

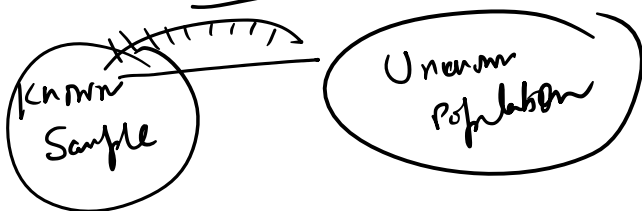
Statistical Inference



STATISTICA
L INFERE...

STATISTICAL INFERENCE

Sample known on Sample → Predict for the data



Formally, let x be a random variable describing the population under investigation. Suppose X has p.m.f $f_{\theta}(x) = P(x = x)$ or p.d.f $f_{\theta}(x)$ which depend on some unknown parameter θ (single or vector valued) that may have any value in a set Ω (called the parameters space). We assume that the functional form of $f_{\theta}(x)$ is known but not the parameter θ (except that $\theta \in \Omega$). For example, the family of distributions $\{f_{\theta}(x), \theta \in \Omega\}$ may be the family of Poisson distribution $\{P(\lambda), \lambda \geq 0\}$ or normal distribution $\{N(\mu, \sigma^2), -\infty < \mu < \infty, \sigma \geq 0\}$

Two problem of statistical inference are-

1. Estimate value of θ

2. Hypothesis about θ $\left\{ \begin{array}{l} H_0 \\ H_1 \end{array} \right.$

$\int \int$ Continuous
 \sum Discrete

Same (⊖)

POINT ESTIMATION

Definition: A random sample of size 'n' from the distribution of X is a set of independent and identically distributed random variables $\{x_1, x_2, \dots, x_n\}$ each of which has the same distribution as that of X. The probability of the sample is given by

$$f_0(x_1, x_2, \dots, x_n) = f_0(x_1)f_0(x_2) \dots f_0(x_n)$$

Definition: A statistic $T = T(x_1, x_2, \dots, x_n)$ is any function of the sample values, which does not depend on the unknown parameter θ . Evidently, T is a random variable which has its own probability distribution (called the 'Sampling distribution' of T)

$$s^2 = \frac{1}{(n-1)} \sum (x_i - \bar{x})^2$$
$$X_{(1)} = \min(x_1, x_2, \dots, x_n)$$
$$X_{(n)} = \max(x_1, x_2, \dots, x_n)$$

6:42:13 be
6:42:16 84

Theorem : Let (X_1, X_2, \dots, X_n) be a random sample of 'n' observations on X with mean $E(X) = \mu$ and variance $Var(x) = \sigma^2$. Let the sample mean and sample variance be $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

Then,

(i) $E(\bar{X}) = \mu$ ✓

(ii) $V(\bar{X}) = \frac{\sigma^2}{n}$ ✓

(iii) $E(S^2) = \frac{n-1}{n} \sigma^2$ ✓

Handwritten notes:

- $n > 30$ Large χ^2, Z
- $n < 30$ Small t
- Case specific Statistical Analysis
- Non-parametric methods
- $n = 2$
- $n = 10$
- $n = 60$
- $n = 100$



PROPERTIES OF ESTIMATORS

UNBIASEDNESS:

An estimator T of an unknown parameter θ is called unbiased if

$$E(T) = \theta \text{ for all } \theta \in \Omega$$

Q. if (x_1, x_2, \dots, x_n) is a random sample from a normal distribution $N(\mu, 1)$ show that $T = \frac{1}{n} \sum x_i^2 - 1$ is an unbiased estimator of μ^2 .

