Exam Coaching: Reading Course Inference

Heba Alhosainy

Technische Universität Dortmund

24. February - 25. February, 02. March - 04. March 2022

Table of contents

Method of moments

The Method of Moments (MoM)

It consists of equating sample moments and population moments. If a population has t parameters, the MOM consists of solving the system of equations for the t parameters.

$E(x^k)$ is the k^{th} (theoretical) moment of the distribution (about the origin), for $k = 1, 2, \ldots$

$E((x-\mu)^k)$ is the k^{th} (theoretical) moment of the distribution (about the mean), for $k = 1, 2, \ldots$

$M_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ is the k^{th} sample moment, for $k = 1, 2, \ldots$

$M_{k*} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$ is the k^{th} sample moment about the mean, for $k = 1, 2, \ldots$

Statistical model: Normal distribution

Let $X_1, \ldots, X_n$ normally distributed random variables with parameter μ and σ. What are the method of moments estimators of μ and σ?

the first and second theoretical moments are:

$$E(x) = \mu, \quad E(x^2) = Var(x) + E(x)^2 = \sigma^2 + \mu^2$$

the first and second sample moments are:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad M_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Solving for μ

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

Solving for σ

$$\begin{aligned}
\sigma^2 + \mu^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 \\
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2
\end{aligned}$$

Statistical model: Bernoulli distribution

Let $X_1, \ldots, X_n$ Bernoulli random variables with parameter p. What is the method of moments estimator of p?

the first theoretical moment about the origin is:

$$E(X) = p$$

the first sample moment is:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Solving for p

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

Statistical model: Binomial distribution

Let $X_1, \ldots, X_n$ Binomial random variables with parameters n and p. What are the method of moments estimators of n and p?

the first and second theoretical moments are:

$$E(X) = n\,p, \quad Var(X) = n\,p\,(1-p)$$

the first and second sample moments are:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad S_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Solving for n

$$\begin{aligned} n\,p &= \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \\ \widehat{n} &= \frac{\overline{x}}{\widehat{p}} \end{aligned}$$

Solving for p

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 &= n\,p\,(1-p) = \overline{x}\,(1-p) \\ \widehat{p} &= 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}{\overline{x}} \\ \widehat{n} &= \frac{\overline{x}^2}{\overline{x} - \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \end{aligned}$$

Statistical model: Gamma distribution

Your turn now :)

Let $X_1, \ldots, X_n$ Gamma random variables with parameters α and θ. What are the method of moments estimators of α and θ?

the first and second theoretical moments are:

$$E(X) = \alpha\ \theta, \quad Var(X) = \alpha\ \theta^2$$

the first and second sample moments are:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad S_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Solving for α

$$\begin{aligned}\alpha\,\theta &= \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}\\ \widehat{\alpha} &= \frac{\overline{x}}{\widehat{\theta}}\end{aligned}$$

Solving for θ

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 &= \alpha\,\theta^2 = \overline{x}\,\theta\\ \widehat{\theta} &= \frac{1}{n\,\overline{x}}\sum_{i=1}^{n}(x_i-\overline{x})^2\\ \widehat{\alpha} &= \frac{n\,\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\end{aligned}$$

Statistical model: Arbitrary distribution

Let $X_1, \dots, X_n$ random variables follow a distribution with parameter α, such that the pdf is $f(x|\alpha) = \alpha^{-2}\ x e^{\frac{-x}{\alpha}}$, $x > 0, \alpha > 0$.
What is the method of moments estimator of α?
Hint: $f(x|\theta) = \frac{1}{\theta}\ e^{\frac{-x}{\theta}}$, $\quad E(x^k) = \int_0^\infty \frac{x^k}{\theta}\ e^{\frac{-x}{\theta}}\,\mathrm{d}x = k!\ \ \theta^k$

the first theoretical moment is:

