BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Statistical Inference-I

Course Code: (23ST06CC)

Unit-II

Mean Square Error

Dr. T. Jai Sankar

Associate Professor and Head

Department of Statistics

Ms. N. Saranya

Guest Faculty

Department of Statistics

UNIT-2

Mean Square error:

The mean square error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is the function of θ defined by $E(\hat{\theta}-\theta)^2$ and is denoted as $MSE_{\hat{\theta}}$. This is also called the risk function of an estimator with $(\hat{\theta}-\theta)^2$ called the quadratic loss function.

The expectation is with respect to the random variables $X_1, \dots X_n$ since they are the only random components in the expression.

The MSE measures the average squared difference between the estimator $\hat{\theta}$ and the parameter θ, a somewhat reasonable measure of performance for an estimator. In general, any increasing function for the absolute distance $|\hat{\theta}-\theta|$ would serve to measure the goodness of an estimator (Mean absolute error, $E(|\hat{\theta}-\theta|)$, is a reasonable alternative. But MSE has at least two advantages over other distance measure.

1. Analytically tractable
2. It has intrepretation.

$$M.S.E_{\hat{\theta}} = E(\hat{\theta}-\theta)^2 = Var(\hat{\theta}) + (E(\hat{\theta})-\theta)^2$$
$$= Var(\hat{\theta}) + (\text{Biass of } \hat{\theta})^2$$

This is so because

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

Bias of an estimator:

The Bias of an estimator $\hat{\theta}$ of a parameter θ is the difference between the expected value of $\hat{\theta}$ & θ.

i.e., $\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Unbiased estimator:

An estimator whose bias is identically equal to zero is called unbiased estimator and satisfies.

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

Two Components of MSE:

* Measures the variability of the estimator (precision).
* Measures its bias (accuracy)

An estimator that has good MSE properties has small combined variance and bias. To find an estimator with good MSE properties we need to find estimators that control both variance and bias.

For an unbiased estimator $\hat{\theta}$, we have

$$M.S.E_{\hat{\theta}} = E(\hat{\theta}-\theta)^2 = Var(\hat{\theta})$$

and so if an estimator is unbiased its MSE is equal to its variance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where MSE = Mean Square error
n = no. of data points
y_i = observed values
$\hat{y}_i$ = predicted values

Root Mean Square Error:

The root mean square Error (RMSE) is very frequently used measure of the differences between values predicted by an estimator.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where, RMSE = Root Mean Square Error
n - no. of data points
y_i - observed values
$\hat{y}_i$ - predicted values.

Example 1:

Let $X_1, X_2 \dots X_n$ be iid from $N(\mu, \sigma^2)$ with expected values μ and variance σ^2 then $\bar{X}$ is an unbiased estimator for μ and S^2 is an unbiased estimator for σ^2.

Soln:

we have

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 \dots X_n}{n}\right)$$

$$= \frac{E(X_1) + \dots + E(X_n)}{n}$$

$$= \frac{n\mu}{n}$$

$$= \mu$$

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

The MSE of $\bar{X}$ is

$$MSE_{\bar{X}} = E(\bar{X} - \mu)^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$$

This is because

$$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 \dots + X_n}{n}\right)$$

$$= \frac{Var(X_1) + \dots + Var(X_n)}{n^2}$$

$$= \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

Similarly, $E(S^2) = \sigma^2$

S^2 is an unbiased estimator for σ^2.

The M.S.E of S^2 is given by

$$MSE_{S^2} = E(S^2 - \sigma^2) = Var(S^2)$$

$$= \frac{2\sigma^4}{n-1}$$

Cramer-Rao Inequality:

If T is an unbiased estimator for $\gamma(\theta)$, a function of parameter θ then

$$Var(T) \geq \frac{\left\{\frac{d}{d\theta}\gamma(\theta)\right\}^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} = \frac{\{\gamma'(\theta)\}^2}{I(\theta)}$$

where $I(\theta)$ is the information on θ supplied by the sample and L is maximum likelihood estimator.

Proof: (for cont r.v.)

