Ch.8 Comparison of Estimators

8.1. CRITERIA FOR ESTIMATOR COMPARISON

When a statistical model is specified by assuming that the population distribution is one of the probability density functions $f(x;\theta), \theta \in \Omega$, and a random sample $X_1, \cdots, X_n$ is used to estimate the parameter or a function of the parameter $\eta = \eta(\theta)$, a good estimator is one that can provide an estimate close to the true value.

There are various ways to measure the performance of such an estimator, but the most commonly used one is the mean squared error defined as:

$$E_\theta\left[(\hat{\eta}(X_1,\cdots,X_n) - \eta(\theta))^2\right]$$

which we will denote simply by

$$MSE(\hat{\eta}, \theta) = E_\theta\left[(\hat{\eta} - \eta(\theta))^2\right]$$

Ex. 8.1.1 When comparison is feasible

Using random samples $X_1, \cdots, X_n$, the estimators of θ in $U(0,\theta)$ are:

$$\hat{\theta}^{MLE} = X_{(n)}, \quad \hat{\theta}^{MME} = 2\bar{X}$$

The MSE of these estimators are:

$$MSE(\hat{\theta}^{MLE}, \theta) = E_\theta\left(X_{(n)} - \theta\right)^2$$

$X/\theta \sim U(0,1)$

$X_{(n)}/\theta \sim beta(n,1)$

$$E\left[\frac{X_{(n)}}{\theta}\right] = \frac{a}{a+\beta} = \frac{n}{1+n}, \quad Var\left[\frac{X_{(n)}}{\theta}\right] = \frac{a\beta}{(a+\beta)^2(a+\beta+1)} = \frac{n}{(1+n)^2(2+n)}$$

$$\therefore E[X_{(n)}] = \frac{n}{1+n}\theta, \quad Var(X_{(n)}) = \frac{n}{(1+n)^2(2+n)}\theta^2$$

$$E_\theta[(X_{(n)} - \theta)^2] = E_\theta\left[\left\{\theta \cdot \left(\frac{X_{(n)}}{\theta} - 1\right)\right\}^2\right]$$

$$= \theta^2 E_\theta\left[\left(\frac{X_{(n)}}{\theta} - 1\right)^2\right] = \theta^2 E_\theta\left[(X_{(n)}/\theta)^2\right] \quad E[(U-1)^2] = E[U^2]$$

$$= \theta^2\left\{Var(X_{(n)}/\theta) + E[X_{(n)}/\theta]^2\right\}$$

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

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

* Revision needed

$$MSE(\hat{\theta}^{MME}, \theta) = E_\theta\left[(2\bar{X} - \theta)^2\right] = 4\theta^2 E_\theta\left[\left(\frac{\bar{X}}{\theta} - \frac{1}{2}\right)^2\right]$$

$$X/\theta \sim U(0,1) \quad \therefore \frac{(\bar{X}/\theta) - (1/2)}{\sqrt{1/12}/\sqrt{n}} \sim N(0,1) \quad (\because CLT)$$

$$\left\{\frac{(\bar{X}/\theta) - (1/2)}{\sqrt{1/12}/\sqrt{n}}\right\}^2 \sim \chi^2(1) \quad \therefore E_\theta\left[(\bar{X}/\theta - 1/2)^2\right] = \frac{1}{12n}$$

$$\therefore 4\theta^2 E_\theta\left[\left(\frac{\bar{X}}{\theta} - \frac{1}{2}\right)^2\right] = \frac{1}{3n}\theta^2$$

$$MSE(\hat{\theta}^{MLE}, \theta) = \frac{2}{(n+1)(n+2)}\theta^2$$

$$MSE(\hat{\theta}^{MME}, \theta) = \frac{1}{3n}\theta^2$$

Hence $MSE(\hat{\theta}^{MLE}, \theta) \le MSE(\hat{\theta}^{MME}, \theta) \quad \forall \theta > 0$

and also the rate of converge to 0 are:

$$MSE(\hat{\theta}^{MLE}, \theta) \approx n^{-2}, \quad MSE(\hat{\theta}^{MME}, \theta) \approx n^{-1}$$

In this case, we see that MLE is a better estimator than MME

Ex. 8.1.2. When comparison is difficult

For a random sample $X_1, \cdots, X_n$ from a normal distribution $N(\theta, 1)$, $-\infty < \theta < +\infty$, consider the estimator

$$\hat{\theta}^0 = 0, \quad \hat{\theta}^{MLE} = \bar{X}$$

Their mean squared errors are given by

$$MSE(\hat{\theta}^0, \theta) = E_\theta(0-\theta)^2 = \theta^2, \quad MSE(\hat{\theta}^{MLE}, \theta) = E_\theta(\bar{X} - \theta)^2 = \frac{1}{n}$$

Therefore, if $|\theta| < 1/\sqrt{n}$, the estimator $\hat{\theta}^0 = 0$, which does not use the observed data, has a smaller MSE than the MLE $\hat{\theta}^{MLE} = \bar{X}$.

Hence, based on the MSE criterion, the MLE $\bar{X}$ cannot always be said to be superior to $\hat{\theta}^0 = 0$. However, when the estimators are compared in terms of the rate at which the MSE converges to zero as the sample size increases, $\hat{\theta}^{MLE}$ is a better estimator than $\hat{\theta}^0 = 0$.

Motivated by the comparison with $\hat{\theta}^0 = 0$, consider a hybrid estimator that utilizes the result of a hypothesis test, defined by:

$$\hat{\theta}^H = \begin{cases} 0, & |\bar{X}| < 1.96/\sqrt{n} \\ \bar{X}, & |\bar{X}| > 1.96/\sqrt{n} \end{cases}$$

Its MSE can be expressed, using a random variable Z following the standard normal distribution as

$$E_\theta[(\hat{\theta}^H - \theta)^2] = n^{-1}\left\{|\sqrt{n}\theta|^2 P(|Z + \sqrt{n}\theta| < 1.96) + E[Z^2 I(|Z + \sqrt{n}\theta| \ge 1.96)\right.$$

$$= \underbrace{\theta^2 \cdot P(|\bar{X}| < 1.96/\sqrt{n})}_{①} + \underbrace{\int (\bar{x} - \theta)^2 \cdot pdf_{N(\mu, \sigma^2/n)}(x)\, dx \cdot I(|\bar{x}| > 1.96/\sqrt{n})}_{②}$$

* $\bar{X} \sim N(\theta, 1/n) \quad \therefore (\bar{X} - \theta)\cdot\sqrt{1/n} = \sqrt{n}(\bar{X} - \theta) \overset{d}{\equiv} Z \sim N(0,1), \quad \therefore \sqrt{n}\cdot\bar{X} \overset{d}{\equiv} Z + \sqrt{n}\theta$

$$① = \theta^2 \cdot P(-1.96 < \bar{X}\cdot\sqrt{n} < 1.96) = \theta^2 P(-1.96 < Z + \sqrt{n}\theta < 1.96)$$

$$= \theta^2 P(|Z + \sqrt{n}\theta| < 1.96)$$

$$② \; E\left[(\bar{X} - \theta)^2 I(|\bar{X}| > 1.96/\sqrt{n})\right] = \frac{1}{n} E\left[Z^2 I(|\sqrt{n}\bar{X}| > 1.96)\right]$$

$$\because \bar{X} - \theta \sim N(0, 1/n)$$

$$\Rightarrow \sqrt{n}(\bar{X} - \theta) \sim N(0,1) \overset{d}{\equiv} Z$$

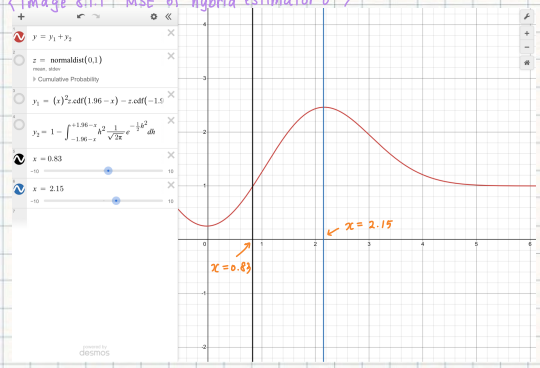
$$\Rightarrow \bar{X} - \theta = \frac{1}{\sqrt{n}} \cdot Z$$

$$\Rightarrow \therefore (\bar{X} - \theta)^2 = n^{-1} Z^2$$

$$= n^{-1} E\left[Z^2 \cdot I(|Z + \sqrt{n}\theta| > 1.96)\right]$$

From this expression, we see that the MSE depends only on $|\sqrt{n}\theta|$, and its graph is given as in <Figure 8.1.1>. Moreover, as the sample size increases, the rate at which the MSE converges to zero is : $MSE(\hat{\theta}^H, \theta) \approx n^{-1}$, which is the same as that of $\hat{\theta}^{MLE} = \bar{x}$. 114)

<Image 8.1.1 MSE of hybrid estimator $\hat{\theta}^H$>



* 114] 표준정규분포의 확률밀도함수 ϕ와 누적분포함수 Φ에 대하여 Mill's ratio라고 불리우는 다음의 근사식이 성립하는 것이 알려져 있다.

$$x(1-\Phi(x)) \sim \phi(x), \quad x \to +\infty$$

이로부터 $MSE(\hat{\theta}^H, \theta) \approx n^{-1}$임을 밝힐 수 있다.

In other words, if the two estimators are compared using the MSE criterion, the hybrid estimator has nearly the same performance as the sample mean when the sample size is large, while for small sample sizes the sample mean is superior in most cases, although there are situations in which the hybrid estimator performs better.

As seen in <Ex. 8.1.2>, it is often difficult to compare two estimators based on MSE alone. Since the estimator that is superior depends on the range of unknown parameter value, such comparison can be problematic.

Therefore, comparison criteria that do not depend on the value of the parameter are often introduced. Among the most commonly used such criteria are the "maximum MSE" and the "Bayesian MSE."

Definition: Maximum MSE and Bayesian MSE

(a) Maximum MSE criteria: based on the maximum mean square error, that is:

$$\max_{\theta \in \Omega} E_\theta[(\hat{\eta} - \eta(\theta))^2],$$

we compare $\hat{\eta}$, the estimator of $\eta = \eta(\theta)$. The value, $\hat{\eta}^*$, that minimizes the above is called the "minimax MSE estimator."

$$\max_{\theta \in \Omega} E_\theta[(\hat{\eta}^* - \eta(\theta))^2] = \min_{\hat{\eta}} \max_{\theta \in \Omega} E_\theta[(\hat{\eta} - \eta(\theta))^2]$$

(b) Bayesian MSE criteria: Based on:

$$r(\pi, \hat{\eta}) = \int_\Omega E_\theta[(\hat{\eta} - \eta(\theta))^2]\, \pi(\theta)\, d\theta$$

Which is the weighted mean of MSE for π, a function of the parameter space, we compare $\hat{\eta}$, the estimator of $\eta = \eta(\theta)$. The estimator $\hat{\eta}^\pi$, which minimizes $r(\pi, \hat{\eta})$, is called the "Bayesian MSE estimator" for prior density π. The prior density π is either $\pi = c\ (c \geq 0)$ or a pdf that represents the weight.

If the subject of estimation $\eta = \eta(\theta)$ is strictly positive, instead of MSE, we can utilize "Mean Square Relative Error":

$$E_\theta[(\hat{\eta}/\eta - 1)^2]$$

and compare the estimators based on maximum MSE or Bayesian MSE.

Ex. 8.1.3.

Among the estimators (below) of the population variance based on a random sample $X_1, \cdots, X_n$ from $N(\mu, \sigma^2)$, $-\infty < \mu < +\infty$, $\sigma^2 > 0$, find the estimator that minimizes maximum MSE.

$$\hat{\sigma}_c^2 = c \sum_{i=1}^n (X_i - \bar{X})^2, \quad 0 < c < +\infty \qquad (\text{Find } c)$$

criterion: $\max_{\theta \in \Omega} E_\theta[(\hat{\sigma}_c^2/\sigma^2 - 1)^2]$

$$\max_{\theta \in \Omega} E_\theta[(\hat{\sigma}_c^2/\sigma^2 - 1)^2]$$

$$\hat{\sigma}_c^2/\sigma^2 = cV, \quad V \sim \chi^2(n-1)$$

$$\therefore \Rightarrow \max_{\theta \in \Omega} E_\theta[(cV-1)^2] = \max_{\theta \in \Omega} E_\theta[\{cV - c(n-1) + c(n-1) - 1\}^2]$$

$$\Rightarrow \max_{\theta \in \Omega} E_\theta[\{cV - c(n-1)\}^2] + E_\theta[c^2(V-(n-1))(n-1-\tfrac{1}{c})] + \{c(n-1)-1\}^2$$

$$\Rightarrow c^2 \mathrm{Var}(V) + \{c(n-1)-1\}^2$$

$$\Rightarrow c^2 \cdot (n-1) \cdot 2 + c^2(n-1)^2 - 2c(n-1) + 1$$

$$\Rightarrow c^2(n-1)(n+1) - 2c(n-1) + 1 = f(c)$$

$$\frac{\partial}{\partial c} f(c) = 2c(n-1)(n+1) - 2(n-1)$$

$$\text{if } c = \frac{1}{n+1}, \ \frac{\partial}{\partial c} f(c) = 0 \qquad \therefore \hat{\sigma}^2 = \frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\frac{\partial^2}{\partial c^2} f(c) = 2(n-1)(n+1) \geq 0 \ \therefore\ c = (n+1)^{-1} \text{ is the minimum critical point}$$

When comparing estimators using the MSE criterion, the class of estimators under consideration is sometimes restricted to those possessing certain properties. In particular, an estimator that satisfies

$$E_\theta[\hat{\eta}(X_1,\cdots,X_n)] = \eta(\theta), \quad \forall \theta \in \Omega \quad \text{115)}$$

115) 불편추정량임을 나타내는 이러한 항등식을 간단히 $E_\theta \hat{\eta}^{UE} \overset{\theta\in\Omega}{\equiv} \eta(\theta)$ 로 나타내기로한다.

is called an unbiased estimator. One may then seek, among such estimators, the one that minimizes the MSE.