$$
\begin{aligned}
E(x) &= \int_0^\infty x\ \ \alpha^{-2} x\ e^{\frac{-x}{\alpha}}\,\mathrm{d}x \\
&= \frac{1}{\alpha}\int_0^\infty \frac{x^2}{\alpha}\ \ e^{\frac{-x}{\alpha}}\,\mathrm{d}x \\
&= \frac{1}{\alpha}\ (2!)\ \alpha^2 = 2\ \alpha
\end{aligned}
$$

the first sample moment is:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Solving for α

$$2\,\alpha = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

$$\widehat{\alpha} = \frac{1}{2}\,\overline{x}$$

Maximum likelihood

Given a random sample $X_1, X_2, \ldots, X_n$ from a population with parameter θ and density or mass $p(x_i|\theta)$, we have:
The Likelihood, $L(\theta)$,

$$L(\theta) = f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

Denote the unknown parameter by θ.

How should we estimate θ based on the sample data?

Choose the value of θ that yields the greatest probability of getting the observed data. The Maximum Likelihood Estimator, $\widehat{\theta}$

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} L(\theta) = \underset{\theta}{\operatorname{argmax}} \log L(\theta)$$

Likelihood

Assuming independent observations (a “random sample")

$$L(\theta) = \prod_{i=1}^{n} p(y_i|\theta) \text{ or } \prod_{i=1}^{n} f(y_i|\theta)$$

The likelihood is the probability of obtaining the observed data – expressed as a function of the parameter.

This is a standard calculus problem in maximizing a function.

It is usually more convenient to maximize the natural log of the likelihood.

$$\ell(\lambda) = \log L(\theta) = \log \prod_{i=1}^{n} p(x_i|\theta)$$

The answer is the same because $\log(x)$ is an increasing function.

The greater the likelihood, the greater the log likelihood.

The distributive law

$a(b + c) = ab + ac$. You may see this in a form like

$$\theta \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \theta x_i$$

Power of a product is the product of powers

$(ab)^c = a^c\, b^c$. You may see this in a form like

$$\left(\prod_{i=1}^{n} x_i\right)^{\alpha} = \prod_{i=1}^{n} x_i^{\alpha}$$

Multiplication is addition of exponents

$a^b a^c = a^{b+c}$. You may see this in a form like

$$\prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n \exp(-\theta \sum_{i=1}^{n} x_i)$$

Powering is multiplication of exponents

$(a^b)^c = a^{bc}$. You may see this in a form like

$$\left(e^{\mu t+\frac{1}{2}\sigma^2 t^2}\right)^n = e^{n\mu t+\frac{1}{2}n\sigma^2 t^2}$$

Log of a product is sum of logs

log **means natural log, base** e**, possibly denoted** ln **on your calculator**

$\log(ab) = \log(a) + \log(b)$. You may see this in a form like

$$\log \prod_{i=1}^{n} x_i = \sum_{i=1}^{n} \log x_i$$

Log of a power is the exponent times the log

$\log(a^b) = b\ \log(a)$. You may see this in a form like

$$\log(\theta^n) = n \log \theta$$

The log is the inverse of the exponential function

$\log(e^a) = a$. You may see this in a form like

$$\log\left(\theta^n \exp(-\theta \sum_{i=1}^{n} x_i)\right) = n \log \theta - \theta \sum_{i=1}^{n} x_i$$

Statistical model: Exponential distribution

Let $X_1, \ldots, X_n$ be a random sample (that is, independent and identically distributed) from an Exponential distribution with parameter λ

$$f(x) = \lambda e^{-x\lambda}, \quad x = 0, 1, 2, 3, ... \ \lambda > 0$$

Derive a formula for $\widehat{\lambda}$, the maximum likelihood estimate of λ.

Steps to find the MLE estimator: Step 1

Define the likelihood function $L(\lambda)$

The likelihood function of the Exponential distribution is:

$$L(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \lambda e^{-x_i \lambda}$$

Steps to find the MLE estimator: Step 2

Define the log likelihood function $\ell(\lambda)$

Maximizing the likelihood function is equivalent to maximizing the natural log of the likelihood function.