Let X be an r.v having p.d.f $f(x, \theta)$ and L be the likelihood function of the random samples $x_1, x_2, \ldots x_n$ from this population

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

Since L is a joint density function.

So, we have $\int L\, dx = 1$.

where $\int dx = \iint \cdots \int dx_1\, dx_2 \cdots dx_n$

Differentiate w.r.to θ and using regularity condition, we get

$$\int \frac{\partial}{\partial \theta} L\, dx = 0$$

$$\Rightarrow \left(\frac{\partial}{\partial \theta} \log L\right) L\, dx = 0$$

$$\Rightarrow E\left(\frac{\partial}{\partial \theta} \log L\right) = 0 \rightarrow (1) \qquad \because E(x) = \int x\, f(x)$$

Let T be an unbiased estimator $\gamma(\theta)$ such that $E(T) = \gamma(\theta) \rightarrow (2)$

$$\Rightarrow \int T L\, dx = \gamma(\theta)$$

Diff. w.r.to θ, we get,

$$\int T \frac{\partial L}{\partial \theta} dx = \gamma'(\theta)$$

$$\Rightarrow \int T \left(\frac{\partial}{\partial \theta} \log L\right) L\, dx = \gamma'(\theta)$$

$$E\left(T \cdot \frac{\partial}{\partial \theta} \log L\right) = \gamma'(\theta) \rightarrow (3)$$

Now, using covariance

$$\text{Cov}\left(T, \frac{\partial}{\partial \theta} \log L\right) = E\left(T \cdot \frac{\partial}{\partial \theta} \log L\right) - E(T) \cdot E\left(\frac{\partial}{\partial \theta} \log L\right)$$

[using $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$]

$$\Rightarrow \text{Cov}\left(T, \frac{\partial}{\partial \theta} \log L\right) = \gamma'(\theta) \rightarrow \text{From } (4)$$

(from (1), (3))

WKT : $\boxed{(\text{cov}(X,Y))^2 \leq \text{Var}(X)\,\text{Var}(Y)}$

$$\Rightarrow \left[\text{cov}\left(T, \frac{\partial}{\partial\theta}\log L\right)\right]^2 \leq \text{Var}(T)\,\text{Var}\left(\frac{\partial}{\partial\theta}\log L\right)$$

$$\Rightarrow [\gamma'(\theta)]^2 \leq \text{var}(T)\left[E\left[\left(\frac{\partial}{\partial\theta}\log L\right)^2\right] - \left[E\left(\frac{\partial}{\partial\theta}\log L\right)\right]^2\right]$$

$$\Rightarrow [\gamma'(\theta)]^2 \leq \text{var}(T)\left\{E\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}$$

$$\because \left(E\left(\frac{\partial}{\partial\theta}\log L\right)\right)^2 = 0$$

$$\Rightarrow \text{Var}(T) \geq \frac{[\gamma'(\theta)]^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2}$$

Hence proved.

Regularity Conditions for Cramer - Rao inequality:

* The parameter space Θ is a non - degenerable open interval on the real line $R^1(-\infty, \infty)$

* For almost all $x = (x_1, \dots x_n)$ and for all $\theta \in \Theta$, $\frac{\partial}{\partial\theta} L(x, \theta)$ exists, the exceptional set, if any is independent of θ.

* The range of integration is independent of the parameter θ, so that if (x, θ) is differentiable under integral sign. If range is not independent of θ and f is zero at the extremes of the range. i.e.,

$$f(a, \theta) = 0 = f(b, \theta), \text{ then}$$

$$\frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b \frac{\partial f}{\partial\theta}\,dx - f(a,\theta)\frac{\partial a}{\partial\theta} + f(b,\theta)\frac{\partial b}{\partial\theta}$$

$$\Rightarrow \frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b \frac{\partial f}{\partial\theta}\,dx; \quad \text{since } f(a,\theta) = 0 = f(b,\theta)$$

* The condition of uniform convergence of integrals are satisfied so that differentiation under the integral sign is valid.