For an unbiased estimator $\hat{\eta}^{UE}$, since

$$MSE(\hat{\eta}^{UE},\theta) = E_\theta[(\hat{\eta}^{UE} - \eta(\theta))^2] = E_\theta[(\hat{\eta}^{UE} - E_\theta(\hat{\eta}^{UE}))^2] = Var_\theta(\hat{\eta}^{UE})$$

the estimator with the smallest variance among all unbiased estimators is optimal under this comparison criterion. If there exists an unbiased estimator whose variance is minimized for all possible true values of the parameter θ, such an estimator is called a "uniformly minimum variance unbiased estimator." (UMVUE).

Definition : Uniformly minimum variance unbiased estimator

(i) $E_\theta[\hat{\eta}^*] \overset{\theta\in\Omega}{\equiv} \eta(\theta)$

(ii) if $E_\theta[\hat{\eta}^{UE}] \overset{\theta\in\Omega}{\equiv} \eta(\theta)$, $Var_\theta(\hat{\eta}^{UE}) \geq Var_\theta(\hat{\eta}^*)$, $\forall \theta \in \Omega$

If (i) and (ii) are satisfied we call $\hat{\eta}^*$ the UMVUE of $\eta = \eta(\theta)$. It is sometimes denoted as $\hat{\eta}^{UMVUE}$

〈Theorem 8.1.1 Uniqueness of UMVUE〉

If the variance of UMVUE of $\eta = \eta(\theta)$ is real, the UMVUE is unique. In other words, if $\hat{\eta}_1, \hat{\eta}_2$ are the UMVUE of $\eta = \eta(\theta)$ and their variances are real,

$$P_\theta[\hat{\eta}_1 = \hat{\eta}_2] = 1 \quad \forall \theta \in \Omega$$

(proof)

$E_\theta[\hat{\eta}_1] = \eta$ and $E_\theta[\hat{\eta}_2] = \eta$

$\therefore E_\theta[(\hat{\eta}_1+\hat{\eta}_2)/2] = E_\theta[\hat{\eta}_1/2] + E_\theta[\hat{\eta}_2/2] = \eta/2 + \eta/2 = \eta$

$\therefore \hat{\eta}_3 = (\hat{\eta}_1+\hat{\eta}_2)/2$ is also an unbiased estimator(UE)

$Var(\hat{\eta}_1) = Var(\hat{\eta}_2)$ $\because \hat{\eta}_1$ and $\hat{\eta}_2$ are UMVUE

$Var(\hat{\eta}_3) = Var(\hat{\eta}_1/2) + 2Cov(\hat{\eta}_1/2, \hat{\eta}_2/2) + Var(\hat{\eta}_2/2)$

$= Var(\hat{\eta}_1)/4 + 1/2\, Cov(\hat{\eta}_1, \hat{\eta}_2) + Var(\hat{\eta}_2)/4$

we know that $Var(\hat{\eta}_3) \geq Var(\hat{\eta}_1) = Var(\hat{\eta}_2)$

$\therefore \frac{1}{4}\{Var(\hat{\eta}_1) + 2Cov(\hat{\eta}_1, \hat{\eta}_2) + Var(\hat{\eta}_2)\} \geq Var(\hat{\eta}_1) = Var(\hat{\eta}_2)$

$\therefore Cov(\hat{\eta}_1, \hat{\eta}_2) \geq Var(\hat{\eta}_1) = Var(\hat{\eta}_2)$

$\therefore Var(\hat{\eta}_1 - \hat{\eta}_2) = Var(\hat{\eta}_1) - 2Cov(\hat{\eta}_1, \hat{\eta}_2) + Var(\hat{\eta}_2) \leq 0$

$\therefore 0 \leq Var(\hat{\eta}_1 - \hat{\eta}_2) \leq 0$

Meanwhile, $\hat{\eta}_1$ and $\hat{\eta}_2$ are both UEs of $\eta = \eta(\theta)$

Therefore from〈Theorem 1.6.4〉, $P_\theta[\hat{\eta}_1 - \hat{\eta}_2 = 0] = 1 \quad \forall \theta \in \Omega$

Key ideas :

① Introduce an arbitrary UE $\hat{\eta}_2$ s.t. $Var(\hat{\eta}_1) = Var(\hat{\eta}_2)$
② Find the variance of $\hat{\eta}_3 = \frac{1}{2}\hat{\eta}_1 + \frac{1}{2}\hat{\eta}_2$
③ Show that $Cov(\hat{\eta}_1, \hat{\eta}_2) \geq Var(\hat{\eta}_1) = Var(\hat{\eta}_2)$
④ Find the variance of $\hat{\eta}_4 = \frac{1}{2}\hat{\eta}_1 + \frac{1}{2}\hat{\eta}_2$
⑤ Show that $Var(\hat{\eta}_1 - \hat{\eta}_2) = 0$

8.2. SUFFICIENT STATISTICS

When making inferences about a parameter based on a random sample $X_1, \cdots, X_n$ if the same level of precision can be achieved using only a part of the sample rather than all n observations, then there would be an advantage in recording and managing only that subset of observations instead of the entire sample. As a concept related to such data reduction, let us consider sufficiency.

Ex. 8.2.1.

Instead of the observation of two iid Bernoulli(θ) $(0<\theta<1)$ trials, $(X_1, X_2)^t$, Suppose that we only have the observation $Y = X_1 + X_2$. Can we restore the original data regardless of θ?

(Solution)

X_1, X_2 are either 0 or 1 and thus $Y = X_1 + X_2$ is one of 0, 1, or 2.

It is clear that if $Y=0$, $(X_1, X_2)^t = (0,0)^t$ and if $Y=2$, $(X_1, X_2)^t = (1,1)^t$. If $Y=1$, either $(X_1, X_2)^t = (1,0)$ or $(X_1, X_2)^t = (0,1)$. Thus, we must restore $(X_1, X_2)^t$ by their respective conditional probabilities:

$$P_\theta(X_1=0, X_2=1 \mid Y=1), \quad P_\theta(X_1=1, X_2=0 \mid Y=1)$$

However,

$$P_\theta(X_1=0, X_2=1 \mid Y=1) = \frac{P_\theta(X_1=0, X_2=1)}{P_\theta(X_1+X_2=1)} = \frac{(1-\theta)\theta}{\binom{2}{1}\theta^1(1-\theta)^{2-1}} = \frac{1}{2}$$

$$P_\theta(X_1=1, X_2=0 \mid Y=1) = \frac{P_\theta(X_1=1, X_2=0)}{P_\theta(X_1+X_2=1)} = \frac{\theta(1-\theta)}{\binom{2}{1}\theta^1(1-\theta)^{2-1}} = \frac{1}{2}$$

thus, restoration can be carried out regardless of the value of θ.

Definition: Sufficient Statistic

When a statistical model is specified by assuming that the population distribution is one of the pdf $f(x;\theta), \theta \in \Omega$, consider a statistic $Y = u(X_1, \cdots, X_n)$ based on the random sample $X_1, \cdots, X_n$. If:

$$P_\theta\{(X_1, \cdots, X_n)^t \in A \mid Y=y\} = P_{\theta_2}\{(X_1, \cdots, X_n)^t \in A \mid Y=y\} \quad \forall A, \forall y, \forall \theta_1 \in \Omega, \theta_2 \in \Omega$$

in other words, if the conditional probability of $(X_1, \cdots, X_n)^t$ given $Y = u(X_1, \cdots, X_n) = y$ does not depend on the unknown parameter $\theta \in \Omega$, the statistic $Y = u(X_1, \cdots, X_n)$ is called the "sufficient statistic for $\theta \in \Omega$"

Ex. 8.2.2. Sufficient Statistics in Bernoulli trials.

Let $X_1, \cdots, X_n$ be random samples from independent Bernoulli trials, Bernoulli(θ), $0<\theta<1$.

$$Y = X_1 + \cdots + X_n \sim B(n, \theta)$$

Hence, the conditional probability of $(X_1, \cdots, X_n)^t$ given $Y = X_1 + \cdots + X_n = y$ is as follows, and we see that it does not depend on parameter $\theta \in \Omega$

$$P_\theta(X_1 = x_1, \cdots, X_n = x_n \mid X_1 + \cdots + X_n = y) = \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} I_{(X_1+\cdots+x_n=y)}$$

$$= \binom{n}{y}^{-1} I_{(X_1+\cdots+x_n=y)} \quad \forall \theta : 0<\theta<1$$

Hence, $Y = X_1 + \cdots + X_n$ is a sufficient statistic of $\theta \in (0,1)$

〈Theorem 8.2.1 Factorization Theorem〉

If the population distribution is one of pdf $f(x;\theta), \theta \in \Omega$, the necessary and sufficient condition for the statistic $Y = u(X_1, \cdots, X_n)$, a function of random sample $X_1, \cdots, X_n$, to be a sufficient statistic is that there exists functions K_1 and K_2 satisfying the following inequality:

$$\prod_{i=1}^n f(x_i;\theta) = K_1(u(x), \theta) K_2(x) \quad \forall x = (x_1, \cdots, x_n), \forall \theta \in \Omega$$

(proof)

The proof holds for the general case but here will be prove for the discrete case.

(proof for sufficient condition)

From the given conditions,

$$P_\theta(X=x) = \prod_{i=1}^n f(x_i;\theta) = K_1(u(x), \theta) K_2(x) \quad \forall x = (x_1, \cdots, x_n), \forall \theta \in \Omega$$

Hence for all $\theta \in \Omega$, $x = (x_1, \cdots, x_n)^t$ and y, the following holds:

$$P_\theta(X=x \mid Y=y) = P_\theta(X=x, Y=u(x)=y) / P_\theta(Y=u(x)=y)$$

$$= P_\theta(X=x) I_{(u(x)=y)} / \sum_{z:u(z)=y} P_\theta(X=z), \quad z = (z_1, \cdots, z_n)^t$$

$$= \frac{K_1(u(x), \theta) K_2(x) I_{(u(x)=y)}}{\sum_{z:u(z)=y} K_1(u(z), \theta) \cdot K_2(z)} = \frac{K_1(y, \theta) K_2(x) I_{(u(x)=y)}}{K_1(y, \theta) \sum_{z:u(z)=y} K_2(z)}$$

$$= \frac{K_2(x) I_{(u(x)=y)}}{\sum_{z:u(z)=y} K_2(z)}$$

This conditional probability does not rely on the parameter $\theta \in \Omega$. Hence $Y = u(X_1, \cdots, X_n)$ is a sufficient statistic for θ.

(proof for necessary condition)

If $Y=u(X_1,\cdots,X_n)$ is a Sufficient Statistic for $\theta\in\Omega$, for all x, y, the conditional probability $P_\theta(X=x|Y=y)$ does not rely on the parameter $\theta\in\Omega$. Also,

$$P_\theta(Y=u(x)) = K_1(u(x),\theta)$$

Thus for all $\theta\in\Omega$ and $x=(x_1,\cdots,x_n)^t$, the following is true:

$$P_\theta(X=x) = P_\theta(X=x, Y=u(x)) = P(X=x|Y=u(x))P_\theta(Y=u(x))$$

$$\therefore \prod_{i=1}^n f(x_i;\theta) = P(X=x|Y=u(x))P_\theta(Y=u(x)) = K_1(u(x),\theta)K_2(x)$$

Ex. 8.2.3 Poisson(θ), $\theta>0$

$$f_X(x;\theta)=\frac{e^{-\lambda}\lambda^x}{x!},\quad \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}=\frac{e^{-n\lambda}\lambda^{n\bar{x}}}{x_1!\cdots x_n!}=\left\{e^{-n\lambda}\lambda^{n\bar{x}}\right\}\left\{x_1!\cdots x_n!\right\}^{-1}$$

$\therefore$ by the factorization theorem, $\sum x_i$ is a SS (Sufficient Statistic) for $\theta>0$

Ex. 8.2.4 Exp(θ), $\theta>0$

$$f(x;\theta)=\frac{1}{\theta}e^{-x/\theta}I_{(0,+\infty)}(x),\quad \prod_{i=1}^n f(x_i;\theta)=\underbrace{\theta^{-n}e^{-\frac{1}{\theta}(x_1+\cdots+x_n)}}_{K_1(x;\theta)}\underbrace{I_{(0<x_{(1)})}}_{K_2(x)}$$

$\therefore$ By the factorization theorem, $\sum x_i$ is a SS for $\theta>0$

Ex. 8.2.5 Gamma(α,β), $\alpha>0$, $\beta>0$

$$f(x;\theta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp\left\{-x/\beta\right\}I_{(0<x)}$$

$$\prod_{i=1}^n f(x_i;\theta)=\left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right]^n\left(\prod_{i=1}^n x_i\right)^{\alpha-1}\exp\left\{-\sum_{i=1}^n x_i/\beta\right\}\prod_{i=1}^n I_{(0<x_i)}$$

Let $K_1(x;\theta)=\left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right]^n\left(\prod_{i=1}^n x_i\right)^{\alpha-1}\exp\left\{-\sum_{i=1}^n x_i/\beta\right\}$, $K_2(x)=\prod_{i=1}^n I_{(0<x_i)}$

$\therefore \prod_{i=1}^n x_i$ and $\sum_{i=1}^n x_i$ are SS for $\alpha>0$, $\beta>0$ $\rightarrow (\alpha,\beta)^t\in(0,\infty)\times(0,\infty)$

〈Theorem 8.2.2. Sufficient Statistic in Multivariate exponential family〉

If the pdf is given as:

$$f(x;\theta)=\exp\left\{\sum_{j=1}^k g_j(\theta)T_j(x)-A(g(\theta))+S(x)\right\},\quad x\in\mathcal{X},\ \theta\in\Omega$$

and the support of the distribution $\mathcal{X}=\{x: f(x;\theta)>0\}$ is invariant of the parameter, let $X_1,\cdots,X_n$ be the random samples from the population. Then,

$$\sum_{i=1}^n T(x_i)=\left(\sum_{i=1}^n T_1(x_i),\cdots,\sum_{i=1}^n T_k(x_i)\right)^t$$ are the SS for $\theta\in\Omega$.