The natural log of the likelihood function of the Exponential distribution is:

$$
\begin{aligned}
\ell(\lambda) &= \log L(\lambda) = \log \prod_{i=1}^{n} f(x_i|\lambda) \\
&= \log \prod_{i=1}^{n} \lambda e^{-x_i \lambda} \\
&= \log \left(\lambda^n \;\; e^{-\lambda \sum_{i=1}^{n} x_i}\right) \\
&= n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i
\end{aligned}
$$

Steps to find the MLE estimator: Step 3

Take the derivative of the log likelihood function

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter

$$
\begin{aligned}
\frac{\partial \ell(\lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \log\left(\prod_{i=1}^{n} f(x_i|\lambda)\right) \\
&= \frac{\partial}{\partial \lambda}\left(n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i\right) \\
&= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i
\end{aligned}
$$

Steps to find the MLE estimator: Step 4

Set the derivative of the log likelihood function $\overset{!}{=} 0$

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter

$$\frac{\partial \ell(\lambda)}{\partial \lambda} \overset{!}{=} 0$$

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Steps to find the MLE estimator: Checking step

Take the second derivative of the log likelihood function

Check that the estimator is the maximum of the likelihood by taking the second derivative w.r.t the distribution parameter, It SHOULD be < 0

$$\begin{aligned}\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} &= \frac{\partial^2 \ell(\lambda)}{\partial \lambda^2}\left(\frac{n}{\lambda} - \sum_{i=1}^{n} x_i\right) \\ &= \frac{-n}{\lambda^2} < 0\end{aligned}$$

Statistical model: Poisson distribution

Let $X_1, \ldots, X_n$ be a random sample (that is, independent and identically distributed) from a Poisson distribution with parameter λ

$$p(x_i|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, 3, ... \ \lambda > 0$$

Derive a formula for $\widehat{\lambda}$, the maximum likelihood estimate of λ.

Steps to find the MLE estimator: Step 1

Define the likelihood function $L(\lambda)$

The likelihood function of a Poisson distribution is:

$$L(\lambda) = \prod_{i=1}^{n} p(x_i|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Steps to find the MLE estimator: Step 2

Define the log likelihood function $\ell(\lambda)$

Maximizing the likelihood function is equivalent to maximizing the natural log of the likelihood function.

The natural log of the likelihood function of a Poisson distribution is:

$$\begin{aligned}
\ell(\lambda) &= \log L(\lambda) = \log \prod_{i=1}^{n} p(x_i|\lambda) \\
&= \log \prod_{i=1}^{n} \frac{e^{-\lambda}\,\lambda^{x_i}}{x_i!} \\
&= \log \frac{(e^{-n\lambda})\;\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \log(e^{-n\lambda}) + \log(\lambda^{\sum_{i=1}^{n} x_i}) - \log(\prod_{i=1}^{n} x_i!)
\end{aligned}$$

n n

Steps to find the MLE estimator: Step 3

Take the derivative of the log likelihood function

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter

$$\begin{aligned}\frac{\partial \ell(\lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \log \left(\prod_{i=1}^{n} p(x_i|\lambda) \right) \\ &= \frac{\partial}{\partial \lambda} \left(-n\lambda + \sum_{i=1}^{n} x_i \log(\lambda) - \sum_{i=1}^{n} \log(x_i!) \right) \\ &= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}\end{aligned}$$

Steps to find the MLE estimator: Step 4

Set the derivative of the log likelihood function $\stackrel{!}{=} 0$

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter

$$\frac{\partial \ell(\lambda)}{\partial \lambda} \stackrel{!}{=} 0$$

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Steps to find the MLE estimator: Checking step

Take the second derivative of the log likelihood function

Check that the estimator is the maximum of the likelihood by taking the second derivative w.r.t the distribution parameter, It SHOULD be < 0

$$\begin{aligned}\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} &= \frac{\partial^2 \ell(\lambda)}{\partial \lambda^2}\left(-n + \frac{\sum_{i=1}^{n} x_i}{\lambda}\right) \\ &= \frac{-\sum_{i=1}^{n} x_i}{\lambda^2} < 0\end{aligned}$$

Numerical estimate

Find the maximum likelihood estimator of $P[X = 4]$, call it $\widehat{P}[X = 4]$.

