WEEK 11

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 $Var(T_1) \leq Var(T_2)$ 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

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

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

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

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 $E_\theta(W)$.

Example 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

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

and the mean of the secon derivative

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

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

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

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

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

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

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

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

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 "$\geq$" 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

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

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

Theorem *(Rao-Blackwell)* Let $X_1, X_2, \dots, X_n$ 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 $\varphi(T) = E(W \mid T)$ 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 $\varphi(T) = E(W \mid T)$ does not depend on θ. That is, $\varphi(T)$ is an estimator. On the other hand, the mean of $\varphi(T)$ is calculated as

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

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

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

and therefore $Var_\theta\left(\varphi\left(T\right)\right) \leq Var_\theta\left(W\right)$ 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 $\tau(\theta)$ 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 $\varphi(T) = E(W|T)$ is the unique UMVUE estimator for θ (completeness of $X_{(n)}$ is verified) : That is, the UMVUE estimator for θ is

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

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

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

That is W is unbiased for $\tau(\theta)=e^{-\theta}$. By Rao-Blackwell Theorem, $\varphi(T)=E(W\,|\,T)$ 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

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

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

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

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

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

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

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

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,

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

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

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

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

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

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

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.

Asymptotic Normality:

As we have mentioned in many times, the normality is the most important property in order to make any statistical inference. As we have studied in the previous section, according to the central limit theorem the sample mean is distributed as normal in the limit. Let us remember the central limit theorem one more time.

The Central Limit Theorem: Let $X_1, X_2 \ldots, X_n$ be a sequence of independent and identically distributed random variables with probability (or probability density) function $f(x;\theta)$, $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ such that $\sigma^2 < \infty$. Then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0,1) \text{ as } n \to \infty .$$

or

$$\frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \xrightarrow{D} N(0,1) \text{ as } n \to \infty .$$

Therefore, the sample mean is asymptotically distributed as normal and it is denoted as

$$\bar{X}_n \sim AN(\mu, \sigma^2 / n) .$$

Here, the notation "AN" stands for "asymptotic normal". In general, if $\bar{X}_n \sim AN(\mu, \sigma^2 / n)$ then $E(\bar{X}_n)$ may or may not μ (in general $E(\bar{X}_n) \neq \mu$) and similarly the variance of $\bar{X}_n$ may not be σ^2 / n. Tthat is, in general $Var(\bar{X}_n) \neq \sigma^2 / n$. We have a sample from $\bar{X}_n \sim AN(\mu, \sigma^2)$ population, it is obvious that $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2 / n$. Asymptotic normality is usually verified by the central limit theorem but there are cases that the estimators may asymptotically distributed as normal which can be verified by the "Taylor Series expansion" that we are not going to study in this section.

There are more statistical properties of estimators (like sufficiency, completeness and many more) but we are not going to study here. However we need to explain some of the methods of finding estimators. Again there are many methods of finding estimators but here we are going to study two methods (method of moments and maximum likelihood) in this part. Later we are going to study one more method (ordinary least squares) in the regression analysis.

Method of Finding Estimators:

a) Method of Moments Estimators:

Let $X_1, X_2 \dots, X_n$ be a random sample with probability (or probability density function $f(x;\underset{\sim}{\theta})$. Here, $\underset{\sim}{\theta}$ is the vector of unknown of the population. If there are k unknowns (say then θ_i, $i = 1,2,...,k$) then $\underset{\sim}{\theta} = (\theta_1, \theta_2, ..., \theta_k)'$. Usually $k = 1$ or $k = 2$ and rarely $k = 3$. In order to find the method of moment estimators of the parameters θ_i, we calculate k population moments which are functions of the parameters θ_i and k sample moments (we calculate as many moments as the number of parameters, here we assumed that the population has k parameters)

Population Moments	Sample Moments
$E(X) = g_1(\underset{\sim}{\theta}) = g_1(\theta_1, \theta_2, ..., \theta_k)'$	$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$
$E(X^2) = g_2(\underset{\sim}{\theta}) = g_2(\theta_1, \theta_2, ..., \theta_k)'$	$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$
$E(X^3) = g_3(\underset{\sim}{\theta}) = g_3(\theta_1, \theta_2, ..., \theta_k)'$	$m_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3$
.	.
.	.
$E(X^k) = g_k(\underset{\sim}{\theta}) = g_k(\theta_1, \theta_2, ..., \theta_k)'$	$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.