So $E(T) = E\left[\frac{1}{n} \sum x_i^2 - 1\right] = \frac{1}{n} \sum E(x_i^2) - 1$

$$E(x_i^2) = V(x) + E(x)^2 = 1 + \mu^2$$
$$= \frac{1}{n} \sum (\mu^2 + 1) - 1 = \mu^2$$

T_{hr} $\frac{6.30}{12/13}$ $\frac{7.30}{1}$ $\frac{8.30}{1}$

$\frac{7}{8} \text{ km}$

Q2

Example Let (x_1, x_2, \dots, x_n) be a random sample of observation from a Bernoulli distribution

$f_\theta(x) = \theta^x (1-\theta)^{1-x}$ ($x = 0, 1$) show that $T = \frac{y(y-1)}{n(n-1)}$ is an unbiased estimator of θ^2 where $y = \sum_{i=1}^n x_i$

$E(x) = \theta$, $v(x) = \theta(1-\theta) = \theta - \theta^2$

$E(y) = n\theta$, $v(y) = n\theta(1-\theta)$

$E[y(y-1)] = E(y^2) - E(y) = v(y) + [E(y)]^2 - E(y)$

$= n\theta(1-\theta) + n^2\theta^2 - n\theta$

$= n(n-1)\theta^2$

$E(T) = E\left[\frac{y(y-1)}{n(n-1)}\right] = \theta^2$

Unbiased Estimator of θ^2 .

Example: Show that the mean \bar{x} of a random sample of size n from the exponential distribution $f_\theta(x) = \frac{1}{\theta} e^{-x/\theta}$ ($x > 0$) is an unbiased estimator of θ and has variance θ^2/n

$E(x) = \theta$ $V(x) = \theta^2$
 $E(\bar{x}) = \frac{\theta n}{n} = \theta$ $V(\bar{x}) = \theta^2/n$

$E(x) = \theta/2$ $V(x) = \theta^2/n$
 $E(\tau_1) = E\left(\frac{\sum x_i}{n}\right) = \theta$ $V(\tau_1) = \frac{\theta^2}{3n}$

Example Let (x_1, x_2, \dots, x_n) be a random sample from the rectangular distribution $R(0, \theta)$ having

b. d. f $f_\theta(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \ (\theta > 0) \\ 0, & \text{otherwise} \end{cases}$

Show that $T_1 = 2\bar{x}$, $T_2 = \frac{n+1}{n}Y_n$ and $T_3 = (n+1)Y_1$ are all unbiased for θ , where $Y_1 = \min(x_1, x_2, \dots, x_n)$ and $Y_n = \max(x_1, x_2, \dots, x_n)$

To obtain the expectation of T_1, T_2 we need the pdf of Y_n

pdf of Y_n $F_y(y) = P(Y_n \leq y)$
 $= P[\max(x_1, x_2, \dots, x_n) \leq y]$

$= P(x_1 \leq y, \dots, x_n \leq y)$
 $= [P(x \leq y)]^n$

$\Rightarrow \frac{y}{\theta} = \frac{y^n}{\theta^n}$

pdf of Y_n $g_{Y_n}(y) = \frac{ny^{n-1}}{\theta^n}$ $0 \leq y \leq \theta$

$$= P(x_1 \leq y, \dots, x_n \leq y)$$

$$= [P(x \leq y)]^n$$

$$\Rightarrow \frac{y}{\theta} = \frac{y^n}{\theta^n}$$

pdf of Y_n

$$g_{Y_n}(y) = \frac{ny^{n-1}}{\theta^n} \quad \underline{0 \leq y \leq \theta}$$

$$E(Y_n) = \int_0^{\theta} \frac{ny^n}{\theta^n} \cdot y = \frac{(n)\theta}{(n+1)}$$

$$\text{or, } E\left(\frac{n+1}{n} Y_n\right) = \theta$$

T_2 is an UE of θ

Similar for T_3 ...

Q $T \rightarrow$ UFE of θ
 Does it imply that T^2 & \sqrt{T} are also UFE of θ

$\theta^2, \sqrt{\theta}$??

$Y(n)$

0906 2395723

!

$E(T^2) = \theta^2$
 $\text{Var}(T) = 0$

So, $P(T = \theta) = 1$

which is impossible since T has to be independent of θ

$\text{Var}(\sqrt{T}) = E(T) - (E(\sqrt{T}))^2$
 $E(\sqrt{T}) = \sqrt{\theta}$ so $\text{Var}(\sqrt{T}) = 0$

$P(\sqrt{T}) = P(\theta) = 1$
 $= P(T = \theta)$

\rightarrow Hence no \leftarrow

Example: Let (x_1, x_2, \dots, x_n) be a random variable from the Rectangular distribution $R(\theta, 2\theta)$ having p, d, f

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & \theta \leq x \leq 2\theta \\ 0, & \text{elsewhere} \end{cases}$$

Show that

$$T_1 = \frac{n+1}{2n+1} x_{(n)}, T_2 = \frac{n+1}{n+2} x_{(1)}$$

And

$$T_3 = \frac{n+1}{5n+4} [2x_{(n)} + x_{(1)}] \text{ and } T_4 = \frac{2}{3} \bar{x} \text{ are all unbiased}$$

Unbiased estimators.

