قسم الرياضيات

الاحصاء الرياضي2

المرحلة الثالثة

الفصل الدراسي الثاني

ا.م.د تسنيم حسن كاظم ا.م.د صادق ناجي ناصر

Lecture 1

Point Estimation

Consider random variables for which the functional form of the pdf is known, but the distribution depends on an unknown parameter (say, θ) that may have any value in a set (say, Ω) called the parameter space.

A point estimate of a parameter θ is a single number that can be regarded as a good guess for the true value of θ.

Recall that a statistic is a function of a random sample $X_1, X_2, \ldots, X_n$ that is free of the population parameter θ.

Estimator and Estimate

Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x; \theta)$. Any statistic that can be used to estimate the value of θ is called an estimator and is denoted by $\hat{\theta}$. The numerical value of this statistic is called an estimate of θ.

Example

Let the random variable $X \sim N(\theta_1, \theta_2)$, then $\hat{\theta}_1 = \bar{X}$ is an estimator of θ_1 and $\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an estimator for θ_2. But $\bar{x}$ and $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ are the observed values of θ_1 and θ_2.

One of the basic problems is how to find an estimator of the parameter θ.

Method of Moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population X with pdf $f(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta}=\{\theta_1, \theta_2, \ldots, \theta_k\}$ are unknown parameters. Define

$E(X^r) = \int_{-\infty}^{\infty} x^r f(x; \boldsymbol{\theta})dx$, the rth population moment about 0. And $m_r = \frac{1}{n}\sum x_i^r$, the rth sample moment about 0.

In the method of moments, the first k sample moments are set equal to the first k population moments that are given in terms of the unknown parameters. That is

$E(X \quad) = m_1$, $E(X^2) = m_2$, $E(X^3) = m_3$, … , $E(X^k) = m_k$

These k equations are then solved simultaneously for the unknown parameters.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the exponential distribution with parameter θ. Estimate θ by the method of moments.

$$f(x; \theta) = \frac{1}{\theta} e^{\frac{-x}{\theta}} \qquad 0 < x < \infty, \theta > 0$$

$$E(X) = \int_0^{\infty} x \frac{1}{\theta} e^{\frac{-x}{\theta}} \, dx = \theta$$

Set $E(X) = m_1$, where $m_1 = \bar{X}$

Hence, $\hat{\theta} \quad = \bar{X}$ is an estimate of θ.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from N (μ, σ^2). What are the moment estimators for the population parameters μ and σ^2.

We know that for the normal distribution $E(X) = \mu$ *and* $E(X^2) = \sigma^2 + \mu^2$

Set $E(X) = m_1$, where $m_1 = \bar{X}$

Hence, $\hat{\mu} = \bar{X}$

Next, we find the estimator of σ^2 by setting $E(X^2) = m_2$

$\sigma^2 = E(X^2) - [E(X)]^2 = m_2 - \mu^2$

Thus, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Example

Let $X_1, X_2, \ldots, X_n$, be a random sample from a distribution with pdf

$f(x;\, \theta) = \theta x^{\theta-1}, \;\; 0 < x < 1 \; and \; 0 < \theta < \infty$

1. Using the method of moments, find an estimator of θ
2. If $x_1 = 0.2, x_2 = 0.6, x_3 = 0.5, x_4 = 0.3$ is a random sample of size n=4, what is the estimate of θ.

Solution

1. $E(X) = \int_0^1 xf(x;\theta)dx = \int_0^1 x\theta x^{\theta-1}dx = \theta \int_0^1 x^{\theta}dx = \frac{\theta}{\theta+1}\left[x^{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$

   Set $E(X) = m_1$, where $m_1 = \bar{X}$

   We get $\frac{\theta}{\theta+1} = \bar{X}$ and solving this equation for θ, the moment estimator is $\hat{\theta} = \frac{\bar{X}}{1-\bar{X}}$

2. $\bar{x} = \frac{\sum_1^n x_i}{n} = \frac{0.2+0.6+0.5+0.3}{4} = 0.4$, then $\hat{\theta} = \frac{0.4}{1-0.4} = \frac{2}{3}$ is an estimate of θ.

H. W.

1. Let $X_1, X_2, \ldots, X_n$, be a random sample from U(0, θ). Find an estimator of θ by the method of moments.
2. Let $X_1, X_2, \ldots, X_n$, be a random sample from Gamma(α, β). Estimate α and β by the method of moments.
3. Let $X_1, X_2, \ldots, X_n$, be a random sample from b(n, p). What is the moment estimator of p.

Lecture 2

Maximum Likelihood Estimation

Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution having pdf $f(x; \boldsymbol{\theta})$. The joint pdf of $X_1, X_2, \ldots, X_n$ is

$f(x_1, x_2, \ldots, x_n ; \boldsymbol{\theta}) = f(x_1; \boldsymbol{\theta})\, f(x_2; \boldsymbol{\theta}) \ldots f(x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$

This joint pdf when regarded as a function of $\boldsymbol{\theta}$ is called the likelihood function of the random sample.

$L(\boldsymbol{\theta}) = f(x_1; \boldsymbol{\theta})\, f(x_2; \boldsymbol{\theta}) \ldots f(x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$

When $\boldsymbol{\theta}$ is replaced by a statistic $u(X_1, X_2, \ldots, X_n)$ in the likelihood function L, such that

$L[u(X_1, X_2, \ldots, X_n)]$ is at least as great as $L(\boldsymbol{\theta})$, then the statistic $u(X_1, X_2, \ldots, X_n)$ will be called a maximum likelihood estimator(mle) of $\boldsymbol{\theta}$ and will be denoted by $\hat{\boldsymbol{\theta}} = u(X_1, X_2, \ldots, X_n)$.

In many instances there will be a unique mle $\hat{\boldsymbol{\theta}}$ of a parameter θ, and often it may be obtained by the process of differentiation.

Example

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with pdf

$f(x; \theta) = \theta^x (1-\theta)^{1-x} \quad x = 0, 1, \quad 0 \le \theta \le 1$

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Since L(θ) and its logarithm ln L(θ) have their maxima at the same value θ, it is sometimes easier to maximize ln L(θ). Here

$$\ln L(\theta) = \left(\sum_{i=1}^{n} x_i\right)\ln\theta + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-\theta)$$

$$\frac{\partial \ln L(\theta)}{\partial\,\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} = 0 \qquad ,\theta \neq 0,1$$

$$(1-\theta)\sum_{i=1}^{n} x_i = \theta\left(n - \sum_{i=1}^{n} x_i\right)$$

$$\sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i = n\theta - \theta\sum_{i=1}^{n} x_i$$

Hence, $\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$ is the value of θ that maximizes L(θ). That is the statistic $\hat{\theta} = \bar{X}$ is the mle of θ.

Note that

$$\frac{\partial^2}{\partial\theta^2}\ln L(\theta)|_{\theta=\hat{\theta}} = -\frac{\sum_{i=1}^{n} x_i}{\hat{\theta}^2} - \frac{n - \sum_{i=1}^{n} x_i}{\left(1-\hat{\theta}^2\right)} < 0$$

Because

1. θ^2 and $(1-\theta)^2$ are non-negative numbers
2. X is a Bernoulli random variable taking the values 0 or 1and $\sum_{i=1}^{n} x_i$ is positive.
3. The largest value that $\sum_{i=1}^{n} x_i$ takes is n. So, (n-$\sum_{i=1}^{n} x_i$) is always positive.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta_1, \theta_2)$. Find the mle of $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \exp\,[-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2]$$

$$= (2\pi\theta_2)^{-\frac{n}{2}} \exp\,[-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2]$$

$$\ln L(\boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2} ln\theta_2 - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

$$\frac{\partial lnL(\boldsymbol{\theta})}{\partial\theta_1} = -\frac{1}{2\theta_2}(-2)\sum_{i=1}^{n}(x_i-\theta_1) \quad = \frac{1}{\theta_2}\sum_{i=1}^{n}(x_i-\theta_1) \quad = 0$$

$$\sum_{i=1}^{n}\left(x_i-\hat{\theta}_1\right) \quad = 0$$

$$\sum_{i=1}^{n} x_i = n\hat{\theta}_1$$

or $\hat{\theta}_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$ is the mle of θ_1

Now

$$\frac{\partial lnL(\boldsymbol{\theta})}{\partial\theta_2} = -\frac{n}{2\theta_2} + \frac{1}{\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2 = 0$$

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{2\hat{\theta}_2^2} = \frac{n}{2\hat{\theta}_2}$$

$$n\hat{\theta}_2 = \sum_{i=1}^{n}(X_i-\bar{X})^2$$

Hence, $\hat{\theta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$ is the mle of θ_2.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a geometric distribution with pdf

$f(x;p) = (1-p)^{x-1}p \quad x = 1, 2, 3, \ldots, \; 0 \le p \le 1$. Find the mle of p.

$$L(p) = p^n (1-p)^{\sum_{i=1}^{n} x_i - n}$$

$$\ln L(p) = n \ln p + \left(\sum_{i=1}^{n} x_i - n \right) ln(1-p)$$

$$\frac{\partial lnL(p)}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p} = 0$$

$$\frac{n}{\hat{p}} = \frac{\sum_{i=1}^{n} x_i - n}{1-\hat{p}}$$

$$n - n\hat{p} = \hat{p} \sum_{i=1}^{n} x_i - n\hat{p}$$

Hence, $\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$ is the mle of p.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution with pdf

$f(x;\theta) = \frac{1}{\theta}$, $0 \le x < \theta$, $0 < \theta < \infty$

Here $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n}$

Lecture 3

Since L(θ) is a decreasing function of θ and as such the maximum occurs at the left endpoint of the interval (y_n, ∞). Thus L(θ) can be made no larger than $\frac{1}{(Y_n)^n}$, where $Y_n = \text{Max}(X_i)$.

Hence, the mle of θ is $\hat{\theta} = Y_n$

Invariance Property of Maximum Likelihood Estimators

Theorem

Let $\hat{\theta}$ be a maximum likelihood estimator of a parameter θ and let g(θ) be a function of θ. Then the maximum likelihood estimator of g(θ) is given by $g(\hat{\theta})$.

In our last example the maximum likelihood estimator of $\frac{\theta}{2}$, the mean of the uniform distribution is $\frac{Y_n}{2}$

Example

Let $X \sim N(\mu, \sigma^2)$. The mle of μ is $\hat{\mu} = \bar{X}$ and of σ^2 is , $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$

Then the mle of σ is $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$

And the mle of μ- σ is $\widehat{\mu - \sigma} = \bar{X} - \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution with pdf

$f(x; \alpha, \beta) = \frac{1}{\beta - \alpha}$, $\alpha < x < \beta$. Find the mle of $\sqrt{\alpha^2 + \beta^2}$.

The likelihood function is $L(\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\beta-\alpha} = \left(\frac{1}{\beta-\alpha}\right)^n$ for all $\alpha \le x_i$ (i=1, 2, …, n) and for all $\beta \ge x_i$ (i=1, 2, …, n). Hence the domain of L(α, β) is $\Omega = \{(\alpha,\beta); 0 < \alpha < Y_1 \, and \, Y_n < \beta < \infty\}$, where Y_1=Min(x_i) and Y_n=Max(x_i).

Therefore, the mle's of α and β are $\hat{\alpha} = Y_1$ and $\hat{\beta} = Y_n$. And using the invariance property of the mle, the mle of $\sqrt{\alpha^2 + \beta^2}$ is $\sqrt{Y_1^2 + Y_n^2}$.

Lecture 4

Properties of Estimators

The Unbiased Estimator

Definition: Any statistic whose mathematical expectation is equal to the parameter θ is called unbiased estimator. Otherwise it is said to be biased. That is

$$E(\hat{\theta}) = E[u(X_1, X_2, \dots, X_n)] = \theta$$

The bias of $\hat{\theta}$ is $b_{\hat{\theta}}(\theta) = E(\hat{\theta}) - \theta$

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from N(μ, σ^2).