Calculate an estimate using this estimator when $x_1 = 1, x_2 = 2, x_3 = 4, x_4 = 2$.

$$\widehat{\lambda} = 2.25,$$

$$\begin{aligned} \widehat{P}[X = 4] &= \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \frac{e^{(-2.25)}\ (2.25)^4}{4!} \\ &= 0.1126 \end{aligned}$$

Statistical model: Bernoulli Distribution

Your turn now :)

Let $Y_1, \ldots, Y_n$ as a random sample from a Bernoulli distribution. That is, independently for $i = 1, \ldots, n$,

$$p(y_i|\theta) = \theta^y (1-\theta)^{1-y}, \quad 0 < \theta < 1$$

for $y = 0$ or $y = 1$, and zero otherwise.

Derive a formula for $\widehat{\theta}$, the maximum likelihood estimate of θ.

Carry out the second derivative test.

The sample mean for a sample of $n = 49$ is $\overline{y} = 4.2$. Give a point estimate of θ. Your answer is a number.

Find the MLE of θ

Denoting the likelihood by $L(\theta)$ and the log likelihood by $\ell(\theta) = \log L(\theta)$, maximize the log likelihood by taking the first derivative w.r.t the parameter of the Bernoulli distribution.

$$\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \log \left(\prod_{i=1}^{n} p(y_i|\theta) \right) \\
&= \frac{\partial}{\partial \theta} \log \left(\prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i} \right) \\
&= \frac{\partial}{\partial \theta} \log \left(\theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{n-\sum_{i=1}^{n} y_i} \right) \\
&= \frac{\partial}{\partial \theta} \left((\sum_{i=1}^{n} y_i) \log \theta + (n - \sum_{i=1}^{n} y_i) \log(1-\theta) \right) \\
&= \frac{\sum_{i=1}^{n} y_i}{\theta} - \frac{n - \sum_{i=1}^{n} y_i}{1-\theta}
\end{aligned}$$

Setting the derivative to zero and solving

$\theta = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$

Second derivative test: $\frac{\partial^2 \log \ell}{\partial \theta^2} = -n \left(\frac{1-\overline{y}}{(1-\theta)^2} + \frac{\overline{y}}{\theta^2} \right) < 0$

Concave down, maximum, and the MLE is the sample proportion:
$\widehat{\theta} = \overline{y} = p$

Statistical model: Standard Normal Distribution

Let $X_1, \ldots, X_n$ as a random sample from a normal distribution. That is, independently for $i = 1, \ldots, n$, with mean μ and standard deviation σ:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

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

Derive a formula for $\widehat{\mu}$ and $\widehat{\sigma}$, the maximum likelihood estimates of μ and σ.

Steps to find the MLE estimator: Step 1

Define the likelihood function $L(\mu, \sigma)$

The likelihood function of the standard normal distribution is:

$$L(\mu, \sigma) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Steps to find the MLE estimator: Step 2

Define the log likelihood function $\ell(\mu, \sigma)$

Maximizing the likelihood function is equivalent to maximizing the natural log of the likelihood function.
The natural log of the likelihood function of a Poisson distribution is:

$$
\begin{aligned}
\ell(\mu, \sigma) &= \log L(\mu, \sigma) = \log \prod_{i=1}^{n} f(x_i \mid \mu, \sigma) \\
&= \log \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n - \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right) \\
&= n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)
\end{aligned}
$$