Lovey love home.

$y_2 = \theta/2$ $y_3 = \theta$ $y_1 = \theta/4$
 $E_1 = \theta/4 = y_1$

Example: Let y_1, y_2, y_3 be the order statistics of a random sample of size 3 from a uniform distribution having p.d.f. $f(x, \theta) = \frac{1}{\theta} (0 \leq x \leq \theta)$ show that $4y_1, 2y_2, \frac{4}{3}y_3$ are all unbiased estimator of θ . Also obtain their variance.

$f_{y_1}(y) = \frac{3(\theta - y)^2}{\theta^3} \Rightarrow 0 \leq y \leq \theta$
 $f_{y_2}(y) = \frac{6y(\theta - y)}{\theta^3} \Rightarrow 0 \leq y \leq \theta$
 $f_{y_3}(y) = \frac{3y^2}{\theta^3} \Rightarrow 0 \leq y \leq \theta$

$E(y_1) = \theta/4, \quad E(y_2) = \theta/2, \quad E(y_3) = \frac{3}{4}\theta$
 $V(y_1) = 3\theta^2/80, \quad V(y_2) = \theta^2/20, \quad V(y_3) = 3\theta^2/80$
 y_1, y_2, y_3 are UE with minimum variance

$Y = k_1 y_1 + k_2 y_2 + k_3 y_3$

$k_i = \frac{1}{\theta_i^2} / \sum (1/\theta_i^2)$
 i.e. $y = \frac{\left(\frac{y_1}{\theta_1^2} + \frac{y_2}{\theta_2^2} + \dots + \frac{y_n}{\theta_n^2} \right)}{\left(\frac{1}{\theta_1^2} + \frac{1}{\theta_2^2} + \dots + \frac{1}{\theta_n^2} \right)}$

Example: Let 'T' be an unbiased estimator of θ . Does it imply that T^2 and \sqrt{T} , are unbiased for θ^2 and $\sqrt{\theta}$ respectively?

Already solved

Already solved above



MVUE

$$P\{|\tau - \theta| \leq \varepsilon\} \geq 1 - \frac{\text{Var}(\tau)}{\varepsilon^2}$$

Example let y_1, y_2 , be independent unbiased estimator of θ , having finite variance (σ_1^2, σ_2^2 , say). Obtain a linear combination of y_1, y_2 which is unbiased and has the smallest variance.

Remarks: (i) An unbiased estimator may not exist. Let x be a random variable with Bernoulli distribution.

$$f_x(x) = \theta^x(1-\theta)^{1-x}, x = 0, 1$$

It can be shown that no unbiased estimator exists for θ^2 .

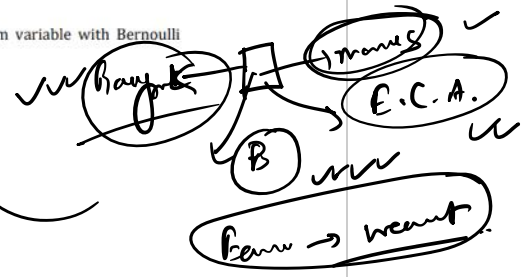
(ii) Unbiased estimator may not exist.

Let X be a random variable having Poisson distribution $P(x)$ and suppose we want estimator $g(\lambda) = e^{-3\lambda}$. Consider a sample of one observation and the estimator $T = (-2)^x$. Then $E(T) = e^{-3\lambda}$ so that T is an unbiased estimator of $e^{-3\lambda}$ but $T(x) = (-2)^x$ for x even and $T(x) < 0$ for x odd, which is absurd since $e^{-3\lambda}$ is always positive.

(iii) Instead of the parameter θ we may be interested in estimating a function $g(\theta)$. $g(\theta)$ is said to be 'estimable' if there exists an estimator T such that $E(T) = g(\theta), \theta \in \Omega$.