(proof) The joint pdf of random samples can be expressed as:

$$\prod_{i=1}^n f(x_i;\theta)=\exp\left\{\sum_{j=1}^k g_j(\theta)\sum_{i=1}^n T_j(x_i)-nA(g(\theta))\right\}\exp\left\{\sum_{i=1}^n S(x_i)\right\},$$

Hence, from the factorization theorem, $\sum_{i=1}^n T(x_i)=\left(\sum_{i=1}^n T_1(x_i),\cdots,\sum_{i=1}^n T_k(x_i)\right)^t$ is a SS for $\theta\in\Omega$

〈Theorem 8.2.3. one-to-one function of Sufficient Statistics〉

If the population distribution is one of the pdf $f(x;\theta)$, $\theta\in\Omega$, and the Statistic $Y=u(X_1,\cdots,X_n)$, which is a function of random samples $X_1,\cdots,X_n$, is a Sufficient Statistic for $\theta\in\Omega$, the one-to-one function of Y, $W=g(Y)$, is also a Sufficient Statistic.

(proof)

Since $Y=u(X_1,\cdots,X_n)$ is a sufficient statistic for $\theta\in\Omega$, the conditional probability

$$P_\theta\left\{(X_1,\cdots,X_n)^t\in A|Y=y\right\}$$

does not depend on θ for all values of y. However, since $W=g(Y)$ is a one-to-one function, for an arbitrary value w in the range of $W=g(Y)$, $(W=w)$ corresponds to $(Y=g^{-1}(w))$. Hence, the conditional probability of $X=(X_1,\cdots,X_n)^t$ given $W=g(Y)$:

$$P_\theta\left\{(X_1,\cdots,X_n)^t\in A|W=w\right\}=P_\theta\left\{(X_1,\cdots,X_n)^t\in A|Y=g^{-1}(w)\right\}$$

does not depend on $\theta\in\Omega$. i.e., the statistic $W=g(Y)$ is also a SS for $\theta\in\Omega$.

(proof by chatGPT)

Suppose that $T(x)$ is a SS and $S(x)=g(T(x))$ is one-to-one.

① Since $T(x)$ is a SS, by the factorization theorem,

$f(x)=h(x)K(T(x),\theta)$

② Because $g(\cdot)$ is 1-1, $T(x)=g^{-1}(S(x))$

$\therefore f(x)=h(x)K(T(x),\theta)=h(x)K(g^{-1}(S(x)),\theta)$

③ Let $K^*(S(x),\theta)=K(g^{-1}(S(x)),\theta)$

Then $f(x;\theta)=h(x)\cdot K^*(S(x),\theta)$

④ By the factorization theorem $S(x)$ is a SS for $\theta\in\Omega$

Ex. 8.2.6. Normal Distribution case

$f(x;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$

$\prod_{i=1}^{n} f(x_i;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$

$= (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2)\right\}$

$\therefore (\Sigma x_i^2, \Sigma x_i)^t$ are SS for $\theta = (\mu, \sigma^2)^t$

let $g(t_1,t_2) = \left(\frac{1}{n-1}(t_1 - t_2^2),\ \frac{1}{n}t_2\right)^t$ $\quad g(t_1,t_2)$ is a 1-1 function

$g(\Sigma x_i^2, \Sigma x_i) = \left(\frac{1}{n-1}(\Sigma x_i^2 - (\Sigma x_i)^2),\ \frac{1}{n}\Sigma x_i\right)^t$

$= \left(\frac{1}{n-1}\Sigma(x_i - \bar{x})^2,\ \bar{x}\right)^t$ is also a SS for θ

$f(x;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{\sigma^2}\right\} = \exp\left\{\frac{\mu}{\sigma^2}x + \frac{-1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\}$

the support of the distribution is $\mathbb{R}$. Thus, from ⟨theorem 8.2.2⟩,

$Y = \left(\sum_{i=1}^{n} x_i,\ \sum_{i=1}^{n} x_i^2\right)^t$ is a SS for $(\mu,\sigma^2)^t \in (-\infty,+\infty)\times(0,+\infty)$

also, the 1-1 function $\left(\bar{x},\ \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)\right)$ is also a SS for $(\mu,\sigma^2)^t \in (-\infty,+\infty)\times(0,+\infty)$

Ex. Beta(α,1) case

(Approach #1)

$f(x;\theta) = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(1)} x^{\alpha-1}(1-x)^{1-1} I_{(0,1)}(x) = \alpha x^{\alpha-1} I_{(0,1)}(x)$

$\prod_{i=1}^{n} f(x_i;\theta) = \alpha^n \prod_{i=1}^{n}(x_i)^{\alpha-1} I_{(0,1)}(x_i)$

$\therefore \prod_{i=1}^{n} x_i$ is a ss for θ and $\Sigma \log x_i$ and $\overline{\log x}$, a 1-1 transformation of $\prod_{i=1}^{n} x_i$, is are also SS for θ.

(Approach #2)

$f(x;\theta) = \alpha x^{\alpha-1} = \exp\{\log(\alpha) + (\alpha-1)\log x\} I_{(0,1)}(x)$

$\prod_{i=1}^{n} f(x_i;\theta) = \exp\left[n\log(\alpha) + (\alpha-1)\Sigma\log x_i\right]$, support of the distribution does not depend on α. $\therefore$ apply ⟨Thm 8.2.2⟩

$\therefore Y = \sum_{i=1}^{n} \log x_i$ is a SS for α and its 1-1 function $(\overline{\log x}) = \frac{1}{n}\sum_{i=1}^{n}\log x_i$ is also a SS for α

If the support of the population is dependent of the parameter, ⟨Theorem 8.2.2⟩ cannot be applied. In this case, consider the indicator function showing the support of the distribution and utilize the factorization theorem to find the sufficient statistic.

Ex. 8.2.8. Exponential distribution with two parameters[116]

$f(x;\theta) = \frac{1}{\sigma}\exp\left\{\frac{x-\mu}{\sigma}\right\} I_{[\mu,\infty)}(x)$, $\quad -\infty < \mu < +\infty$, $\sigma > 0$

The support of the distribution depends on μ. Hence, ⟨theorem 8.2.2⟩ cannot be applied.

$\prod_{i=1}^{n} f(x_i;\theta) = \sigma^{-n}\exp\left\{\frac{1}{\sigma}\sum_{i=1}^{n}(x_i-\mu)\right\} I_{[\mu,\infty)}(x_{(1)})$

$\therefore (\Sigma x_i,\ x_{(1)})^t$ are SS for $\theta = (\mu,\sigma)^t \in (-\infty,+\infty)\times(0,+\infty)$

116) 이와 같이 두 개의 모수를 갖는 지수분포를 기호로는 Exp(μ,σ)로 나타낸다.

Ex. 8.2.9. $U[\theta_1,\theta_2]$, $-\infty < \theta_1 < \theta_2 < +\infty$

$f(x;\theta) = \frac{1}{\theta_2-\theta_1} I_{[\theta_1,\theta_2]}(x)$, $\prod_{i=1}^{n} f(x_i;\theta) = (\theta_2-\theta_1)^{-n} I_{[\theta_1 \le x_{(1)},\, x_{(n)} \le \theta_2]}$

$\therefore$ By the factorization theorem, $(x_{(1)}, x_{(n)})^t$ is a SS for $\theta = (\theta_1,\theta_2)^t$ $(-\infty < \theta_1 < \theta_2 < +\infty)$

⟨Theorem 8.2.4 MLE and SS⟩

If the population distribution is one of $f(x;\theta)$, $\theta \in \Omega$, and the MLE of θ, $\hat{\theta}^{MLE}(X_1,\cdots,X_n)$ - a function of random samples $X_1,\cdots,X_n$ -, exists uniquely, the MLE is a function of a SS for θ. i.e.,

$$\hat{\theta}^{MLE}(X_1,\cdots,X_n) = \text{function}(S)$$

(proof)

From the definition of MLE and the factorization theorem for the SS $S = u(X_1,\cdots,X_n)$,

$$\hat{\theta}^{MLE}(x_1,\cdots,x_n) = \underset{\theta\in\Omega}{\text{argmax}} \prod_{i=1}^{n} f(x_i;\theta) = \underset{\theta\in\Omega}{\text{argmax}}\ K_1(u(x),\theta)\cdot K_2(x)$$

$$= \underset{\theta\in\Omega}{\text{argmax}}\ K_1(u(x),\theta)$$

$$\therefore Y = \hat{\theta}(X_1,\cdots,X_n) = \text{function}(u(X_1,\cdots,X_n)) = \text{function}(S)$$

From ⟨Theorem 8.2.4⟩, on top of the unique existence of the MLE $Y = \hat{\theta}^{MLE}(X_1,\cdots,X_n)$, if the MLE is a SS for $\theta \in \Omega$, MLE is then a function of other arbitrary SS $S = u(X_1,\cdots,X_n)$. This type of SS is called the "minimal sufficient statistic".

* MLE is unique & a SS ⇒ The # of total SS cannot be lesser than 1. Hence MLE is minimal.

The MLE of models in ⟨Ex. 8.2.2⟩ to ⟨Ex. 8.2.9⟩ are unique and are SS. Hence they are the MSS(Minimal Sufficient Statistic) for θ

☆ ⟨Theorem 8.2.5. Rao-Blackwell Theorem⟩

Suppose that the population model is one of the pdf $f(x;\theta)$, $\theta\in\Omega$ and the statistic $Y=u(X_1,\cdots,X_n)$ is a SS for $\theta\in\Omega$. For $\hat\eta$, an estimator of $\eta=\eta(\theta)$, if we define:

$$\hat\eta^{RB}(Y)=E[\hat\eta(X_1,\cdots,X_n)|Y],$$

$\hat\eta^{RB}$ is an estimator, and the following holds:

$$MSE(\hat\eta^{RB},\theta)=E_\theta[(\hat\eta^{RB}-\eta(\theta))^2]\le E_\theta[(\hat\eta-\eta(\theta))^2]=MSE(\hat\eta,\theta),\quad\forall\theta\in\Omega$$

(proof)

$Y=u(X_1,\cdots,X_n)$ is a SS for $\theta\in\Omega$. Thus, the conditional probability of the random sample $(X_1,\cdots,X_n)^t$ does not depend on the parameter $\theta\in\Omega$ and consequently, the conditional expectation $E_\theta[\hat\eta(X_1,\cdots,X_n)|Y]$ does not depend on $\theta\in\Omega$. Meaning,

$$E_\theta[\hat\eta(X_1,\cdots,X_n)|Y]\overset{\theta\in\Omega}{\equiv}E[\hat\eta(X_1,\cdots,X_n)|Y]$$

and therefore, $\hat\eta^{RB}(Y)=E[\hat\eta(X_1,\cdots,X_n)|Y]$ is a statistic.

Meanwhile from ⟨Theorem 2.3.4⟩ and ⟨Theorem 2.3.6⟩,

$E[E[V|W]]=E[V]$, $Var[V]=E[Var(V|W)]+Var[E[V|W]]$

Here, let $V=\hat\eta$ and $W=Y$. $E[V|W]=E[\hat\eta|Y]=\hat\eta^{RB}$

① $E_\theta[\hat\eta^{RB}]=E_\theta[E[\hat\eta|Y]]=E_\theta(\hat\eta)$

② $Var(\hat\eta)=Var(E[\hat\eta|Y])+E[Var(\hat\eta|Y)]=Var(\hat\eta^{RB})+E[Var(\hat\eta|Y)]$

$\therefore Var(\hat\eta^{RB})\le Var(\hat\eta)$

③ $MSE(\hat\eta,\theta)=E_\theta[(\hat\eta-\eta(\theta))^2]=E_\theta[(\hat\eta-E_\theta[\hat\eta]+E_\theta[\hat\eta]-\eta(\theta))^2]$

$=E_\theta[(\hat\eta-E_\theta[\hat\eta])^2]+E_\theta[(\hat\eta-E_\theta[\hat\eta])(E_\theta[\hat\eta]-\eta(\theta))]+E_\theta[(E_\theta[\hat\eta]-\eta(\theta))^2]$

$=Var(\hat\eta)+E_\theta[(\hat\eta-E_\theta(\hat\eta))](E_\theta(\hat\eta)-\eta(\theta))+(E_\theta[\hat\eta]-\eta(\theta))^2$

$=Var(\hat\eta)+0+(E_\theta[\hat\eta]-\eta(\theta))^2$

$MSE(\hat\eta^{RB},\theta)=Var(\hat\eta^{RB})+(E_\theta[\hat\eta^{RB}]-\eta(\theta))^2$

$=Var(\hat\eta^{RB})+(E_\theta[\hat\eta]-\eta(\theta))^2$ ①

$\le Var(\hat\eta)+(E_\theta[\hat\eta]-\eta(\theta))^2$ ②

$=MSE(\hat\eta,\theta)$ ③

$\therefore MSE(\hat\eta^{RB},\theta)\le MSE(\hat\eta,\theta)\quad\forall\theta\in\Omega$ □

☆ Ex. 8.2.10. $U[0,\theta]$, $\theta>0$ case

$f(x;\theta)=\frac{1}{\theta}I_{(0,\theta)}(x_i)$, $\prod_{i=1}^n f(x_i;\theta)=\theta^{-n}I_{[0\le x_{(1)},x_{(n)}\le\theta]}$

$\therefore X_{(n)}$ is a SS for $\theta\in(0,+\infty)$

① $E_\theta[X_i]=\theta/2$ $\therefore\hat\theta^{MME}=2\sum_{i=1}^n x_i/n=2\bar x$

The MSE of $\hat\theta^{MME}$ is: $MSE(\hat\theta,\theta)=E_\theta[(2\bar x-\theta)^2]=4E_\theta[(\bar x-\theta/2)^2]$

$=4Var(\bar x)=4\cdot\frac{\theta^2}{12n}=\frac{\theta^2}{3n}$

* $X\sim U(a,b)$. $Var(X)=\frac{(b-a)^2}{12}$, $\bar x\sim N(\mu,\frac{Var(X)}{n})$ by CLT.