1. Is the mle $\bar{X}$ unbiased estimator of μ?

   Since $X \sim N(\mu, \sigma^2)$, we have E(X)= μ and Var(X)= σ^2

   Then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Hence E($\bar{X}$) =μ

   $\therefore \bar{X}$ is an unbiased estimator of μ

2. Is $\hat{\sigma}^2$ unbiased estimator of σ^2?

   $$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

   $$E(\hat{\sigma}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = E\left[\frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

   $$= \frac{n-1}{n}E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2] = \frac{n-1}{n}E(S^2)$$

   $$= \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

   Therefore, the mle of σ^2 is a biased estimator.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from U(0, θ). Find an unbiased estimator of θ.

We have seen that for the uniform distribution the mle of θ is $Y_n = Max(X_i)$

$$g_n(y_n) = n[F(y_n)]^{n-1} f(y_n) = n[\frac{y}{\theta}]^{n-1} \frac{1}{\theta}$$

$$E(Y_n) = \int_0^\theta y_n g_n(y_n) dy_n = \int_0^\theta \frac{n}{\theta^n} y^n dy_n = \frac{n}{\theta^n} \frac{y^{n+1}}{n+1}]_0^\theta = \frac{n}{n+1} \frac{\theta^{n+1}}{\theta^n}$$

$$E(Y_n) = \frac{n}{n+1} \theta$$

Hence, the unbiased estimator of θ is $\frac{n+1}{n} Y_n$. Check!

Minimum Mean Square Error (MSE)

Definition

Let $Y = u(X_1, X_2, \ldots, X_n)$ be an estimator of an unknown parameter θ. The mathematical expectation

$$E(Y-\theta)^2 = E\left[u(X_1, X_2, \ldots, X_n) - \theta \ \ \right]^2 = E\left(\hat{\theta} - \theta\right)^2$$

is defined to be the mean square error of the estimator Y. That is

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

The mean square error is a measure of goodness or closeness of the estimator Y to the true value of θ.

Lemma

$$MSE\left(\hat{\theta}\right) = Var\left(\hat{\theta}\right) + b_{\hat{\theta}}^2(\theta)$$

When $\hat{\theta}$ is unbiased estimator to θ, that is $b_{\hat{\theta}}^2(\theta) = 0$, then

$$MSE\left(\hat{\theta}\right) = Var\left(\hat{\theta}\right)$$

Proof

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2$$

$$= E[(\hat{\theta} - E(\hat{\theta}))^2 + 2(\hat{\theta} - E(\hat{\theta})) \ (E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2]$$

$$= E(\hat{\theta} - E(\hat{\theta}))^2 + 2(E(\hat{\theta}) - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2$$

$$= Var(\hat{\theta}) + 0 + b_{\hat{\theta}}^2(\theta)$$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with unknown mean μ, -∞< μ<∞ and unknown variance $\sigma^2>0$. Show that the statistics $\bar{X}$ and $Y = \frac{X_1 + 2X_2 + \cdots + n X_n}{\frac{n(n+1)}{2}}$ are both unbiased estimators of μ.

Further, show that Var($\bar{X}$)< Var(Y).

Solution

We know that $\bar{X}$ is an unbiased estimator of μ irrespective of the distribution of μ.

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$$

Now

$$E(Y) = E\left(\frac{X_1 + 2X_2 + \cdots + \ nX_n}{\frac{n(n+1)}{2}}\right) = \frac{2}{n(n+1)} E\left(\sum_{i=1}^{n} iX_i\right)$$

$$= \frac{2}{n(n+1)}\sum_{i=1}^{n} iE(X_i) = \frac{2}{n(n+1)}\sum_{i=1}^{n} i\mu = \frac{2}{n(n+1)}\frac{n(n+1)}{2}\mu = \mu \text{ (unbiased)}$$

The variance of $\bar{X}$ is

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

And the variance of Y is

$$Var(Y) = Var\left[\frac{X_1 + 2X_2 + \cdots + nX_n}{\frac{n(n+1)}{2}}\right] = \frac{4}{n^2(n+1)^2}Var\left(\sum_{i=1}^{n} iX_i\right)$$

$$= \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} i^2 Var(X_i) = \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} i^2\sigma^2 = \frac{4\sigma^2}{n^2(n+1)^2}\sum_{i=1}^{n} i^2$$

$$= \frac{4\sigma^2}{n^2(n+1)^2}\frac{n(n+1)(2n+1)}{6} = \frac{2}{3}\frac{2n+1}{n+1}\frac{\sigma^2}{n}$$

$$= \frac{2(2n+1)}{3(n+1)}Var(\bar{X})$$

Since $\frac{2(2n+1)}{3(n+1)} > 1$ for n≥2, we see that Var($\bar{X}$)<Var(Y).

In statistics, between two unbiased estimators one prefers the estimator with the minimum variance.

Definition

The statistic Y that minimizes $MSE(Y) = E[(Y - \theta)^2]$ is the one with minimum mean square error. And if E(Y)=θ (unbiased) then $Var(Y) = E[(Y - \theta)^2]$ is said to be the minimum variance unbiased estimator of θ (MVUE).

Consistency

Definition: Convergence in Probability

Let $\{X_n\}$ be a sequence of random variables and let X be a random variable defined on a sample space. We say that X_n converges in probability to X $(X_n \xrightarrow{P} X)$ if for all ϵ >0,

$$\lim_{n\to\infty} P[|X_n - X| \geq \epsilon] = 0$$

Lecture 5

or equivalently,

$$\lim_{n\to\infty} P[|X_n - X| < \epsilon] = 1$$

In statistics, often the limiting of random variable X is a constant.

Theorem (Weak Law of Large Numbers)

Let $\{X_n\}$be a sequence of iid random variables having common mean μ and variance $\sigma^2<\infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\bar{X}_n \xrightarrow{P} \mu$

Proof

We already know that the mean and variance of $\bar{X}_n$ is μ and $\frac{\sigma^2}{n}$, respectively.

Hence, by Chebychev's Theorem, we have for every $\epsilon > 0$,

$$P[|\bar{X}_n - \mu| \geq \epsilon] = P\left[|\bar{X}_n - \mu| \geq \left(\epsilon\frac{\sqrt{n}}{\sigma}\right)\left(\frac{\sigma}{\sqrt{n}}\right)\right] \leq \frac{\sigma^2}{n\epsilon^2}$$

Taking the limit as $n\to\infty$ yields

$$\lim_{n\to\infty} P[|\bar{X}_n - \mu| \geq \epsilon] = 0$$

Hence, $\bar{X}_n \xrightarrow{P} \mu$

This theorem says that if n is large, then $\bar{X}_n$ becomes close to μ.

Definition: Consistency

An estimator $Y_n = u(X_1, X_2, \dots, X_n)$ is said to be consistent for θ if it converges in probability to θ. That is, if for all $\epsilon>0$,

$$\lim_{n\to\infty} P[|Y_n - \theta| \geq \epsilon] = 0$$

And we write $Y_n \xrightarrow{P} \theta$

If Y_n is consistent, then the probability that the estimator $Y_n$differs from the true θ becomes small as the sample size n increases.

Definition

An estimator $Y_n = u(X_1, X_2, \ldots, X_n)$ is said to be mean squared error consistent for θ if $\lim_{n\to\infty} E[(Y_n - \theta)^2] = 0$

A sequence of estimators $\{Y_n\}$ is MSE consistent estimator of θ if

1. $\lim_{n\to\infty} E(Y_n) = \theta$ (asymptotically unbiased)
2. $\lim_{n\to\infty} Var(Y_n) = 0$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from N(μ, σ²). Show that $\bar{X}$ and $\hat{\sigma}^2$ are consistent estimators for μ and σ².

1. $\lim_{n\to\infty} E(\bar{X}) = \lim_{n\to\infty} \mu = \mu$
2. $\lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = \sigma^2 \lim_{n\to\infty} \frac{1}{n} = 0$

$\therefore \bar{X}$ is a consistent estimator for μ.

Now

1. $\lim_{n\to\infty} E(\hat{\sigma}^2) = \lim_{n\to\infty} E(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \lim_{n\to\infty} E\left(\frac{(n-1)S^2}{n}\right) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2$

$= \sigma^2 \lim_{n\to\infty}\left(1 - \frac{1}{n}\right) = \sigma^2$

2. $\lim_{n\to\infty} Var(\hat{\sigma}^2) = \lim_{n\to\infty} Var(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2) = \lim_{n\to\infty} \frac{1}{n^2} Var[(n-1)S^2]$

$= \lim_{n\to\infty} \frac{\sigma^4}{n^2} Var[\frac{(n-1)S^2}{\sigma^2}] = \lim_{n\to\infty} \frac{\sigma^4}{n^2}[2(n-1)] = 2\sigma^4 \lim_{n\to\infty} \frac{n-1}{n^2}$

$= 2\sigma^2 \lim_{n\to\infty} [\frac{1}{n} - \frac{1}{n^2}] = 0$

Hence, $\hat{\sigma}^2$ is a consistent estimator for σ².

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from U(0, θ). Show that the mle $\hat{\theta} = Y_n$= Max(X_i) is consistent estimators for θ.

We have seen earlier that

$$g_n(y_n) = n[F(y_n)]^{n-1} f(y_n) = n[\tfrac{y}{\theta}]^{n-1}\tfrac{1}{\theta} = \tfrac{n}{\theta^n} y^{n-1} \qquad 0<y<\theta$$

And

$$E(Y_n) = \frac{n}{n+1}\theta$$

$$E(Y_n^2) = \int_0^\theta y_n^2 g_n(y_n)dy_n = \int_0^\theta y^2 \frac{n}{\theta^n} y^{n-1} dy_n = \frac{n}{\theta^n}\int_0^\theta y^{n+1}dy$$

$$= \frac{n}{\theta^n}\frac{y^{n+2}}{n+2}]_0^\theta = \frac{n}{n+2}\frac{\theta^{n+2}}{\theta^n} = \frac{n}{n+2}\theta^2$$

Hence,

$$Var(Y_n) = E(Y_n^2) - [E(Y_n)]^2 = \tfrac{n}{n+2}\theta^2 - \tfrac{n^2}{(n+1)^2}\theta^2$$

$$= n\theta^2\left(\frac{1}{n+2} - \frac{n}{(n+1)^2}\right) = n\theta^2\left(\frac{(n+1)^2 - n(n+2)}{(n+2)(n+1)^2}\right)$$

$$= n\theta^2\left(\frac{n^2+2n+1-n^2-2n}{(n+2)(n+1)^2}\right) = \frac{n}{(n+2)(n+1)^2}\theta^2$$

Now

1. $\lim_{n\to\infty} E(Y_n) = \lim_{n\to\infty}\left[\left(\frac{n}{n+1}\right)\theta\right] = \theta\, lim_{n\to\infty}\frac{1}{1+\frac{1}{n}} = \theta$

2. $\lim_{n\to\infty} Var(Y_n) = lim_{n\to\infty}\left[\frac{n}{(n+2)(n+1)^2}\theta^2\right] = \theta^2 \lim_{n\to\infty}\left[\frac{n}{(n+2)(n+1)^2}\right]$

$$= \theta^2 \lim_{n\to\infty}\frac{\frac{1}{n^2}}{\left(1+\frac{2}{n}\right)\left(1+\frac{1}{n}\right)^2} = 0$$

$\therefore Y_n$ is consistent estimator for θ.

H.W.

1. Let $X_1, X_2, \dots, X_n$ be a random sample from exponential distribution with pdf $f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, x>0.

a. Find the mle of the parameter θ.

b. Show that $\hat{\theta}$ is unbiased and has variance $\frac{\theta^2}{n}$.

2. Let $X_1, X_2, \dots, X_n$ be a random sample from Poisson distribution with pdf $f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$ x=0, 1, 2, … , λ>0. Find the mle of λ.

3. Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size n=3 from a uniform distribution with pdf $f(x;\theta) = \frac{1}{\theta}$ $, 0 < x < \theta$. Show that $4Y_1, 2Y_2, \frac{4}{3}Y_3$ are all unbiased estimators of θ. Find the variance of each of these unbiased estimators.

4. Let X_1, X_2, X_3 be a random sample of size n=3 from a distribution with unknown mean μ, -∞<μ<∞ and the variance is a known positive number. Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = \frac{2X_1+X_2+5X_3}{8}$ are unbiased estimators for μ. Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Lecture 6

Sufficiency

Definition: Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that has pdf $f(x;\theta), \theta \in \Omega$. Let $Y = u(X_1, X_2, \ldots, X_n)$ be a statistic whose pdf is $g(y;\theta)$, then Y is a sufficient statistic for θ if the conditional distribution of $X_1, X_2, \ldots, X_n$, given Y= y, does not depend upon θ. That is

$$\frac{f(x_1, x_2, \ldots, x_n; \theta)}{g(y;\theta)} = H(x_1, x_2, \ldots, x_n)$$

Where $H(x_1, x_2, \ldots, x_n)$ does not depend upon $\theta \in \Omega$.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from b(1, θ). Show that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

$f(x;\theta) = \theta^x (1-\theta)^{1-x}$ x=0, 1; $0 < \theta < 1$

The distribution of the sample is

$$f(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$= \theta^y (1-\theta)^{n-y}$$

The Statistic $Y = \sum_{i=1}^{n} X_i \sim$ b (n, θ) has the pdf

$g(y;\theta) = C_y^n \, \theta^y (1-\theta)^{n-y}$ y=0, 1, 2, …, n

The conditional probability

$$f(x_1, x_2, \dots, x_n | Y = y) = \frac{\theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2} \dots \theta^{x_n}(1-\theta)^{1-x_n}}{C_y^n \, \theta^y(1-\theta)^{n-y}}$$

$$= \frac{\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}}{C_{\sum_{i=1}^n x_i}^n \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}} = \frac{1}{C_{\sum_{i=1}^n x_i}^n}$$

Note that this conditional probability does not depend upon the parameter θ and the statistic $Y = \sum_{i=1}^n X_i$ is a sufficient statistic for θ.

Remark

If we are to show by means of the definition that a certain statistic Y is or is not a sufficient statistic for a parameterθ, we must first of all know the pdf of Y, $g(y;\theta)$. In many instances it may be quite difficult to find this pdf. Fortunately, this problem can be avoided by applying the following factorization theorem of Neyman.

The Factorization Theorem (Neyman)

Let X_1, X_2, …, X_n be a random sample from a distribution that has pdf $f(x;\theta), \theta \in \Omega$. The statistic $Y=u(X_1, X_2, \dots, X_n)$ is a sufficient statistic for θ iff we can find two non negative functions, K_1 and K_2, such that

$$f(x_1, x_2, \dots, x_n; \theta) = K_1[u(x_1, x_2, \dots, x_n); \theta] \, K_2(x_1, x_2, \dots, x_n)$$

where $K_2(x_1, x_2, \dots, x_n)$ does not depend upon θ.

Example

Let X_1, X_2, …, X_n be a random sample from Poisson(λ), where $\lambda > 0$. Show that $\bar{X}$ is a sufficient statistic for λ.

By using the factorization theorem

$$f(x_1, x_2, \dots, x_n; \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\mathrm{x_1!\, x_2! \dots x_n!}} = \left(\lambda^{n\bar{x}} e^{-n\lambda}\right)\left(\frac{1}{\mathrm{x_1!\, x_2! \dots x_n!}}\right)$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Hence $\bar{X}$ is a sufficient statistic for λ.

Remark:

In the previous example, if we replace $n\bar{x}$ by $\sum_{i=1}^{n} x_i$, it is obvious that $\sum_{i=1}^{n} x_i$ is also a sufficient statistic for λ. In general if Y is sufficient for a parameter θ, then every single–valued function of Y not involving θ, but with a single valued inverse, is also a sufficient statistic for θ.

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution with pdf

$f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$. Find a sufficient statistic for θ.

The joint pdf of $X_1, X_2, \dots, X_n$ is

$$f(x_1, x_2, \dots, x_n; \theta) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta} \left(\frac{1}{\prod_{i=1}^{n} x_i}\right)$$

In the factorization theorem, let

$$K_1[u(\mathrm{x_1, x_2, \dots, x_n}); \theta] = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}$$

and

$$K_2(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n}) = \frac{1}{\prod_{i=1}^{n} x_i}$$

Since $K_2(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$ does not depend upon θ, the product $\prod_{i=1}^{n} x_i$ is a sufficient statistic for θ.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution that is $N(\theta, \sigma^2)$,

$-\infty < \theta < \infty$ and the variance $\sigma^2 > 0$ is known. Show that $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is a sufficient statistic for θ.

The joint pdf of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n; \theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

Since

$$\sum_{i=1}^{n}(x_i - \theta)^2 = \sum_{i=1}^{n}[(x_i - \bar{x}) + (\bar{x} - \theta)]^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \theta)^2$$

Then

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\theta)^2]} = \left[e^{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2}\right]\left[\frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}}{\left(\sqrt{2\pi\sigma^2}\right)^n}\right]$$

Lecture 7

Because the first factor depends on $x_1, x_2, \dots, x_n$ only through $\bar{x}$, and the second factor does not depend upon θ, hence, according to the factorization theorem, the sample mean $\bar{X}$ is a sufficient statistic for θ, the mean of the normal distribution.

We now state a theorem that tells us to restrict our search for an unbiased minimum variance estimator to functions of the sufficient statistics if it exists.

Theorem: (Rao - Blackwell)

Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution with pdf $f(x;\theta), \theta \in \Omega$. Let $Y_1 = u_1(X_1, X_2, \dots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, X_2, \dots, X_n)$ be an unbiased estimator of θ, where Y_2 is not a function of Y_1 alone. Then $E(Y_2 | y_1) = u(y_1)$ defines a statistic $u(Y_1)$, a function of the sufficient statistic Y_1, which is an unbiased estimator of θ, and its variance is less than that of Y_2. That is

1. $E[u(Y_1)] = \theta$

2. $Var[u(Y_1)] \leq Var(Y_2)$

Proof

Let $g(y_1, y_2; \theta)$ be the joint pdf of Y_1 and Y_2. Let $g_1(y_1; \theta)$ be the marginal of Y_1. The conditional pdf of Y_2 given $Y_1 = y_1$ is

$$h(y_2|y_1) = \frac{g(y_1, y_2; \theta)}{g_1(y_1; \theta)}$$

This equation does not depend upon θ, since Y_1 is a sufficient statistic for θ.

In the continuous case

$$u(y_1) = E(Y_2|y_1) = \int_{S_2} y_2\, h(y_2|y_1)\, dy_2 = \int_{S_2} y_2 \frac{g(y_1, y_2; \theta)}{g_1(y_1; \theta)}\, dy_2$$

$$E[u(Y_1)] = \int_{S_1} u(y_1) g_1(y_1; \theta)\ dy_1$$

$$= \int_{S_1} \left(\int_{S_2} y_2 \frac{g(y_1, y_2; \theta)}{g_1(y_1; \theta)}\, dy_2 \right) g_1(y_1; \theta)\, dy_1$$

$$= \int_{S_1} \int_{S_2} y_2 g(y_1, y_2; \theta)\, dy_2 dy_1$$

$$= \int_{S_2} y_2 \int_{S_1} g(y_1, y_2; \theta)\ dy_1\, dy_2 = \int_{S_2} y_2\, g_2(y_2; \theta) dy_2 = \theta$$

because Y_2 is an unbiased estimator of θ. Thus $u(Y_1)$ is also an unbiased estimator of θ.

Now, consider

$$Var(Y_2) = E(Y_2 - \theta)^2 = E[(Y_2 - u(Y_1) + u(Y_1) - \theta)^2]$$

$$= E[(Y_2 - u(Y_1)]^2 + E[u(Y_1) - \theta)]^2 + 2E[(Y_2 - u(Y_1))(u(Y_1) - \theta)]$$

But the third expression is equal to

$$2\int_{S_1}\int_{S_2}(Y_2-u(y_1))(u(y_1)-\theta)g(y_{1,}\,y_2;\theta)dy_1\,dy_2$$

$$=2\int_{S_1}\int_{S_2}(Y_2-u(y_1))(u(y_1)-\theta)h(y_2|y_1)g_1(y_1\,;\theta)\;dy_1dy_2$$

$$=2\int_{S_1}(u(y_1)-\theta)\left\{\int_{S_2}(y_2-u(y_1))h(y_2|y_1)\,dy_2\right\}g_1(y_1;\theta)dy_1=0$$

because $u(y_1)$ is the conditional mean of Y_2 in the conditional distribution given by $h(y_2|y_1)$. That is

$$\int_{S_2}(y_2-u(y_1))h(y_2|y_1)\,dy_2\;=E(Y_2|y_1)-u(y_1)=0$$

Thus

$$Var(Y_2)=E[(Y_2-u(Y_1)]^2+E[u(Y_1)-\theta)]^2$$

$$=E[(Y_2-u(Y_1)]^2+Var[u(Y_1)]$$

The first term is the expected value of a positive expression. Therefore

$Var[u(Y_1)]\leq Var(Y_2)$ □

Furthermore, there is a connection between sufficient statistics and maximum likelihood estimates, as shown in the following theorem:

Theorem

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has pdf $f(x;\theta), \theta \in \Omega$, If a sufficient statistic $Y = u(X_1, X_2, \ldots, X_n)$ for θ exists, and if a maximum likelihood estimator $\hat{\theta}$ of θ also exists uniquely, then $\hat{\theta}$ is a function of the sufficient statistic $Y = u(X_1, X_2, \ldots, X_n)$.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the exponential distribution with pdf $f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$ $0 < x < \infty, \theta > 0$.

1. Find a sufficient statistic Y for θ and show that the mle for θ is a function of Y.

2. Determine an unbiased estimator of θ that is a function of the sufficient statistic alone.

Solution

1. The joint pdf (likelihood function) is

$$L(\theta; x_1, x_2, \ldots, x_n) = \theta^{-n} e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$$

Hence, by the factorization theorem $Y = \sum_{i=1}^{n} X_i$ is sufficient for θ, since

$K_1[u(x_1, x_2, \ldots, x_n); \theta] = \theta^{-n} e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$ and $K_2(x_1, x_2, \ldots, x_n) = 1$

The log likelihood function is

$$\log L(\theta; x_1, x_2, \dots, x_n) = -n\log\theta - \frac{\sum_{i=1}^{n} x_i}{\theta}$$

Taking the partial with respect to θ and setting it to zero results in the mle of θ

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Hence, the mle $\bar{X}$ is a function of the sufficient statistic $Y = \sum_{i=1}^{n} X_i$

2. $E(Y) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = n\theta$

Thus

$$E\left(\frac{Y}{n}\right) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{n\theta}{n} = \theta$$

That is $\bar{X}$ is that function of the sufficient statistic for the parameter θ that is unbiased. Thus the statistic $\bar{X}$ is an MVUE of θ.

H.W.

1. Let $X_1, X_2, \dots, X_n$ be a random sample from $N(0, \sigma^2)$.

a. Show that $Y = \sum_{i=1}^{n} X_i^2$ is a sufficient statistic for σ^2.

b. Show that the mle for σ^2 is a function of Y.

c. Is the mle for σ^2 unbiased?

2. Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a distribution whose pdf $f(x;\theta) = \theta x^{\theta-1}$ $0 < x < 1$.

1. Show that $Y = \prod_{i=1}^{n} X_i$ is a sufficient statistic for θ.

2. Show that the mle is a function of Y.

Lecture 8

Fisher Information and the Rao – Crame´r Inequality

Definition

Let X be a random variable from a distribution with pdf $f(x;\theta)$ of a continuous type such that the parameter θ does not appear in the endpoints of the interval in which $f(x;\theta) > 0$ and that we can interchange integration and differentiation with respect to θ. The Fisher information, I(θ), in a single observation X about θ is given by

$$I(\theta) = \int_{-\infty}^{\infty} \left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2 f(x;\theta)\, dx = E\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right]$$

That is, I(θ) is the expected value of the square of the random variable $\frac{\partial \ln f(X;\theta)}{\partial\theta}$.