Minimum Variance Unbiased (MVU) estimators: The class of unbiased estimators may, in general, be quite large and we would like to choose the best estimator from this class. Among two estimators of θ which are both unbiased, we would choose the one with smaller variance. The reason for doing this rests on the interpretation of variance as a measure of concentration about the mean. Thus, if T is unbiased for θ , then by Chebyshev's inequality-

$$P\left\{\left|T - \theta\right| \leq \epsilon\right\} \geq 1 - \frac{\text{var}(T)}{\epsilon^2}$$

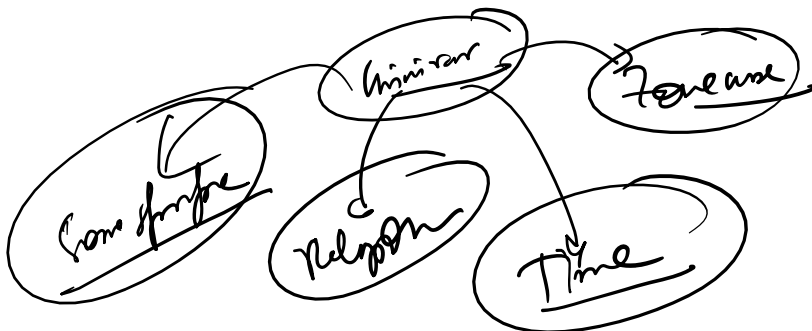
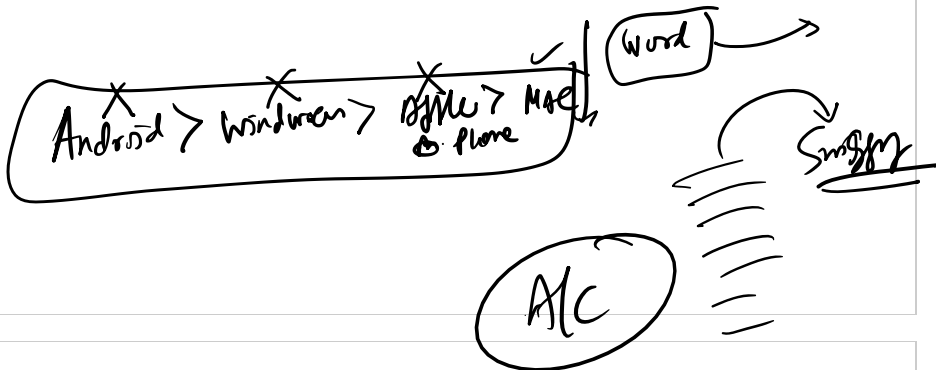
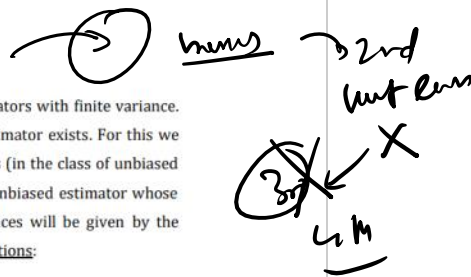


Therefore, the smaller $Var(T)$ is, the larger the lower bound of the probability of concentration of T about θ becomes. Consequently, within the restricted class of unbiased estimators we would choose the estimator with the smallest variance.

Definition: An estimator $T = T(X_1, \dots, X_n)$ is said to be a uniformly minimum variance unbiased

(UMVU) estimator of θ (or an estimator for $g(\theta)$) if it is unbiased and has the smallest variance within the class of unbiased estimators of θ (or $g(\theta)$), of all $\theta \in \Omega$. That is if T is any other unbiased estimator of θ , then-

Suppose we decide to restrict ourselves to the class of all unbiased estimators with finite variance. The problem arises as to how we find an UMVU estimator, if such an estimator exists. For this we would first determine a lower bound for the variances of all estimators (in the class of unbiased estimators under consideration) and then would try to determine an unbiased estimator whose variance is equal to this lower bound. The lower bound for the variances will be given by the Cramer-Rao inequality for which we assume the following regularity conditions:



