WEEK 10

Efficiency:

For any parameter θ we may have many different unbiased estimators as well as any consistent estimators. Among those estimators, we may want to have more efficient estimators. If it is possible, we want the most efficient estimator. Actually, among the class of unbiased estimators, it there is one that the smallest variance this is the most efficient estimators. Such an estimator may not exist. In some cases, we may find such estimators but we are not going to discuss these types of estimators.

Definition: Let $X_1, X_2 \dots, X_n$ be a random sample with probability (or probability density) function $f(x;\theta)$ and consider two estimators T_1 and T_2 to estimate the parameter θ. If

$$Var(T_1) \le Var(T_2) \text{ for all } \theta$$

then the estimator T_1 is said to be more efficient that T_2.

Example: Let $X_1, X_2 \dots, X_n$ be a random sample from $Poisson(\theta)$ population. We know that the mean and variance for the Poisson distribution are the same. That is, $E(X) = \theta$ and $Var(X) = \theta$ since the sample mean and sample variance are unbiased for the population mean and variance (population mean $E(X) = \mu = \theta$ and $Var(X) = \sigma^2 = \theta$) we have

$$E(\bar{X}_n) = \theta \text{ and } E(S_n^2) = \theta .$$

Moreover, for any fixed reel number a a class of estimators $T_a = a\bar{X}_n + (1-a)S_n^2$ are all unbiased for the population parameter θ. That is for a parameter we can find infinitely many unbiased estimators for the population parameter. The sample mean is unbiased for θ with the variance θ / n . That is, $Var(\bar{X}_n) = \theta / n$. It can also be shown that $Var(\bar{X}_n) \le Var(S_n^2)$ (the calculation of $Var(S_n^2)$ is really difficult and the inequality can be shown from the Cramer-Rao's inequality that we are not going to cover here). That is, the sample mean is more efficient that the sample variance for the Poisson parameter θ. Again, from Cramer-Rao's inequality, it can be shown that the sample mean is the most efficient estimator among all these unbiased estimators (This is known as the Uniformly Minimum variance Unbiased Estimator, the UMVUE).

Sufficiency:

Sufficiency property of the estimators are very important especially for statistical inference. In estimation, we usually look for the most efficient unbiased estimators. If we can find a sufficient estimator for a parameter, based on this sufficient estimator we can write and unbiased estimator for the parameter that we are interested in.

Definition: Let $X_1, X_2 \ldots, X_n$ be a random sample from a population with probability (or probability density) function $f(x;\theta)$ and T be any estimator for the parameter θ. If the conditional probability density function of $X_1, X_2 \ldots, X_n$ given $T = t$ then we say that the estimator T is sufficient for θ.

A sufficient estimator is the one that summarized all the information in the sample about the parameter. If and estimator T is sufficient for a parameter θ, then any statistical inference can be made based on the value of the sufficient estimator T.

Let $X_1, X_2 \ldots, X_n$ be a random sample from a population with a probability function $f(x,\theta)$ and T be any estimator for θ. As we know, the estimator is a function of the sample and it is a random variable. For discrete case, if the conditional probability

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T = t)$$

does not depend on the parameter θ the T is sufficient for θ. Similar argument can be carried out for continuous case. For simplicity let us denote $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)'$ and the estimator as $T(\underset{\sim}{X})$. As it is understood from the definition a sufficient estimator is not unique.

Let T be a sufficient estimator for θ then the conditional probability $P(\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = t)$ does not depend on the parameter θ. On the other hand, since $\{\underset{\sim}{X} = \underset{\sim}{x}\} \subset \{T(\underset{\sim}{X}) = T(\underset{\sim}{x})\}$ the conditional probability $P\left(\underset{\sim}{X} = \underset{\sim}{x} | T = t\right)$ can be calculated as

$$P\left(\underset{\sim}{X} = \underset{\sim}{x} | T = t\right) = \frac{P_\theta(\underset{\sim}{X} = \underset{\sim}{x} \ , T = t)}{P_\theta(T = t)} = \frac{P_\theta(\underset{\sim}{X} = \underset{\sim}{x})}{P_\theta(T = t)} = \frac{p(\underset{\sim}{x};\theta)}{q(T(\underset{\sim}{x});\theta)}$$

and therefore if the ratio $p(\underset{\sim}{x};\theta) / q(T(\underset{\sim}{x});\theta)$ does not depend on the parameter the estimator T is sufficient for θ. Therefore we can state the following theorem without the proof.

Theorem: An estimator T is sufficient for θ if and only if the ration $p(\underset{\sim}{x};\theta) / q(T(\underset{\sim}{x});\theta)$ does not depen on the parameter.