Lemma

The Fisher information I(θ) contained in a single observation about the unknown parameter can be given alternatively as

$$I(\theta) = \int_{-\infty}^{\infty} \frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} f(x;\theta)\, dx = -E\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right]$$

Proof

Since $f(x;\theta)$ is a pdf, we have that

$$\int_{-\infty}^{\infty} f(x;\theta)dx = 1$$

By taking the derivative with respect to θ

$$\int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{d\theta} dx = 0$$

which can be rewritten as

$$\int_{-\infty}^{\infty} \frac{\frac{\partial f(x;\theta)}{d\theta}}{f(x;\theta)} f(x;\theta)dx = \int_{-\infty}^{\infty} \frac{\partial lnf(x;\theta)}{d\theta} f(x;\theta)dx = 0$$

If we differentiate again it follows that

$$\int_{-\infty}^{\infty} \left[\frac{\partial^2 lnf(x;\theta)}{d\theta^2} f(x;\theta) + \frac{\partial f(x;\theta)}{d\theta} \frac{\partial lnf(x;\theta)}{d\theta}\right] dx = 0$$

Rewriting the last equality, we have

$$\int_{-\infty}^{\infty} \left[\frac{\partial^2 lnf(x;\theta)}{d\theta^2} f(x;\theta) + \frac{\partial \ln f(x;\theta)}{\partial\theta} \frac{\frac{\partial f(x;\theta)}{\partial\theta}}{f(x;\theta)} f(x;\theta)\right] dx = 0$$

which is

$$\int_{-\infty}^{\infty} \left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} f(x;\theta) + \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)\right] dx = 0$$

and we see that

$$\int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)dx = -\int_{-\infty}^{\infty} \frac{\partial^2 lnf(x;\theta)}{d\theta^2} f(x;\theta)dx$$

Hence

$$I(\theta) = E\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right] \blacksquare$$

Example 1

Let X be a single observation taken from $N(\theta, \sigma^2)$, where $-\infty < \theta < \infty$ and σ^2 is known. Find the Fisher information in X about θ.

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \qquad -\infty < x < \infty$$

$$\ln f(x;\theta) = -\frac{1}{2} ln(2\pi\sigma^2) - \frac{(x-\theta)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = -\frac{2(x-\theta)(-1)}{2\sigma^2} = \frac{x-\theta}{\sigma^2}$$

$$\therefore I(\theta) = E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{x-\theta}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4} E[(x-\theta)^2] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Or

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = -\frac{1}{\sigma^2}$$

$$\therefore I(\theta) = -E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right] = -\left(-\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2}$$

Sometimes, one expression is easier to compute than the other, but often we prefer the second expression.

Example 2

Let X~ b(1, θ). Find the Fisher information in X about θ.

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} \qquad x = 0, 1$$

$$ln f(x;\theta) = x ln\theta + (1-x)\ln(1-\theta)$$

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

$$\therefore I(\theta) = -E\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right] = -\left[-\frac{E(X)}{\theta^2} - \frac{1-E(X)}{(1-\theta)^2}\right]$$

Lecture 9

$$= \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1-\theta+\theta}{\theta(1-\theta)} = \frac{1}{\theta(1-\theta)}$$

which means that the information is larger for θ values close to zero or one.

Fisher Information of a Random Sample

We have already seen that

$$\int_{-\infty}^{\infty} \frac{\partial \ln f(x;\theta)}{\partial \theta} f(x;\theta)dx = 0$$

This equation represents the following expectation

$$E\left[\frac{\partial \ln f(X;\theta)}{\partial \theta}\right] = 0$$

That is, the mean of the random variable $\frac{\partial \ln f(X;\theta)}{\partial \theta}$ is 0, and the variance of this random variable is the Fisher information I(θ);i.e.,

$$Var\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right) = E\left[\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right] = I(\theta)$$

Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution having pdf $f(x;\theta)$. The likelihood function L(θ) is

$$L(\theta) = f(x_1;\theta).f(x_2,\theta)\dots f(x_n;\theta)$$

$$lnL(\theta) = \ln f(x_1;\theta) + \ln f(x_2,\theta) + \dots + \ln f(x_n;\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i;\theta)}{\partial \theta}$$

The summands are iid with common variance I(θ). Hence the information in the sample is

$$Var\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right) = Var\left(\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial \theta}\right) = \sum_{i=1}^{n} Var\left(\frac{\partial \ln f(X_i;\theta)}{\partial \theta}\right)$$

$$\therefore Var\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right) = nI(\theta)$$

Thus, the information in a random sample of size n is n times the information in a single observation.

So, in examples 1 and 2, the Fisher information in a random sample of size n is $\frac{n}{\sigma^2}$ and $\frac{n}{\theta(1-\theta)}$ respectively.

Theorem (Rao – Cramer Lower Bound)

Let $X_1, X_2, \dots, X_n$ be iid with common pdf $f(x;\theta)$ for $\theta \in \Omega$. Assume that the parameter θ does not appear in the endpoints of the interval in which $f(x;\theta) > 0$ so that we can interchange integration and differentiation with respect to θ. Let $Y = u(X_1, X_2, \dots, X_n)$ be a statistic (an estimator of θ) with mean $E(Y) = E[u(X_1, X_2, \dots, X_n)] = k(\theta)$. Then

$$Var(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

Proof:

For the continuous case, the mean of Y can be written as

$$k(\theta) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} u(x_1, x_2, \dots, x_n)\, f(x_1;\theta) f(x_2;\theta) \dots f(x_n;\theta) dx_1 \dots dx_n$$

Differentiating with respect to θ, we obtain

$$k'(\theta) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} u(x_1, x_2, \dots, x_n) \left[\sum_{i=1}^{n} \frac{1}{f(x_i;\theta)} \frac{\partial f(x_i;\theta)}{\partial \theta}\right]$$

$$\times f(x_1;\theta) f(x_2;\theta) \dots f(x_n;\theta) dx_1 \dots dx_n.$$

$$k'(\theta) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} u(x_1, x_2, \dots, x_n) \left[\sum_{i=1}^{n} \frac{\partial \ln f(x_i; \theta)}{\partial \theta} \right]$$

$$\times f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n. \qquad (*)$$

Define the random variable $Z = \sum_{i=1}^{n} \frac{\partial \ln f(X_i; \theta)}{\partial \theta}$. We know that

$$\int_{-\infty}^{\infty} \frac{\partial \ln f(x; \theta)}{\partial \theta} f(x; \theta) dx = E\left[\frac{\partial \ln f(X; \theta)}{\partial \theta} \right] = 0$$

And

$$Var\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right) = E\left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right] = I(\theta)$$

Hence, $E(Z) = \sum_{i=1}^{n} E\left[\frac{\partial \ln f(X; \theta)}{\partial \theta} \right] = 0$

And

$$Var(Z) = \sum_{i=1}^{n} Var\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right) = nI(\theta)$$

So, we can express equation (*) as $k'(\theta) = E(YZ)$

Recall that the correlation coefficient between Y and Z is

$$\rho = \frac{E(YZ) - E(Y)E(Z)}{\sigma_Y \sigma_Z}$$

Hence

$$k'(\theta) = E(YZ) = E(Y)E(Z) + \rho \sigma_Y \sigma_Z$$

Using $E(Z) = 0$ and $Var(Z) = nI(\theta)$

$$\rho = \frac{k'(\theta)}{\sigma_Y \sqrt{nI(\theta)}}$$

Because $\rho^2 \leq 1$. We have

$$\frac{[k'(\theta)]^2}{\sigma_Y^2 nI(\theta)} \leq 1$$

Hence

$$Var(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)} \blacksquare$$

Corollary

Under the assumptions of the theorem, if $Y = u(X_1, X_2, \dots, X_n)$ is an unbiased estimator of θ, so that $E(Y) = k(\theta) = \theta$,

then the Rao – Cramer inequality becomes

$$Var(Y) \geq \frac{1}{nI(\theta)}$$

Back to example 1, we have $I(\theta) = \frac{1}{\sigma^2}$ and $nI(\theta) = \frac{n}{\sigma^2}$. Let $Y = \bar{X}$, then $E(\bar{X}) = \theta$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$. So $Var(\bar{X}) = \frac{1}{nI(\theta)} = \frac{\sigma^2}{n}$.

Back to example 2, we have $I(\theta) = \frac{1}{\theta(1-\theta)}$ and $nI(\theta) = \frac{n}{\theta(1-\theta)}$. Let $Y = \bar{X}$, then $E(\bar{X}) = \theta$ and $Var(\bar{X}) = \frac{\theta(1-\theta)}{n}$.

So $Var(\bar{X}) = \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$.

In both cases, the unbiased estimator $\bar{X}$ of θ, which is based upon the sufficient statistic for θ, have a variance that is equal to the Rao– Cramer lower bound. That is $Var(\bar{X}) = \frac{1}{nI(\theta)}$.

Lecture 10

Efficiency

Definition (Efficient Estimator)

Let Y be an unbiased estimator of a parameter θ. Then the statistic Y is called an efficient estimator of θ if and only if the variance of Y attains the Rao - Cramer lower bound. That is

$$Var(Y) = \frac{1}{nI(\theta)}$$

Definition (Efficiency)

The ratio of the Rao - Cramer lower bound to the actual variance of any unbiased estimator of a parameter is called the efficiency of that estimator.

$$Eff(Y) = \frac{RCLB}{Var(Y)} \leq 1$$

If equality holds, then the statistic Y is an efficient estimator by definition.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution. Show that $\bar{X}$ is an efficient estimator of θ.

$$f(x;\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \quad ,x = 0, 1.2, \ldots$$

$$lnf(x;\theta) = -\theta + \mathrm{x}\ln\theta - \ln \mathrm{x}!$$

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = -1 + \frac{\mathrm{x}}{\theta} = \frac{\mathrm{x}-\theta}{\theta}$$

$$I(\theta) = E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right] = \frac{E(X-\theta)^2}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

Or

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = -\frac{x}{\theta^2}$$

$$I(\theta) = -E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right] = -\left[\frac{-E(X)}{\theta^2}\right] = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

$\therefore I(\theta) = \frac{1}{\theta}$ and $nI(\theta) = \frac{n}{\theta}$

Thus the RCLB is $\frac{1}{nI(\theta)} = \frac{\theta}{n}$

We know that $\hat{\theta} = \bar{X}$ is the mle of θ and $E(\hat{\theta}) = E(\bar{X}) = \theta$ (unbiased).

The variance of $\bar{X}$ is $\frac{\theta}{n} = \frac{1}{nI(\theta)}$ the RCLB.

Hence, $\bar{X}$ is an efficient estimator of θ.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the exponential distribution with pdf $f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$ $0 < x < \infty$, $\theta > 0$. Show that $\bar{X}$ is an efficient estimator of θ.

$$\ln f(x;\theta) = -\ln\theta - \frac{x}{\theta}$$

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

$$I(\theta) = E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{-\theta+x}{\theta^2}\right)^2\right] = \frac{1}{\theta^4}E(X-\theta)^2 = \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2}$$

Or

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$$

$$I(\theta) = -E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right] = -E\left[\frac{\theta-2x}{\theta^3}\right] = -\left[\frac{\theta-2E(x)}{\theta^3}\right] = \frac{-\theta+2\theta}{\theta^3} = \frac{1}{\theta^2}$$

Thus the RCLB is $\frac{1}{nI(\theta)} = \frac{\theta^2}{n}$

We know that $\hat{\theta} = \bar{X}$ is the mle of θ and $E(\hat{\theta}) = E(\bar{X}) = \theta$ (unbiased).