Let $X_{(1)}<\cdots<X_{(n)}$ denote the order statistics.

$E[2\bar x|X_{(n)}=y]=\frac{2}{n}E[X_{(1)}+\cdots+X_{(n)}|X_{(n)}=y]=\frac{2}{n}E[X_{(1)}+\cdots+X_{(n-1)}|X_{(n)}=y]+\frac{2}{n}y$

$X_{(1)}+\cdots+X_{(n-1)}=X_1'+\cdots+X_{n-1}'$, $X_i'\sim U(0,y)$, $i=(1,\cdots,n-1)$

$E[X_1'+\cdots+X_{n-1}']=n-1\,E[X_i']=\frac{n-1}{2}y$

$\therefore E[2\bar x|X_{(n)}=y]=\frac{2}{n}\cdot\frac{n-1}{2}\cdot y+\frac{2}{n}y=\frac{n-1}{n}y+\frac{2}{n}y=\frac{n+1}{n}y$ ✓✓

[Textbook Approach]

$E[2\bar x|X_{(n)}=y]=\frac{2}{n}E[X_{(1)}+\cdots+X_{(n-1)}+X_{(n)}|X_{(n)}]=\frac{2}{n}\left[\sum_{r=1}^{n-1}E[X_{(r)}|X_{(n)}=y]+y\right]$

The conditional pdf of $X_{(r)}$ given $X_{(n)}=y$ is:

$$pdf_{r|n}(x|y)=\frac{(n-1)!}{(r-1)!(n-r-1)!}\left(\frac{x}{y}\right)^{r-1}\left(1-\frac{x}{y}\right)^{n-r-1}\frac{1}{y}I_{(0,y)}(x)$$

$\therefore E[X_{(r)}|X_{(n)}=y]=\int_0^y x\cdot pdf_{r|n}(x|y)\,dx=\frac{r}{n}y\quad(1\le r\le n-1)$

$\therefore$ if $\hat\theta^{RB}=E[\hat\theta|X_{(n)}]$, $\hat\theta^{RB}=E[2\bar x|X_{(n)}]=\frac{2}{n}\left[\sum_{r=1}^{n-1}\frac{r}{n}X_{(n)}+X_{(n)}\right]=\frac{n+1}{n}X_{(n)}$

$MSE(\hat\theta^{RB},\theta)=E[(\hat\theta^{RB}-\theta)^2]=Var(\hat\theta^{RB})=\left(\frac{n+1}{n}\right)^2Var(X_{(n)})$

$X_{(n)}/\theta\sim Beta(1,n)$ $Var(X_{(n)}/\theta)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=\frac{n}{(1+n)^2(2+n)}$

$\therefore\left(\frac{n+1}{n}\right)^2Var(X_{(n)})=\left(\frac{n+1}{n}\right)^2\cdot\frac{n}{(n+1)^2(n+2)}\cdot\theta^2=\frac{1}{n(n+2)}\theta^2$

[Textbook Approach]

$MSE(\hat\theta^{RB},\theta)=E_\theta\left[\left\{\left(\frac{n+1}{n}\right)X_{(n)}-\theta\right\}^2\right]$

$\therefore MSE(\hat\theta^{RB},\theta)=\int_0^\theta\left[\frac{n+1}{n}x-\theta\right]^2\cdot\frac{1}{\theta^n}nx^{n-1}dx=\frac{\theta^2}{n(n+2)}$

8.3. MINIMUM VARIANCE UNBIASED ESTIMATOR

According to 〈Theorem 8.2.5〉, we know that by taking the conditional expectation of a given estimator w.r.t a sufficient statistic, we can obtain an estimator that improves upon the original one. In particular, when the given estimator is an unbiased estimator of the quantity of interest, and the sufficient statistic additionally satisfies the following property, it is known that this method yields a uniformly minimum variance unbiased estimator (UMVUE)

Definition: Complete Statistic & Complete Sufficient Statistic

Suppose that the population distribution is one of pdf $f(x;\theta), \theta\in\Omega$ and that the following is true for the statistic $Y=u(X_1,\cdots,X_n)$

$$\text{if } E_\theta[g(Y)]=0 \ \forall\theta\in\Omega,\ g(Y)=0 \text{ . i.e., } P_\theta(g(Y)=0)=1 \ \forall\theta\in\Omega$$

then $Y=u(X_1,\cdots,X_n)$ is called the complete statistic for $\theta\in\Omega$.
Also, if $Y=u(X_1,\cdots,X_n)$ is a complete statistic and a sufficient statistic for θ, $Y=u(X_1,\cdots,X_n)$ is called a complete sufficient statistic for θ.

〈Theorem 8.3.1. UMVUE using CSS (Comple Sufficient Statistics)〉

Suppose that the population's pdf is one of the pdf $f(x;\theta), \theta\in\Omega$ and the statistic $Y=u(X_1,\cdots,X_n)$, a function of random samples $X_1,\cdots,X_n$ from this distribution, is a CSS. The following holds for $Y=u(X_1,\cdots,X_n)$

(a) For $\hat\eta_0^{UE}=\hat\eta_0^{UE}(X_1,\cdots,X_n)$, an unbiased estimator of $\eta=\eta(\theta)$, let

$$\hat\eta^{RB}(Y)=E[\hat\eta_0^{UE}(X_1,\cdots,X_n)|Y]$$

Then, $\hat\eta_0^{RB}=\hat\eta_0^{RB}(Y)$ is an UMVUE of $\eta=\eta(\theta)$

(b) If $\delta(Y)$, a function of the statistic $Y=u(X_1,\cdots,X_n)$, is an unbiased estimator of $\eta=\eta(\theta)$, $\delta(Y)$ is a UMVUE of $\eta=\eta(\theta)$

(proof)

(a) (i) From the properties of conditional expectation:

$$E_\theta[\hat\eta_0^{RB}(Y)]=E_\theta[E(\hat\eta_0^{UE}|Y)]=E_\theta[\hat\eta^{UE}]\overset{\theta\in\Omega}{\equiv}\eta(\theta)$$

notice how there is no θ

(ii) Let $\hat\eta_1^{UE}=\hat\eta_1^{UE}(X_1,\cdots,X_n)$ be an arbitrary unbiased estimator of $\eta=\eta(\theta)$.

Let $\hat\eta_1^{RB}(Y)=E[\hat\eta_1^{UE}(X_1,\cdots,X_n)|Y]$

We know from 〈Theorem 8.2.5〉 that

① $E_\theta[\hat\eta_1^{RB}(Y)]=E_\theta\hat\eta^{UE}\overset{\theta\in\Omega}{\equiv}\eta(\theta)$

② $\mathrm{Var}(\hat\eta_1^{RB})\le MSE(\hat\eta_1^{RB},\theta)\le MSE(\hat\eta_1^{UE},\theta)=\mathrm{Var}_\theta(\hat\eta_1^{UE}) \ \forall\theta\in\Omega$

Hence, $\hat\eta_0^{RB}(Y)$ and $\hat\eta_1^{RB}(Y)$ are both unbiased estimators.

$$\Rightarrow E_\theta[\hat\eta_0^{RB}(Y)-\hat\eta_1^{RB}(Y)]\overset{\theta\in\Omega}{\equiv}\eta(\theta)-\eta(\theta)=0$$

and since Y is a CSS for $\theta\in\Omega$, $\underbrace{P_\theta[\hat\eta_0^{RB}(Y)-\hat\eta_1^{RB}(Y)=0]\overset{\theta\in\Omega}{\equiv}1}_{*\text{ CSS의 정의}}$

$$\therefore \mathrm{Var}_\theta(\hat\eta_0^{RB})=\mathrm{Var}_\theta(\hat\eta_1^{RB})\le\mathrm{Var}_\theta(\hat\eta_1^{UE}) \ \forall\theta\in\Omega$$

Therefore, from (i) and (ii), $\hat\eta_0^{RB}=\hat\eta_0^{RB}(Y)$ is an UMVUE of $\eta=\eta(\theta)$.

(b) in (a)'s proof, substitute $\hat\eta_0^{RB}(\theta)$ with $\delta(Y)$ and we see that (b) is true

* Q): why use $\hat\eta_0^{RB}$ and $\hat\eta_1^{RB}$ separately? Aren't they the same?
A): There can be multiple unbiased estimators for a parameter θ. Since there is no guarantee that MME or MLE are unique. So, we choose one UE and find its conditional expectation for Y, which is $\hat\eta_0^{RB}$. Then, we show that other conditional expectation of $\hat\eta_1^{UE}$ - arbitrary - given Y has the same minimum variance, using the completeness property. (In fact, the completeness property shows that $\hat\eta_0^{RB}$ and $\hat\eta_1^{RB}$ are equivalent) This essentially proves that whichever unbiased estimator you use, the conditional expectation of that UE given Y, outputs the same statistic which has the minimum variance among other UEs.

Ex. 8.3.1 CSS in independent Bernoulli trials.

$X\sim\text{Bernoulli}(\theta)$, $0<\theta<1$ $\qquad f(x;\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$

$$\prod_{i=1}^n f(x_i;\theta)=\theta^{\sum x_i}(1-\theta)^{n(1-\bar x)}$$

$\therefore$ By the factorization theorem, $\sum X_i$ is a SS

$$\sum X_i\sim B(n,p) \ \therefore E_p[g(\textstyle\sum X_i)]=\sum_{x=0}^n g(x)\cdot\binom{n}{x}p^x(1-p)^{n-x}=0$$

$$\Leftrightarrow g(x)=0$$

$\therefore \sum X_i$ is a complete sufficient statistic

[Textbook Approach]

$$E_\theta[g(Y)]\overset{\theta\in(0,1)}{\equiv}0 \Rightarrow \sum_{y=0}^n\binom{n}{y}\theta^y(1-\theta)^{n-y}\overset{\theta\in(0,1)}{\equiv}0$$

$$\Rightarrow \sum_{y=0}^n g(y)\cdot\binom{n}{y}\lambda^y\overset{\lambda>0}{\equiv}0 \quad \left(\lambda=\frac{\theta}{1-\theta}\right)$$

$$\Rightarrow g(y)\binom{n}{y}\overset{y}{\equiv}0 \quad (\because \text{uniqueness of polynomial expansion})$$

$$\Rightarrow g(y)=0$$

$E[\bar{X}] = E[\Sigma X_i/n] = \theta \quad \therefore \hat{\theta}^{UE} = \bar{X}$

$\hat{\theta}^{RB} = E_\theta[\hat{\theta}^{UE} \mid \Sigma X_i = y] = y/n$

$\therefore Y/n = \Sigma X_i/n$ is an UMVUE of θ ✓

Ex. 8.3.2. CSS of $\theta>0$ in Poisson(θ)

$$f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad \prod_{i=1}^n f(x_i;\theta) = \frac{e^{-n\theta}\theta^{\Sigma x_i}}{\prod_{i=1}^n x_i!}$$

SS : $(\Sigma X_i) = Y$

$\Sigma X_i \sim \text{Poisson}(n\theta) \quad \therefore E_\theta[g(Y)] \overset{\theta>0}{\equiv} 0 \Rightarrow \sum_{y=0}^{\infty} g(y) \frac{e^{-n\theta}(n\theta)^y}{y!} \overset{\theta>0}{\equiv} 0$

$$\Rightarrow \sum_{y=0}^{\infty} g(y) \frac{e^{-\lambda}\lambda^y}{y!} \overset{\lambda>0}{\equiv} 0 \quad (\lambda = n\theta)$$

$$\Rightarrow \sum_{y=0}^{\infty} g(y) \frac{\lambda^y}{y!} \overset{\lambda>0}{\equiv} 0 \quad \Rightarrow g(y)\frac{\lambda^y}{y!} \equiv 0 \quad (\because \text{uniqueness of power series})$$

$$\Rightarrow g(y) = 0$$

$\therefore \Sigma X_i = Y$ is a CSS

$\hat{\theta}^{UE} = \bar{X}$, $\hat{\theta}^{RB} = E_\theta[\Sigma X_i/n \mid \Sigma X_i = y] = \frac{Y}{n} = \frac{\Sigma X_i}{n}$ is the UMVUE of θ □

Ex. 8.3.3 CSS of $\theta>0$ in Exp(θ)

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}, \quad \prod_{i=1}^n f(x_i;\theta) = \theta^{-n}\exp\left\{-\frac{1}{\theta}\Sigma X_i\right\}$$

$\therefore$ SS of θ is ΣX_i

$X \sim \text{Exp}(\theta)$, $X/\theta \sim \text{Exp}(1) \overset{d}{\equiv} \text{Gamma}(1,1)$

$\therefore \Sigma X_i \sim \text{Gamma}(n,\theta)$. Let $Y = \Sigma X_i$

$$E_\theta[g(Y)] = \int_0^\infty g(y)\cdot\frac{1}{\Gamma(n)\theta^n} y^{n-1}\exp\left\{-y/\theta\right\} dy \overset{\theta>0}{\equiv} 0$$

$$= \int_0^\infty g(y)\cdot y^{n-1}\exp\left\{-y/\theta\right\} dy \overset{\theta>0}{\equiv} 0$$

$$\Rightarrow \int_0^\infty g(y)\, y^{n-1} e^{-\lambda y} dy \overset{\lambda>0}{\equiv} 0 \quad (\lambda = \theta^{-1})$$

$$\Rightarrow g(y)\, y^{n-1} \overset{a.e.}{\equiv} 0 \quad (\because \text{uniqueness of Laplace transformation})$$

$$\Rightarrow g(Y) = 0$$

☆☆ a.e.) 정의역이 양수 전체 집합이고 적분가능한 함수 f에 대하여

$$Lf(\lambda) = \int_0^\infty f(x)e^{-\lambda x}dx$$

을 f의 라플라스(Laplace) 변환이라고 하며, 라플라스변환은 함수 f와 일대일로 대응하는 것이 알려져 있다.