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with the parameter p. let us try to check whether the estimator $T = X_1 + X_2 + \ldots + X_n$ is sufficient or not. Note that $T \sim Binom(n, p)$ with the probability function,

$$P_p(T = t) = \binom{n}{t} p^t (1-p)^{n-t} \; , \; t = 0,1,2,...,n \, .$$

Therefore if the ratio $p(\underset{\sim}{x}; p) / q(T(\underset{\sim}{x}); p)$ for $t = x_1 + x_2 + ... + x_n$ does not depen on the parameter p then T will be sufficient. Note that the ratio for $t = x_1 + x_2 + ... + x_n$ can be written as

$$\frac{p(\underset{\sim}{x}; p)}{q(T(\underset{\sim}{x}); p)} = \frac{P_p(\underset{\sim}{X} = \underset{\sim}{x})}{P_p(T = t)} = \frac{\prod_{i=1}^{n} P_p(X_i = x_i)}{P_p(T = t)} = \frac{\prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{1}{\binom{n}{t}}$$

and which is independent from θ and therefore the estimator $T = X_1 + X_2 + \ldots + X_n$ is sufficient for p.

The following theorem is very important to find a sufficient estimatıor for a parameter. We state the theorem without the proof. The proof of the theorem can be found in any textbook retaled to estimation theory (e.g Casella and Berger, 2002).

Theorem *(Factorization Theorem, very important)* Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with probability or probability density function $f(x;\theta)$. The estimator T is sufficient for θ if and only if there are functions $g(t;\theta)$ and $h(x)$ such that the joint probability or probability density function of $X_1, X_2, \ldots, X_n$ can be written as

$$f(\underset{\sim}{x};\theta) = g(T(\underset{\sim}{x});\theta)\, h(\underset{\sim}{x}) \, .$$

Here g is a function of $T(\underset{\sim}{x})$ and θ and h is only a function of $\underset{\sim}{x}$ which does not depend on the parameter.

Example: a) Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with the parameter p. In the previous example we showed that $T = X_1 + X_2 + \ldots + X_n$ is sufficient for p. By using the factorization theorem the joint probability distribution of X 's can be written as

$$f(\underset{\sim}{x}; p) = P_p(\underset{\sim}{X} = \underset{\sim}{x}) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i} = p^{T(\underset{\sim}{x})} (1-p)^{n - T(\underset{\sim}{x})}$$

and therefore for the functions $g(T(\underset{\sim}{x}); p) = p^{T(\underset{\sim}{x})}(1-p)^{n-T(\underset{\sim}{x})}$ and $h(\underset{\sim}{x}) = 1$ the joint probability function can be written as $f(\underset{\sim}{x}; p) = g(T(\underset{\sim}{x}); p)\, h(\underset{\sim}{x})$ and thus by the fuctorization theorem T is sufficient for p.

b) let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution. The joint probability density function of $X_1, X_2, \ldots, X_n$ can be written as for $\underset{\sim}{T}(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$,

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

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

$$= \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \exp\left[\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)\begin{pmatrix}\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i\end{pmatrix}\right]$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \exp\left[\underset{\sim}{w}'\left(\mu, \sigma^2\right)\underset{\sim}{T}(\underset{\sim}{x})\right].$$

Therefore for the functions

$$h(\underset{\sim}{x}) = (1/2\pi)^{n/2} \text{ and } g(T(\underset{\sim}{x}); \mu, \sigma^2) = \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \exp\left[\underset{\sim}{w}'\left(\mu, \sigma^2\right)\underset{\sim}{T}(\underset{\sim}{x})\right]$$

the joint probability density function can be written as

$$f(\underset{\sim}{x}; \mu, \sigma^2) = g(\underset{\sim}{T}(\underset{\sim}{x}); \mu, \sigma^2)\, h(\underset{\sim}{x})$$

and thus by the factorization theorem the bivariate estimator $\underset{\sim}{T}(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient for (μ, σ^2).

c) Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution with the parameter θ. The probability density function of the uniform distribution is given by

$$f(x; \theta) = \begin{cases} 1/\theta & , \quad 0 < x < \theta \\ 0 & , \quad d.y. \end{cases}$$

This probability density function can also be written with the following indicator function

$$I_A(x) = \begin{cases} 1 & , \quad x \in A \\ 0 & , \quad x \notin A \end{cases}$$

as $f(x;\theta)=\theta^{-1} I_{(0<x<\theta)}(x)$. Therefore the joint probability density function of the sample can be written for $x_{(n)}=\max\{x_1,x_2,...,x_n\}$ as

$$f(\underset{\sim}{x};\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\frac{1}{\theta}I_{\{0<x_{(1)}<\theta\}}(x_1)\frac{1}{\theta}I_{\{0<x_{(2)}<\theta\}}(x_2)...\frac{1}{\theta}I_{\{0<x_{(n)}<\theta\}}(x_n)=\frac{1}{\theta^n}I_{\{0<x_{(n)}<\theta\}}(\underset{\sim}{x}).$$

By defining the functions $g(T(\underset{\sim}{x});\theta)=\theta^{-n}I_{\{0<x_{(n)}<\theta\}}(\underset{\sim}{x})$ and $h(\underset{\sim}{x})=1$ allows us to write the joint probability density function as $f(\underset{\sim}{x};\theta)=g(T(\underset{\sim}{x});\theta)\,h(\underset{\sim}{x})$ and thus by the factorization theorem $T(\underset{\sim}{X})=X_{(n)}$ is sufficient for θ.