The variance of $\bar{X}$ is $\frac{\theta^2}{n} = \frac{1}{nI(\theta)}$ the RCLB. Hence, $\bar{X}$ is an efficient estimator of θ.

Definition (Asymptotic Efficiency)

If $\lim_{n\to\infty} Eff(Y) = 1$, we say that Y is asymptotically efficient estimator of θ.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from N(μ, θ). Assuming μ is known; show that $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an asymptotically efficient estimator of θ.

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-\mu)^2} \qquad -\infty < x < \infty$$

$$\ln f(x;\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\theta - \frac{(x-\mu)^2}{2\theta}$$

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = -\frac{1}{2\theta} + \frac{(x-\mu)^2}{2\theta^2}$$

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = \frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}$$

$$I(\theta) = -E\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right] = -\left[\frac{1}{2\theta^2} - \frac{E(X-\mu)^2}{\theta^3}\right] = -\frac{1}{2\theta^2} + \frac{\theta}{\theta^3} = \frac{1}{2\theta^2}$$

Thus the RCLB is $\frac{1}{nI(\theta)} = \frac{2\theta^2}{n}$

We know that $E(S^2) = \sigma^2 = \theta$ (unbiased), and

$$Var(S^2) = Var\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{\sigma^4}{(n-1)^2} Var\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\right]$$

Since $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$$Var(S^2) = \frac{\sigma^4}{(n-1)^2} Var\left(\chi^2_{(n-1)}\right) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1} = \frac{2\theta^2}{n-1}$$

Hence,

$$Var(S^2) = \frac{2\theta^2}{n-1} > RCLB = \frac{2\theta^2}{n}$$

and $Eff(S^2) = \frac{RCLB}{Var(S^2)} = \frac{2\theta^2/n}{2\theta^2/n-1} = \frac{n-1}{n}$

$$\lim_{n\to\infty} Eff(S^2) = \lim_{n\to\infty} \frac{n-1}{n} = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right) = 1$$

$\therefore S^2$ is an asymptotically efficient estimator of θ.

H. W.

1. Show that $\bar{X}$, the mean of a random sample of size n from N(θ, σ^2) is an efficient estimator of θ.

2. Show that $\bar{X}$, the mean of a random sample of size n from b(1, θ), 0 < θ < 1 is an efficient estimator of θ.

Lecture 11

Interval Estimation (Confidence Intervals)

In point estimation we find a value for the parameter θ given a sample data. It is sometimes more valuable to be able to specify an interval for which we have a certain degree of confidence that θ lies within. To obtain such an interval estimator, we make use of the probability distribution of the point estimator.

Definition :(Confidence Interval)

Let $X_1, X_2, \dots, X_n$ be a random sample from a population with pdf $f(x;\theta)$, $\theta \in \Omega$. Let $0 < \alpha < 1$ be specified. Let L $(X_1, X_2, \dots, X_n)$ and U $(X_1, X_2, \dots, X_n)$ be two statistics with L$\leq$ U. We say that the interval (L, U) is a (1-α) 100% confidence interval for θ if

$(1-\alpha) = P_\theta[\theta \in (L, U)]$ .That is, the probability that the interval includes θ is

$P(L \leq \theta \leq U) = 1 - \alpha$

The random variable L is called the lower confidence limit and U is called the upper confidence limit. The number (1-α) is called the confidence coefficient or the confidence level of the interval.

Let $x_1, x_2, \dots, x_n$ be a set of sample data, then the value of the confidence interval is (l, u), an interval of real numbers.

A very wide confidence interval gives a message that there is a great deal of uncertainty concerning the value of θ. A shorter confidence interval gives more precise estimate of θ.

A procedure for obtaining confidence intervals is based on a pivot random variable. The pivot is a function of an estimator of θ and the parameter and, further, the distribution of the pivot is known.

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Then a pivot random variable for μ is $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, since it is a function of the maximum likelihood estimator $\bar{X}$ and the parameter μ and has a distribution that is N(0, 1).

In general, if W is a pivot random variable, then a (1-α) 100% confidence interval for θ may be constructed as follows

1. Find two values a and b such that

$P(a \leq W \leq b) = 1 - \alpha$

2. Convert the inequality $a \leq W \leq b$ into the form $L \leq \theta \leq U$.

Confidence Intervals for Means

- Confidence Interval for the Mean μ of a Normal Distribution when σ^2 is known.

Let $X_1, X_2, \dots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$, where the variance σ^2 is known. Construct a confidence interval for the unknown parameter μ.

The pivot random variable $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

For the probability 1-α, we can find a number $z_{\alpha/2}$ from the standard normal table such that

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left[-z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \bar{X} - \mu \leq z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$

$$P\left[-\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq -\mu \leq -\bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$

$$P\left[\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$

So, the probability that the random interval

$\left[\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right]$ includes the unknown mean μ is 1-α.

For example, if 1-α =0.95, then α=0.05 and $z_{\alpha/2} = z_{0.025} = 1.96$. Then a 95% confidence interval for μ is $\left[\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$.

Also if 1-α =0.90, then α=0.10 and$z_{\alpha/2} = z_{0.05} = 1.645$. Then a 90% confidence interval for μ is $\left[\bar{x} - 1.645\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + 1.645\left(\frac{\sigma}{\sqrt{n}}\right)\right]$.

It is useful to mention that higher the confidence level, bigger is the length of the confidence interval.

We now consider the case when the distribution of the pivot random variable $W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is not normal. We can still obtain an approximate confidence interval for μ. The Central Limit Theorem shows that the distribution of W is approximately N(0, 1).

Theorem: (Central Limit Theorem)

Let $X_1, X_2, \dots, X_n$ denote the observations of a random sample from a distribution that has mean μ and finite variance σ^2. Then the distribution function of the random variable $W_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ converges to Φ, the distribution function of the N(0, 1) distribution, as $n \to \infty$.

In this case

$$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) \approx 1-\alpha$$

And that $\left[\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right]$ is an approximate (1-α)100% confidence interval for μ.

- Confidence Interval for the Mean μ of a Normal Distribution when σ^2 is unknown.

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$, where μ and σ^2 are unknown. Find a confidence interval for μ.

Let $\bar{X}$ and S^2 denote the sample mean and the sample variance, respectively. The pivot random variable $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a t distribution with n-1 degrees of freedom. For 0 < α < 1, select $t_{\alpha/2(n-1)}$ such that

$$P\left(-t_{\alpha/_2(n-1)} \le \frac{\bar{X}-\mu}{S/_{\sqrt{n}}} \le t_{\alpha/_2(n-1)}\right) = 1-\alpha$$

$$P\left(-t_{\alpha/_2(n-1)}\left(\frac{S}{\sqrt{n}}\right) \le \bar{X}-\mu \le t_{\alpha/_2(n-1)}\left(\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha$$

$$P\left(\bar{X}-t_{\alpha/_2(n-1)}\left(\frac{S}{\sqrt{n}}\right) \le \mu \le \bar{X}+t_{\alpha/_2(n-1)}\left(\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha$$

Thus, if the observations of a random sample $x_1, x_2, \dots, x_n$ provide $\bar{x}$ and s^2, then a (1-α)100% confidence interval for μ is given by $\left[\bar{x}-t_{\alpha/_2(n-1)}\left(\frac{s}{\sqrt{n}}\right), \bar{x}+t_{\alpha/_2(n-1)}\left(\frac{s}{\sqrt{n}}\right)\right]$, where $\frac{s}{\sqrt{n}}$ is the estimate of the standard deviation of $\bar{X}$, which is referred to as the standard error of $\bar{X}$.

In the case where the population is not normal, but μ and σ are both unknown, approximate confidence interval for μ can still be constructed using $T = \frac{\bar{X}-\mu}{S/_{\sqrt{n}}}$ which now has an approximate t distribution. This approximation is quite good if the distribution is symmetric, unimodal and of continuous type.

H.W.

1. Let X equals the length of life of a certain kind of light bulbs. Assume that X~$N(\mu, 1296)$. A random sample of n=27 light bulbs were tested until they burned out, yielding a sample mean $\bar{x}$ =1478 hours. Construct a 95% confidence interval for μ.

2. Let $X_1, X_2, \dots, X_n$ be a random sample of size n=11 from $N(\mu, 9.9)$. If $\sum_{i=1}^{n} x_i = 132$. Find respectively the 95% and 90% confidence intervals for μ.

3. Let $X_1, X_2, \dots, X_n$ be a random sample of size n=40 from a distribution with known variance and unknown mean μ. If $\sum_{i=1}^{n} x_i = 286.56$ and $\sigma^2 = 10$. What is the 90% confidence interval for μ.

- Confidence Intervals for the Difference of Means of Two Normal Distributions when the Variances are known

Suppose that we are interested in comparing the means of two normal distributions. Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_m$ be, respectively two random samples of sizes n and m from two independent normal distributions $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$. Suppose that σ_x^2 *and* σ_y^2 are known. The sample means $\bar{X}$ and $\bar{Y}$ are also independent and have distributions $N(\mu_x, \frac{\sigma_x^2}{n})$ and $N(\mu_y, \frac{\sigma_y^2}{m})$, respectively.

Then the distribution of $\bar{X} - \bar{Y}$ is

$N(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m})$, and that $Z = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}} \sim N(0,1)$.

Thus

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(-z_{\alpha/2} \le \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}} \le z_{\alpha/2}\right) = 1 - \alpha$$

which can be written as

$$P\left(-z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}} \le (\bar{X}-\bar{Y})-(\mu_x-\mu_y) \le z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}\right) = 1 - \alpha$$

$$P\left((\bar{X}-\bar{Y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}} \le \mu_x - \mu_y \le (\bar{X}-\bar{Y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}\right) = 1 - \alpha$$

Thus, if the observed sample means $\bar{x}\, and\ \bar{y}$ have been computed, then

$$\left[(\bar{x}-\bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}, (\bar{x}-\bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}\right]$$

provides a (1-α)100% confidence interval for $\mu_x - \mu_y$.

Confidence Intervals for the Difference of Means of Two Normal Distributions when the Variances are Unknown and the sample sizes are small

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be, respectively two random samples of sizes n and m from two independent normal distributions $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$. The sample means $\bar{X}$ and $\bar{Y}$ are also independent and have distributions $N(\mu_x, \frac{\sigma_x^2}{n})$ and $N(\mu_y, \frac{\sigma_y^2}{m})$, respectively. Then the distribution of $\bar{X} - \bar{Y}$ is $N(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m})$ and that $Z = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}} \sim N(0,1)$

We know that $\frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi_{n-1}^2$ and $\frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi_{m-1}^2$ and because X and Y are independent, these chi-square random variables are independent. Hence the distribution of $\frac{(n-1)S_x^2}{\sigma_x^2} + \frac{(m-1)S_y^2}{\sigma_y^2}$ $is\ \chi_{n+m-2}^2$

A random variable $T = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}}}{\sqrt{\frac{\frac{(n-1)S_x^2}{\sigma_x^2}+\frac{(m-1)S_y^2}{\sigma_y^2}}{n+m-2}}} \sim t_{n+m-2}$

Assuming that $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (common variance), then

$$T = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y) \Big/ \sqrt{\frac{1}{n}+\frac{1}{m}}}{\sqrt{(n-1)S_x^2+(m-1)S_y^2 \Big/ n+m-2}} \sim t_{n+m-2}$$

which is independent of σ^2. Thus

$$P\left(-t_{\alpha/2}(n+m-2) \le T \le t_{\alpha/2}(n+m-2)\right) = 1-\alpha$$

Let $S_p = \sqrt{\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}}$ be the pooled estimator of the common standard deviation. Then

$$P\left[-t_{\alpha/2}(n+m-2) \le \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \le t_{\alpha/2}(n+m-2)\right] = 1-\alpha$$

Solving for $\mu_x-\mu_y$ yields

$$P\left((\bar{X}-\bar{Y}) - t_{\alpha/2}(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \le \mu_x-\mu_y \le (\bar{X}-\bar{Y}) + t_{\alpha/2}(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right) = 1-\alpha$$

If $\bar{x}$, $\bar{y}$, and s_p are observed values, then

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

is a (1-α)100% confidence interval for $\mu_x-\mu_y$.