Let $\hat{\theta}_i$ be the method of moment estimator of θ_i for $i = 1, 2, ..., k$ and equate the population moments to the sample moments as

$$g_1(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k)' = m_1, \qquad g_2(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k)' = m_2, \ . \ . \ . \ , \ g_k(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k)' = m_k$$

and solve k unknowns from k equations. The solutions $\hat{\theta}_i$ will be a function of m_i for $i = 1, 2, ..., k$ namely,

$$\hat{\theta}_i = h_i(m_1, m_2, ..., m_k) \text{ for } i = 1, 2, ..., k .$$

These solutions are the method of moment estimators of the parameter θ_i. As it is mentioned before, usually there are one or two parameters to be interested in.

Example: **a)** Let $X_1, X_2 \dots, X_n$ be a random sample from a $Poisson(\theta)$ distribution. Since there is only one parameter to be estimated, in order to find the method of moment estimator for the parameter θ, we need to calculate first population moment and sample moment. As it is well known, $E(X) = \theta$ and the first sample moment is the sample mean and therefore the method of moment estimator of θ is $\overline{X}_n$, that is, $\hat{\theta} = \overline{X}_n$.

b) Let $X_1, X_2 \dots, X_n$ be a random sample from a $U(0, \theta)$ distribution. Again, there is only one parameter to be estimated. Therefore, we need one population moment and one sample moment in order to find the method of moment estimator of the parameter θ. As it is known, $E(X) = \theta / 2$ and the first sample moment is the sample mean and therefore the method of moment estimator of θ is $\overline{X}_n$. Thus, the method of moment estimator of θ will be the solution to the equation $\hat{\theta} / 2 = \overline{X}_n$. And the solution and therefore the method of moment estimator of θ is $\hat{\theta} = 2\overline{X}_n$.

c) Let $X_1, X_2 \dots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Now, there are two parameters to be estimated and therefore we need to calculate two population moments and two sample moments. Obviously, the population moments are $E(X) = \mu$, $E(X^2) = \mu^2 + \sigma^2$ and the sample moments are

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n \text{ and } m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 .$$

From the equations

$$\hat{\mu} = \overline{X}_n \quad , \quad \hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

the solutions are

$$\hat{\mu} = \bar{X}_n \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 .$$

Therefore, the method of moment estimators for the normal parameters μ and σ^2 are

$$\hat{\mu} = \bar{X}_n \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 .$$

b) <u>Maximum Likelihood estimators</u>:

The maximum likelihood estimators of the parameters are the ones that maximizes the likelihood function. Let $X_1, X_2 \dots, X_n$ be a random sample from a population with a probability (or probability density) function $f(x;\underset{\sim}{\theta})$. The likelihood function of $\underset{\sim}{\theta}$ is

$$L(\underset{\sim}{\theta};\underset{\sim}{X}) = f(\underset{\sim}{X};\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(X_i;\underset{\sim}{\theta}) .$$

The maximum likelihood estimator of $\underset{\sim}{\theta}$ is the one that maximizes $L(\underset{\sim}{\theta};\underset{\sim}{X})$. That is, if $\hat{\underset{\sim}{\theta}}$ is the maximum likelihood estimator of $\underset{\sim}{\theta}$ we can write

$$\hat{\underset{\sim}{\theta}} = \arg\max_{\underset{\sim}{\theta}\in\Theta} L(\underset{\sim}{\theta};\underset{\sim}{X}) .$$

In many cases, the maximization of the likelihood function is difficult and therefore instead of maximization of the likelihood function, the maximum likelihood estimators are found by maximizing the log-likelihood function ($\ell(\underset{\sim}{\theta}) = \log(L(\underset{\sim}{\theta};\underset{\sim}{X}))$). Because of the properties of the logarithm function, it is easy to see that

$$\hat{\underset{\sim}{\theta}} = \arg\max_{\underset{\sim}{\theta}\in\Theta} L(\underset{\sim}{\theta};\underset{\sim}{X}) = \arg\max_{\underset{\sim}{\theta}\in\Theta} \ell(\underset{\sim}{\theta}) .$$

Another important property of the maximum likelihood estimator (we will call MLE) is that if $\hat{\theta}$ is the MLE of a parameter θ the MLE of $h(\theta)$ is $h(\hat{\theta})$. That is, if $\bar{X}_n$ is the MLE of μ, the MLE of μ^2 is $\bar{X}_n^2$.