d) Let $X_1,X_2,\ldots,X_n$ be a random sample from a population with the following probability density function:

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1} & , \quad 0<x<1\\ 0 & , \quad d.y.\end{cases}$$

Note that the joint probability density function of $X_1,X_2,\ldots,X_n$ can be written as

$$f(\underset{\sim}{x};\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\theta^n\prod_{i=1}^{n} x_i^{\theta-1}=\theta^n\left(\prod_{i=1}^{n} x_i\right)^{\theta}\prod_{i=1}^{n}\frac{1}{x_i}=g\left(T(\underset{\sim}{x});\theta\right)h(\underset{\sim}{x})$$

and thefore by the factorization theorem the estimator $T(\underset{\sim}{X})=\prod_{i=1}^{n} X_i$ is sufficient for θ. Here the functions g and h are defined as

$$g\left(T(\underset{\sim}{x});\theta\right)=\theta^n\left(\prod_{i=1}^{n} x_i\right)^{\theta} \quad \text{and} \quad h(\underset{\sim}{x})=\prod_{i=1}^{n}\frac{1}{x_i}.$$

The application of the factorization theorem does not requires the independence. That is, the theorem is valid for non-independent and identically distributed sample.

As it is noted before, the sufficient estimator is not unique. Any one-to-one function of a sufficient estimator is also sufficient. To show that, let T be any sufficient estimator for a parameter θ and r be any one-to-one function. then $T^*=r(T)$ is also sufficient for the same parameter θ. Since T is sufficient for θ by the factorization theorem the joint probability density function can be written as $f(\underset{\sim}{x};\theta=g(T(\underset{\sim}{x});\theta)\,h(\underset{\sim}{x})$ for specified function $g(t;\theta)$ and $h(x)$. Moreover, since r is a one-to-one $T=r^{-1}(T^*)$ and therefore the joint probability density function of the sample can also be written as

$$\begin{aligned}f(\underset{\sim}{x};\theta)&=g(T(\underset{\sim}{x});\theta)\,h(\underset{\sim}{x}=g(r^{-1}(T^*(\underset{\sim}{x}));\theta)\ h(\underset{\sim}{x})\\&=(g\circ r^{-1})(T^*(\underset{\sim}{x});\theta)h(\underset{\sim}{x})=g^*(T^*(\underset{\sim}{x});\theta)\,h(\underset{\sim}{x})\end{aligned}$$

and thus by the factorization theorem $T^* = r(T)$ is also sufficient for θ. According to this result, since $T = \sum_{i=1}^{n} X_i$ is sufficient for p in the example (a), $\bar{X}_n$ is also sufficient for the same parameter. In the example (b) we showed that $\underset{\sim}{T}(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient for (μ, σ^2) and for the same reason the estimator $\underset{\sim}{T}^*(\underset{\sim}{X}) = (\bar{X}_n, S_n^2)$ is also sufficent for the same parameter (μ, σ^2). In the example (d) since $T = \prod_{i=1}^{n} X_i$ is sufficient for θ, the estimator $T^* = -\sum_{i=1}^{n} \ln(X_i)$ is also sufficient for the same parameter θ.

UMVUE Estimators :

As we have noted earlier, we want to use the most efficient unbiased estimators in the estimation. Usually such estimators may not exist. We have defined the efficient estimators before. That is we say that and estimator T_1 is relatively more efficient than T_2 if $Var(T_1) \le Var(T_2)$ for all θ. That is, it is possible to consider as many estimators as for a parameter θ. Consider a class of unbiased estimators for a parameter $\tau(\theta)$ as $\mathcal{C} = \{T : E_\theta(T) = \tau(\theta)\}$. If there exists and estimator $T \in \mathcal{C}$ such that

$$Var_\theta(T) \le Var\theta(T^*) \text{ for all } T^* \in \mathcal{C} \text{ and for all } \theta$$

the estimator T is said to be the most unbiased estimator for $\tau(\theta)$. That is the estimator T has the smallest variance among all unbiased estimators of $\tau(\theta)$. In other word, the estimator T is the **U**niformly **M**inimum **V**ariance **U**nbiased **E**stimator for $\tau(\theta)$ (the UMVUE estimator). Finding such an estimator may be difficult. In many cases the UMVUE estimators do not exist. The following theorem allows us to find the UMVUE estimators under some certain cases.

Theorem *(Cramer Rao's Inequality)*: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with probability or probability density function $f(x;\theta)$ and W be any estimator such that $E_\theta(W)$ is differentiable with respect to θ. If for a function $h(\underset{\sim}{x})$ with $E_\theta(|h(\underset{\sim}{X})|) < \infty$ satisfies

$$\frac{d}{d\theta}\int\int\ldots\int\left[h(\underset{\sim}{x})\, f(\underset{\sim}{x};\theta)\right] d\underset{\sim}{x} = \int\int\ldots\int\left[h(\underset{\sim}{x})\frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)\right] d\underset{\sim}{x}$$

then

$$Var_\theta(W) \geq \frac{\left[\frac{d}{d\theta}E_\theta(W)\right]^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}[\ln(f(\underset{\sim}{X};\theta))]\right)^2\right)} .$$

dir.