If the sample sizes are large, and σ_x^2 *and* σ_y^2 are unknown, we can replace them with S_x^2 *and* S_y^2 the values of the unbiased estimates of the variances. This means that

$$\left[(\bar{x}-\bar{y}) - z_{\alpha/2}\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}, (\bar{x}-\bar{y}) + z_{\alpha/2}\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}\right]$$

serves as an approximate (1-α)100% confidence interval for $\mu_x-\mu_y$.

Example

To reach maximum efficiency in performing an assembly operation in a manufacturing plant, new employees require approximately a one month training period. Two training methods were applied on two groups of nine new employees each for three weeks period. A test was conducted to compare the two methods. The length of time (in minutes) required for each employee to assemble the device was recorded:

Method 1: 32 37 35 28 41 44 35 31 34

Method 2: 35 31 29 25 34 40 27 32 31

Assume that the assembly times are approximately normally distributed with approximately equal variances. And that the samples are independent. Construct a 95% confidence interval for the difference of means $\mu_1 - \mu_2$.

We have $n = m = 9, \bar{x}_1 = 35.22, \bar{x}_2 = 31.56$

$$s_1^2 = \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}{n-1} = \frac{195.56}{8} = 24.445$$

$$s_2^2 = \frac{\sum_{i=1}^{m}(x_{2i} - \bar{x}_2)^2}{n-1} = \frac{160.22}{8} = 20.027$$

Hence

$$s_p^2 = \frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-2} = \frac{8(24.445)+8(20.027)}{9+9-2} = \frac{195.56+160.22}{16} = 22.236$$

and $s_p = \sqrt{22.236} = 4.716$

Notice that, because $n = m = 9$, s_p^2 is a simple average of s_1^2 and s_2^2.

Since $t_{\alpha/2(n+m-2)} = t_{0.025(16)} = 2.12$, the observed confidence interval is

$$l = (\bar{x}_1 - \bar{x}_2) - t_{0.025(16)}\, s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$= (35.22 - 31.56) - (2.12)(4.716)\sqrt{\frac{1}{9} + \frac{1}{9}} = 3.66 - 4.71 = -1.05$$

and

$$u = (\bar{x}_1 - \bar{x}_2) + t_{0.025(16)}\, s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$=(35.22 - 31.56) + (2.12)(4.716)\sqrt{\frac{1}{9} + \frac{1}{9}} = 3.66 + 4.71 = 8.37$$

The 95% confidence interval is $[-1.05,\ 8.37]$

If $\mu_1 - \mu_2$ is positive, then $\mu_1 > \mu_2$ and the first method has larger assembly time than the second method. If $\mu_1 - \mu_2$ is negative the reverse is true. Because the interval contains both positive and negative values, neither training method can be said to produce a mean assembly time that differs from the other.

بما ان القيمة $\mu_1 - \mu_2 = 0$ التي تعني ان $\mu_1 = \mu_2$ داخلة ضمن فترة الثقة لايمكننا الجزم اي الطريقتين افضل.

H.W.

Scores on a test in math. taken by college students joining morning and evening studies are $N(\mu_x, \sigma^2)$ and $N(\mu_y, \sigma^2)$ respectively. Assume the common variance σ^2 is unknown. A random sample of n=9 students from morning studies yielded $\bar{x} = 81.31$ and $s_x^2 = 60.76$ and a random sample of m=15 students from evening studies yielded $\bar{y} = 78.61$ and $s_y^2 = 48.24$. Construct a 95% confidence interval for $\mu_x - \mu_y$.

Confidence Interval for Population Variance

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Construct a (1-α)100% confidence interval for the population variance σ^2. Let S^2 be the sample variance. We know that $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2_{n-1}distribution with (n-1) degrees of freedom. Thus we can find constants a and b so that the probability

$$P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right) = 1 - \alpha$$

where $a = \chi^2_{1-\alpha/2}(n-1)$ and $b = \chi^2_{\alpha/2}(n-1)$

Rewrite the above inequality to obtain

$$P\left(\frac{1}{b} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{a}\right) = 1 - \alpha$$

or

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) = 1 - \alpha$$

That is, if s^2 is the observed sample variance from a sample of n observations, a (1-α)100% confidence interval for σ^2 is

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

or equivalently

$$\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{\alpha/2}(n-1)}, \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{1-\alpha/2}(n-1)}\right]$$

Lecture 13

Example

If n=13 and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 128.41$, a 90% confidence interval for the variance σ^2 is [l, u] such that

$$l = \frac{(n-1)s^2}{\chi^2_{0.05}(12)} = \frac{128.41}{21.03} = 6.11$$

and

$$u = \frac{(n-1)s^2}{\chi^2_{0.95}(12)} = \frac{128.41}{5.23} = 24.57$$

Accordingly, the 90% confidence interval for σ^2 is [6.11, 24.57] and for σ is [2.47, 4.96].

Confidence Interval for the Ratio σ_1^2/σ_2^2

There are occasions when it is of interest to compare the variances of two independent normal distributions.

Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_m$ be two independent random samples from the respective normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where the four parameters are unknown. Construct a (1-α)100% confidence interval for the ratio of the variances σ_1^2/σ_2^2. Let S_1^2 and S_2^2 be the respective sample

variances of the two independent random samples of sizes n and m. We know that the random variable $\frac{(n-1)S_1^2}{\sigma_1^2}$ has a χ^2_{n-1}distribution and the random variable $\frac{(m-1)S_2^2}{\sigma_2^2}$ has a χ^2_{m-1} distribution and the two chi-square random variables are independent. Define the random variable

$$F = \frac{\frac{(m-1)S_2^2}{\sigma_2^2}\Big/_{m-1}}{\frac{(n-1)S_1^2}{\sigma_1^2}\Big/_{n-1}} = \frac{S_2^2/_{\sigma_2^2}}{S_1^2/_{\sigma_1^2}}$$ has an F distribution with (m-1) and (n-1) degrees

of freedom. Thus we can find constants a and b such that

$P(a \le F \le b) = 1 - \alpha$

where $a = F_{1-\alpha/_2}(m-1)(n-1)$ and $b = F\alpha_{/_2}(m-1)(n-1)$. So that

$$P\left(a \le \frac{S_2^2/_{\sigma_2^2}}{S_1^2/_{\sigma_1^2}} \le b\right) = 1 - \alpha$$

$$P\left(a\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le b\frac{S_1^2}{S_2^2}\right) = 1 - \alpha$$

or

$$P\left(\frac{1}{F\alpha_{/_2(n-1)(m-1)}}\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2}F\alpha_{/_2}(m-1)(n-1)\right) = 1 - \alpha$$

where

$$F_{1-\alpha/_2}(m-1)(n-1) = \frac{1}{F\alpha_{/_2}(n-1)(m-1)}$$

Thus, if s_1^2 and s_2^2 are the observed sample variances, a (1-α)100% confidence interval for $\sigma_1^2/_{\sigma_2^2}$ is

$$\left[\frac{1}{F\alpha_{/_2}(n-1)(m-1)}\frac{s_1^2}{s_2^2}, F\alpha_{/_2}(m-1)(n-1)\frac{s_1^2}{s_2^2}\right]$$

Example

Construct a 98% confidence interval for the ratio of variances σ_1^2 / σ_2^2

if the observed values are $n = 13,\ m = 9,\ s_1^2 = 10.7\ and\ s_2^2 = 4.59$.

$1 - \alpha = 0.98,\ \frac{\alpha}{2} = 0.01$

$F_{\alpha/2}(n-1)(m-1) = F_{0.01}(12)(8) = 5.67$

$F_{\alpha/2}(m-1)(n-1) = F_{0.01}(8)(12) = 4.50$

The 98% confidence interval for the ratio of variances σ_1^2 / σ_2^2 is

$$\left[\frac{1}{5.67}\frac{10.7}{4.59}, 4.50\frac{10.7}{4.59}\right] = [0.41, 10.49]$$

And the 98% confidence interval for σ_1 / σ_2 is [0.64, 3.24].

H.W.

1. A random sample of size n=9 from $N(\mu, \sigma^2)$ yielded the observed statistics $\sum_{i=1}^{n} x_i^2 = 313\ and\ \sum_{i=1}^{n} x_i = 45$. Construct a 95% confidence interval for σ^2.

Confidence Interval for the Parameter P of the Binomial Distribution

Let X be a Bernoulli random variable with probability of success p and x=0, 1. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the distribution of X. Then the number of successes $Y = \sum_{i=1}^{n} X_i$ has a binomial distribution b(n, p). Let $\hat{p} = \frac{Y}{n} = \bar{X}$ be the sample proportion of success, an unbiased estimator of p. That is $\frac{Y}{n} = \bar{X}$ is the sample mean and that $Var\left(\frac{Y}{n}\right) = \frac{p(1-p)}{n}$. By the central limit theorem the distribution of $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}$ has an approximate normal distribution N(0, 1), provided that n is large enough. This means that an approximate

(1-α)100% confidence interval for p is given by

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) \approx 1-\alpha$$

$$P\left(-z_{\alpha/2} \le \frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right) \approx 1-\alpha$$

$$P\left(-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \frac{Y}{n}-p \le z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \approx 1-\alpha$$

$$P\left(-\frac{Y}{n}-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\le -p\le -\frac{Y}{n}+z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)\approx 1-\alpha$$

$$P\left(\frac{Y}{n}-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\le p\le \frac{Y}{n}+z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)\approx 1-\alpha$$

In order to avoid the appearance of p in the endpoints of the inequality, we replace p(1-p) by its estimate $\frac{Y}{n}\left(1-\frac{Y}{n}\right)$. Hence

$$P\left(\frac{Y}{n}-z_{\alpha/2}\sqrt{\frac{\frac{Y}{n}\left(1-\frac{Y}{n}\right)}{n}}\le p\le \frac{Y}{n}+z_{\alpha/2}\sqrt{\frac{\frac{Y}{n}\left(1-\frac{Y}{n}\right)}{n}}\right)\approx 1-\alpha$$

serves as an approximate (1-α)100% confidence interval for p.

Example

In a certain political campaign, one candidate has a poll taken at random among the voting population. The results are y=185 out of n=351voters favor this candidate. Should the candidate feel very confident of winning?

The point estimate of the proportion of voters who favor the candidate is

$\hat{p}=\frac{y}{n}=\frac{185}{351}=0.527$

An approximate 95% confidence interval for the fraction p of the voting population who favor this candidate is

$$l = \frac{y}{n} - z_{\alpha/2}\sqrt{\frac{y}{n}\left(1 - \frac{y}{n}\right)} = 0.527 - 1.96\sqrt{(0.527)(0.473)} = 0.475$$

$$u = \frac{y}{n} - z_{\alpha/2}\sqrt{\frac{y}{n}\left(1 - \frac{y}{n}\right)} = 0.527 + 1.96\sqrt{(0.527)(0.473)} = 0.579$$

or equivalently [0.475, 0.579].Thus there is a good possibility that p is less than 50% and the candidate should take this into account.

Confidence Intervals for the Difference of Proportions $p_1 - p_2$

Let X_1 and X_2 be two random variables with Bernoulli distributions b(1, p_1) and b(1, p_2), respectively. A random sample of size n_1 is drawn from the distribution of X_1 and a random sample of size n_2 is drawn from the distribution of X_2. Let us say that they result in Y_1 *and* Y_2 successes, respectively. Assume that the samples are independent and that $n_1 + n_2 = n$ the total sample size.

Since the independent random variables $\frac{Y_1}{n_1}$ *and* $\frac{Y_2}{n_2}$ have respective means p_1 and p_2 and variances $\frac{p_1(1-p_1)}{n_1}$ and $\frac{p_2(1-p_2)}{n_2}$, then the difference $\frac{Y_1}{n_1} - \frac{Y_2}{n_2}$ must have mean $p_1 - p_2$ and variance $\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$.