Cramer-Rao inequality: Let (X_1, \dots, X_n) be a random sample of n observations on X with p.d.f $f(x; \theta)$ and suppose the above regularity conditions hold. If T is any unbiased estimator of θ , then-

$$\text{Var}(T) \leq \frac{1}{nE\left[\frac{\partial}{\partial\theta} \log f(x; \theta)\right]^2}$$















Remark: (i)

(ii) If $g(\theta)$ is an estimable function for which an unbiased estimator is T (i.e. $E(T) = g(\theta)$) then C.R Inequality becomes-

$$V(T) \geq \frac{[g'(\theta)]^2}{nE \left[\frac{\partial}{\partial \theta} \log f(x; \theta) \right]^2}$$

(iii) It can be show that

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

(iv) If an unbiased estimator exists which is such that its variance is equal to the lower bound $CRB = \frac{1}{nE \left[\frac{\partial}{\partial \theta} \log f(x; \theta) \right]^2}$ then it will be UMVUE.

(v) If there is no unbiased estimator whose variance equals the C.R.B it does not mean that UMVUE will not exist. Such estimators can be found (if these exists) by other methods.

(vi) In case of distributions not satisfying the regularity conditions (e.g.: Rectangular distribution) UMVU estimators, if these exists can be found by other methods. For such cases UMVU estimator may have variance less than CRB.

Example: Let (x_1, \dots, x_n) be a random sample from a Bernoulli distribution $f(x; \theta) = \theta^x(1-\theta)^{1-x}$ ($x = 0, 1, 0 < \theta < 1$)

Show that $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is a UMVU of θ

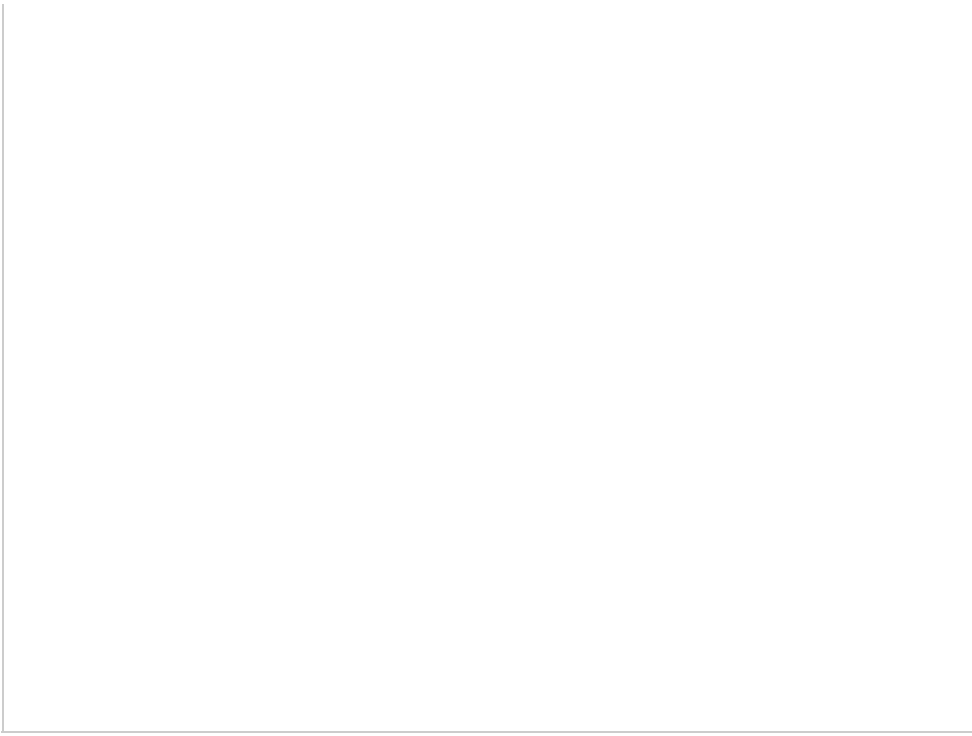
Let $f(x; \theta) = \theta^x + (1-\theta)^{1-x}$

$$\frac{\partial}{\partial \theta} \log f(x; \theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$
$$E \left[\frac{\partial}{\partial \theta} \log f(x; \theta) \right]^2 = \frac{E(x-\theta)^2}{\theta^2(1-\theta)^2} = \frac{\theta - \theta}{\theta(1-\theta)} = \frac{0}{\theta(1-\theta)}$$
$$= \frac{\theta(1-\theta)}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

$$CRB = \frac{\theta(1-\theta)}{n}$$

$$E(\bar{x}) = \theta \quad \text{Var}(\bar{x}) = \frac{\theta(1-\theta)}{n}$$

Here, \bar{x} is UMVUE of θ .