In this inequality, the sample does not have to be independent and identically distributed random variables. If we have independent and identically distributed random sample then, the inequality can be written as

$$Var_\theta(W) \geq \frac{\left[\frac{d}{d\theta}E_\theta(W)\right]^2}{-n\,E_\theta\left(\frac{\partial^2}{\partial\theta^2}\left[\ln\left(f(X;\theta)\right)\right]\right)} = (\text{say } CRLB).$$

According to this inequality the variance of any estimator W is greater than CRLB. Therefore if we can find an unbiased estimator for $\tau(\theta)$ (usually θ or $E_\theta(W)$) such that its variance is equal to CRLB, then the estimator is the UMVUE estimator for $E_\theta(W)$. That is, if "=" holds in the Cramer-Rao's inequality, the estimator is the UMVUE estimator for $E_\theta(W)$.

<u>Example</u> a) Let $X_1, X_2, \ldots, X_n$ be a random sample from Poisson ditribution with the parameter θ. The probability function of Poisson distribution is

$P_\theta(X = x) = e^{-\theta}\theta^x / x!$, $x = 0,1,2,\ldots$

Note that $E_\theta(\bar{X}_n) = \theta$ and $Var_\theta(\bar{X}_n) = \theta / n$. Moreover $E_\theta(\bar{X}_n)$ is fdifferentiable with respect to θ (which is 1). Now let us calculate Cramer Rao's Lower Bound (CRLB). The numerator is 1. In order to calculate the denominator the probability function

$\ln(f(X;\theta)) = -\theta + X\ln(\theta)) - \ln(X!)$

and the derivatives are

$$\frac{\partial}{\partial\theta}\left[\ln\left(f(X;\theta)\right)\right] = -1 + \frac{X}{\theta} \quad \text{and} \quad \frac{\partial^2}{\partial\theta^2}\left[\ln\left(f(X;\theta)\right)\right] = -\frac{X}{\theta^2}.$$

and the mean of the secon derivative

$$-n\,E_\theta\left(\frac{\partial^2}{\partial\theta^2}\left[\ln\left(f(X;\theta)\right)\right]\right) = -n\,E_\theta\left(-\frac{X}{\theta^2}\right) = \frac{n}{\theta^2}E_\theta(X) = \frac{n}{\theta^2}\theta = \frac{n}{\theta}$$

and thus for $W = \bar{X}_n$ Cramer-Rao's inequality can be written as

$$\frac{\theta}{n} = Var_\theta\left(\bar{X}_n\right) \geq \frac{\left[\frac{d}{d\theta} E_\theta\left(\bar{X}_n\right)\right]^2}{-n\; E_\theta\left(\frac{\partial^2}{\partial\theta^2}\left[\ln\left(f\left(X;\theta\right)\right)\right]\right)} = \frac{1}{\left(\frac{n}{\theta}\right)} = \frac{\theta}{n}.$$

Since the equality holds, the estimator $\bar{X}_n$ is the UMVUE estimator for θ. because the variance of any estimator can not be smaller than CRLB.

b) Let $X_1, X_2, \ldots, X_n$ be a random sample from Bernoulli distribution with the parameter p. We know that $E_p(\bar{X}_n) = p$ and it is differentiable wit respect to p (which is 1) and $Var_p(\bar{X}_n) = p(1-p)/n$. The probability function of the distribution for $x = 0,1$ is

$$f(x;p) = P_p(X = x) = p^x(1-p)^{1-x}$$

and

$$\ln(f(X;p)) = \ln[p^X(1-p)^{1-X}] = X\ln(p) + (1-X)\ln(1-p).$$

The second derivative and its expected value is calculated as

$$\frac{\partial^2}{\partial p^2}\left[\ln\left(f\left(X;p\right)\right)\right] = -\frac{X}{p^2} - \frac{1-X}{(1-p)^2}$$

$$\Rightarrow -n\,E_p\left(\frac{\partial^2}{\partial p^2}\left[\ln\left(f\left(X;p\right)\right)\right]\right) = -n\left[-\frac{p}{p^2} - \frac{1-p}{(1-p)^2}\right] = \frac{n}{p(1-p)}$$

and therefore the Cramer-Rao's inequality for the estimator $W = \bar{X}_n$ can be written as

$$\frac{p(1-p)}{n} = Var_p\left(\bar{X}_n\right) \geq \frac{\left[\frac{d}{dp} E_p\left(\bar{X}_n\right)\right]^2}{-n\; E_p\left(\frac{\partial^2}{\partial p^2}\left[\ln\left(f\left(X;p\right)\right)\right]\right)} = \frac{1}{\left(\frac{n}{p(1-p)}\right)} = \frac{p(1-p)}{n}.$$