Moreover, since $\frac{Y_1}{n_1}$ *and* $\frac{Y_2}{n_2}$ have approximate normal distributions, then the difference $\frac{Y_1}{n_1} - \frac{Y_2}{n_2}$ would have an approximate normal distribution. That is

Lecture 14

$$Z=\frac{\left(\frac{Y_1}{n_1}-\frac{Y_2}{n_2}\right)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\sim N(0,1)$$

for large n_1 and n_2. If we replace p_1 *and* p_2 in the denominator by $\frac{Y_1}{n_1}$ *and* $\frac{Y_2}{n_2}$, respectively, it is still true that the new ratio will be approximately N(0, 1). Thus the approximate (1-α)100% confidence interval for p_1-p_2 is

$$P\left(-z_{\alpha/2}\le\frac{\left(\frac{Y_1}{n_1}-\frac{Y_2}{n_2}\right)-(p_1-p_2)}{\sqrt{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}}\le z_{\alpha/2}\right)\approx 1-\alpha$$

Hence, the random interval will take the following form

$$\left[\left(\frac{Y_1}{n_1}-\frac{Y_2}{n_2}\right)\pm z_{\alpha/2}\sqrt{\frac{\frac{Y_1}{n_1}\left(1-\frac{Y_1}{n_1}\right)}{n_1}+\frac{\frac{Y_2}{n_2}\left(1-\frac{Y_2}{n_2}\right)}{n_2}}\right]$$

Let y_1 *and* y_2 be the observed number of successes in the two independent samples, then the approximate (1-α)100% confidence interval for the difference p_1-p_2 is

$$\left[\left(\frac{y_1}{n_1}-\frac{y_2}{n_2}\right)\pm z_{\alpha/2}\sqrt{\frac{\frac{y_1}{n_1}\left(1-\frac{y_1}{n_1}\right)}{n_1}+\frac{\frac{y_2}{n_2}\left(1-\frac{y_2}{n_2}\right)}{n_2}}\right]$$

Example

Two detergents were tested for their ability to remove stiff stains. An inspector judged the first one to be successful on 63 out of 91 independent trials. The second one was successful on 42 out of 79 trials. Construct an approximate 90% confidence interval for the difference $p_1 - p_2$.

$$\hat{p}_1 = \frac{y_1}{n_1} = \frac{63}{91} = 0.692$$

and

$$\hat{p}_2 = \frac{y_2}{n_2} = \frac{42}{79} = 0.532$$

An approximate 90% confidence interval for $p_1 - p_2$ is

$$l = \left(\frac{y_1}{n_1} - \frac{y_2}{n_2}\right) - z\alpha_{/2}\sqrt{\frac{\frac{y_1}{n_1}\left(1 - \frac{y_1}{n_1}\right)}{n_1} + \frac{\frac{y_2}{n_2}\left(1 - \frac{y_2}{n_2}\right)}{n_2}}$$

$$= (0.692 - 0.532) - 1.645\sqrt{\frac{(0.692)(0.308)}{91} + \frac{(0.532)(468)}{79}} = 0.038$$

$$u = \left(\frac{y_1}{n_1} - \frac{y_2}{n_2}\right) + z\alpha_{/2}\sqrt{\frac{\frac{y_1}{n_1}\left(1 - \frac{y_1}{n_1}\right)}{n_1} + \frac{\frac{y_2}{n_2}\left(1 - \frac{y_2}{n_2}\right)}{n_2}}$$

$$= (0.692 - 0.532) + 1.645\sqrt{\frac{(0.692)(0.308)}{91} + \frac{(0.532)(468)}{79}} = 0.282$$

or equivalently [0.038, 0.282]. Accordingly, it seems that the first detergent is better than the second one for removing these stains.

Tests of Statistical Hypotheses

A statistical hypothesis H is a statement or a conjecture about the distribution $f(x;\theta)$ of a population X. This conjecture is usually about the parameter θ.

Definition (Simple and Composite Hypotheses)

A hypothesis H is said to be simple hypothesis if it completely specify the density $f(x;\theta)$ of the population; otherwise it is called a composite hypothesis.

Definition (Null and Alternative Hypotheses)

A statistical hypothesis involves a separation of the parameter space $\Omega = \{\theta: a \le \theta \le b\}$ into two disjoint regions, ω and ω'. The hypothesis to be tested is called the null hypothesis and is denoted by H_o. The negation of the null hypothesis is called the alternative hypothesis and is denoted by H_1.

$H_o: \theta \in \omega$
$H_o: \theta \in \omega'$
where $\omega \cap \omega' = \emptyset \; and \; \omega \cup \omega' = \Omega$

Foe example $H_o: \theta = \theta_o$ is a simple hypothesis whereas

$H_1: \theta >, <, or \; \ne \theta_o$ is a composite hypothesis

Since we do not know the true value of the parameter, we must base our decision on the observed value of X.

Definition

A test of a statistical hypothesis is a procedure based on determining a partition of the sample space into two sets, the critical region or rejection region C and its compliment C' which is called the acceptance region. If $\boldsymbol{x} \in C$, then we reject H_o, and if $\boldsymbol{x} \in C'$, then we do not reject (accept) H_o.

الفشل في الحصول على دليل لرفض فرضية العدم لايعني بالضرورة وجود دليل على قبولها

Definition (Critical Region)

The critical region for a test of hypothesis is the subset of the sample space that corresponds to rejecting H_o.

Definition (Test Statistic)

A statistic used to define the critical region is called a test statistic. It is a summary of the data used to help make the decision.

Definition (Types of Errors)

In testing hypotheses, two types of errors we can make

Type I Error: Reject H_o when H_o is true.

Type II Error: Do not reject H_o when H_o is false (H_1 is true).

Definition (Significance Level)

The probability of type I error is called the significance level of the test or the size of the critical region.

$\alpha = P(Type\ I\ Error) = P(Reject\ H_o|H_o \text{ is true})$

$= P(\boldsymbol{x} \in C|H_o \text{ is true}) = P(\boldsymbol{x} \in C|\theta \in \omega)$

Definition

The probability of type II error is defined as

$\beta = P(Type\ II\ Error) = P(Accept\ H_o|H_1 is\ true)$

$= P(\boldsymbol{x} \in C'|H_1) = P(\boldsymbol{x} \in C'|\theta \in \omega')$

Decision	H_o is True	H_1 is True
Reject H_o	α	Correct Decision
Accept H_o	Correct Decision	β

Definition (Power Function)

The power function of a hypothesis test H_o: $\theta \in \omega$ versus H_1: $\theta \in \omega'$ is a function $k: \Omega \to [0,1]$ defined by

$$k(\theta) = \begin{cases} \alpha(\theta) & \theta \in \omega \ \ (H_o \text{ is true}) \\ 1-\beta(\theta) & \theta \in \omega' \ \ (H_1 \text{ is true}) \end{cases}$$

That is k(θ) is a function that gives the probability of rejecting H_o for each parameter point in H_o and H_1. The value of the power function at a parameter point is the power of the test at that point.

Remark

A good test is the one with small α(θ) and β(θ), but as α(θ) increases β(θ) decreases and as α(θ) decreases β(θ) increases. Under certain assumptions because both probabilities cannot be minimized, so we fix one, usually α(θ) and try to find that test such that β(θ) is small.

Definition

We say that the critical region is of size α if $\alpha = \max_{\theta\in\omega} P(\boldsymbol{x} \in C)$

The power of the test is $1-\beta = P(\boldsymbol{x} \in C | \theta \in \omega')$

Remark

It is useful to point out that significance level, size of the critical region, power function when H_o is true and the probability of type I error are all equivalent.

Example

Let $X \sim b(1,p)$, where $0 < p \leq ½$. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n=20 from the distribution of X and let $Y = \sum_{i=1}^{n} X_i$ be the total number of successes in the sample. A hypothesis $H_o: p = \frac{1}{2}$ is tested against $H_1: p < \frac{1}{2}$. If H_o is rejected when $y = \sum_{i=1}^{n} x_i \leq 6$. What is the probability of type I error?

The random variable $Y = \sum_{i=1}^{n} X_i \sim b(20, p)$

$\alpha = P(Type\ I\ error) = P(Reject\ H_o | H_o \text{ is true})$

$= P\left(Y \leq 6 | p = \frac{1}{2}\right) = \sum_{y=0}^{6} C_y^{20} \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right)^{20-y} = 0.0577$

Definition (P- Value)

The P- value of a test is defined as the smallest size α at which H_o can be rejected based on the observed value of the test statistic. If the P- value is small, we tend to reject H_o.

Example

Let $X \sim N(\mu, 100)$. Suppose that we are testing $H_o: \mu = 60$ against $H_1: \mu > 60$. The observed sample mean $\bar{x} = 63$ based on n=52 observations. What is the P-value of this test?

$P - value = P(\bar{X} \geq \bar{x} | \mu = 60)$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{63 - 60}{10/\sqrt{52}} | \mu = 60\right) = P\left(Z \geq \frac{63 - 60}{10/\sqrt{52}}\right)$$

$$= 1 - P\left(Z \leq \frac{63 - 60}{10/\sqrt{52}}\right) = 1 - \Phi\left(\frac{63 - 60}{10/\sqrt{52}}\right)$$

$= 1 - \Phi(2.163) = 1 - 0.9843 = 0.0157$

Since the P-value is small < 0.05, we reject $H_o: \mu = 60$ and conclude that $\mu > 60$.

H.W.

1. Let $X \sim b(1, p)$, where $0 < p \leq ½$. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n=10 from the distribution of X and let $Y = \sum_{i=1}^{n} X_i$ be the total number of successes in the sample. A hypothesis $H_o: p = \frac{1}{2}$ is tested against $H_1: p < \frac{1}{2}$. If H_o is rejected when $y = \sum_{i=1}^{n} x_i \leq 2$.

   a. What is the probability of type I error α?
   b. What is the power of the test when $p = \frac{1}{4}$?

Best Critical Region

Let $f(x; \theta)$ denote the pdf of a random variable X. Let $X_1, X_2, \ldots, X_n$ denote a random sample from this distribution. Consider the test of the simple null hypothesis $H_o: \theta = \theta_o$ against the simple alternative hypothesis $H_1: \theta = \theta_1$.

Let C be a critical region of size α, that is, $\alpha = P(\boldsymbol{x} \in C|H_o)$. Then C is called a best critical region of size α if for every other critical region A of size α

$\alpha = P(\boldsymbol{x} \in A|H_o)$, we have that $P(\boldsymbol{x} \in C|H_1) \geq P(\boldsymbol{x} \in A|H_1)$.

Thus a best critical region of size α is the critical region that has the greatest power among all critical regions of size α. An important theorem that provides a systematic method of determining a best critical region follows

Theorem (Neyman – Pearson)

Let $X_1, X_2, \dots, X_n$ be a random sample from a population with pdf $f(x;\theta)$. The joint pdf of $X_1, X_2, \dots, X_n$ is

$L(\theta; x_1, x_2, \dots, x_n) = f(x_1;\theta).f(x_2;\theta)\dots f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

Let θ_o and $\theta_1$be distinct fixed values of θ so that $\Omega = \{\theta: \theta = \theta_o, \theta_1\}$. If there exist a positive constant k and a subset C of the sample space such that

1. $P[(x_1, x_2, \dots, x_n) \in C|H_o] = \alpha$
2. $\frac{L(\theta_0;\ x_1, x_2, \dots x_n)}{L(\theta_1;\ x_1, x_2, \dots x_n)} \leq k$ for $(x_1, x_2, \dots, x_n \in C)$
3. $\frac{L(\theta_0;\ x_1, x_2, \dots x_n)}{L(\theta_1;\ x_1, x_2, \dots x_n)} \geq k$ for $(x_1, x_2, \dots, x_n \in C')$

Then C is a best critical region of size α for testing $H_o: \theta = \theta_o$ against $H_1: \theta = \theta_1$.