Steps to find the MLE estimator: Step 3

Take the derivative of the log likelihood function

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter μ, σ is considered as a constant

$$
\begin{aligned}
\frac{\partial \ell(\mu,\sigma)}{\partial \mu} &= \frac{\partial}{\partial \mu} \log \left(\prod_{i=1}^{n} f(x_i | \mu, \sigma) \right) \\
&= \frac{\partial}{\partial \mu} \left(n \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \left(\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right) \right) \\
&= \frac{\sum_{i=1}^{n} (x_i - \mu)}{\sigma^2}
\end{aligned}
$$

Steps to find the MLE estimator: Step 4

Set the derivative of the log likelihood function $\overset{!}{=} 0$

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter μ

$$\frac{\partial \ell(\mu, \sigma)}{\partial \mu} \overset{!}{=} 0$$

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Steps to find the MLE estimator: Checking step

Take the second derivative of the log likelihood function

Check that the estimator is the maximum of the likelihood by taking the second derivative w.r.t the distribution parameter, It SHOULD be < 0

$$\begin{aligned} \frac{\partial^2 \ell(\mu,\sigma)}{\partial \mu^2} &= \frac{\partial^2 \ell(\mu,\sigma)}{\partial \mu^2}\left(\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}\right) \\ &= -1 < 0 \end{aligned}$$

Steps to find the MLE estimator: Step 3

Take the derivative of the log likelihood function

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter σ, μ is considered as a constant

$$
\begin{aligned}
\frac{\partial \ell(\mu,\sigma)}{\partial \sigma} &= \frac{\partial}{\partial \sigma} \log\left(\prod_{i=1}^{n} f(x_i|\,\mu,\sigma)\right) \\
&= \frac{\partial}{\partial \sigma}\left(n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)\right) \\
&= \frac{\partial}{\partial \sigma}\left(n \log(1) - n\log(\sigma) - \frac{n}{2}\log(2\pi) - \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)\right) \\
&= -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}
\end{aligned}
$$

Steps to find the MLE estimator: Step 4

Set the derivative of the log likelihood function $\stackrel{!}{=} 0$

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter μ

$$\frac{\partial \ell(\mu, \sigma)}{\partial \sigma} \stackrel{!}{=} 0$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Steps to find the MLE estimator: Checking step

Take the second derivative of the log likelihood function

Check that the estimator is the maximum of the likelihood by taking the second derivative w.r.t the distribution parameter, It SHOULD be < 0

$$\begin{aligned} \frac{\partial^2 \ell(\mu,\sigma)}{\partial \sigma^2} &= \frac{\partial^2 \ell(\mu,\sigma)}{\partial \sigma^2}\left(-\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}\right) \\ &< 0 \end{aligned}$$

The MLE estimators

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Examples

Let $X_1, \ldots, X_n$ iid, a random sample from a population with pdf,

$$1.\ p(x_i|\theta) = \frac{2}{\theta}\ \ y\ \exp\frac{-y^2}{\theta}, \quad y > 0,\ \theta > 0$$

$$2.\ p(x_i|\theta) = \frac{\theta}{x^2}, \quad 0 < \theta \leq x < \infty$$

Find the maximum likelihood estimator of θ.

Bias

Definition (Bias)

The difference between this estimator's expected value and the true value of the parameter being estimated.

$$\mathrm{bias}(\hat{\theta}, \theta) = \mathbb{E}_\theta[\hat{\theta} - \theta]$$
$$\mathrm{bias}(\widehat{\theta}) = E(\widehat{\theta}) - \theta$$

where $E[\ ]$ denotes the expected value over the distribution $p(x|\theta)$, i.e. averaging over all possible observations x

When the parameter of the distribution and estimator θ are obvious from the context, we write $\mathrm{bias}(\hat{\theta})$ for ease of notation.

Bias

Positive bias indicates that the estimator provides an over-estimate the target parameter on average. We typically prefer to have zero bias, which motivates the following definition.

Definition (Unbiasedness)

An estimator $\hat{\theta}$ is said to be unbiased if $\mathrm{bias}(\hat{\theta}, \theta) = 0$ for all $\theta \in \theta$. Then $E(\widehat{\theta}) = \theta$.