Since "=" holds instead of "$\geq$" for the estimator $\bar{X}_n$, the sample mean $\bar{X}_n$ is the UMVUE for the parameter p .

c) Let $X_1, X_2, \ldots, X_n$ be a random sample from exponential distribution with the parameter θ. Again $\bar{X}_n$ is the UMVUE estimator for θ. Note that $E_\theta(\bar{X}_n) = \theta$ and $Var_\theta(\bar{X}_n) = \theta/n$ and the expected value is differentiable with respect to θ (which is 1) The probability density function of the distribution is $f(x;\theta) = (1/\theta)e^{-x/\theta}$ for $x > 0$ and the derivatives in the inequality are

$$\frac{\partial}{\partial\theta}\left[\ln\left(f\left(X;\theta\right)\right)\right] = -\ln(\theta) - \frac{X}{\theta} \quad \text{and} \quad \frac{\partial^2}{\partial\theta^2}\left[\ln\left(f\left(X;\theta\right)\right)\right] = \frac{1}{\theta^2} - \frac{2X}{\theta^3}\ .$$

Therefore the expected value of the denominator at the Cramer-Rao's inequality is

$$-n\,E_\theta\left(\frac{\partial^2}{\partial\theta^2}[\ln(f(X;\theta))]\right)=-n\,E_\theta\left(\frac{1}{\theta^2}-\frac{2X}{\theta^3}\right)=-\frac{n}{\theta^2}+\frac{2n\,E_\theta(X)}{\theta^3}=-\frac{n}{\theta^2}+\frac{2n\theta}{\theta^3}=\frac{n}{\theta^2}.$$

Thus, The Cramer-Rao's inequality for the estimator $\bar{X}_n$ is

$$\frac{\theta^2}{n}=Var_\theta\left(\bar{X}_n\right)\geq\frac{\left[\frac{d}{d\theta}E_\theta\left(\bar{X}_n\right)\right]^2}{-n\;E_\theta\left(\frac{\partial^2}{\partial\theta^2}\left[\ln\left(f\left(X;\theta\right)\right)\right]\right)}=\frac{1}{\left(\frac{n}{\theta^2}\right)}=\frac{\theta^2}{n}.$$

That is, in the Cramer-Rao's inequality "=" holds instead of "≥" and therefore the estimator $\bar{X}_n$ is the UMVUE estimator for θ.

Cramer-Rao's inequality is not applicable for some distributions especially if the range of the distribution depend on the parameter (like uniform distribution) we can not apply Cramer-Rao's inequality. The probability density function of the uniform distribution $f(x;\theta)=(1/\theta)\,I_{(0<x<\theta)}(x)$ and logarithm (and therefore the derivative) is undefined. Thus, we can not find the UMVUE estimator for uniform parameter θ by using Cramer-Rao's inequality.

For the case where Cramer-Rao's inequality is not applicable, we use the following method to find the UMVUE estimator. First remember that for the random variables X and Y we have

$E(X)=E(X\,|\,Y))$ and $Var(X)=E(Var(X\,|\,Y))+Var(E(X\,|\,Y))$.

The following theorem is a useful tool to find the UMVUE estimator.

Theorem *(Rao-Blackwell)* Let $X_1, X_2,\ldots, X_n$ be a random sample from a population with parameter θ. Let W be any unbisaed estimator for $\tau(\theta)$ and T be a sufficient estimator for θ. In this case the estimator $\varphi(T)=E(W\,|\,T)$ is a better unbiased estimator for $\tau(\theta)$. That is,

$E_\theta(\varphi(T))=\tau(\theta)$ and $Var_\theta(\varphi(T))\leq Var_\theta(W)$.

Proof: Since T is sufficient the conditional probability of X's given T does not depen on the parameter and thefore the conditional expectation $\varphi(T)=E(W\,|\,T)$ does not depend on θ. That is, $\varphi(T)$ is an estimator. On the other hand, the mean of $\varphi(T)$ is calculated as

$$E_\theta(\varphi(T))=E_\theta\left[E(W\,|\,T)\right]=E_\theta(W)=\tau(\theta).$$

That is, $\varphi(T)$ is unbiased for $\tau(\theta)$. Since, $E_\theta(Var(W\,|\,T))\geq 0$ the variance of W can be written as

$$Var_\theta(W)=Var_\theta(\varphi(T))+E_\theta(Var(W\,|\,T))\geq Var_\theta(\varphi(T))$$

and therefore $Var_\theta\left(\varphi(T)\right) \le Var_\theta\left(W\right)$ which completes the proof.

According to this theorem, when we find a sufficient estimator (say T) for θ (by factorization theorem) and an unbiased estimator (say W) for $\tau(\theta)$ we can always find a "better" unbiased estimator for $\tau(\theta)$. Of course, we can find may unbiased estimator for $\tau(\theta)$. Our goal is to find a unique unbiased estimator among all unbiased astimator of $\tau(\theta)$. The following theorem guarantees the unique best unbiased (the most efficien unbiased) estimator. This requires the completeness of the sufficient estimator that we are not going to discuss here.