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\theta, 1)$. It is desired to test $H_o: \theta = \theta_o = 0$ against $H_1: \theta = \theta_1 = 1$. A random sample of size n=25 is observed. Find a best critical region c of size α, and the power of the test 1- β.

$f(x;\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$, $-\infty < x < \infty$

$$\frac{L(\theta_o;\ x_1, x_2, \ldots, x_n)}{L(\theta_1;\ x_1, x_2, \ldots, x_n)} = \frac{(2\pi)^{-n/2}\, e^{-\frac{\sum_{i=1}^{n} x_i^2}{2}}}{(2\pi)^{-n/2}\, e^{-\frac{\sum_{i=1}^{n}(x_i-1)^2}{2}}}$$

$$= \frac{e^{-\frac{\sum_{i=1}^{n} x_i^2}{2}}}{e^{-\frac{\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i + n}{2}}} = \frac{1}{e^{\left(\sum_{i=1}^{n} x_i - \frac{n}{2}\right)}} = e^{\left(-\sum_{i=1}^{n} x_i + \frac{n}{2}\right)} \le k$$

By taking natural logarithm, we get

$$-\sum_{i=1}^{n} x_i + \frac{n}{2} \le \ln k$$

$$-\sum_{i=1}^{n} x_i \le \ln k - \frac{n}{2}$$

or equivalently

$$\sum_{i=1}^{n} X_i \ge\ -\ln k + \frac{n}{2} = c$$

In this case the best critical region is the set

$$C = \left\{(x_1, x_2, \ldots, x_n) ; \sum_{i=1}^{n} x_i \ge c\right\}$$

The event $\sum_{i=1}^{n} x_i \ge c$ is equivalent to the event $\bar{x} \ge \frac{c}{n} = c_1$. Then if H_o is true, the statistic $\bar{X} \sim N(0, \frac{1}{n})$. So that $P(\bar{X} \ge c_1 | H_o) = \alpha$, where c_1 can be found from standard normal tables for a given sample size n and a given significance level α. Now

$$P(\bar{X} \ge c_1 | H_o) = P\left(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \ge \frac{c_1 - 0}{1/\sqrt{25}}\right) = P(Z \ge 5c_1) = 0.05$$

$1 - \Phi(5c_1) = 0.05$, or $\Phi(5c_1) = 0.95$. Hence

$5c_1=1.645$ and $c_1=0.329$.

That is, we reject $H_o: \theta = 0$ with significance level $\alpha=0.05$ if the sample mean $\bar{x} \geq 0.329$. The power of the test is

$$1 - \beta = P(\bar{X} \geq c_1 | H_1) = P\left(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \geq \frac{c_1 - 1}{1/\sqrt{25}}\right)$$

$$= P\big(Z \geq 5(0.329 - 1)\big) = 1 - P(Z \leq -3.355) = 1 - \Phi(-3.355)$$

Hence, $1 - \beta = 1 - [1 - \Phi(3.355)] = \Phi(3.355) = 0.999$.

Now for the same example let the hypothesis be $H_o: \theta = \theta_o = 0$ against

$H_1: \theta = \theta_1 = -1$. To find the best critical region we proceed as follows

$$\frac{L(\theta_o;\ x_1, x_2, \ldots, x_n)}{L(\theta_1;\ x_1, x_2, \ldots, x_n)} = \frac{(2\pi)^{-n/2} e^{-\frac{\sum_{i=1}^{n} x_i^2}{2}}}{(2\pi)^{-n/2} e^{-\frac{\sum_{i=1}^{n} (x_i+1)^2}{2}}} \leq k$$

$$= \frac{e^{\frac{-\sum_{i=1}^{n} x_i^2}{2}}}{e^{\frac{-\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i - \frac{n}{2}}{2}}} = e^{\sum_{i=1}^{n} x_i + \frac{n}{2}} \leq k$$

$$\sum_{i=1}^{n} x_i + \frac{n}{2} \leq \ln k$$

$$\sum_{i=1}^{n} x_i \leq \ln k - \frac{n}{2}$$

$$\frac{\sum_{i=1}^{n} x_i}{n} \leq \frac{\ln k}{n} - \frac{1}{2}$$

$\therefore$ The best critical region is $\bar{X} \le c$. That is $P(\bar{X} \le c|H_o) = \alpha$.

$$\alpha = P\left(Z \le \frac{c-0}{1/\sqrt{25}}\right) = \Phi(5c) = 0.05$$

From the standard normal tables $5c = -1.645$ and $c = -0.329$

The power of the test is

$$1-\beta = P(\bar{X} \le c|H_1) = P\left(\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} \le \frac{c+1}{1/\sqrt{25}}\right)$$

$$1-\beta = P\big(Z \le 5(c+1)\big) = P(Z \le 5(-0.329+1)) = \Phi(3.355) = 0.999$$

H.W

1. Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, 36)$. To test the hypothesis $H_o: \mu = 50$ versus $H_1: \mu = 55$.

a. Show that $\bar{x} \ge c$ is the best critical region of size α.

b. If a random sample of size is n=16 is observed. Determine c for a significance level α=0.05 and the power of the test.

2. Let X~ Poisson (λ). To test $H_o: \lambda = \lambda_o$ versus $H_1: \lambda = \lambda_1$.

a. Determine the best critical region of size α according to the Neyman- Pearson theorem.

b. Let the test be $H_o: \lambda = 0.3$ versus $H_1: \lambda = 0.6$, if n=5 and c=13, determine the significance level α and the power of the test 1-β.

3. Let the random variable X have a pdf $f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x > 0$. To test $H_o: \theta = \theta' = 2$ versus $H_1: \theta = \theta'' = 4$, let X_1, X_2 be a random sample of size two from this distribution. Show that the best critical region is known by the use of the statistic $X_1 + X_2$.

Lecture 15

Uniformly Most Powerful Tests

In a test where $H_o: \theta = \theta_o$ and $H_1: \theta = \theta_1$ are both simple hypotheses, a critical region of size α is a best critical region if the probability of rejecting H_o when H_1 is true is a maximum when compared with all critical regions of size α.

The test using the best critical region is called a most powerful test because it has the greatest value of the power function at $\theta = \theta_1$ when compared with that of other tests of significance level α.

If H_1 is a composite hypothesis, the power of a test depends on each simple alternative in H_1.

Definition (Uniformly Most Powerful Test)

A test defined by a critical region C of size α, is called a uniformly most powerful test if it is a most powerful test against each simple alternative in H_1. The critical region C is called a uniformly most powerful critical region of size α.

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, 36)$. Consider testing $H_o: \mu = 50$ against $H_1: \mu > 50$. Find the uniformly most powerful critical region of size α.

For each simple alternative hypothesis, let $H_1: \mu = \mu_1$, where $\mu_1 > 50$. The quotient

$$\frac{L(50;\ x_1, x_2, \dots, x_n)}{L(\mu_1;\ x_1, x_2, \dots, x_n)} = \frac{\left(2\pi(36)\right)^{-n/2} e^{-\frac{\sum_{i=1}^{n}(x_i-50)^2}{72}}}{\left(2\pi(36)\right)^{-n/2} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{72}}} \leq k$$

$$= exp\left[-\frac{1}{72}\left(\sum_{i=1}^{n} x_i^2 - 100\sum_{i=1}^{n} x_i + n(50)^2 - \sum_{i=1}^{n} x_i^2 + 2\mu_1 \sum_{i=1}^{n} x_i - n(\mu_1)^2\right)\right] \leq k$$

$$= exp\left[-\frac{1}{72}\left(2(\mu_1 - 50)\sum_{i=1}^{n} x_i + n[(50)^2 - \mu_1^2]\right)\right] \leq k$$

$$= -\frac{1}{72}\left[2(\mu_1 - 50)\sum_{i=1}^{n} x_i + n((50)^2 - \mu_1^2)\right] \leq \ln k$$

$$= -2(\mu_1 - 50)\sum_{i=1}^{n} x_i \leq 72\ln k + n((50)^2 - \mu_1^2)$$

$$= \sum_{i=1}^{n} x_i \geq \frac{-72\ln k}{2(\mu_1 - 50)} + \frac{n}{2}(50 + \mu_1)$$

$$= \bar{x} \geq \frac{-72\ln k}{2n(\mu_1 - 50)} + \frac{50 + \mu_1}{2} = c$$

Thus the best critical region of size α is $C = \{(x_1, x_2, \dots, x_n); \bar{X} \geq c\}$, where c is selected such that $P(\bar{X} \geq c|\mu = 50) = \alpha$

Since the critical region C defines a test that is most powerful against each simple alternative $\mu_1 > 50$, this is a uniformly most powerful test of size α. If n=16 and α=0.05, then c≈52.47.

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(0, \theta)$, where θ is an unknown positive number. Find a uniformly most powerful test of $H_o: \theta = \theta'$ against $H_1: \theta > \theta'$.

The joint pdf of $X_1, X_2, \dots, X_n$ is

$$L(\theta; x_1, x_2, \dots, x_n) = (2\pi\theta)^{-n/2} e^{\frac{-\sum_{i=1}^{n} x_i^2}{2\theta}}$$

Let $H_1: \theta = \theta''$ where $\theta'' > \theta'$, and let k denote a positive number. Let C be a set of points where

$$\frac{L(\theta';\ x_1, x_2,\ \dots, x_n)}{L(\theta'';\ x_1, x_2,\ \dots, x_n)} = \left(\frac{\theta''}{\theta'}\right)^{n/2} exp\left[-\left(\frac{\sum_{i=1}^{n} x_i^2}{2\theta'} - \frac{\sum_{i=1}^{n} x_i^2}{2\theta''}\right)\right] \leq k$$

$$= \left(\frac{\theta''}{\theta'}\right)^{n/2} exp\left[-\frac{\sum_{i=1}^{n} x_i^2}{2}\left(\frac{\theta'' - \theta'}{\theta'\theta''}\right)\right] \leq k$$

$$= -\sum_{i=1}^{n} x_i^2 \left(\frac{\theta'' - \theta'}{2\theta'\theta''}\right) \leq \ln k - \frac{n}{2} ln\left(\frac{\theta''}{\theta'}\right)$$

$$= \sum_{i=1}^{n} x_i^2 \geq \left(\frac{2\theta'\theta''}{\theta'' - \theta'}\right)\left[\frac{n}{2} ln\left(\frac{\theta''}{\theta'}\right) - \ln k\right] = c$$

The set $C = \{(x_1, x_2, \ldots, x_n); \sum_{i=1}^{n} x_i^2 \geq c\}$ is then a best critical region for testing $H_o: \theta = \theta'$ against $H_1: \theta = \theta''$.

It remains to determine c, so that this critical region has the desired size α. If H_o is true, the random variable $\frac{\sum_{i=1}^{n} x_i^2}{\theta'} \sim \chi^2_{(n)}$, recalling that $\left[\sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}\right]$.

Now

$P\left(\frac{\sum_{i=1}^{n} x_i^2}{\theta'} \geq \frac{c}{\theta'} | H_o\right) = \alpha$, where $^{c}/_{\theta'}$ may be obtained from chi-square tables and c determined.

Moreover, the foregoing argument holds for each number $\theta'' > \theta'$. Accordingly, the critical region $C = \{(x_1, x_2, \ldots, x_n); \sum_{i=1}^{n} x_i^2 \geq c\}$ is a uniformly most powerful critical region of size α.

If n=15 experimental values were observed and the hypothesis $H_o: \theta = 3$ versus $H_1: \theta > 3$ was tested. Then from chi-square tables with α=0.05, we have $\frac{c}{3} = 25$ and hence c= 75. That is H_o is rejected if $\sum_{i=1}^{n} x_i^2 \geq 75$ with significance level α=0.05.

Remark

It is interesting to note that if the alternative hypothesis is $H_1: \theta < \theta'$, then a uniformly most powerful critical region is of the form $\sum_{i=1}^{n} x_i^2 \leq c$.