* $I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta}\log L(x,\theta)\right\}^2\right]$, exists and is positive $\forall\ \theta \in \Theta$

Fisher Information:

(The Fisher Information is the amount of information that an observable random variable, X carries about an unknown parameter - θ upon which the likelihood function of θ, $L(\theta) = f(x;\theta)$ depends). The likelihood function is the joint probability of the data, the X's conditional on the value of θ, as a function of B.

Since the expectation of the score is 0, the variance is simply the second moment of the score, the derivative of log of the likelihood function w.r.t θ.

$$I(\theta) = E\left\{\left[\frac{\partial}{\partial\theta}\log f(x,\theta)\right]^2 \Big| \theta\right\},$$

which implies $0 \le I(\theta) < \infty$

The fisher information is thus the expectation of the squared score.

If t is an unbiased estimator of parameter θ. i.e., $E(t) = \theta \Rightarrow \gamma(\theta) = \theta \Rightarrow \gamma'(\theta) = 1$

then from Cramer Rao inequality, we get,

$$Var(t) \ge \frac{1}{E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}} = \frac{1}{I(\theta)}$$

where $I(\theta) = E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}$

is called by R.A. Fisher at the amount of information on θ supplied by the sample $(x_1, x_2 \ldots x_n)$ and its reciprocal $1/I(\theta)$ as the information limit to the variance of estimator $t = t(x_1, x_2 \ldots x_n)$

Fisher Information for θ contained in the random Variable X :

$$I(\theta) = E(\theta) \{ [l'(x|\theta)]^2 \}$$

$$\int [(l'[x|\theta))]^2 f(x)|\theta \, dx \rightarrow (1)$$

We assume that we can exchange the order of differentiation and integration then

$$\int f'(x|\theta) \, dx = \frac{\partial}{\partial \theta} \int f(x|\theta) \, dx = 0$$

Similarly,

$$\int f''(x|\theta) \, dx = \frac{\partial^2}{\partial \theta^2} \int f(x|\theta) \, dx = 0$$

It is easy to see that

$$E_\theta [l'(x|\theta)] = \int l'(x|\theta) f(x|\theta) \, dx$$

$$= \int \frac{f'(x|\theta)}{f(x|\theta)} f(x|\theta) \, dx$$

$$= \int f'(x|\theta) \, dx = 0$$

The defn of Fisher information (1) can be written as

$$I(\theta) = Var_\theta [l''(x|\theta)] \rightarrow (2)$$

Also

$$l''(x|\theta) = \frac{\partial}{\partial \theta} \left[\frac{f'(x|\theta)}{f(x|\theta)} \right]$$

$$= \frac{f''(x|\theta) f(x|\theta) - [f'(x|\theta)]^2}{[f(x|\theta)]^2}$$

$$= \frac{f''(x|\theta)}{f(x|\theta)} - [l'(x|\theta)]^2$$

$$\therefore E_\theta [l''(x|\theta)] = \int \left[\frac{f''(x|\theta)}{f(x|\theta)} - [l'[(x|\theta)]^2 \right] f(x|\theta) \, dx$$

$$= \int f''(x|\theta) \, dx - E_\theta \{ [l'(x|\theta)]^2 \}$$

$$= -I(\theta)$$

Finally we have another formula to calculate Fisher Information

$$I(\theta) = -E_\theta[l''(x|\theta)]$$

$$= -\int \left[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta)\right] f(x|\theta)\, dx \rightarrow (3)$$

$\therefore$ We have 3 methods to calulate the Fisher information [eqn 1, 2 & 3]

Cramer Rao Lower Bound:

Theorem:

Consider a parametric model $\{f(x|\theta): \theta \in \Omega\}$ (satisfying certain mild regularity assumptions) where $\theta \in \mathbb{R}$ is a single parameter. Let T be any unbiased estimator of θ based on data $X_1 \ldots X_n \overset{iid}{\sim} f(x|\theta)$. Then

$$Var_\theta[T] \geq \frac{1}{nI(\theta)}$$

Proof:

The score function

$$z(x, \theta) = \frac{\partial}{\partial\theta} \log f(x|\theta) = \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}$$

and let $Z = Z(X_1 \ldots X_n, \theta) = \sum_{i=1}^{n} z(X_i, \theta)$

By definition of correlation and the fact that the correlation of two random variables is always between -1 and 1.

$$Cov_\theta[Z, T]^2 \leq Var_\theta[Z] \times Var_\theta[T]$$

The random variables $z(X_1, \theta) \ldots z(X_n, \theta)$ are iid and they have mean 0 and variance $I(\theta)$.

