Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 41 (2) 01-06 (2009)

SURJ

Sindh Univ. Res. Jour.

VARIANCE UNBIASED ESTIMATION OF PARAMETERS IN NORMAL DISTRIBUTION

Abdul Wasim Shaikh

Institute of Mathematics and Computer Science, University of Sindh, Jamshoro,
(Received 02nd January 2009 and Revised 12th March 2009)

Abstract

Any method of estimation is subject to uncertainty and therefore we cannot expect to find any method of estimation for giving us a close estimate of the unknown parameters on every occasion and for every sample. Under such circumstances in this paper we discuss the estimation of statistical parameters of Normal distribution, which must be, the best estimated in the sense that it satisfies the criteria, which are intuitively desirable. We also discuss and suggest the theorem, which helps us to deal greatly in the existing research for functions of minimum variance unbiased estimators.

Keywords: Estimation for parameters which are intuitively desirable.

1. Introduction

The most important continuous distributions are Normal, Log normal, Weibull, Negative exponential, Beta, Gamma and Rayleigh distributions.The Cauchy distribution is also an important distribution from the theoretical point of view because most of theorems in the statistical theory can not be applied to this distribution. From the practical point of view the Normal distribution is very important distribution because many distribution tend to Normal for large number for example the Binomial, Poisson and the derived distribution t and X^2 tend to normality for large degree of freedom. In real life situations most of the data follows the normal distributions or its some transformation followers the normal distribution.

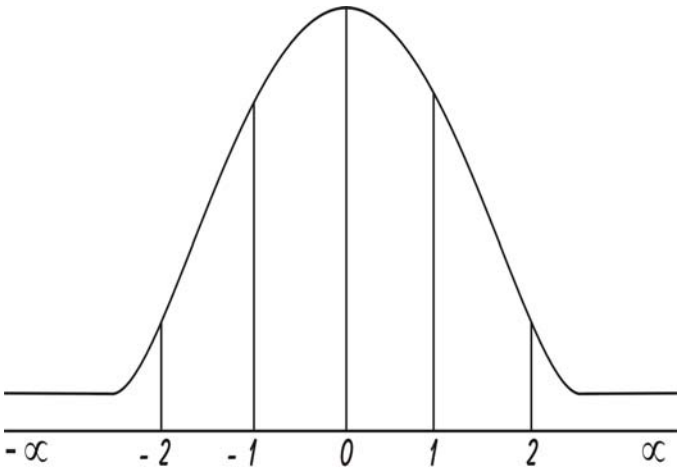
The probability density function (*p.d..f*) of the normal distribution is given as:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\left[(x-\mu)^2/2\sigma^2\right]}, \qquad \text{for } -\infty<x<\infty$$

or

$$=\frac{1}{\sigma\sqrt{2\pi}}e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad \text{for } -\infty<x<\infty$$

where μ is the mean, σ is the standard deviation obviously a normal distribution is characterized by two parameters μ and σ, its mean and standard deviation. The graph is as follows:



It is a symmetric distribution.

2. Problem of Estimation

One important problem of statistics id the estimation of population parameters such as population mean, median, variance etc. The theory of estimation was founded by R.A Fisher (1921,1925),Weiss (1961,1963), Chaudhry (1985) and Kotz, Samuel and Johnson, Norman L. (1982) roposed to estimate σ.The authors Cook, J.R. and Stefanski, L.A.(1994) have worked in estimating the simulation of parametric error models There are two types of estimation: (i) Point estimation (ii) Interval estimation.

Point estimations are particularly required when the estimate itself is to be inserted in large mathematical structure. In internal estimation, generally two functions of the sample data are constructed within which in a certain probability sense the unknown value of the parameter lines.
We may, for instance be able to say on the basis of sample data that the best estimate we can make about the population parameter is $5.3^{"}$ and we feel 95% confident that the true value of the parameter lies between $5.0^{"}$ and $5.5^{"}$. To approximate the means of order statistics in the extreme value distribution and also order statistics are discussed by Blom (1958) and Barlow (1972).