Theorem *(Lehmann-Scheffe Uniqueness Theorem)*: If W is any UMVUE estimator $\tau\left(\theta\right)$ then it is unique.

Theorem Under the conditions of Rao-Blackwell Theorem if the sufficient estimator T is also complete then the estimator $\varphi(T) = E(W \mid T)$ is the unique UMVUE estimator for $E_\theta(W)$.

(For the proof See Casella ve Berger, 2002, page 347).

Example a) Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution with parameter θ. As we remember, we could not apply Cramer-rao's inequality for the uniform population. Note that the estimator $T = X_{(n)}$ is sufficient for θ (se the example in the sufficieny part (c) above). On the other hand the estimator $W = (n+1)X_{(n)}/n$ is unbiased for θ. Therefore by Rao-Blackwell Theorem $\varphi(T) = E\left(W|T\right)$ is the unique UMVUE estimator for θ (completeness of $X_{(n)}$ is verified) : That is, the UMVUE estimator for θ is

$$\varphi(T) = E\left(W|T\right) = E\left(\frac{n+1}{n} X_{(n)} \middle| X_{(n)}\right) = \frac{n+1}{n} X_{(n)}.$$

b) Let $X_1, X_2, \ldots, X_n$ be a random sample from Bernoulli distribution with the parameter p and let us try to find the UMVUE estimator for the variance ($\tau(p) = p(1-p)$) of Bernoulli distribution.

Note that the estimator $T = \sum_{i=1}^{n} X_i$ is sufficient for p and the it distribution of T is Binomial (that is, $T \sim Binom(n, p)$, it is also complete). That is T is complete and sufficient estimator for p. Since $E(S_n^2) = p(1-p)$, S_n^2 is unbiased for $p(1-p)$. By Rao-Blackwell Theorem, $E(S_n^2 \mid T)$ is the UMVUE estimator for $p(1-p)$. Now, we need to calculate the conditional expectation. Since

X distributed as Bernoulli, the random variable takes the values only 0 or 1 and therefore we have $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i^2$. Therefore the conditional expectation is

$$E\left(S_n^2 \middle| T\right) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 \middle| \sum_{i=1}^{n} X_i\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2 \middle| \sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n-1}E\left(\sum_{i=1}^{n} X_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2 \middle| \sum_{i=1}^{n} X_i\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = S_n^2.$$

According to Rao-Blackwell Theorem S_n^2 is the unique UMVUE estimator for $\tau(p) = p(1-p)$.

Now, for the sample sample, let us try to find the UMVUE estimator for $\tau(p) = p^2$. Note that the estimator T is sufficient and complete. Moreover, if we define an estimator W as

$$W = \frac{1}{n(n-1)}\left[\left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i\right]$$

it is unbiased for p^2 because

$$E_p(W) = \frac{1}{n(n-1)}E_p\left[\left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i\right] = \frac{1}{n(n-1)}E_p\left[E_p\left(\sum_{i=1}^{n} X_i\right)^2 - E_p\left(\sum_{i=1}^{n} X_i\right)\right]$$

$$= \frac{1}{n(n-1)}\left[Var_p\left(\sum_{i=1}^{n} X_i\right) + \left[E_p\left(\sum_{i=1}^{n} X_i\right)\right]^2 - E_p\left(\sum_{i=1}^{n} X_i\right)\right]$$

$$= \frac{1}{n(n-1)}\left[np(1-p) + n^2p^2 - np\right] = \frac{1}{n(n-1)}\left[np - np^2 + n^2p^2 - np\right]$$

$$= \frac{1}{n(n-1)}\left[-np^2 + n^2p^2\right] = \frac{p^2(n^2-n)}{n(n-1)} = \frac{p^2 n(n-1)}{n(n-1)} = p^2.$$

According to Rao -Blackwell Theorem

$$E\left(W \middle| T\right) = E\left(\frac{1}{n(n-1)}\left[\left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i\right] \middle| \sum_{i=1}^{n} X_i\right) = \frac{1}{n(n-1)}\left[\left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i\right] = W$$

is the unique UMVUE estimator for p^2.

We use Rao-Blackwell Theorem to find the UMVUE estimators of a parameter or its functions.