$\therefore \Sigma X_i$ is a CSS of θ

$\hat{\theta}^{UE} = \bar{X}$, $\hat{\theta}^{RB} = E_\theta[\bar{X} \mid \Sigma X_i = y] = \frac{y}{n} = \frac{\Sigma X_i}{n}$ is a UMVUE of θ.

〈Ex. 8.3.1〉 through 〈Ex. 8.3.3〉 are all cases of one-parameter exponential families, and it is known in general that, in case of multi-parameter exponential families, the sufficient statistic given in 〈Thm. 8.2.2〉 is a complete statistic under appropriate conditions.

Since this property of multi-parameter exponential families relies on the uniqueness of the Laplace transform and goes beyond the level of this book, the proof is omitted and only the result is presented below.

〈Theorem 8.3.2. CSS in multi-parameter exponential family〉

If the pdf is given as:

$$f(x;\eta) = \exp\left\{\sum_{j=1}^k \eta_j T_j(x) - A(\eta) + S(x)\right\}, \quad x \in \mathcal{X}, \ \eta = (\eta_1, \cdots, \eta_k)^t \in N$$

and if random samples $X_1, \cdots, X_n$ were drawn from the population satisfying conditions (i), (ii) and (iii),

$$\sum_{i=1}^n T(X_i) = \left(\sum_{i=1}^n T_1(X_i), \cdots, \sum_{i=1}^n T_k(X_i)\right)^t$$

are complete sufficient statistics for $\eta \in N$.

(i) The support of the distribution, $\mathcal{X}$, does not depend on the parameters. $\mathcal{X} := \{x : f(x;\eta) > 0\}$

(ii) The parameter space contains an open-box $(a_1,b_1)\times\cdots\times(a_k,b_k)$ in $\mathbb{R}^k$

(iii) For any non-zero arbitrary constants $c_1, \cdots, c_k$, $c^t T(X) = c_1T_1(X) + \cdots + c_kT_k(X)$ is not a constant.

In 〈Theorem 8.3.2〉, the parameter η is defined as:

$$\eta = g(\theta) = (g_1(\theta), \cdots, g_k(\theta))^t, \quad \theta \in \Omega$$

, a one-to-one transformation of θ. Meaning, the pdf in 〈Theorem 8.3.2〉 is expressed as:

$$pdf(x;\theta) = \exp\left\{\sum_{j=1}^k g_j(\theta)T_j(x) - A(g(\theta)) + S(x)\right\}, x \in \mathcal{X}, \theta \in \Omega$$

Hence condition (ii) implies that there exists an open-box $(a_1,b_1)\times\cdots\times(a_k,b_k)$ such that:

$$\left\{(g_1(\theta), \cdots, g_k(\theta))^t : \theta \in \Omega\right\} \supseteq (a_1,b_1)\times\cdots\times(a_k,b_k)$$

Ex. 8.3.4. CSS from a random sample of a normal population.

$X \sim N(\mu, \sigma^2)$, $f(x;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$

$= \exp\left\{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left\{x_i^2 - 2x_i\cdot\mu + \mu^2\right\}\right.$

$\prod_{i=1}^n f(x_i;\theta) = \exp\left\{-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right\}\right.$

(i) The support of the distribution $\mathcal{X} := \{x : f(x;\theta) > 0\} = (-\infty, +\infty)$ does not depend on the parameter $\theta = (\mu, \sigma^2)^t$

(ii) The parameter space $\theta = (\mu, \sigma^2)^t \in (-\infty, +\infty) \times (0, +\infty)$ contains an open box $(a_1, b_1) \times (a_2, b_2)$.

(iii) $\mathrm{Var}(T(x)) \neq 0$ (x) $C^t T(x)$ is not a constant $\forall\, C \in \mathbb{R}^K$, $C_1, \cdots, C_K \neq 0$ $C = (C_1, \cdots, C_K)^t$

↳ 1차함수 지수족의 경우

$\therefore$ we can apply ⟨Theorem 8.3.2⟩. $\therefore (\sum X_i, \sum X_i^2)^t$ are CSS

The one-to-one transformation $Y = (\bar{X}, \frac{1}{n-1}\sum(X_i - \bar{X})^2)^t$ are also CSS.

$\hat{\theta}^{UE} = (\bar{X}, \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2)$ $\therefore \hat{\theta}^{RB} = E_\theta[\hat{\theta}^{UE} | Y] = (\bar{X}, \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2)$

are UMVUE of θ.

☆ Ex 8.3.5. In the case of a normal distribution with a restricted parameter space.

$X \sim N(\mu, \mu^2)$

$f(x;\theta) = (2\pi\mu^2)^{-1/2} \exp\left\{-\frac{1}{2\mu^2}(x-\mu)^2\right\}$

$\prod_{i=1}^n f(x_i;\theta) = (2\pi\mu^2)^{-n/2} \exp\left\{-\frac{1}{2\mu^2}\sum_{i=1}^n x_i^2 - 2\mu x_i + \mu^2\right\}$

$\therefore$ By the factorization theorem $(\sum X_i, \sum X_i^2)^t$ are SS.

$\Rightarrow$ one-to-one transformation $(\bar{X}, \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2)^t$ are also SS.

$\bar{X} \sim N(\mu, \frac{\mu^2}{n})$, $\sum(X_i - \bar{X})^2/\mu^2 \sim \chi^2(n-1) \overset{d}{=} \mathrm{Gamma}(\frac{n-1}{2}, 2)$

$\therefore \sum(X_i - \bar{X}) \sim \mathrm{Gamma}(\frac{n-1}{2}, 2\mu^2)$

Let $Y = (Y_1, Y_2)^t = (\bar{X}, \frac{1}{n-1}\sum(X_i - \bar{X})^2)^t$

$E_\theta[Y_1] = \mu_1$, $E_\theta[Y_2] = \mu^2$

$E_\theta[(Y_1 - \mu)^2/(\mu^2/n)] = 1$

$\therefore E_\theta[(Y_1 - \mu)^2] = \mu^2/n \Rightarrow E_\theta[Y_1^2 - 2Y_1\mu + \mu^2] = \mu^2/n$

$\Rightarrow E_\theta[Y_1^2] = \mu^2/n - \mu^2 + 2\mu^2 = (\frac{n+1}{n})\mu^2$

$\therefore E_\theta[Y_2 - \frac{n}{n+1}Y_1] = 0$ $\therefore \exists\, g(y) = (y_2 - \frac{n}{n+1}y_1)$ such that $E_\theta[g(y)] = 0$

$\therefore Y = (\bar{X}, \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2)$ is NOT complete

$\bar{X}$ is the UMVUE of μ

Ex. 8.3.6 CSS from a random sample from $U[0,\theta]$

$f(x;\theta) = \frac{1}{\theta} I_{[0,\theta]}(x)$

The support of the distribution $\mathcal{X} := \{x : 0 < x < \theta,\ x \in \mathbb{R}\}$ depends on the parameter θ. Hence ⟨Theorem 8.3.2⟩ is not applicable.

$\prod_{i=1}^n f(x_i;\theta) = \theta^{-n} I_{[0 < x_{(1)}, x_{(n)} < \theta]}$

$\therefore$ By the factorization theorem, $\max_{1\le i\le n} x_i = x_{(n)}$ is a SS

$pdf(y;\theta) = \frac{n}{\theta^n} y^{n-1} I_{[0,\theta]}(y)$

$E_\theta[g(Y)] \overset{\theta>0}{\equiv} 0 \Rightarrow \int_0^\theta g(y)\cdot\frac{n}{\theta^n} y^{n-1} \overset{\theta>0}{\equiv} 0$

$\Rightarrow \frac{n}{\theta^n}\int_0^\theta g(y)\cdot y^{n-1} \overset{\theta>0}{\equiv} 0$

$\Rightarrow \frac{\partial}{\partial y}\left\{\frac{n}{\theta^n}\int_0^\theta g(y)\cdot y^{n-1} dy\right\} \overset{\theta>0}{\equiv} \frac{d}{dy} 0 = 0$

$\Rightarrow \frac{n}{\theta^n}\int_0^\theta \frac{d}{dy}(g(y)\cdot y^{n-1}) dy \overset{\theta>0}{\equiv} 0$

$\Rightarrow \frac{n}{\theta^n}\left[g(y)\cdot y^{n-1}\right]_0^\theta = \frac{n}{\theta^n} g(\theta)\theta^{n-1} \overset{\theta>0}{\equiv} 0$

$\Rightarrow g(\theta)\theta^{n-1} \overset{\theta>0}{\equiv} 0 \Rightarrow g(Y) = 0$

Hence $X_{(n)}$ is a CSS for θ

$X_{(n)}/\theta \sim \mathrm{Beta}(n, 1)$, $E[X_{(n)}/\theta] = \frac{n}{1+n}$ $\therefore E[X_{(n)}] = \frac{n}{1+n}\theta$

$\therefore \hat{\theta}^{RB} = E_\theta[\hat{\theta}^{UE} | X_{(n)} = y] = \frac{1+n}{n} y$ is the UMVUE of θ

Ex. 8.3.7 Unbiased estimation of population variance in Bernoulli distribution.

Let $X_1, \cdots, X_n$ $(n \ge 2)$ denote the iid random samples from the Bernoulli trial Bernoulli(θ) $0 < \theta < 1$. We know from ⟨Ex. 8.3.1⟩ that $Y = X_1 + \cdots + X_n$ is a complete sufficient statistic for θ. Let us think about the biased estimation of $\eta = \theta(1-\theta)$.

$$E_\theta[X_1(1-X_2)] = E_\theta[X_1]E_\theta[1-X_2] = \theta(1-\theta)$$

Therefore, $\hat{\eta}_0^{UE} = X_1(1-X_2)$ is an unbiased estimator for $\eta = \theta(1-\theta)$ and from ⟨Theorem 8.3.1⟩,

$$\hat{\eta}_0^{RB}(Y) = E[X_1(1-X_2)|Y]$$

is a UMVUE for $\eta = \theta(1-\theta)$. Meanwhile, the specific shape of $\hat{\eta}_0^{RB}(Y)$ can be derived as:

$$E[X_1(1-X_2)|Y=y] = P(X_1=1, X_2=0 \,|\, X_1+\cdots+X_n = y)$$

$$= \frac{P(X_1=1, X_2=0, X_3+\cdots+X_n = y-1)}{P(X_1+\cdots+X_n = y)}$$

$$= \binom{n-2}{n-1}/\binom{n}{y} = y(n-y)/n(n-1)$$

Let the sample ratio (표본비율) $\hat{\theta} = \frac{Y}{n}$.

$$\hat{\eta}_0^{RB} = \frac{n}{n-1}\hat{\theta}(1-\hat{\theta})$$

☆ Ex. 8.3.8. Unbiased estimation of the function of population mean in a Poisson distribution

Let $X_1, \cdots, X_n$ be iid random samples from Poisson(θ). We know from ⟨Ex. 8.3.2⟩ that $Y = X_1 + \cdots + X_n$ is a complete sufficient statistic for $\theta > 0$. Let us think about the unbiased estimation of

$$\eta = e^{-2\theta}$$

using ⟨Theorem 8.3.1⟩.

