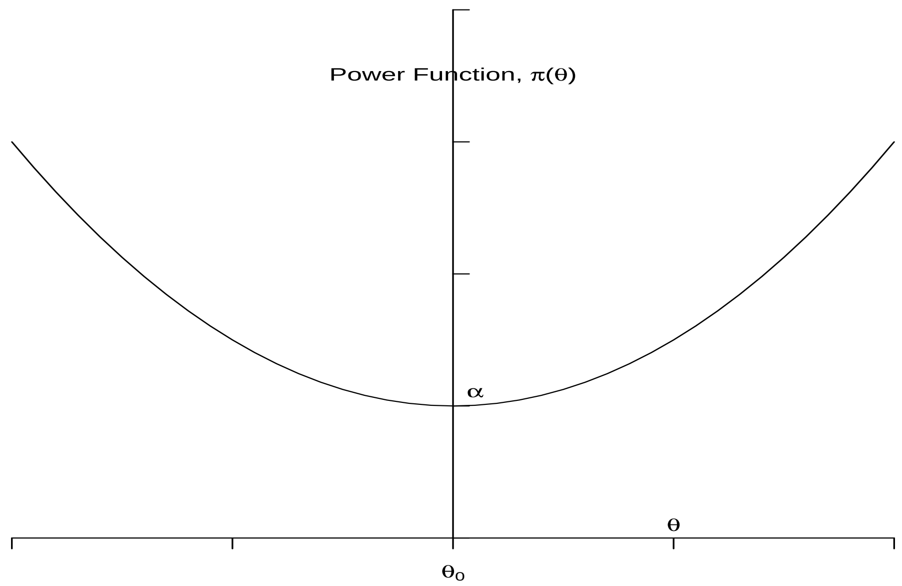
or equivalently

$$H_1 : \theta \in \Omega_1 = \Omega \backslash \{\theta_0\}$$

. Suppose we use test statistics T and reject H_0 at significance level α if T takes values in the rejection region. For any $\theta \in \Omega$, the power function is defined by

$$\pi(\theta) = P(T \in \text{ Rejection Region } \mid \theta), \tag{133}$$

i.e. $\pi(\theta)$ is the probability that the test rejects H_0 given the true value of the parameter is θ. Clearly, $\pi(\theta_0) = \alpha$. A typical power function curve will look like



so that as the true value of θ becomes more distant from θ_0 the power increases. For a fixed sample size n, the strategy is to firstly fix α and then choose that test statistic T which has the largest power for all $\theta \in \Omega_1$. The question then is how do we find such a test statistic?

Neyman–Pearson Lemma

Suppose we wish to test

$$H_0 : \theta = \theta_0$$

versus

$$H_1 : \theta = \theta_1$$

based on a random sample $x_1, x_2, \ldots, x_n$ from a distribution with parameter θ. Let $L(\theta)$ denote the likelihood of the sample. Then, for a given significance level α, the test that maximizes the power at $\theta = \theta_1$ has the rejection region determined by

$$\frac{L(\theta_0)}{L(\theta_1)} < k. \tag{134}$$