Properties of estimators

Unbiased estimators:
Let $\hat{\theta}$ be an estimator of a parameter θ. We say that $\hat{\theta}$ is an unbiased estimator of θ if

$$E(\hat{\theta}) = \theta$$

Examples:
Let $X_1, X_2, \cdots, X_n$ be an i.i.d. sample from a population with mean μ and standard deviation σ. Show that $\bar{X}$ and S^2 are unbiased estimators of μ and σ^2 respectively.

Minimum variance unbiased estimators (MVUE):

Cramer-Rao inequality:

Let $X_1, X_2, \cdots, X_n$ be an i.i.d. sample from a distribution that has pdf $f(x)$ and let $\hat{\theta}$ be an estimator of a parameter θ of this distribution. We will show that the variance of $\hat{\theta}$ is at least:

$$var(\hat{\theta}) \geq \frac{1}{nE(\frac{\partial ln f(x)}{\partial \theta})^2} \qquad \text{or} \qquad var(\hat{\theta}) \geq \frac{1}{-nE(\frac{\partial^2 ln f(x)}{\partial \theta^2})}$$

Theorem:

If $\hat{\theta}$ is an unbiased estimator of θ and if

$$var(\hat{\theta}) = \frac{1}{nE(\frac{\partial ln f(x)}{\partial \theta})^2}$$

In other words, if the variance of $\hat{\theta}$ attains the minimum variance of the Cramer-Rao inequality we say that $\hat{\theta}$ is a minimum variance unbiased estmator of θ (MVUE).

Example:

Let $X_1, X_2, \cdots, X_n$ be an i.i.d. sample from a normal population with mean μ and standard deviation σ. Show that $\bar{X}$ is a minimum variance unbiased estimator of μ.

Relative efficiency:

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of a parameter θ we say that $\hat{\theta}_1$ is relatively more efficient if $var(\hat{\theta}_1) < var(\hat{\theta}_2)$. We use the ratio

$$\frac{var(\hat{\theta}_1)}{var(\hat{\theta}_2)}$$

as a measure of the relative efficiency of $\hat{\theta}_2$ w.r.t $\hat{\theta}_1$.

Example:

Suppose $X_1, X_2, \cdots, X_n$ is an i.i.d. random sample from a Poisson distribution with parameter λ. Let $\hat{\lambda}_1 = \bar{X}$ and $\hat{\lambda}_2 = \frac{X_1+X_2}{2}$ be two unbiased estimators of λ. Find the relative efficiency of $\hat{\lambda}_2$ w.r.t. $\hat{\lambda}_1$.

Consistent estimators:

Definition:

The estimator $\hat{\theta}$ of a parameter θ is said to be consistent estimator if for any positive ϵ

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| \le \epsilon) = 1$$

or

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$$

We say that $\hat{\theta}$ converges in probability to θ (also known as the weak law of large numbers). In other words: the average of many independent random variables should be very close to the true mean μ with high probability.

Theorem:

An unbiased estimator $\hat{\theta}$ of a parameter θ is consistent if $var(\hat{\theta}) = 0$ as $n \to \infty$.

Bias:

The bias B of an estimator $\hat{\theta}$ is given by

$$B = E(\hat{\theta}) - \theta$$

In general, given two unbiased estimators we would choose the estimator with the smaller variance. However this is not always possible (there may exist biased estimators with smaller variance). We use the mean square error (MSE)

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

as a measure of the goodness of an estimator. We can show that

$$MSE = var(\hat{\theta}) + B^2$$

Example:

The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval $(\theta, \theta+1)$, where θ is the true but unknown voltage of he circuit. Suppose that $X_1, X_2, \cdots, X_n$ denotes a random sample of such readings.

  a. Show that $\bar{X}$ is a biased estimator of θ, and compute the bias.

  b. Find a function of $\bar{X}$ that is an unbiased estimator of θ.

  c. Find the MSE when $\bar{X}$ is used as an estimator of θ.

Fisher information and Cramér-Rao inequality

Let X be a random variable with pdf $f(x;\theta)$. Then

$$\int_{-\infty}^{\infty} f(x;\theta)dx = 1 \quad \text{take derivatives w.r.t. } \theta \text{ on both sides}$$

$$\int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{\partial \theta} dx = 0 \quad \text{this is the same as:}$$

$$\int_{-\infty}^{\infty} \frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial \theta} f(x;\theta)dx = 0 \quad \text{or}$$

$$\int_{-\infty}^{\infty} \frac{\partial ln f(x;\theta)}{\partial \theta} f(x;\theta)dx = 0 \quad \text{differentiate again w.r.t. } \theta$$

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

$$\int_{-\infty}^{\infty} \left[\frac{\partial^2 ln f(x;\theta)}{\partial \theta^2} f(x;\theta) + \frac{\partial ln f(x;\theta)}{\partial \theta} \frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial \theta} f(x;\theta) \right] dx = 0 \quad \text{or}$$

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

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

$$E\left(\frac{\partial^2 ln f(X;\theta)}{\partial \theta^2} \right) + E\left(\frac{\partial ln f(X;\theta)}{\partial \theta} \right)^2 = 0 \quad \text{or}$$

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

The expression

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

is the so called Fisher information for one observation.