Example: **a)** Assume thatthe number of customers enter a store in Kızılay for a certain time period is distributed as Poisson with parameter θ. On the same time period the number of customers in n different day are $X_1, X_2, \ldots, X_n$. That is we have a random sample from $Poisson(\theta)$ distribution. The probability function of the sample is $P_\theta(X = x) = e^{-\theta}\theta^x / x!$ for $x = 0,1,2,\ldots$. Using the Cramer-Rao's inequality the UMVUE estimator for θ is $\bar{X}_n$. Suppose, we want to estimate the probability that no customers will come to store at the same time period. That is, we want to estimate $\tau(\theta) = P_\theta(X = 0) = e^{-\theta}$. Note that $T = \sum_{i=1}^{n} X_i$ is sufficient and complete for the parameter θ. An unbiased estimator for $\tau(\theta)$ can be choosen as

$$W = \begin{cases} 1 & , \quad X_1 = 0 \\ 0 & , \quad d.y. \end{cases}$$

Note that the randomn variable W takes only the values 0 or 1 and therefore W is distributed as Bernoulli and therefore the espected value of W is

$$E_\theta(W) = 1P(W = 1) + 0\,P(W = 0) = P(W = 1) = P(X_1 = 0) = e^{-\theta}.$$

That is W is unbiased for $\tau(\theta) = e^{-\theta}$. By Rao-Blackwell Theorem, $\varphi(T) = E(W \mid T)$ is the unique UMVUE estimator for $\tau(\theta) = e^{-\theta}$. Now, we need to calculate this conditional expectation. Remember that $T \sim Poisson(n\theta)$ for $t = 0,1,2,\ldots$. That is the probability function of T is

$$P_\theta(T = t) = e^{-n\theta}(n\theta)^x / t!.$$

Note also that since X_1 is independent wiy-th the random variables $X_2, X_3, \ldots, X_n$ we also have $X_2 + X_3 + \ldots + X_n \sim Poisson((n-1)\theta)$. Therefore the conditional expectation is

$$\begin{aligned}
\varphi(t) &= E\left(W \middle| T = t\right) = 0\,P\left(W = 0 \middle| T = t\right) + 1P\left(W = 1 \middle| T = t\right) = P\left(W = 1 \middle| T = t\right) \\
&= P\left(X_1 = 0 \middle| T = t\right) = \frac{P_\theta\left(X_1 = 0,\, T = t\right)}{P_\theta\left(T = t\right)} = \frac{P_\theta\left(X_1 = 0,\, X_1 + X_2 + \ldots + X_n = t\right)}{P_\theta\left(X_1 + X_2 + \ldots + X_n = t\right)} \\
&= \frac{P_\theta\left(X_1 = 0\right) P_\theta\left(X_2 + \ldots + X_n = t\right)}{P_\theta\left(X_1 + X_2 + \ldots + X_n = t\right)} = \frac{\left[e^{-\theta}\right]\left[e^{-(n-1)\theta}\left((n-1)\theta\right)^T / t!\right]}{\left[e^{-n\theta}\left(n\theta\right)^t / t!\right]} \\
&= \frac{e^{-n\theta}\theta^t / t!}{e^{-n\theta}\theta^t / t!} \frac{(n-1)^t}{n^t} = \frac{(n-1)^t}{n^t} = \left(1 - \frac{1}{n}\right)^t.
\end{aligned}$$

Therefore the UMVUE estimator for $\tau(\theta) = e^{-\theta}$ is

$$\varphi(T) = E(W|T) = \left(1-\frac{1}{n}\right)^T = \left(1-\frac{1}{n}\right)^{\sum_{i=1}^{n} x_i}.$$

Since this estimator can be approximated for large number of observations as $(1-a/n)^n \approx e^{-a}$ the UMVUE estimator for $e^{-\theta}$ can be approximated as

$$\varphi(T) = E(W|T) = \left(1-\frac{1}{n}\right)^T = \left(1-\frac{1}{n}\right)^{\sum_{i=1}^{n} X_i} = \left(1-\frac{1}{n}\right)^{n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)} = \left(1-\frac{1}{n}\right)^{n\,\bar{X}_n} \approx e^{-\bar{X}_n}$$

for large n. Since the UMVUE estimator of θ is $\bar{X}_n$, the UMVUE estimator of θ is approximately $e^{-\bar{X}_n}$ for large n.

Suppose the owner of the store wants to decide the openning hour in the morning. He/She counts the number of customers between 8:00- 9:00 o'clock for 10 days. Suppose he/she counts the number of customers for 10 days are

2 3 2 1 0 1 2 0 2 1.

based on these observed values the probability that no customers will come to the store between 8:00-9:00 o'clock in the morning is estimated as $(1-1/n)^{n\bar{x}_n} \cong 0.229$. The mean number of customers will come to the store between 8:00-9:00 is estimated as $\bar{x}_n = 1.4$. Moreover, the estimated probability is $(1-1/n)^{n\bar{x}_n} \cong 0.229$ and if we assume that 10 number of customers is large enough it can be estimated as $e^{-\bar{x}_n} \cong 0.246$. Note that these estimated probabilities are very close to each other. If we had more number of observations we get much closer estimated probabilities.