The problem is that we are given a random sample $X_1, X_2, X_3, \ldots, X_n$ of size *n* from a population, the distribution of which has known mathematical form but in values a certain number of unknown parameters. The observation are random variables. Any estimate of a parameter is the function of the sample observations and any function of the observation id also a random variable and there can be and infinite number of function of the given observations. For example, it might be assumed that the distribution is of the form $F\left(\frac{x-\mu}{\theta}\right)$, where μ and θ are the unknown location and scale parameters respectively. Suppose that the distribution under consideration is normal their the unknown parameters μ and θ are the mean and standard deviation respectively. The problem is to devise estimates for μ and θ, as functions of the elements of a random sample from $F(X)$. These functions of the observations are known statistic and these can an infinite number of functions of the given observations. If the sample is not random and nothing precise is known about the mature of the bi as operating when it was chosen then very little can be inferred about these unknown parameters. We always therefore assume that the samples are simple or random. With the help of these *X*'s we want to estimate the unknown parameters and if these parameters are known then distribution is completely specified. We call the rule or method of estimation an estimator and particular value it gives is the estimate, example , the arithmetic mean is an estimator and the particular value of the arithmetic mean for the specified values of the *X*'s is the estimate. Our problem is to find estimate and to judge its merits by the properties of its sampling distribution. We should assess the merits of an estimator by the values it takes as the sample size increases *i ,e* there is little discrepancy between the true value of the parameters and their estimates as the sample size increases. Thus the goodness of an estimate can not be judged from individual values. Any method of estimation is subject to uncertainty and therefore we can not expect to find any method of estimation to give us a close estimate of the unknown parameters on every occasion and for every sample. The best possible evaluation of the parameter/parameters has meaning only if it satisfies certain axioms or postulates which have got some theoretical justification on the basis of probability or convenience or both.

3. Properties of an Estimator

As we have seen that there can be infinite number of estimators then the problem is to determine which is the best estimator in the sense that it satisfies criteria which are intuitively desirable. It is therefore, essential to discriminate between the various statistics in order to obtain the best estimator. It will be a good estimator in the sense that the distribution of the estimator is concentrated near the true parameter. The following are the properties of a good estimator.

(i) Unbiasedness

An estimator $\hat{\theta}$ is called unbiased estimator of a parameter θ of $E(\hat{\theta}) = \theta$

otherwise it is said to be biased, if $E(\hat{\theta}) \neq \theta$. The estimator is defined to be positively biased when $E(\hat{\theta}) > \theta$ and it is said to be negatively biased when $E(\hat{\theta}) < \theta$. Unbiasedness is a property which requires that the probability distribution of $\hat{\theta}$ be necessarily centered at the parameter θ, irrespective of the value *n*, but sometimes the calculation of an unbiased estimator may be come cumber some and we may prefer to use a biased estimator. The chief merit of unbiased estimator is that they can easily be combined. Thus if $t_1, t_2, \ldots, t_k$ are all unbiased for θ and are obtained from the different sources than $\sum a_i t_i$ is another unbiased estimator such that $\sum a_i = 1$.

(ii) Consistency

An estimate is said to be consistent estimate of any parameter, when calculated form an indefinitely large sample it tends to be accurately equal to that parameter. Thus the estimator $\hat{\theta}$ is said to be consistent if as the sample size increases indefinitely $\hat{\theta}$ tends stochastically to the true value *i. e* for any given small positive quality

$$P\left\{\theta-\delta \leq \hat{\theta} \leq \theta+\delta\right\} \to 1 \quad \text{as} \quad n \to \infty$$

or

$$P\left\{\left|\hat{\theta}-\theta\right| > \delta\right\} \to 1 \quad \text{as} \quad n \to \infty$$

The property of consistency is the most desirable property of an estimation but this is a limiting property. It is an asympototic property and we are not interested in this property, when the sample is small. A consistent estimate is unbiased in the limit but an unbiased estimator may or may not be consistent estimator. If there exists one consistent estimator we may construct any others. If t_k is a consistent estimator then $\frac{n-a}{n-b} t_n$ is also a consistent estimator.