Let's find the Fisher information in a sample: Let $X_1, X_2, \cdots, X_n$ be an i.i.d. random sample from a distribution with pdf $f(x; \theta)$. The joint pdf of $X_1, X_2, \cdots, X_n$ is

$$L(\theta) = f(x_1; \theta) f(x_2; \theta) \cdots \cdots f(x_n; \theta)$$

Take logarithms on both sides $\cdots$

$$lnL(\theta) = lnf(x_1; \theta) + lnf(x_2; \theta) + \cdots + lnf(x_n; \theta)$$

Take derivatives w.r.t θ on both sides

$$\frac{\partial lnL(\theta)}{\partial \theta} = \frac{\partial lnf(x_1; \theta)}{\partial \theta} + \frac{\partial lnf(x_2; \theta)}{\partial \theta} + \cdots + \frac{\partial lnf(x_n; \theta)}{\partial \theta}$$

When one observation was involved (see previous page) the Fisher information was $E\left(\frac{\partial lnf(X;\theta)}{\partial \theta}\right)^2$. Now we are dealing with a random sample $X_1, X_2, \cdots, X_n$ and $f(x; \theta)$ is replaced by $L(\theta)$ (the joint pdf). Therefore, the Fisher information in the sample will be $E\left(\frac{\partial lnL(\theta)}{\partial \theta}\right)^2$.

$$\left(\frac{\partial lnL(\theta)}{\partial \theta}\right)^2 = \left(\frac{\partial lnf(x_1; \theta)}{\partial \theta} + \frac{\partial lnf(x_2; \theta)}{\partial \theta} + \cdots + \frac{\partial lnf(x_n; \theta)}{\partial \theta}\right)^2 \quad \text{or}$$

$$\left(\frac{\partial lnL(\theta)}{\partial \theta}\right)^2 = \left(\frac{\partial lnf(x_1; \theta)}{\partial \theta}\right)^2 + \left(\frac{\partial lnf(x_2; \theta)}{\partial \theta}\right)^2 + \cdots + \left(\frac{\partial lnf(x_n; \theta)}{\partial \theta}\right)^2$$

$$+ \quad 2\frac{\partial lnf(x_1; \theta)}{\partial \theta} \frac{\partial lnf(x_2; \theta)}{\partial \theta} + \cdots$$

Take expected values on both sides

$$E\left(\frac{\partial lnL(\theta)}{\partial \theta}\right)^2 = E\left(\frac{\partial lnf(X_1; \theta)}{\partial \theta}\right)^2 + E\left(\frac{\partial lnf(X_2; \theta)}{\partial \theta}\right)^2 + \cdots + E\left(\frac{\partial lnf(X_n; \theta)}{\partial \theta}\right)^2$$

The expected value of the cross-product terms is equal to zero. Why?

We conclude that the Fisher information in the sample is:

$$E\left(\frac{\partial lnL(\theta)}{\partial \theta}\right)^2 = I(\theta) + I(\theta) + \cdots + I(\theta)$$

or

$$I_n(\theta) = nI(\theta)$$

The Fisher information in the sample is equal to n times the Fisher information for one observation.

Cramér-Rao inequality:

$$var(\hat{\theta}) \geq \frac{1}{nI(\theta)} \quad \text{or} \quad var(\hat{\theta}) \geq \frac{1}{nE\left(\frac{\partial ln f(X;\theta)}{\partial\theta}\right)^2} \quad \text{or} \quad var(\hat{\theta}) \geq \frac{1}{-nE\left(\frac{\partial^2 ln f(X;\theta)}{\partial\theta^2}\right)}$$

Let $X_1, X_2, \cdots, X_n$ be an i.i.d. random sample from a distribution with pdf $f(x;\theta)$, and let $\hat{\theta} = g(X_1, X_2, \cdots, X_n)$ be an unbiased estimator of the unknown parameter θ. Since $\hat{\theta}$ is unbiased, it is true that $E(\hat{\theta}) = \theta$, or

$$\int_{-\infty}^{\infty} \cdots \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) dx_1 dx_2 \cdots dx_n = \theta$$

Take derivatives w.r.t θ on both sides

$$\int_{-\infty}^{\infty} \cdots \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) \left[\sum_{i=1}^{n} \frac{1}{f(x_i;\theta)} \frac{\partial f(x_i;\theta)}{\partial\theta}\right] f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) dx_1 dx_2 \cdots dx_n = 1$$

Since

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

we can write the previous expression as

$$\int_{-\infty}^{\infty} \cdots \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) \left[\sum_{i=1}^{n} \frac{\partial ln f(x_i;\theta)}{\partial\theta}\right] f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) dx_1 dx_2 \cdots dx_n = 1$$

or

$$\int_{-\infty}^{\infty} \cdots \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) Q f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) dx_1 dx_2 \cdots dx_n = 1$$

where

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

But also, $\hat{\theta} = g(X_1, X_2, \cdots, X_n)$.