Then

$$Var_\theta(Z) = n\, Var_\theta[z(X_1, \theta)] = nI(\theta)$$

Since T is unbiased,

$$\theta = E_\theta[T] = \int_{\mathbb{R}^n} T(x_1 \ldots x_n) f(x_1|\theta) \ldots f(x_n|\theta)\, dx_1 \ldots dx_n.$$

Differentiating both sides with respect to θ and applying the product rule of differentiation,

$$1 = \int_{\mathbb{R}^n} T(x_1 \dots x_n)\Big(\frac{\partial}{\partial\theta} f(x_1|\theta) \times \dots f(x_n|\theta) + f(x_1|\theta) \times \frac{\partial}{\partial\theta} f(x_2|\theta) \times \dots f(x_n|\theta) + f(x_1|\theta) \times f(x_2|\theta) \times \dots \frac{\partial}{\partial\theta} f(x_n|\theta)\Big)$$

$$= \int_{\mathbb{R}^n} T(x_1 \dots x_n)\, z(x_1, \dots x_n, \theta)\, f(x_1|\theta) \times \dots f(x_n|\theta)\, dx_1 \dots dx_n$$

$$= E_\theta [TZ]$$

Since $E_\theta[Z] = 0$

this implies

$$Cov_\theta[T, Z] = E_\theta[T, Z] = 1$$

So, $Var_\theta[T] = \frac{1}{nI(\theta)}$ as desired.

Example:

Suppose random variable X has a Bernoulli distribution for which the parameter θ is unknown $(0 < \theta < 1)$. Determine the fisher Information $I(\theta)$ is X.

Soln:

The point mass function of X is

$$f(x|\theta) = \theta^x (1-\theta)^{1-x} \quad \text{for } x = 1, \text{ or } 0.$$

$$\therefore \ l(x|\theta) = \log f(x|\theta) = x \log\theta + (1-x)\log(1-\theta)$$

and $l'(x|\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta}$

$$l''(x|\theta) = \frac{-x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

Since $E(X) = \theta$, the fisher information is

$$I(x|\theta) = -E[l''(x|\theta)]$$

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

$$= \frac{1}{\theta} + \frac{1}{1-\theta}$$

$$I(x|\theta) = \frac{1}{\theta(1-\theta)}$$

Minimum Variance Bound estimator (MVB)

An unbiased estimator t of $\gamma(\theta)$ for which Cramer-Rao lower Bound in

$$Var(t) \geq \frac{\left\{\frac{d}{d\theta}\gamma(\theta)\right\}^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} = \frac{\{\gamma'(\theta)\}^2}{I(\theta)}$$

is attained is called a Minimum Variance Bound (MVB) estimator.

Bhattacharya's Bound (Bhattacharya Inequality)

Let $X_1, \dots X_n$ be a random sample from a population with pdf (pmf) $f(x,\theta), \theta \in \Omega$ (Ω is any open interval on the real line)

Let $S_i = \frac{\partial^i}{\partial\theta^i}\left\{\prod_{j=1}^{n} f(x_j,\theta)\right\} \Big/ \left\{\prod_{j=1}^{n} f(x_j;\theta)\right\}$; $i=1,2\dots k$

Let the following conditions hold

i) $\frac{\partial^i}{\partial\theta^i} f(x,\theta)$ exists $\forall\, \theta \in \Omega$ for almost all x, $i=1\dots k$.

ii) $\int\int\dots\int \prod_{j=1}^{n} f(x_j,\theta)\, d\mu(\underline{x})$ can be differentiated under the integral sign i times $(i=1\dots k)$

iii) $S_1, \dots S_k$ are linearly independent.

iv) $\int\int\dots\int S(\underline{x}) \prod_{j=1}^{n} f(x_j,\theta)\, d\mu(x)$ can be differentiated under the integral sign i times $(i=1\dots k)$ for any integral function S.