<u>Example</u>: **a)** Let $X_1, X_2 \dots, X_n$ be a random sample from a $Poisson(\theta)$ distribution. The probability function of the Poisson distribution is

$$P(X = x) = e^{-\theta}\theta^x / x! \quad , \quad x = 0,1,2,... \text{ and } \theta > 0$$

and therefore, the likelihood function of θ is

$$L(\underset{\sim}{\theta};\underset{\sim}{X}) = \prod_{i=1}^{n} P(X_i = x;\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{X_i}}{X_i!} = e^{-n\theta}\,\theta^{\sum_{i=1}^{n} X_i} \prod_{i=1}^{n} \frac{1}{X_i!} .$$

In a direct way maximization of this function seems to be difficult and therefore we use the log-likelihood function

$$\ell(\theta) = \log(L(\underset{\sim}{\theta}; \underset{\sim}{X})) = -n\theta + \log(\theta) \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \log(X_i!)$$

and the MLE of θ can be found by maximizing the log-likelihood function. Since it is a differentiable function of θ, the maximum can be found by differentiated with respect to θ and equating zero as

$$\frac{\partial \ell(\theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^{n} X_i}{\theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n .$$

That is, the log likelihood (therefore the likelihood) function has either a maximum or a minimum at the point $\hat{\theta} = \bar{X}_n$. In order to verify that it is a maximum we look at the second derivative at this point. If the second derivative is negative at this point then it is a maximum. Since,

$$\left. \frac{\partial^2 \ell(\theta)}{\partial \theta^2} \right|_{\theta=\bar{X}_n} = \left(-\left. \frac{\sum_{i=1}^{n} X_i}{\theta^2} \right| \right)_{\theta=\bar{X}_n} = -\frac{n\bar{X}_n}{\bar{X}_n^2} = -\frac{n}{\bar{X}_n} < 0$$

and therefore, the maximum likelihood estimator of θ is the sample mean. That is, $\hat{\theta} = \bar{X}_n$. Mathematically, in order to check that the function has a maximum at the point, it is important to look at the second derivative. From now on, we are not going to check the second derivative whenever the function is differentiable and we will assume that the solution is the maximum.

b) Let $X_1, X_2 \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution and let us try to find the maximum likelihood estimators of the normal parameters. The probability density function of $N(\mu, \sigma^2)$ distribution is given by

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

and therefore, the likelihood function

$$L(\mu, \sigma^2; \underset{\sim}{X}) = f(\underset{\sim}{x}; \mu, \sigma^2) = \prod_{i=1}^{n} f(X_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2} \right)$$

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

again, instead of maximizing this function, we maximize the log likelihood function. The log-likelihood function is given by,

$$\ell(\mu,\sigma^2) = \log(L\mu,\sigma^2;\underset{\sim}{X})) = -\frac{n}{2}\ln(2\pi - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

and the derivatives of the log-likelihood function (with respect to μ and σ^2 are

$$\frac{\partial\ell(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad , \quad \frac{\partial\ell(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

and equating these first derivatives to zero

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \qquad \text{and} \qquad -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

the solutions are

$$\hat{\mu}_n = \bar{X}_n \text{ and } \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 .$$

After checking the second derivative, the maximum likelihood estimators of μ and σ^2 are found to be

$$\hat{\mu}_n = \bar{X}_n \text{ and } \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 .$$

c) Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution with parameter θ. That is, $X_1, X_2, \ldots, X_n$ is a random sample from $U(0,\theta)$ population. The probability density function of the uniform distribution is given by

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

The probability density function can also be written as

$$f(x;\theta) = \frac{1}{\theta} I_{0<x\le\theta}(x) \text{ where } I_{\{0<x\le\theta\}}(x) = \begin{cases} 1 & , \quad 0 < x \le \theta \\ 0 & , \quad \textit{elsewhere.} \end{cases} \text{ and } x_{(n)} = \max\{x_1, x_2, ..., x_n\} .$$

Therefore the likelihood function of θ is

$$L(\theta \mid \underset{\sim}{X} = \underset{\sim}{x}) = \theta^{-n} I_{\{0<x_{(n)}\le\theta\}}(\underset{\sim}{x})$$

and as it is seen the likelihood function is not differentiable with respect to θ. However, it is a decreasing function of θ for $\theta \ge x_{(n)}$ and therefore the likelihood function takes the maximum value at the point $\theta = x_{(n)}$ and therefore the maximum likelihood estimator of θ is $\hat{\theta}_n = X_{(n)}$.