UNIT – IV

Measures of Location and Scale.

Introduction:

In order to restrict the class of estimators having minimum mean – squarred error within the restricted class, an alternative property 'called invariance' is used in this section. Our study is limited only two types namely location variance and scale variance.

Location Invariance.

If the observations $X_1, X_2 \ldots X_n$ represented measurements of some sort and the parameter being estimated was also measured in the same units, one might reasonably require that an estimator $t(\cdot, \ldots, \cdot)$ satisfy the property $t(x_1+c, x_2+c \ldots x_n+c) = t(x_1, x_2 \ldots x_n)+c$ for every constant c. The idea is that if a constant c is added to each of the measurements $x_1, x_2 \ldots x_n$, then the estimator evaluated at the adjusted measurements $x_1+c, x_2+c \ldots x_n+c$ ought to adjust the estimated values $t(x_1, x_2 \ldots x_n)$ by adding the same constant to it.

For example, suppose that it is desired to estimate the average weight of a group of pigs when the only method available for weighing is for a person to stand on a scale holding a pig; person and pig are weighed. If one person were to hold the pigs, the measurements (weights) $x_1+c, \ldots x_n+c$ would be obtained, where x_i is the weight of the i^{th} pigs and c is the person's weight.

If on the other hand, someone else were to hold the measurements $x_1+c', \ldots x_n+c'$ would be obtained, where c' is the other person's weight.

It seems reasonable that the estimate of the average weight of the group of pigs obtained should not depend on which person held the pig; that is the estimate should not vary with c, the weight of the pig holder.

Def - Location Invariant.

An estimator $T = t(x_1, x_2 \ldots x_n)$ is defined to be location - invariant if and only if $t(x_1+c, x_2+c \ldots x_n+c) = t(x_1, x_2 \ldots x_n) + c$, for all values $x_1, x_2 \ldots x_n$ and c.

Examples 1. $\bar{x}$ is a location invariant estimator

$$t(x_1+c, x_2+c \ldots x_n+c) = \frac{\Sigma(x_i+c)}{n} = \frac{\Sigma x_i}{n} + c = t(x_1 x_2 \ldots x_n) + c$$

for $t(x_1 x_2 \ldots x_n) = \bar{x}$; and

$$t(x_1+c, \ldots x_n+c) = \frac{\min[x_1+c, x_2+c \ldots x_n+c] + \max[x_1+c \ldots, +x_n+c]}{2}$$

$$= \frac{\min[x_1, x_2 \ldots x_n] + c + \max[x_1, x_2 \ldots x_n] + c}{2}$$

$$= \frac{\min(x_1, x_2 \ldots x_n) + \max(x_1 \ldots x_n)}{2} + c$$

$$= t(x_1, x_2 \ldots x_n) + c$$

for $t(x_1, x_2 \ldots x_n) = (y_1 + y_n)/2$

2. S^2 and $Y_n - Y_1$ are not location invariant estimators.

$$T = t(x_1 x_2 \ldots x_n) = S^2 = \Sigma(x_i - \bar{x}_n)^2/n-1$$

$$t(x_1+c, x_2+c \ldots x_n+c) = \Sigma\left[x_i + c - \Sigma(x_i+c)/n\right]^2/n-1$$

$$= t(x_1 \ldots x_n)$$

instead of $t(x_1 \ldots x_n) + c$

$\therefore$ S^2 is not a location invariant estimator.

$$T = t(x_1 \ldots x_n) = Y_n - Y_1$$

$$t(x_1+c, x_2+c \ldots x_n+c) = \max[x_1+c, \ldots x_n+c] - \min[x_1+c \ldots x_n+c]$$

$$= \max[x_1 \ldots x_n] + c - \min[x_1 \ldots x_n] - c$$

$$= t(x_1 \ldots x_n) \text{ instead of } t(x_1 \ldots x_n) + c$$

Location Parameter:

Let $\{f(\cdot\,;\theta),\ \theta \in \Theta\}$ be a family of densities indexed by a parameter θ, where Θ is the real line. The parameter θ is defined to be a location parameter if and only if the density $f(x;\theta)$ can be written as a function of $x-\theta$.

That is $f(x;\theta) = h(x-\theta)$, for some function $h(\cdot)$

Equivalently, θ is a location parameter for the density $f_X(x;\theta)$ of a random variable X if and only if the distribution of $X-\theta$ does not depend on θ.

Example: If $X \sim N(\theta, 1)$, then θ is a location parameter since

$$\phi_{\theta,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} = \phi_{0,1}(x-\theta).$$

or if X is distributed normally with mean θ and variance 1, then $X-\theta$ has a standard normal distribution; hence the distribution of $X-\theta$ is independent of θ.