From ⟨Theorem 8.3.1⟩-(b), if there exists a function of Y, g(Y) such that

$$E_\theta[\delta(Y)] \overset{\theta>0}{\equiv} e^{-2\theta}$$

$\delta(Y)$ is a UMVUE of $\eta = e^{-2\theta}$

$$Y = X_1 + \cdots + X_n \sim \text{Poisson}(n\theta)$$

If we expand the identity(항등식) above using polynomial expansion:

$$\sum_{y=0}^{\infty} \delta(y) \frac{e^{-n\theta}(n\theta)^y}{y!} \overset{\theta>0}{\equiv} e^{-2\theta}$$

$$\Rightarrow \sum_{y=0}^{\infty} \delta(y) \frac{(n\theta)^y}{y!} \overset{\theta>0}{\equiv} e^{(n-2)\theta}$$

$$\Rightarrow \sum_{y=0}^{\infty} \delta(y) \frac{\lambda^y}{y!} \overset{\theta>0}{\equiv} e^{(\lambda - \frac{2}{n}\lambda)} \equiv e^{(1-\frac{2}{n})\lambda} \equiv \sum_{y=0}^{\infty} \frac{\{(1-\frac{2}{n})\lambda\}^y}{y!}$$

$$\therefore \delta(y) = (1-\tfrac{2}{n})^y$$

Therefore, $\hat{\eta}_n^{UMVUE} = (1-2/n)^{n\bar{X}}$ is the UMVUE for $\eta = e^{-2\theta}$. If the sample size n is large,

$$\hat{\eta}_n^{UMVUE} = (1-\tfrac{2}{n})^{n\bar{X}} \simeq e^{-2\bar{X}} = \hat{\eta}_n^{MLE}$$

and this estimator has the same limiting distribution as the MLE of $\eta = e^{-2\theta}$. Meanwhile, if $n=1$, $\hat{\eta}_1^{UMVUE} = (-1)^{X_1}$ which is a bizarre shape. This tells us that blindly insisting on unbiased estimation may sometimes lead to peculiar results.

☆☆ Ex. 8.3.9 Unbiased estimation of reliability in normal distribution.

Let $X_1, \cdots, X_n$ ($n \geq 2$) be iid random samples from $N(\theta, 1)$, $-\infty < \theta < +\infty$. From ⟨Theorem 8.3.2⟩ we know that the sample mean $\bar{X}$ is a complete sufficient statistic for θ. Now, let us think about the unbiased estimation of "reliability at a" (a에서의 신뢰도):

$$\eta = P_\theta(X_1 > a)$$

using ⟨Theorem 8.3.1⟩.

$$\eta = P_\theta(X_1 > a) = E_\theta I_{(a,\infty)}(X_1)$$

Thus, $\hat{\eta}_0^{UE} = I_{(a,\infty)}(X_1)$ is an unbiased estimator, and from ⟨Theorem 8.3.1⟩,

$$\hat{\eta}_0^{RB}(\bar{x}) = E[I_{(a,+\infty)}(X_1) \mid \bar{x}] = P(X_1 > a \mid \bar{x})$$

is the UMVUE of $\eta = P_\theta(X_1 > a)$. Meanwhile, we can find $\hat{\eta}_0^{RB}(\bar{x})$ using the independence of $X_1 - \bar{X}$ and $\bar{X}$:

$$P(X_1 > a \mid \bar{x} = y) = P(X_1 - \bar{X} > a - y \mid \bar{x} = y)$$
$$= P(X_1 - \bar{X} > a - y) \quad (\because X_1 - \bar{X} \text{ and } \bar{X} \text{ are independent}^{*})$$
$$= P\left\{\sqrt{\tfrac{n-1}{n}} \cdot z > a - y\right\}, \; z \sim N(0,1) \quad (\because X_1 - \bar{X} \sim N(0, \tfrac{n-1}{n})^{**})$$

Let Φ denote the cdf of a standard normal distribution.

$$P(X_1 > a \mid \bar{x} = y) = 1 - \Phi\left(\sqrt{\tfrac{n}{n-1}}(a-y)\right)$$

$$\therefore \hat{\eta}_0^{RB} = 1 - \Phi\left(\sqrt{\tfrac{n}{n-1}}(a - \bar{x})\right)$$

* $X_1 - \bar{X} = \dfrac{(n-1)X_1 - (X_2 + \cdots + X_n)}{n}$, $\bar{X} = \dfrac{(X_1 + \cdots + X_n)}{n}$

$$E[(X_1 - \bar{x})\bar{x}] = \frac{1}{n^2}\left\{(n-1)E[X_1^2] - E[X_2^2] - \cdots - E[X_n^2]\right\} = 0$$
$$E[X_1 - \bar{x}]E[\bar{x}] = (E[X_1] - E[\bar{x}]) \cdot E[\bar{x}] = (\mu - \mu)\cdot\mu = 0$$
$$\therefore E[(X_1 - \bar{x})\bar{x}] = E[X_1 - \bar{x}]E[\bar{x}]$$

** $X_1 - \bar{x} = \dfrac{(n-1)}{n}X_1 - \dfrac{1}{n}(X_2 + \cdots + X_n)$

$$\text{Var}(X_1 - \bar{X}) = \left(\frac{n-1}{n}\right)^2 \text{Var}(X_1) + \frac{n-1}{n^2}\text{Var}(X_1)$$
$$= \frac{(n-1)^2 + (n-1)}{n^2} \cdot 1 = \frac{n(n-1)}{n^2} = \frac{n-1}{n}$$

⟨Theorem 8.3.3. Independence between CSS and ancillary Statistics.⟩ [19]

Suppose that the population pdf is one of $f(x;\theta)$, $\theta\in\Omega$. If the statistic $Z=v(X_1,\cdots,X_n)$, a function of random samples $X_1,\cdots,X_n$, does not depend on the parameter $\theta\in\Omega$, this statistic is independent of the complete sufficient statistic for $\theta\in\Omega$, $Y=u(X_1,\cdots,X_n)$.

(proof)

$Y=u(X_1,\cdots,X_n)$ is a CSS for $\theta\in\Omega$. Thus, the conditional probability given Y:

$$P(Z\in B|Y)=P(v(X_1,\cdots,X_n)\in B|Y) \text{ does not depend on } \theta\in\Omega.$$

Also, from the property of conditional expectation,

$$E_\theta[P(Z\in B|Y)]=E_\theta[E(I_B(Z)|Y)]=E_\theta I_B(Z)=P_\theta(Z\in B)\quad\forall\theta\in\Omega$$

$\because$ independence of θ; $P(Z\in B|Y)=E[I_B(Z)|Y]$; $E[E[X|Y]]=E[X]$

And the distribution of $Z=v(X_1,\cdots,X_n)$ does not depend on θ. Thus:

$$P_\theta(Z\in B)=P(Z\in B)\quad\forall\theta\in\Omega$$

$$\therefore \underbrace{E_\theta[\underbrace{P(Z\in B|Y)}_{\text{constant}}-P(Z\in B)]}_{g(Y)}\overset{\theta\in\Omega}{\equiv}0$$

However, $Y=u(X_1,\cdots,X_n)$ is a complete statistic $\theta\in\Omega$, thus:

$$E_\theta[g(Y)]\overset{\theta\in\Omega}{\equiv}0 \Leftrightarrow g(Y)=0$$

$$P(Z\in B|Y)-P(Z\in B)=0 \Leftrightarrow P(Z\in B|Y)=P(Z\in B)$$

Thus, Z and Y are independent.

Q) $P(Z\in B|Y)$ (r.v.; $Z\in B$: X, number) $=E_\theta[P(Z\in B|Y)]=E_\theta[E[I_B(Z)|Y]]=E_\theta[I_B(Z)]$

$=P_\theta[Z\in B]=P[Z\in B]$

$\therefore P(Z\in B|Y)=P(Z\in B)$?

⭐ Summary Table

Distribution	Parameter	Complete Sufficient Statistic	Common Ancillary Statistic
Bernoulli	p	$\sum X_i$	None (non-trivial)
Binomial	p	X	None
Geometric	p	$\sum X_i$	None
Negative Binomial	p	$\sum X_i$	None
Exponential	λ	$\sum X_i$	$X_i/\sum X_i$
Gamma (shape known)	β	$\sum X_i$	$X_i/\sum X_i$
Chi-square	–	$\sum X_i$	$X_i/\sum X_i$
Normal (σ known)	μ	$\bar{X}$	Sample variance / residuals
Normal (σ unknown)	μ,σ²	$(\bar{X},S^2)$	Standardized residual vector

Ex. 8.3.10. Unbiased estimation of reliability in exponential distribution

Let $X_1,\cdots,X_n$ be iid random samples from $\text{Exp}(\theta)$, $\theta>0$. From ⟨Ex. 8.3.3⟩ we know that $Y=X_1+\cdots+X_n$ is a CSS for $\theta>0$. Now, we will take a look into the unbiased estimation of "reliability at a" using ⟨Theorem 8.3.3⟩.

$$\eta=P_\theta(X_1>a)=e^{-a/\theta}$$ (reliability at a) — a에서의 신뢰도

$$\eta=P_\theta(X_1>a)=E_\theta[I_{(a,+\infty)}(X_1)]$$

Thus, $\hat{\eta}_0^{UE}=I_{(a,\infty)}(X_1)$ is an unbiased estimator of $\eta=P(X_1>a)$ and from ⟨Theorem 8.3.1⟩,

$$\hat{\eta}_0^{RB}(Y)=E_\theta[I_{(a,\infty)}(X_1)|Y]=P(X_1>a|Y)$$

is the UMVUE of $\eta=P_\theta(X_1>a)$

Meanwhile, from the representational definition of exponential distribution,

$$(X_i)_{1\le i\le n}\overset{d}{\equiv}(\theta Z_i)_{1\le i\le n},\quad Z_i\overset{iid}{\sim}\text{Exp}(1)\ (i=1,2,\cdots,n)$$

$$\therefore \frac{X_1}{X_1+\cdots+X_n}\overset{d}{\equiv}\frac{Z_1}{Z_1+\cdots+Z_n},\quad Z_i\overset{iid}{\sim}\text{Exp}(1)\ (i=1,\cdots,n)$$

Therefore, $X_1/\sum_{i=1}^n X_i$ is an ancillary statistic for $\theta>0$ and by ⟨Theorem 8.3.3⟩, it is independent to $Y=\sum_{i=1}^n X_i$. We can find $\hat{\eta}_0^{RB}(Y)$ using this independence.

$$\begin{aligned}&P(X_1>a|Y=y)\\&=P(X_1/\textstyle\sum X_i>a/y\,|\,Y=y)\\&=P(X_1/\textstyle\sum_{i=1}^n X_i>a/y)\quad[\because X_1/\sum_{i=1}^n X_i \text{ and } Y=\sum_{i=1}^n X_i \text{ are independent}]\\&=P(Z>a/y),\ Z\sim\text{Beta}(1,n-1)\quad[\because X_1/\textstyle\sum_{i=1}^n X_i\sim\text{Beta}(1,n-1)]\end{aligned}$$

$$\therefore P(X_1>a|Y=y)=\int_{a/y}^1\frac{\Gamma(n)}{\Gamma(1)\Gamma(n-1)}z^{1-1}(1-z)^{n-1-1}dz\,I_{(a/y<1)}$$

$$=(1-a/y)^{n-1}I_{(a,\infty)}(y)$$

Hence, the UMVUE of $\eta=P_\theta(X_1>a)=e^{-a/\theta}$ is

$$\hat{\eta}^{RB}=\left(1-\frac{a/\bar{X}}{n}\right)^{n-1}I_{(a/n,+\infty)}(\bar{X})$$

If sample size n is large,

$$\hat{\eta}_n^{UMVUE}=\left(1-\frac{a/\bar{X}}{n}\right)^{n-1}I_{(a/n,\infty)}(\bar{X})\simeq e^{-a/\bar{X}}=\hat{\eta}_n^{MLE}$$

and this estimator has the same limiting distribution as the MLE of $\eta=e^{-a/\theta}$.

8.4. Asymptotic Comparison of Estimators

When comparing estimators that possess the property of consistency, meaning that the estimator approaches the true value of the parameter as the sample size increases, a smaller dispersion in the limiting distribution indicates a smaller estimation error. Therefore, the variance of the limiting distribution is commonly used as a criterion for comparing consistent estimators.[20] In this section, we examine the comparison of estimators using the variance of the limiting distribution.

[20] 크기 n인 표본에 기초한 추정량의 평균제곱오차는

$$MSE(\hat{\eta}_n, \theta) = E_\theta[(\hat{\eta}_n - \eta(\theta))^2] = Var_\theta(\hat{\eta}_n) + [E_\theta(\hat{\eta}_n) - \eta(\theta)]^2$$

와 같이 추정량의 분산 $Var(\hat{\eta}_n(\theta))$과 치우침 또는 편의(bias)라고 불리우는

$$Bias_\theta(\hat{\eta}_n) = E_\theta(\hat{\eta}_n) - \eta(\theta)$$

의 제곱의 합으로 주어진다. 그런데 많은 일치추정량에 대하여 편의의 제곱 $[Bias_\theta(\hat{\eta}_n)]^2$은 n^{-2}의 속도로 0에 가까워지고 분산 $Var_\theta(\hat{\eta}_n)$은 n^{-1}의 속도로 0에 가까워지므로 표본 크기가 커질 때 추정량을 비교하는 경우에는 극한분포의 분산을 비교 기준으로 하는 것이다.

Ex. 8.4.1 Comparison of Estimators in Logistic distribution

In a Logistic distribution $L(\theta, 1)$, θ is both the median and the mean. Therefore, the estimator for θ can be: Sample median[21] $\hat{\theta}_n = X_{[(n+1)/2]}$, Sample mean $\hat{\theta}_n^{mean} = \bar{X}$ or the MLE $\hat{\theta}_n^{MLE}$ that we have seen in ⟨Ex. 6.4.2⟩. However, from the pdf of $L(\theta,1)$ in ⟨Ex. 4.1.5⟩ and the limiting distribution or sample quantile in ⟨Ex. 5.3.8⟩,

$$\sqrt{n}(\hat{\theta}_n^{med} - \theta) \xrightarrow[n\to\infty]{d} N(0, \alpha(1-\alpha)/[f(F^{-1}(\alpha))]^2) \overset{d}{=} N(0,4)$$

$* \alpha = 0.5,\ F^{-1}(\alpha) = \theta,\ f(\theta) = \frac{1}{4}$

From the CLT and the variance of Logistic distribution given in ⟨Ex. 5.4.3⟩,

$$\sqrt{n}(\hat{\theta}_n^{mean} - \theta) \xrightarrow[n\to\infty]{d} N(0, \pi^2/3)$$

Also, from the limiting distribution of MLE given in ⟨Ex. 6.4.4⟩,

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta) \xrightarrow[n\to\infty]{d} N(0,3)$$

Therefore, as the sample size increases, the order of efficient statistics is: MLE, Sample mean, and Sample median. (descending order: best ↔ worst)

[21] 여기에서 $[(n+1)/2]$는 $(n+1)/2$ 이하의 가장 큰 수를 뜻한다. 즉 표본크기가 큰 경우에 표본중앙값은 $n = 2m$ 이면 $X_{(m)}$, $n = 2m+1$ 이면 $X_{(m+1)}$로 하게로 한다.