(iii) Invariance

If $\hat{\theta}$ is estimator of θ then the some method leads to $g(\hat{\theta})$ as the estimator of $g(\theta)$. This is also an important and desirable property can not be satisfied in general.

(iv) Efficiency

Efficiency of two consistent estimators of θ say $\hat{\theta}_1$ and $\hat{\theta}_2$. We naturally want to get the estimator with the minimum variance and the relative efficiency of an estimator is defined by

$$E = \frac{\text{var}(\hat{\theta}_1)}{\text{var}(\hat{\theta}_2)}, \qquad \text{if} \quad \text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$$

sometimes efficiency is restricted to large sample concept. A statistic with 100% efficiency is termed as most efficiency.

Example: If the population sample is normal then

$$\text{var}\left(\overline{X}\right) = \sigma^2/n$$

$$\text{var}\left(median\right) = \frac{\pi}{2}\sigma^2/n \quad \text{for large sample}$$

therefore

$$E = \frac{\text{var}\left(\overline{X}\right)}{\text{var}\left(median\right)} = \frac{\pi}{2}$$

$$= 63.7\%$$

Other things being equal, estimator with the greater efficiency is undoubtedly the one to use. But sometimes other things are not equal. The efficient estimators may be difficult to construct and may therefore be uneconomical. If there large number of observations available then it may sometimes be convenient to use in efficient estimators.

(v) Sufficiency

An estimator is defined to be sufficient, if the statistic used as estimator uses all the information that is contained in the sample.

Let $\left(X_1, X_2, \ldots, X_n\right)$ be a random sample from the distribution having $p.d.f.\ \left(X, \theta\right)$.

The statistic $\hat{\theta}_1 = T_1(X_1, X_2, \ldots, X_n)$ is called a sufficient statistic if and if for any other statistic $\hat{\theta}_2 = T_2(X_1, X_2, \ldots, X_n) \cdots \hat{\theta}_n = T_n(X_1, X_2, \ldots, X_n)$ the conditional of the probability density function of $g(\hat{\theta}_2, \hat{\theta}_3, \cdots, \hat{\theta}_n)$ is independent of the parameter θ, Ignacy I.(1972).

4. Neyman Criteria of sufficiency

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution having $p.d.f.\ f(X, \theta)$. Where the range of X is independent of θ, then the likelihood function *L* of a sample of *n* values from a population is defined as

$$L(X_1, X_2, \ldots, X_n) = f(X_1, \theta) \times f(X_2, \theta) \times \cdots \times f(X_n, \theta)$$

. Let $\hat{\theta}_n = T_n(X_1, X_2, \ldots, X_n)$ a statistic whose $p.d.f.$ is $g(\hat{\theta}, \theta)$ then $\hat{\theta}_1$ is a sufficient statistic for θ if and only if

$L = g(\hat{\theta}, \theta) \times H(X_1, X_2, \ldots, X_n)$ and $H(X_1, X_2, \ldots, X_n)$ does not depend upon θ. If a sufficient estimator exists it is the most efficient estimator.

5. Cramer Rao Inequality

Cramer (1947), Rao (1945) and Ghosh, J.K., Singh, Rajinder (1966) within the general class of all unbiased estimators of θ and under fairly brand condition *i ,e* the assumption concerning the regularity of the estimate $\hat{\theta}$ are satisfied then no estimator can have variance which is less than a quantity which depends on $f(X, \theta)$ and *n*. If $\hat{\theta}$ is an unbiased estimator of θ or some function say $\Im(\theta)$ then

$$\operatorname{var}(\hat{\theta}) \geq -\frac{\left(T'(\theta)\right)^2}{E\left(\frac{\partial^2 \log l}{\partial \theta^2}\right)}$$

if $T'(\theta) = 1$ then we have

$$\operatorname{var}(\hat{\theta}) = -\frac{1}{E\left(\frac{\partial^2 \log l}{\partial \theta^2}\right)}$$