Theorem: Let $X_1, X_2 \ldots X_n$ denote a random sample from the density $f(\cdot\,;\theta)$ where θ is a location parameter and Θ is the real line. The estimator

$$t(X_1, X_2 \ldots X_n) = \frac{\int \theta \prod_{i=1}^{n} f(x_i;\theta)\,d\theta}{\int \prod_{i=1}^{n} f(x_i;\theta)\,d\theta}$$

is the estimator of θ which has uniformly smallest mean-squared error within the class of location-invariant estimators.

Pitman estimator for location.

The estimator given in the following equation is called Pitman estimator for location.

$$t(X_1, X_2 \ldots X_n) = \frac{\int \theta \prod_{i=1}^{n} f(x_i;\theta)\,d\theta}{\int \prod_{i=1}^{n} f(x_i;\theta)\,d\theta}.$$

Scale invariance.

An estimator will be scale - invariant if the estimator does not depend on the scale of the measurement.

Scale invariant.

An estimator $T = t(x_1 x_2 \ldots x_n)$ is defined to be scale invariant if and only if $t(cx_1, cx_2 \ldots cx_n) = c\, t(x_1 \ldots x_n)$ for all values $x_1 \ldots x_n$ and all $c > 0$.

Example : A number of estimators that we have considered are scale - invariant, including $\bar{x}_n$, $\sqrt{s^2}$, $(Y_1 + Y_n)/2$ and $Y_n - Y_1$.

Theorem.

Let $X_1 X_2 \ldots X_n$ be a random sample from the density $f(\cdot, \theta)$, where $\theta > 0$ is a scale parameter. Assume that $f(x; \theta) = 0$ for $x \leq 0$; that is the random variable X_i assume only positive values. Within the class of scale - invariant estimators, the estimator

$$t(x_1, x_2 \ldots x_n) = \frac{\int_0^\infty (1/\theta^2) \prod_{i=1}^n f(x_i, \theta)\, d\theta}{\int_0^\infty (1/\theta^3) \prod_{i=1}^n f(x_i, \theta)\, d\theta}$$

has uniformly smallest risk for the loss function $l(t; \theta) = (t - \theta)^2/\theta^2$.

Pitman estimator for scale.

The estimator given in the following equation is defined to be the Pitman estimator for scale.

$$t(x_1, x_2 \ldots x_n) = \frac{\int_0^\infty \left(\frac{1}{\theta^2}\right) \prod_{i=1}^n f(x_i, \theta)\, d\theta}{\int_0^\infty \left(\frac{1}{\theta^3}\right) \prod_{i=1}^n f(x_i, \theta)\, d\theta}.$$

Remark : The Pitman estimator for scale is a function of

Bayes Estimators.

In point estimation, we have assumed that θ was our random sample came from some density $f(\cdot,\theta)$ where the function $f(\cdot,\cdot)$ was assumed known. We have also assumed that θ was some fixed, though unknown to us. In some real world situations with the density $f(\cdot,\theta)$ represents, there is often additional information about θ. For example, the experimenter may have evidence that θ itself acts as a random variable for which he may be able to postulate a realistic density function.

An important question is: How can this additional information about θ be used to estimate θ_0 where θ_0 is the value that Θ was equal to on the day the sample was drawn.

To examine this problem, we will assume, in addition to the assumption that our random sample came from a density $f(\cdot,\theta)$ that the unknown parameter θ is the value of some random variable, say Θ. We are still be interested in estimating some function of θ say $\tau(\theta)$. If Θ is a random variable, it has a distribution. Let $G(\cdot) = G_{\Theta}(\cdot)$ denote the cumulative distribution function of Θ and $g(\cdot) = g_{\Theta}(\cdot)$ denote the density function of Θ and we assume these functions contain no unknown parameters.

If we assume that the distribution of Θ is known, we have additional information. How can this additional information be used in estimation? This question is answered below.

Posterior Distribution.

Prior and Posterior Distributions.

The density $g_{\Theta}(\cdot)$ is called the prior distribution of Θ. The conditional density of Θ given $X_1 = x_1 \ldots X_n = x_n$ denoted by $f_{\Theta / X_1 = x_1, X_2 = x_2 \ldots X_n = x_n}(\theta / x_1, x_2 \ldots x_n)$ is called the posterior distribution of Θ.

Posterior Bayes estimator.