Let $\lambda_{ij} = Cov(S_i, S_j)$; $i \neq j = 1, \dots k$

$\lambda_{ii} = Var(S_i)$, $i=1,\dots k$

Let $\Delta = [\lambda_{ij}]_{i \neq j = 1,2\dots k}$

Let λ^{rs} denote r-sth term of the matrix Δ^{-1} and $\eta_i = Cov(T, S_i) = \frac{d^i g}{d\theta^i}$ where $E_\theta T(\underline{X}) = g(\theta)$ and $\eta' = (\eta_1, \dots \eta_k)$

The Bhattacharya's Bound is

$$Var_\theta(T) \geq \eta' \Lambda^{-1} \eta = \sum\sum \lambda^{rs} \frac{d^r g}{d\theta^r} \cdot \frac{d^s g}{d\theta^s}$$

[For $k=1$, this will reduce to FRCLB]

Fritchet Rao Cramer lower Bound

Proof:

$$E_\theta T(x) = g(\theta) \ \forall \ \theta \in \Theta$$

$$\Rightarrow \int\int \cdots \int T(\underline{x}) \left\{ \prod_{j=1}^{n} f(x_j, \theta) \right\} d\mu(\underline{x}) = g(\theta) \ \forall \ \theta \in \Theta$$

Differentiating ① w.r.to θ, i times

$$\int\int \cdots \int T(\underline{x}) \left\{ \frac{\partial^i}{\partial\theta^i} \prod_{j=1}^{n} f(x_j,\theta) \right\} d\mu(\underline{x}) = \frac{d^i g}{d\theta^i}$$

$$\int\int \cdots \int T(\underline{x})\, S_i \left\{ \prod_{j=1}^{n} f(x_j,\theta) \right\} d\mu(\underline{x}) = \frac{d^i g}{d\theta^i}$$

$$E(TS_i) = \frac{d^i s}{d\theta^i}$$

Now $E(S_i) = 0$; $i = 1 \dots k$

$$Cov(T, S_i) = \frac{d^i g}{d\theta^i}$$

Multiple Correlation coefficient between T & $(S_1 \dots S_k)$ is

$$R = \frac{\eta' \Lambda^{-1} \eta}{Var(T)} \leq 1$$

where Λ - Dispersion matrix of $\underline{S}$ $= (S_1 \dots S_k)$

So $\eta' \Lambda^{-1} \eta \leq Var(T)$

which is the required lower Bound.

Chapman - Robbins (Inequality):

Let $X = (X_1, X_2, \dots X_n)$ be a random sample from $f_\theta(\cdot)$ where $\theta \in \Theta$. Let $T = t(X)$ be an unbiased estimator of $g(\theta)$. Consider $f_\theta(\cdot)$ is a probability density function. Let $\theta_0 \in \Theta$ be any fixed value of θ, such that for sufficiently small $h \neq 0$, $\theta_0 + h \in \Theta$ and $f_{\theta_0}(x) = 0 \Rightarrow f_{\theta_0+h}(x) = 0$ Then

$$\text{Var}_{\theta_0}(T) \geq \sup_h \left[\frac{\{g(\theta_0+h) - g(\theta_0)\}^2}{E_{\theta_0}\left\{\frac{f_{\theta_0+h}(x)}{f_{\theta_0}(x)} - 1\right\}^2} \right]$$

The quantity in the R.H.S is called the Chapman-Robbins lower bound for the variance of an unbiased estimator of $g(\theta)$.