b) Now for the same example suppose we want to estimate the probability that only one customer will come to the store between 8:00 to 9:00 o'clock in the morning. That is we want to estimate $\tau(\theta) = \theta e^{-\theta}$ or we want to find the UMVUE estimator for $\tau(\theta) = \theta e^{-\theta}$. Again, T is a sufficient and complete estimator for θ and an unbiased estimator for $\tau(\theta)$ can be choosen

$$W = \begin{cases} 1 & , \quad X_1 = 1 \\ 0 & , \quad d.y. \end{cases}$$

Again, the random variable (the unbiased estimator for $\tau(\theta)$) W takes only the values 0 or 1 which is a bernoulli random variable and the expected value of W can be calculated as

$$E_\theta(W) = 1P(W=1) + 0P(W=0) = P(W=1) = P(X_1=1) = \theta e^{-\theta}.$$

That is, the estimator W is unbiased for $\tau(\theta)$. next we need to calculate the conditional expectation. as This contditional expectation can be calculated as,

$$\varphi(t) = E\left(W \middle| T = t\right) = 0\,P\left(W = 0 \middle| T = t\right) + 1P\left(W = 1 \middle| T = t\right) = P\left(W = 1 \middle| T = t\right)$$

$$= P\left(X_1 = 1 \middle| T = t\right) = \frac{P_\theta\left(X_1 = 1, T = t\right)}{P_\theta\left(T = t\right)} = \frac{P_\theta\left(X_1 = 1,\, X_1 + X_2 + \ldots + X_n = t\right)}{P_\theta\left(X_1 + X_2 + \ldots + X_n = t\right)}$$

$$= \frac{P_\theta\left(X_1 = 1\right) P_\theta\left(X_2 + \ldots + X_n = t-1\right)}{P_\theta\left(X_1 + X_2 + \ldots + X_n = t\right)} = \frac{\left[\theta e^{-\theta}\right]\left[e^{-(n-1)\theta}\left((n-1)\theta\right)^{t-1} / (t-1)!\right]}{\left[e^{-n\theta}\left(n\theta\right)^{t} / t!\right]}$$

$$= \frac{e^{-n\theta}\theta^{T} \ / (t-1)!}{e^{-n\theta}\theta^{T} \ / t!} \frac{(n-1)^{t-1}}{n^T} = \frac{t!}{(t-1)!} \frac{(n-1)^{t-1}}{n^T} = \frac{t}{n-1}\left(1 - \frac{1}{n}\right)^t .$$

Therefore the UMVUE estimator for $\tau(\theta)$ can be written as,

$$\varphi(T) = E\left(W \middle| T\right) = \frac{T}{n-1}\left(1 - \frac{1}{n}\right)^T = \frac{\sum_{i=1}^{n} X_i}{n-1}\left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i} .$$

In a similar way, for large n the estimator can be approximated as

$$\varphi(T) = E\left(W \middle| T\right) = \frac{\sum_{i=1}^{n} X_i}{n-1}\left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i} = \left(\frac{n}{n-1}\right)\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\left(1 - \frac{1}{n}\right)^{n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)} \approx \bar{X}_n e^{-\bar{X}_n} .$$

Note that the UMVUE estimator of θ is $\bar{X}_n$ and the UMVUE estimator for $\tau(\theta)$ is approximated as $\tau(\bar{X}_n)$. For the same observations given in (a), the probability that only one customer will come to the store between 8:00-9:00 o'clock in the morning can be approximated as

$$\bar{x}_n e^{-\bar{x}_n} = (1.4)e^{-1.4} \cong 0.345 .$$

As it is seen the above examples, if the UMVUE estimator for θ is $\hat{\theta}_n$ then $g(\hat{\theta}_n)$ is approximately the UMVUE estimator for $g(\theta)$. This can not be true in general. However, this property is always valid for the MLE estimation that we are going to see next.

For example let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with the parameter θ. As we have seen earlier, the UMVUE estimator for θ is $\bar{X}_n$ (the sample mean). On the other hand, the UMVUE estimator for θ^2 is $W = (T^2 - T)/n^2$ where $T = \sum_{i=1}^{n} X_i$. Note that since $E_\theta[(T^2 - T)/n^2] = \theta^2$ the estimator $W = (T^2 - T)/n^2$ is unbiased for θ^2. Moreover, since T is suffivient and complete statistic, according to Rao-Blackwell Theorem the conditional expectation is calculated as

$$\varphi(T) = E\left(W \mid T\right) = E\left(\left(T^2 - T\right)/n^2 \mid T\right) = \left(T^2 - T\right)/n^2 = W$$

and therefore the UMVUE estimator for θ^2 is

$$W_n = \left(T^2 - T\right)/n^2 = \frac{T^2}{n^2} - \frac{T}{n^2} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \bar{X}_n^2 - \frac{1}{n}\bar{X}_n .$$

That is, when the UMVUE estimator for θ is $\bar{X}_n$, the UMVUE estimator for $\tau(\theta) = \theta^2$ is not $\tau(\bar{X}_n)$. But for large n the UMVUE estimator of $\tau(\theta)$ is $\tau(\bar{X}_n)$. On the other hand, since $\bar{X}_n = O_P(1/\sqrt{n})$ for large n the estimator W_n can be approximated as $W_n \approx \bar{X}_n^2$ dir.