Rao Black Well theorem: Let $X_1, X_2, \ldots, X_n$ denote a random sample from distribution $f(X, \theta)$. If it is given that $Y_1 = t(X_1, X_2, \ldots, X_n)$ is sufficient statistic for θ. Let $Y_2 = r(X_1, X_2, \ldots, X_n)$ be another statistics (but not a function of Y_1 alone) such that $E(Y_2) = \theta$ .Let $g(Y_1, Y_2)$ be the joint distribution of Y_1 and Y_2. Also let $g_1(Y_1)$ and $g_2(Y_2)$ be respectively the marginal distribution of Y_1 and Y_2. Let $f(y_2, y_1)$ be the conditional distribution of Y_2 given $Y_1 = Y_2$. As Y_1 is a sufficient statistic for θ, $f(Y_2, Y_1)$ is independent of θ, $E(Y_2 / Y_1 = y_1)$ is a function of y_1 say $\psi(Y_1)$ and is also independent of θ then we have

(i) $E\{\psi(Y_1)\} = \theta$

(ii) $\operatorname{var}\{\psi(Y_1)\} < \operatorname{var}(Y_2)$

Proof: We prove the theorem for the case that the marginal distribution of Y_1 and Y_2 is continuous with density function $f(y_2, y_1; \theta) = f(r, t; \theta)$. The marginal density for Y_1 is

$$f(y_1; \theta) = f(t; \theta)$$

$$= \int_{-\infty}^{\infty} f(r, t; \theta)\, dr \qquad (1)$$

The condition distribution of *r* given *t* is

$$f(r/t) = f(r, t; \theta) / f(t; \theta) \qquad (2)$$

and it does not depend on θ because we are assuming that *t* is a sufficient statistic since *r* is an unbiased estimate of θ, we must have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r\, f\left(r,t;\theta\right)\, drdt = \theta \qquad (3)$$

As an improved estimate, consider $\psi\left(t\right)$ obtained by averaging *r* for fixed *t*.

$$\psi\left(t\right) = \int_{-\infty}^{\infty} r\, f\left(r/t\right)\, dr \qquad (4)$$

It is easily seen that $\psi\left(t\right)$ is an unbiased estimate:

$$E\left\{\psi\left(t\right)\right\} = \int_{-\infty}^{\infty}\psi\left(t\right)\, f\left(t;\theta\right)\, dt$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r\, f\left(r/t\right)\, f\left(t;\theta\right)\, drdt$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r\, f\left(r,t;\theta\right)\, drdt = \theta$$

therefore $E\left\{\psi\left(Y_1\right)\right\} = \theta$

The last two steps following from (3) and (4)

We now show that the variance of the adjusted estimate is smaller than the variance of the original estimate

i,e $\text{var}\left\{\psi\left(Y_1\right)\right\} < \text{var}\left\{Y_2\right\}$

$$\text{var}(r) = E\left\{\left(r-\theta\right)^2\right\}$$

$$= E\left\{\left[\left(\psi-\theta\right)+\left(r-\psi\right)\right]^2\right\}$$

$$= \text{var}\left\{\psi\right\} + E\left\{\left(r-\psi\right)^2\right\} + 2E\left\{\left(r-\psi\right)\left(\psi-\theta\right)\right\}$$