Let $X_1, X_2 \ldots X_n$ be a random sample from a density $f(x/\theta)$, where θ is a value of the random variable Θ with known density $g_{\Theta}(\cdot)$. The posterior Bayes estimator of $\tau(\theta)$ with respect to the prior $g_{\Theta}(\cdot)$ is defined to be

$$E[\tau(\theta) / X_1, X_2 \ldots X_n]$$

where

$$E\left[\tau(\theta) / X_1 = x_1, X_2 = x_2 \ldots X_n = x_n\right] = \int \tau(\theta) f_{\Theta / X_1 = x_1 \ldots X_n = x_n}(\theta / x_1 \ldots x_n)\, d\theta$$

$$= \frac{\int \tau(\theta) \left[\prod_{i=1}^{n} f(x_i/\theta)\right] g_{\Theta}(\theta)\, d\theta}{\int \left[\prod_{i=1}^{n} f(x_i/\theta)\right] g_{\Theta}(\theta)\, d\theta}$$

Bayes Risk.

Let $X_1, X_2 \ldots X_n$ be a random sample from a density $f(x/\theta)$, where θ is the value of a random variable Θ with cumulative distribution function $G(\cdot) = G_{\Theta}(\cdot)$ and corresponding density $g(\cdot) = g_{\Theta}(\cdot)$.

In estimating $\tau(\theta)$, let $\ell(t;\theta)$ be the loss function. The risk of estimator $T = t(X_1, X_2 \ldots X_n)$ is denoted by $R_t(\theta)$. The Bayes risk of estimator $T = t(X_1, X_2 \ldots X_n)$ with respect to the loss function $\ell(\cdot, \cdot)$ and prior cumulative distribution $G(\cdot)$ denoted by $r(t) = r_{\ell, G}(t)$ is defined to be

$$r(t) = r_{\ell, G}(t) = \int_{\Theta} R_t(\theta) \cdot g(\theta) \cdot d\theta.$$

Bayes estimator.

The Bayes estimator of $\tau(\theta)$, denoted by $T^*_{l,G}$, ie $T^*_{l,G} = t^*_{l,G}(x_1, x_2 \ldots x_n)$, with respect to the loss function $l(\cdot,\cdot)$ and prior cumulative distribution $G(\cdot)$ is defined to be that estimator with smallest Bayes risk. Or the Bayes estimator of $\tau(\theta)$ is that estimator $t^*_{l,G}$ satisfying

$$r_{l,G}(t^*) = r_{l,G}(t^*_{l,G}) \leq r_{t,G}(t)$$

for every other estimator $T = t(X_1 X_2 \ldots X_n)$ of $\tau(\theta)$.

Theorem: Let $X_1, X_2 \ldots X_n$ be a random sample from the density $f(x/\theta)$ and let $g(\theta)$ be the density of Θ. Further let $l(t;\theta)$ be the loss function for estimating $\tau(\theta)$. The Bayes estimator of $\tau(\theta)$ is that estimator $t^*(\cdot,\ldots,\cdot)$ which minimizes

$$\int_{\Theta} l[t(x_1 \ldots x_n);\theta] f_{\Theta/X_1=x_1 \ldots X_n=x_n}(\theta/x_1 \ldots x_n)\, d\theta$$

as a function of $l(\cdot,\ldots)$.

Remarks.

Under the above assumptions, the Bayes estimator of θ is given by the median of the posterior distribution of Θ for a loss function equal to the absolute deviation.

Example: Let $X_1, X_2 \ldots X_n$ be a random sample from the normal density with mean θ and variance 1. Consider estimating θ with a squared-error loss function. Assume that Θ has a normal density with mean μ_0 and variance 1. Let $\mu_0 = x_0$.

under the assumptions given in the above theorem, the Bayes estimator of $\tau(\theta)$ is given by

$$g\left[\tau(\Theta) / X_1 = x_1 \ldots X_n = x_n\right] = \frac{\int \tau(\theta) \left[\prod f(x_i/\theta)\right] g(\theta)\, d\theta}{\int \left[\prod_{i=1}^{n} f(x_i/\theta)\right] g(\theta)\, d\theta}.$$

for a squared - error loss function.

Based on this the Bayes estimator is given as the mean of posterior distribution of Θ.

$$f_{\Theta / X_1 = x_1 \ldots X_n = x_n}(\theta / x_1 \ldots x_n)$$

$$= \frac{f_{x_1 \ldots x_n/\theta}(x_1 \ldots x_n/\theta)\, g(\theta)}{f_{x_1 \ldots x_n}(x_1 \ldots x_n)} = \frac{\left[\prod f(x_i/\theta)\right] g(\theta)}{\int \left[\prod f(x_i/\theta)\right] g(\theta)\, d\theta}$$

$$= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2} \sum (x_i - \theta)^2\right] \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta - \mu_0)^2\right]}{\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2} \sum (x_i - \theta)^2\right] \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta - \mu_0)^2\right] d\theta}$$

On simplification

$$\frac{\mu_0 + \sum_{1}^{n} x_i}{n+1}$$ is also the Bayes estimator with respect to a loss function equal to the absolute deviation.