$$C^m_x \quad \binom{m}{x} \quad C(m, x)$$

Example: Let x be a random sample having Binomial distribution

$$f(x, \theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x}; \quad x = 0, 1, \dots, m \quad (0 < \theta < 1)$$

Show that \bar{x}/m is UMVUE of θ .

$$\ln f(x, \theta) = \ln \binom{m}{x} + x \ln \theta + (m-x) \ln(1-\theta)$$

$$\frac{\partial}{\partial \theta} \ln f(x, \theta) = \frac{x}{\theta} - \frac{m-x}{1-\theta}$$

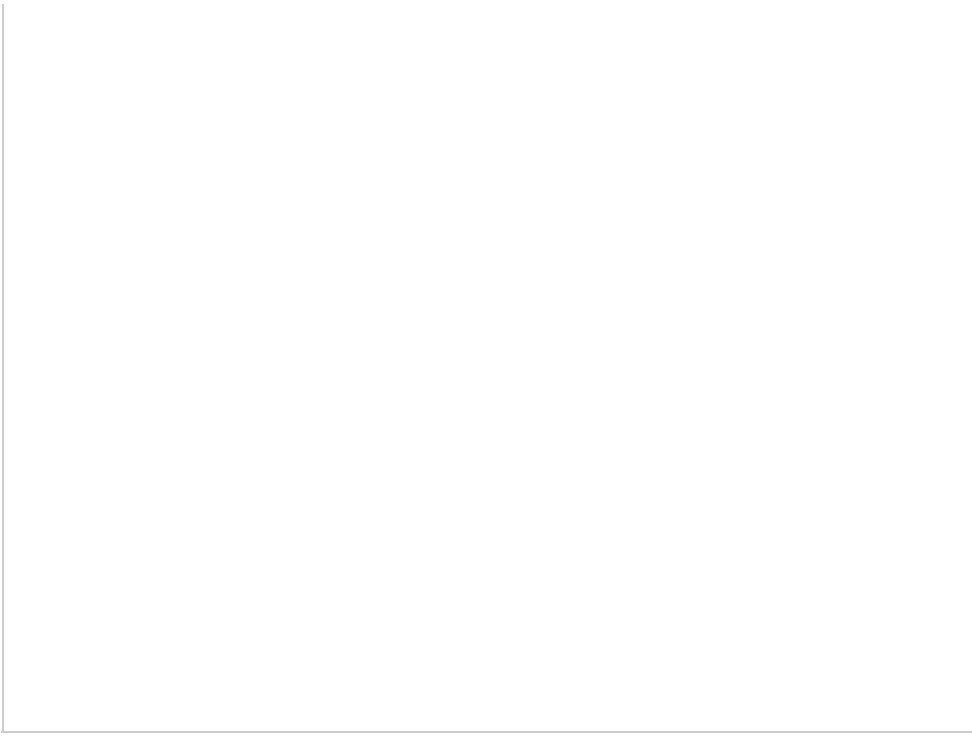
$$= \frac{x - m\theta}{\theta(1-\theta)}$$

$$\text{So, that, } E \left[\frac{\partial}{\partial \theta} \ln f(x, \theta) \right]^2 = \frac{E(x - m\theta)^2}{\theta^2(1-\theta)^2}$$

$$= \frac{m\theta(1-\theta)}{\theta^2(1-\theta)^2} = \frac{m}{\theta(1-\theta)}$$

$$E(\bar{X}_m) = \theta, \quad \text{Var}(\bar{X}_m) = \frac{\theta(1-\theta)}{m}$$

So, that, \bar{X}/m is UMVUE of θ ..



Example: Let (x_1, \dots, x_n) be a random sample from a Poisson distribution

$$f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}; \quad x = 0, 1, \dots \dots (\theta > 0)$$

Show that \bar{x} is UMVUE of θ .