The middle term is positive. The last term is shown below to be equal to zero. This then proves

$$\text{var}\left\{\psi\left(Y_1\right)\right\} > \text{var}\left\{Y_2\right\}$$

i,e $\text{var}\left\{\psi\left(Y_1\right)\right\} < \text{var}\left\{Y_2\right\}$

we complete the proof by showing

$$E\left\{\left(r-\psi\right)\left(\psi-\theta\right)\right\} = 0$$

$$E\left\{\left(r-\psi\right)\left(\psi-\theta\right)\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[r-\psi(t)\right]\left[\psi(t)-\theta\right] f(r,t;\theta)\, drd\theta$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\left[r-\psi(t)\right] f\left(r/t\right)\, dr\right]\left[\psi(t)-\theta\right] f\left(t;\theta\right)\, dt$$

which is equal to zero because the expression in the square bracket is identically zero,

$$\int_{-\infty}^{\infty}\left[r-\psi(t)\right] f\left(r/t\right)\, dr = \int_{-\infty}^{\infty} r\, f\left(r/t\right)\, dr - \psi(t)\int_{-\infty}^{\infty} f\left(r/t\right)\, dr$$

$$= \int_{-\infty}^{\infty}\left\{ f\left(r,t;\theta\right)/ f\left(t;\theta\right)\right\}\, dr$$

$$= \psi\left(t\right) - \frac{\psi\left(t\right)}{f\left(t;\theta\right)}\int_{-\infty}^{\infty} f\left(r,t;\theta\right)\, dr$$

$$= \psi\left(t\right) - \frac{\psi\left(t\right)}{f\left(t;\theta\right)}\, f\left(t;\theta\right)$$

= 0

by virtue of (4), (2) and (1)..

6. Conclusion

(1) The Cramer Rao result provides a method of identifying certain minimum variance unbiased statistic. Thus the inequality holds good unbiased estimate based upon sample of any size.

(2) The theorem helps us a great deal in our research for minimum variance unbiased estimators, since it essentially tells us that we need to look only at unbiased estimators that are functions of sufficient statistics. However, the theorem does not give the final answer, since there may possibly be estimators based on sufficient statistic and each unbiased.

The theorem does not tell us which of these have got the minimum variance.

Let $X_1, X_2, \ldots, X_n$ be random sample from $f\left(X,\theta\right)$ and let $\hat{\theta} = H\left(X_1, X_2, \ldots, X_n\right)$ be a sufficient statistic. Let the density of $\hat{\theta}$ be complete. If a function of $\hat{\theta}$ say $g(\hat{\theta})$ exists such that $E\left\{g(\hat{\theta})\right\} = T(\theta)$, then $g(\hat{\theta})$ is the minimum variance estimator of $T(\theta)$.

References

Blom, G. (1958) Statistical Estimates and transformed Beta variables, New York John Wiley and Sons, Inc. London.

Barlow, R. E., D. J. Bartholamu, J. M. Brenner, and H. D. Brunk, (1972) *Statistical Inference under Order Restrictions*, New York John Wiley and Sons. Inc. London.

Cook, J. R. and L. A. Stefanski, (1994) Simulation-Extrapolation Estimation in Parametric Measurement Error Models, Journal of the American Statistical Association, (**89**), 1314-1328.

Cramer H. (1946) Mathematical Methods of Statistics, Princeton University Press.

Chaudhry S.M. (1985) Introduction to Statistics Theory Part-II, Al. Hajaz Printing Press Lahore.

Fraser D.A.S. (1958) Statistics an introduction , Toppan company Ltd,. Tokyo, Japan 214-224.

Fisher, R.A. (1921) "On the Mathematical foundations of Theoritical Statistics", Phil. Trans, (A), (**222**): 399 Pp.

Fisher R.A. (1925) "Theory of Statistical estimation", Proc. Phil. Soc. (**22**): 701Pp.

Ghosh, J. K., R. Singh, (1966) Unbiased estimation of location and scale parameters. *Ann. Math. Statist*. (**37**): 1671-1675.

Ignacy I. (1972) On a characterization of some probability distributions by conditional expectations. *Sankhya* (**34**): 461-466.

Norman L. (1982) Encyclopaedia of Statistical Sciences.

Rao, C.R. (1945) Information and accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society, (**37**): 81-91.

Weiss, L. (1961) "On estimation of Scale Parameters", Naval Res. Logistics Quart (**8**): 245-256

Weiss, L. (1963) On the asymptotic distribution of an estimate of a scale parameter. Naval Res. Logistics Quart., (12) 355-370.