Proof:

Since T is an unbiased estimator of $g(\theta)$ we have,

$$E_\theta(T) = g(\theta), \forall \theta$$

$$\Rightarrow \int t f_\theta(x)dx = g(\theta), \forall \theta \rightarrow ①$$

From ② and using the fact that

$\int \{f_{\theta_0+h}(x)dx - f_{\theta_0}(x)\}dx = 0$, we get

$$g(\theta_0+h) - g(\theta_0) = \int \{t f_{\theta_0+h}(x) - t f_{\theta_0}(x)\}dx$$

$$= \int \{t - g(\theta_0)\}\{f_{\theta_0+h}(x) - f_{\theta_0}(x)\}dx$$

$$= \int \{t - g(\theta_0)\} \frac{\{f_{\theta_0+h}(x) - f_{\theta_0}(x)\}}{f_{\theta_0}(x)} f_{\theta_0}(x)dx$$

This means

$$\text{Cov}\left[\{T - g(\theta_0)\}, \frac{\{f_{\theta_0+h}(x) - f_{\theta_0}(x)\}}{f_{\theta_0}(x)}\right]$$

$$= g(\theta_0+h) - g(\theta_0).$$

So, using the well known result that for any two random variables U and V.

$$\{\text{Cov}(U,V)\}^2 \leq \text{Var}(U)\,\text{Var}(V)$$

we see that

$$\{g(\theta_0+h) - g(\theta_0)\}^2 \leq \text{Var}_{\theta_0}(T)\,\text{Var}_{\theta_0}\left[\frac{\{f_{\theta_0+h}(x) - f_{\theta_0}(x)\}}{f_{\theta_0}(x)}\right] \rightarrow ②$$

Since

$$E_{\theta_0}\left[\frac{\{f_{\theta_0+h}(x)-f_{\theta_0}(x)\}}{f_{\theta_0}(x)}\right] = \int \frac{\{f_{\theta_0+h}(x)-f_{\theta_0}(x)\}}{f_{\theta_0}(x)} \times f_{\theta_0}(x)\,dx = 0$$

From ② we get

$$Var_{\theta_0}(T) \geq \frac{\{g(\theta_0+h)-g(\theta_0)\}^2}{E_{\theta_0}\left[\frac{\{f_{\theta_0+h}(x)-f_{\theta_0}(x)\}}{f_{\theta_0}(x)}\right]^2}$$

$$= \frac{\{g(\theta_0+h)-g(\theta_0)\}^2}{E_{\theta_0}\left[\frac{f_{\theta_0+h}(x)}{f_{\theta_0}(x)}-1\right]^2} \rightarrow ③$$

Since ③ holds for all values of h, we get

$$Var_{\theta_0}(T) \geq \sup_h \left[\frac{\{g(\theta_0+h)-g(\theta_0)\}^2}{E_{\theta_0}\left\{\frac{f_{\theta_0+h}(x)}{f_{\theta_0}(x)}-1\right\}^2}\right]$$

Hence proved.

Exponential Family of Distribution:-

Exponential Family:-

A family of distributions $\{P_\theta : \theta \in \Lambda\}$ is said to be an exponential family if its density can be placed in the form

$$f_Y(y;\theta) = C(\theta)\exp\left\{\sum_{l=1}^{m} Q_l(\theta)\,T_l(y)\right\} h(y)$$

where $C, Q_1, \ldots Q_m, T_1, \ldots T_m$ and h are real valued functions.

Completeness of exponential family:

If $\Gamma = R^n$, $\Lambda = R^m$, $c, T_1, \ldots T_m$ and h are real valued functions and P_θ has a density of the form

$$f_Y(y;\theta) = c(\theta)\exp\left\{\sum_{l=1}^{m} \theta_l\, t_l(y)\right\} h(\theta)$$

Then $T(Y) = [T_1(Y), \ldots, T_m(Y)]$ is a Complete sufficient statistics for $\{P_\theta ; \theta \in \Lambda\}$ if Λ contains an m-dimensional rectangle.

Exponential family of Single parameter:

$$f(y,\theta) = c(\theta)\, h(y)\, e^{Q(\theta)\, T(y)}$$

Examples:

1) $X \sim Bin(n, p)$, n is known.

$$f(x,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{n}{x}(1-p)^n \left(\frac{p}{1-p}\right)^x$$

$$= (1-p)^n \binom{n}{x} e^{x \log (p/1-p)}$$

So, binomial dist (with n known) is in exponential family.

2) $X \sim P(\lambda)$

$$f(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \frac{1}{x!} e^{x \log \lambda}$$

$\therefore X \sim P(\lambda) \rightarrow$ exponential family.