Definition: Asymptotic Relative Efficiency

For two estimators $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$,

$$\sqrt{n}(\hat{\theta}_n^i - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma_i^2(\theta)) \quad (i = 1, 2)$$

if the asymptotic normality holds, the ratio of the reciprocal of asymptotic variance $\sigma_1^{-2}(\theta) / \sigma_2^{-2}(\theta)$ is called the ARE (Asymptotic Relative Efficiency) ARE of θ_1 for θ_2* and is denoted as:

$$ARE\left[\{\hat{\theta}_n^1\}, \{\hat{\theta}_n^2\}\right] = \sigma_1^{-2}(\theta) / \sigma_2^{-2}(\theta)$$

(* 추정량 $\hat{\theta}_n^1$의 $\hat{\theta}_n^2$에 대한 점근상대효율성)

An increase in the value of ARE means that, as the sample size grows, the precision of the estimation becomes relatively higher. Therefore, an estimation method with relatively better precision can satisfy the required bound on estimation error even with a smaller sample size.

Ex. 8.4.2. Bound on estimation error and ARE in Logistic distribution

In a Logistic distribution $L(\theta,1)$, $-\infty < \theta < +\infty$, the 95% asymptotic confidence interval of θ using sample median is:

$$\theta \in \left[\hat{\theta}_n^{med} - 1.96\sqrt{4/n},\ \hat{\theta}_n^{med} + 1.96\sqrt{4/n}\right] \quad (\text{look ⟨Ex. 8.4.1⟩})$$

The bound on the difference between the estimator and the parameter subject to estimation is $1.96\sqrt{4/n}$. This is called the "95% Asymptotic bound on estimation error" of $\hat{\theta}^{med}$ - Sample median. Therefore, in order to make the "Asymptotic bound on estimation error" of $\hat{\theta}^{med}$ less than d, the required sample size, n_{med}, is:

$$n_{med} \simeq (1.96/d)^2 \cdot 4$$

In the same manner, for $\hat{\theta}^{mean}$,

$$n_{mean} \simeq (1.96/d)^2 \cdot \pi^2/3$$

Therefore the ratio of the reciprocals of sample size is

$$n_{mean}^{-1} / n_{med}^{-1} \simeq 12\pi^2 \fallingdotseq 1.2 = ARE\left[\{\hat{\theta}_n^{mean}\}, \{\hat{\theta}_n^{med}\}\right]$$

That is, since the estimation precision of the sample mean is about 1.2 times that of the sample median, using the sample median would require approximately 1.2 times as many observations, implying relatively lower efficiency.

Ex. 8.4.3. Estimation of the median in a location parameter model

Similar to ⟨Ex. 8.4.1⟩, consider a model such that its population distribution is continuous and its pdf is $f(x-\theta)$, $-\infty<\theta<+\infty$ that is symmetric about θ. i.e.,

$$f(-x)=f(x)\ \forall x:-\infty<x<+\infty \quad ?\ f(\theta-x)=f(\theta+x)\ \forall x\in(-\infty,+\infty)$$

In this case, the limiting distribution of $\hat{\theta}_n^{med}=X_{([n+1)/2])}$ is

$$\sqrt{n}\,(\hat{\theta}_n^{med}-\theta)\xrightarrow[n\to\infty]{d}N\left(0,\ 1/(4f^2[0])\right) \qquad \cdots \langle \text{Ex. } 5.3.8\rangle$$

For sample mean $\hat{\theta}_n^{mean}=\bar{X}$, given that the population variance exists, i.e.,

$$\sigma_f^2=\int_{-\infty}^{+\infty}x^2f(x)\,dx<+\infty$$

the limiting distribution is given as:

$$\sqrt{n}\,(\hat{\theta}_n^{mean}-\theta)\xrightarrow[n\to\infty]{d}N(0,\sigma_f^2)$$

For the MLE, if the conditions in ⟨Theorem 6.4.4⟩ are satisfied, the information is given as:

$$I(\theta)=\mathrm{Var}_\theta\left\{\frac{\partial}{\partial\theta}\log f(X_i-\theta)\right\}=\int_{-\infty}^{\infty}\left(\frac{f'(x)}{f(x)}\right)^2\cdot f(x)\,dx$$

and under the condition that $I(\theta)>0$, the limiting distribution is:

$$\sqrt{n}\,(\hat{\theta}_n^{MLE}-\theta)\xrightarrow[n\to\infty]{d}N(0,I_f^{-1}),\quad I_f=\int_{-\infty}^{+\infty}\left(\frac{f'(x)}{f(x)}\right)^2f(x)\,dx$$

The variance of asymptotic distributions for different population distributions are organized in ⟨Table 8.4.1⟩ and the ARE between different estimation methods is organized in ⟨Table 8.4.2⟩.

⟨Table 8.1.1. Variance of Asymptotic distribution of med, mean & MLE⟩

	Population dist.	$N(\theta,1)$	$L(\theta,1)$	$DE(\theta,1)$
med	$(4f^2(0))^{-1}$	$\pi/2$	4	1
mean	σ_f^2	1	$\pi^2/3$	2
MLE	I_f^{-1}	1	3	1

⟨Table 8.4.2. ARE between median, mean, and MLE⟩

	Population dist.	$N(\theta,1)$	$L(\theta,1)$	$DE(\theta,1)$
$ARE(\hat{\theta}_n^{mean},\hat{\theta}_n^{med})$		$\pi/2$	$12/\pi^2$	1/2
$ARE(\hat{\theta}_n^{mean},\hat{\theta}_n^{MLE})$		1	$9/\pi^2$	1/2
$ARE(\hat{\theta}_n^{med},\hat{\theta}_n^{MLE})$		$2/\pi$	3/4	1

Ex. 8.4.4. Comparison of estimators in Beta(α,1) (α>0)

The pdf of Beta(α,1) is:

$$\alpha x^{\alpha-1}I_{(0,1)}(x)$$

and its population mean is $\alpha/(1+\alpha)$. Hence the MME of α is:

$$\hat{\alpha}^{MME}=\frac{\bar{X}}{1-\bar{X}}$$

and its limiting distribution is given as the following from ⟨Theorem 6.1.a⟩,

$$\sqrt{n}\,(\hat{\alpha}^{MME}-\alpha)\xrightarrow[n\to\infty]{d}N(0,\dot{g}(m_1)^2\cdot\sigma^2),\ g(m_1)=\frac{m}{1-m},\ \sigma^2=\frac{\alpha}{(\alpha+1)^2(\alpha+2)}$$

$$\therefore\sqrt{n}\,(\hat{\alpha}^{MME}-\alpha)\xrightarrow[n\to\infty]{d}N\left(0,\frac{\alpha(\alpha+1)^2}{\alpha+2}\right)$$

$$\dot{g}(m_1)=\frac{1}{1-m}+\frac{m}{(1-m)^2}=\frac{1-m+m}{(1-m)^2}=\frac{1}{(1-m)^2}$$

$$\dot{g}(m_1)=\dot{g}\left(\frac{\alpha}{1+\alpha}\right)=(1+\alpha)^2$$

$$\therefore\dot{g}(m_1)^2\cdot\sigma^2=(1+\alpha)^4\cdot\frac{\alpha}{(\alpha+1)^2(\alpha+2)}$$

Meanwhile, the pdf of the model can be expressed as:

$$f(x;\alpha)=\alpha x^{\alpha-1}I_{(0,1)}(x)=\exp\left\{(\alpha-1)\log x+\log\alpha\right\}I_{(0,1)}(x)$$

The Beta(α,1) is an exponential family and satisfies the conditions in ⟨Theorem 6.2.4⟩. Thus, the MLE of α is given as the root of the likelihood equation:

$$\sum\log x_i+\frac{n}{\alpha}=0$$

Therefore, the MLE of α is: $\dfrac{1}{(-\log\bar{X})}$ and its limiting distribution is given as the following by the CLT or ⟨Theorem 6.4.4⟩.

$$\sqrt{n}\,(\hat{\alpha}_n^{MLE}-\alpha)\xrightarrow[n\to\infty]{d}N(0,I^{-1}(\alpha)),\ I(\alpha)=E_\alpha\left[-\frac{\partial^2}{\partial\alpha^2}\log f(x_i;\alpha)\right]$$

i.e., $\sqrt{n}\,(\hat{\alpha}_n^{MLE}-\alpha)\xrightarrow[n\to\infty]{d}N(0,\alpha^2)$

Hence, the ARE of MME w.r.t MLE is:

$$ARE\left(\{\hat{\alpha}_n^{MME}\},\{\hat{\alpha}_n^{MLE}\}\right)=\left[\frac{\alpha(\alpha+1)^2}{\alpha+2}\right]^{-1}/\alpha^{-2}=\frac{\alpha(\alpha+2)}{(\alpha+1)^2}$$

From above, we know that MLE is asymptotically more efficient than the MME.

In the examples considered so far, the estimator with the greatest ARE has been the MLE. The asymptotic efficiency of the MLE can also be inferred from the following information inequality.

〈Theorem 8.4.1. Information Inequality〉[122]

[122] 이 부등식을 크라메-라오 (Cramer-Rao) 부등식이라고도 한다.

In a probability model with pdf $f(x;\theta), \theta \in \Omega$, suppose that the conditions (R0)-(R5) in 〈Theorem 6.4.3〉 are satisfied so that the MLE is asymptotically normal. For $\hat{\eta}_n(X_1,\cdots,X_n)$ - an estimator for $\eta=\eta(\theta)$[123] - the following inequality is true: "$\eta=\eta(\theta)$ 표현에 $X_1,\cdots,X_n$을 이용한 실수 값 모수"

$$Var_\theta(\hat{\eta}_n) \geq \left[\frac{\partial}{\partial\theta} E_\theta(\hat{\eta}_n)\right]^t [nI(\theta)]^{-1} \left[\frac{\partial}{\partial\theta} E_\theta(\hat{\eta}_n)\right], \quad \forall \theta \in \Omega$$

(proof)

First, we will prove for the case in which θ is one-dimensional. In this case, we know from the Cauchy schwarz inequality that the follow is true for the variance/covariance of the score function: $\dot{l}_n(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \log f(x_i;\theta)$ and the estimator $\hat{\eta} = \hat{\eta}(X_1,\cdots,X_n)$.

$$Var_\theta(\hat{\eta}_n)\, Var_\theta(\dot{l}_n(\theta)) \geq \left\{Cov_\theta(\hat{\eta}_n, \dot{l}_n(\theta))\right\}^2$$

* Cauchy-Schwartz inequality: $E[XY] \leq \sqrt{E[X^2]E[Y^2]}$

Let $X = T - E[T]$ and $Y = U - E[U]$.

Then, $E[(T-E[T])(U-E[U])] = \sqrt{E[(T-E(T))^2]\, E[(U-E(U))^2]}$

$\Rightarrow Cov(T,U) \leq \sqrt{Var(T)\cdot Var(U)}$

Meanwhile, under conditions given in 〈Theorem 6.4.3〉,

$$E_\theta[\dot{l}_n(\theta)] = E_\theta\left\{\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(X_i;\theta)\right\} = n E_\theta\left\{\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right\} = 0$$

$$Var_\theta(\dot{l}_n(\theta)) = Var_\theta\left\{\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(X_i;\theta)\right\} = n\, Var_\theta\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right) = nI(\theta)$$

and from condition (R4), the following holds:

$$Cov_\theta(\hat{\eta}_n, \dot{l}_n(\theta)) = E_\theta[\hat{\eta}_n \dot{l}_n(\theta)] = E_\theta\left[\hat{\eta}_n \frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i;\theta)\right] = \frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)$$

$$\hookrightarrow E_\theta\left[\hat{\eta}_n \cdot \frac{\frac{\partial}{\partial\theta}\prod_{i=1}^n f(X_i;\theta)}{\prod_{i=1}^n f(X_i;\theta)}\right]$$

$$= \int_{-\infty}^{+\infty} \underbrace{\hat{\eta}_n}_{\hat{\eta}_n(X_1,\cdots,X_n)} \frac{\frac{\partial}{\partial\theta}\prod_{i=1}^n f(X_i;\theta)}{\prod_{i=1}^n f(X_i;\theta)} \prod_{i=1}^n f(X_i;\theta)\, dx$$

$$= \frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} \hat{\eta}_n \prod_{i=1}^n f(X_i;\theta)\, dx = \frac{\partial}{\partial\theta}E_\theta[\hat{\eta}_n]$$

Hence, if we apply this to the Cauchy-inequality above, we get:

$$Var_\theta(\hat{\eta}_n) \geq \left[\frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)\right]^2 / [nI(\theta)], \quad \forall \theta \in \Omega$$

[123] $\eta = \eta(\theta)$가 다차원의 모수, 즉 벡터인 경우에는 같은 방법으로

$$c^t\left\{Var_\theta(\hat{\eta}_n) - \left[\frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)\right]^t [nI(\theta)]^{-1}\left[\frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)\right]\right\}c \geq 0, \ \forall c, \forall \theta \in \Omega$$

따 같은 부등식이 성립하는 것을 알 수 있다.