Multiparameter Exponential family (k parameter)

$$f_Y(y;\theta) = c(\theta)\, h(y) \exp\left\{\sum_{l=1}^{m} Q_l(\theta) T_l(y)\right\}$$

where $\theta \in \mathbb{R}^m$

Examples:

$X \sim N(\mu, \sigma^2)$ both μ & σ^2 are unknown

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$= \frac{e^{-\mu^2/2\sigma^2}}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2\sigma^2 + \frac{\mu x}{\sigma^2}}$$

Two Parameter Exponential family:

$$f(y,\theta) = c(\theta)\, h(y)\, e^{Q(\theta)T(y)}$$

$$\log f(y,\theta) = \log(c(\theta)) + \log h(x) + Q(\theta)\, T(x))$$

$$\frac{\partial \log f(x,\theta)}{\partial \theta} = \frac{c'(\theta)}{c(\theta)} + T(x)\ Q'(\theta)$$

$$S(\underline{x},\theta) = \sum_{i=1}^{n} \frac{\partial \log f(x_i,\theta)}{\partial \theta} = \frac{n\,c'(\theta)}{c(\theta)} + Q'(\theta) \sum_{i=1}^{n} T(x_i)$$

Thus $W = \frac{1}{n} \sum_{i=1}^{n} T(x_i)$ is linearly related with $S(\underline{x},\theta)$ with probability 1. Hence any linear function of W will be attaining the FRCLB for the variance of unbiased estimator of $E(W)$. We also determine $E(W)$ have

$$\int f(x,\theta)\, d\mu(x) = 1$$

$$\Rightarrow \int c(\theta)\, h(x)\, e^{Q(\theta)T(x)}\, d\mu(x) = 1$$

Differentiating under the integral sign, we get

$$\int c'(\theta)\, h(x)\, e^{Q(\theta)T(x)}\, d\mu(x) + \int c(\theta)\, h(x)\, e^{Q(\theta)T(x)} Q'(\theta) T(x)\, d\mu(x) = 0$$

$$\Rightarrow \frac{c'(\theta)}{c(\theta)} + Q'(\theta)\, E_\theta T(x) = 0$$

$$\Rightarrow E_\theta T(x) = -\frac{c'(\theta)}{c(\theta)\, Q'(\theta)}$$

ex: Consider geometric distributions

$$f(x,\theta) = \theta(1-\theta)^x, \quad x = 0,1,\ldots; \quad 0<\theta<1$$

$$= \theta\, e^{x\log(1-\theta)}$$

$C(\theta) = \theta$, $h(x) = 1$, $T(x) = x$, $Q(\theta) = \log(1-\theta)$

$$-\frac{c'(\theta)}{c(\theta)\, Q'(\theta)} = \frac{1-\theta}{\theta} = 1/\theta - 1$$

So $E(W) - E(\bar{x}) = \frac{1}{\theta} - 1$ and

$V(\bar{x})$ will be same as the FRCLB for estimator of $1/\theta - 1$

Then $\bar{x} - 1$ is UMVUE for $1/\theta$.

Asymptotic Normality:

A statistic (or an estimator) $W_n(x)$ is asymptotically normal if

$$\sqrt{n}\,(W_n - J(\theta)) \xrightarrow{d} N(0; V(\theta)) \text{ for all } \theta.$$

where $\xrightarrow{d}$ stands for converges in distribution.

$J(\theta)$ - asymptotic mean

$V(\theta)$ - asymptotic variance

$$W_n \sim AN\left(J(\theta), \frac{V(\theta)}{n}\right)$$

Central Limit Theorem:

Assume $X_i \overset{iid}{\sim} f(x/\theta)$ with finite mean $\mu(\theta)$ and variance $\sigma^2(\theta)$

$$\bar{x} \sim AN\left(\mu(\theta), \frac{\sigma^2(\theta)}{n}\right)$$

$$\Leftrightarrow \sqrt{n}\,(\bar{x} - \mu(\theta)) \xrightarrow{d} N(\theta, \sigma^2(\theta))$$

Dealth Method:

Assume $W_n \sim AN\left(\theta, \frac{\gamma(\theta)}{n}\right)$. If a function g satisfies $g'(\theta) \neq 0$, then

$$g(W_n) \sim AN\left(g(\theta), [g'(\theta)]^2 \frac{\gamma(\theta)}{n}\right)$$

Asymptotic Relative Efficiency:

If two estimator W_n and V_n satisfy

$$\sqrt{n}\,[W_n - T(\theta)] \xrightarrow{d} N(0, \sigma_N^2)$$

$$\sqrt{n}\,[V_n - T(\theta)] \xrightarrow{d} N(0, \sigma_\gamma^2)$$

The asymptotic relative efficiency (ARE) of V_n with respect to W_n is

$$ARE(V_n, W_n) = \frac{\sigma^2 W}{\sigma^2 V}$$

If $ARE(V_n, W_n) \geq 1$ for every $\theta \in \Omega$ then V_n is asymptotically more efficient than W_n.

Example:

Let $X_i \overset{iid}{\sim}$ Poisson (λ). Consider estimating

$$P_r(X=0) = e^{-\lambda}$$

Our estimators are

$$W_n = 1/n \sum_{i=1}^{n} I(X_i = 0)$$

$$V_n = e^{-\bar{X}}$$

Determine which one is more asymptotically efficient estimator.

Soln:

$V_n(x) = e^{-\bar{X}}$, by CLT

$$\bar{X} \sim AN\left(E(X); Var(X)/N\right) \sim AN(\lambda, \lambda/n)$$

Define $g(y) = e^{-y}$ then, $V_n = g(\bar{x})$ and

$$g'(y) = -e^{-y}$$

By deall method,

$$V_n = e^{-\bar{X}} \sim AN\left(g(\lambda), [g'(\lambda)]^2 \frac{\lambda}{n}\right) \sim AN$$

$$\left(e^{-\lambda}, e^{-2\lambda} \frac{\lambda}{n}\right)$$

Define $Z_i = I(X_i = 0)$

$$W_n = 1/n \sum_{i=1}^{n} I(X_i = 0) = \bar{Z}_n$$

$$Z_i \sim \text{Bernoulli } [E(Z)]$$

$$E(Z) = P_r(X=0) = e^{-\lambda}$$

$$Var(Z) = e^{-\lambda}(1-e^{-\lambda})$$

By CLT

$$W_n = \bar{Z}_n \sim AN\left[(E(Z)), Var(Z)/n\right]$$

$$\sim AN\left(e^{-\lambda}, \frac{e^{-\lambda}(1-e^{-\lambda})}{n}\right)$$

$$ARE(W_n, V_n) = \frac{e^{-2\lambda} \lambda/n}{e^{-\lambda}(1-e^{-\lambda})/n}$$

$$= \frac{\lambda}{e^{\lambda}(1-e^{-\lambda})}$$

$$= \frac{\lambda}{e^{\lambda}-1}$$

$$= \frac{\lambda}{\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\dots\right)-1}$$

$$< 1 \qquad (\forall \lambda \geq 0)$$

Therefore, $W_n = \frac{1}{n} \sum I(X_i = 0)$ is less efficient than V_n (MLE) and ARE attains maximum at $\lambda = 0$.

Asymptotic Efficiency (for iid samples):

A sequence of estimators W_n is asymptotically efficiency for $T(\theta)$ if for all $\theta \in \Omega$.

$$\sqrt{n}\,(W_n - T(\theta)) \xrightarrow{d} N\left(0, \frac{[T'(\theta)]^2}{I(\theta)}\right)$$

$$\Leftrightarrow W_n \sim AN\left(T(\theta), \frac{[T'(\theta)]^2}{n I(\theta)}\right)$$

$$I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta} \log f(x|\theta)\right\}^2 \Big/ \theta\right]$$

$$= -E\left[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta) / \theta\right]$$

where $\frac{[T'(\theta)]^2}{n(I(\theta))}$ is C-R Bound for unbiased estimators of $T(\theta)$.