Meanwhile, if θ is multidimensional,

$\Sigma_{11} = Var_\theta(\hat{\eta}_n)$, $\Sigma_{12} = Cov_\theta(\hat{\eta}_n, \dot{l}_n(\theta))$,

$\Sigma_{22} = Var_\theta(\dot{l}_n(\theta))$, $\Sigma_{21} = Cov_\theta(\dot{l}_n(\theta), \hat{\eta}_n)$

and similar to the one-dimensional case,

$\Sigma_{12} = Cov_\theta(\hat{\eta}_n, \dot{l}_n(\theta)) = \left[\frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)\right]^t$, $\Sigma_{22} = Var_\theta(\dot{l}_n(\theta)) = nI(\theta)$

we know that: $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = Var_\theta(\hat{\eta}_n - \Sigma_{12}\Sigma_{22}^{-1}\dot{l}_n(\theta)) \geq 0, \ \forall \theta \in \Omega$

Thus, we see that the following inequality holds:

$$Var_\theta(\hat{\eta}_n) - \left(\frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)\right)^t [nI(\theta)]^{-1}\left[\frac{\partial}{\partial\theta}E_\theta(\hat{\eta}_n)\right] \geq 0 \quad \forall \theta \in \Omega$$

〈Theorem 8.4.2. Information Inequality for Unbiased Estimators〉

Under 〈Theorem 8.4.1〉, the following holds for an unbiased estimator $\hat{\theta}_n^{UE}$

$$c^t Var_\theta(\hat{\theta}_n^{UE})c \geq c^t I^{-1}(\theta)c/n, \quad \forall \theta \in \Omega, \ \forall c$$

If θ is one-dimensional,

$$Var_\theta(\hat{\theta}_n^{UE}) \geq 1/[nI(\theta)]$$

If θ is multi-dimensional,

$$Var_\theta(\hat{\theta}_n^{UE}) - I^{-1}(\theta)/n$$

is a non-negative/positive semidefinite matrix

(proof)

if $\eta = c^t\theta = c_1\theta_1 + \cdots + c_K\theta_K$, for an unbiased estimator $\hat{\eta}_n^{UE} = c^t\hat{\theta}_n^{UE}$,

$$c^t Var_\theta(\hat{\theta}_n^{UE})c = Var_\theta(c^t\hat{\theta}_n^{UE}), \ \frac{\partial}{\partial\theta}E_\theta(c^t\hat{\theta}_n) = c^t$$

Thus, the result comes from applying $\hat{\eta}_n^{UE} = c^t\hat{\theta}^{UE}$ into the inequality in 〈Theorem 8.4.1〉.

Ex. 8.4.5. Unbiased estimation and information inequality in Beta distribution. Beta(a,1)

The pdf of Beta(a,1): $ax^{a-1} = \exp\{\log a + (a-1)\log x_i\}$

Beta distribution is an exponential family and satisfies the conditions in ⟨Theorem 8.3.2⟩.

$$\therefore\ Y = \Sigma \log X_i \text{ is a CSS for } a$$

$f(x;\theta) = a x^{a-1}$, $F(x;\theta) = x^a \sim U(0,1)$

$-\log U \overset{d}{\equiv} -\log(1-U) \sim \text{Exp}(1)$ $\quad\therefore -a\log X_i \sim \text{Exp}(1) \overset{d}{\equiv} \text{Gamma}(1,1)$

$\Rightarrow -a\Sigma\log X_i \sim \text{Gamma}(n,1)$

Let $Z = -a\Sigma\log X_i$, $f(z;\theta) = \frac{1}{\Gamma(n)}\frac{1}{a}\, z^{n-1}\cdot\exp\{-z\}$

$$E\left[\frac{1}{\hat{a}}\right] = \int_0^\infty z^{-1}\frac{1}{\Gamma(n)} z^{n-1}\exp\{-z\}\,dz = \frac{1}{n-1}\int_0^\infty \frac{1}{\Gamma(n-1)} z^{n-2}\exp\{-z\}\,dz$$

$$= (n-1)^{-1}$$

$$\therefore E\left[-\frac{1}{a\Sigma}\right] = (n-1)^{-1} \Leftrightarrow E\left[-\frac{(n-1)}{\Sigma\log X_i}\right] = a$$

$$\therefore \hat{a}^{RB}(X_1,\cdots,X_n) = -\frac{n-1}{\Sigma\log X_i}$$

$$\text{Var}(\hat{a}^{RB}(X_1,\cdots,X_n)) = (n-1)^2\cdot\text{Var}\left(\frac{1}{\Sigma\log X_i}\right)$$

$$\text{Var}\left(\frac{1}{\hat{a}}\right) = E[Z^{-2}] - E[Z^{-1}]^2 = \frac{1}{(n-1)(n-2)} - \frac{1}{(n-1)^2} = \frac{(n-1)-(n-2)}{(n-1)^2(n-2)}$$

$$\therefore \text{Var}\left(\frac{1}{\hat{a}}\right) = \text{Var}\left(\frac{1}{a\Sigma\log X_i}\right) = a^{-2}\text{Var}\left(\frac{1}{\Sigma\log X_i}\right),\quad \therefore \text{Var}\left(\frac{1}{\Sigma\log X_i}\right) = \frac{a^2}{(n-1)^2(n-2)}$$

$$\therefore \text{Var}(\hat{a}^{RB}) = \frac{a^2}{(n-2)}$$

The Rao-Cramer Lower Bound given in ⟨Theorem 8.4.2⟩ is $1/[nI(a)] = a^2/n$. Thus, the variance of the UMVUE is larger than the CRLB, but asymptotically gets closer to it. i.e.,

$$\lim_{n\to\infty} \text{Var}_a(\hat{a}_n^{UMVUE})/[nI(a)]^{-1} = 1$$

Also from ⟨Ex. 8.4.4⟩, $\hat{a}_n^{UMVUE} = \frac{n-1}{n}\hat{a}_n^{MLE}$ and the UMVUE and MLE has the same limiting distribution:

$$\sqrt{n}(\hat{a}_n^{UMVUE} - a) \xrightarrow[n\to\infty]{d} N(0, a^2)$$

and both estimators have the same asymptotic efficiency.

From ⟨Theorem 8.4.2⟩, if the parameter is one-dimensional we know that:

$$\text{Var}_\theta\{\sqrt{n}(\hat{\theta}_n^{UE} - \theta)\} \geq 1/I(\theta)$$

$\text{Var}(\hat{\theta}^{UE}) \geq (nI(\theta))^{-1} \Rightarrow \text{Var}(\hat{\theta}-\theta) \geq (nI(\theta))^{-1} \Rightarrow n\text{Var}(\hat{\theta}_n^{UE}-\theta) \geq I(\theta)^{-1}$
$\Rightarrow \text{Var}(\sqrt{n}(\hat{\theta}_n^{UE}-\theta)) \geq I(\theta)^{-1}$

Hence, for any unbiased statistic that is asymptotically normal:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} N(0, \delta^2(\theta))$$

we can conjecture that $\delta^2(\theta) \geq I(\theta)^{-1}$. Meanwhile, from ⟨Theorem 6.4.4⟩, the limiting distribution of MLE is:

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta) \xrightarrow[n\to\infty]{d} N(0, 1/I(\theta))$$

Thus, we can conjecture that the MLE has the least possible asymptotic variance among estimators that are asymptotically normal.

This conjecture is called the Fisher's conjecture, and under general circumstances, it does not hold. However, under the uniform convergence criteria, it is known to be true.[124] In this respect, we say that the MLE is an asymptotically efficient estimator, and its asymptotic efficiency is true in the estimation of muti-dimensional parameter estimation as well.

[124] 이에 대한 상세한 논의는 참고서적 [P.J. Bickel & K.A. DoKSum] Mathematical Statistics p.233 ~ p.239 에 적어져 있다.

Ex. 8.4.6. Comparision of estimators in Gamma(α,β), α>0, β>0.

The pdf of Gamma(α,β) is:

$$f(x;\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I_{(x>0)},\quad \theta = (\alpha,\beta)^t$$

The Gamma distribution is an exponential family and satisfies the conditions in ⟨Theorem 6.4.4⟩. Using this, in ⟨Ex. 6.4.5⟩, the MLEs, which are given as roots of the likelihood equation:

$$\dot{l}_n(\theta) = \begin{pmatrix} -n\Psi(\alpha) - n\log\beta + n(\overline{\log x}) \\ -n\alpha/\beta + n\bar{x}/\beta^2 \end{pmatrix} = 0$$

are estimated, using the one-step iteration process.
Here, $(\overline{\log x}) = n^{-1}\sum_{i=1}^n \log x_i$, $\Psi(\alpha) = \frac{\partial}{\partial\alpha}\log\Gamma(\alpha)$ and the estimator $\hat{\theta}^{[1]} = (\hat{\alpha}_n^{[1]}, \hat{\beta}_n^{[1]})^t$, estimated via the one-step iteration algorithm, is:

$$\hat{\theta}_n^{[1]} = \hat{\theta}_n^{[0]} + [-\ddot{l}_n(\hat{\theta}_n^{[0]})]^{-1}\dot{l}_n(\hat{\theta}_n^{[0]}),\quad \hat{\theta}_n^{[0]} = \hat{\theta}_n^{MME}$$

Here, the initial estimators (i.e., $\hat{\theta}_n^{[0]}$) are $\hat{\theta}_n^{MME} = (\hat{\alpha}_n, \hat{\beta}_n)^t$ which are given as the roots of the simultaneous equation:

$$\begin{cases} \hat{\alpha}_n\hat{\beta}_n = \bar{x}_n \\ \hat{\alpha}_n\hat{\beta}_n^2 = n^{-1}\sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$$

and from ⟨Theorem 6.1.2⟩ we can find its limiting distribution:

$$\sqrt{n}(\hat{\theta}^{MME} - \theta) \xrightarrow[n\to\infty]{d} N(0, \Sigma(\theta)),\quad \Sigma(\theta) = \begin{pmatrix} 2\alpha(\alpha+1) & -2(\alpha+1)\beta \\ -2(\alpha+1)\beta & (2+3/\alpha)\beta^2 \end{pmatrix}$$

Meanwhile, the information matrix and its inverse is given as:

$$I(\theta) = E_\theta[-\ddot{l}_1(\theta)] = \begin{pmatrix} \dot{\Psi}(\alpha) & \beta^{-1} \\ \beta^{-1} & \alpha\beta^{-2} \end{pmatrix}, \quad I^{-1}(\theta) = \frac{1}{\alpha\dot{\Psi}(\alpha)-1} \begin{pmatrix} \alpha & -\beta \\ -\beta & \dot{\Psi}(\alpha)\beta^2 \end{pmatrix}$$

and for $\hat{\theta}_n^{(1)} = (\hat{\alpha}_n^{(1)}, \hat{\beta}_n^{(1)})^t$,

$$\sqrt{n}(\hat{\theta}_n^{(1)} - \theta) \xrightarrow[n\to\infty]{d} N(0, I^{-1}(\theta))$$

and the limiting distribution of MLE, $\hat{\theta}_n^{MLE} = (\hat{\alpha}_n^{MLE}, \hat{\beta}_n^{MLE})^t$ is also equivalent to the above.

Therefore the ARE of MME w.r.t MLE is:

$$ARE(\{\hat{\alpha}_n^{MME}\}, \{\hat{\alpha}_n^{MLE}\}) = [2\alpha(\alpha+1)]^{-1} / [\alpha/(\alpha\dot{\Psi}(\alpha)-1)]^{-1},$$

$$ARE(\{\hat{\beta}_n^{MME}\}, \{\hat{\beta}_n^{MLE}\}) = [(2+3/\alpha)\beta^2]^{-1} / [\dot{\Psi}(\alpha)\beta^2 / (\alpha\dot{\Psi}(\alpha)-1)]^{-1}$$

and the values calculated[125] are given in ⟨Table 8.4.3⟩ and ⟨Table 8.4.4⟩.

⟨Table 8.4.3. Value of $ARE(\{\hat{\alpha}_n^{MME}\}, \{\hat{\alpha}_n^{MLE}\}) = \frac{1}{2(\alpha+1)(\alpha\dot{\Psi}(\alpha)-1)}$⟩

α	1/10	1/2	1	2	3	4
$ARE(\{\hat{\alpha}_n^{MME}\}, \{\hat{\alpha}_n^{MLE}\})$	0.0497	0.2271	0.3876	0.5749	0.7391	0.8798

⟨Table 8.4.4 Value of $ARE(\{\hat{\beta}_n^{MME}\}, \{\hat{\beta}_n^{MLE}\}) = \frac{\alpha\dot{\Psi}(\alpha)}{(2\alpha+3)(\alpha\dot{\Psi}(\alpha)-1)}$⟩

α	1/10	1/2	1	2	3	4
$ARE(\{\hat{\beta}_n^{MME}\}, \{\hat{\beta}_n^{MLE}\})$	0.3466	0.4203	0.5101	0.6356	0.7628	0.8850

As one can see from the table above, the efficiency of MME is relatively lower than that of MLE. Such phenomenon holds generally and it is known that MLE is asymptotically most efficient.

[125] $\dot{\Psi}(\alpha)$, $\ddot{\Psi}(\alpha)$는 각각 다이감마(digamma), 트라이감마(trigamma) 함수라고 불리우는 함수로서 각각 로그감마함수의 일차·이차 도함수이고 위의 표는 패키지 R에 있는 함수를 이용하여 계산한 것이다.