It is important to note that an unbiased estimator must have zero bias for all possible distributions, for the following reason: for any given estimator we can usually find some distribution under which the estimator has zero bias. For example, for some $c \in \mathbb{R}$, the constant estimator $\hat{\theta} = c$ has zero bias.

Bernoulli distribution

Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p, then

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

is the maximum likelihood estimator (MLE) of p. Is the MLE of p an unbiased estimator of p?

Recall that if x_i is a Bernoulli random variable with parameter p, then $E(x) = p$.

$$E(\widehat{p}) = E(\frac{1}{n} \sum_{i=1}^{n} x_i) = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = \frac{1}{n} \sum_{i=1}^{n} p = \frac{1}{n} (np) = p$$

Therefore, the maximum likelihood estimator is an unbiased estimator of p.

Standard Normal distribution

Let $X_1, \ldots, X_n$ be a random sample from a standard normal distribution with parameters μ and σ^2, then

$$\bar{x} = \widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$$

are the maximum likelihood estimators (MLE) of μ and σ^2. Are the MLE of $\widehat{\mu}$ and $\widehat{\sigma}^2$ are unbiased estimator of μ and σ^2?

Recall that $E(x) = \mu$

$$E(\widehat{\mu}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}(n\mu) = \mu$$

$E(\widehat{\mu}) = \mu$, therefore, the maximum likelihood estimator $\widehat{\mu}$ is an unbiased estimator of μ.

Standard Normal distribution: cont.

$E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right) = \frac{1}{n}\sum^{n} E(x_i^2) - E(\bar{x}^2)$

$= \frac{1}{n}\left[\sum (\sigma^2+\mu^2)\right] - \left(\frac{\sigma^2}{n}+\mu^2\right)$

$= \frac{1}{n}\left[n\sigma^2+n\mu^2\right] - \frac{\sigma^2}{n} - \mu^2$

Recall that $Var(x) = \sigma^2$ $= \sigma^2 - \frac{\sigma^2}{n} = \frac{n\sigma^2-\sigma^2}{n} = \frac{(n-1)\sigma^2}{n}$

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right) = \left[\frac{1}{n}\sum_{i=1}^{n}E(x_i^2)\right] - E(\bar{x}^2) = \frac{(n-1)\sigma^2}{n}$$

$E(\hat{\sigma}^2) \neq \sigma^2$, therefore, the maximum likelihood estimator $\hat{\sigma}^2$ is a biased estimator of σ^2.

$E(x_i^2) = var(x_i) + [E(x_i)]^2$

$= \sigma^2 + \mu^2$

$E(\bar{x}^2) = var(\bar{x}) + [E(\bar{x})]^2$

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

$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$

$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}(n\bar{x}^2)$

$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$

Poisson distribution

Your turn now :)

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, then

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the maximum likelihood estimator (MLE) of λ. Is the MLE of λ an unbiased estimator of λ?

$$E(\hat{\lambda}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\cdot n \cdot \lambda = \lambda$$

$$\text{bias}(\hat{\lambda}) = E(\hat{\lambda}) - \lambda = \lambda - \lambda = 0$$

$\hat{\lambda}_{MLE}$ is <u>unbiased estimator</u>

Poisson distribution

Your turn now :)

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, then

$$\widehat{\lambda} \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the maximum likelihood estimator (MLE) of λ. Is the MLE of λ an unbiased estimator of λ?

Recall that if x_i is a Poisson random variable with parameter λ, then $E(x) = \lambda$.

$$E(\widehat{\lambda}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} \lambda = \frac{1}{n}(n\lambda) = \lambda$$

Therefore, the maximum likelihood estimator is an unbiased estimator of λ.

Mean Squared Error (MES)

Definition (MSE)

The mean squared error (MSE) is the difference between the estimator's expected value and the true value of the parameter being estimated.

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}[\|\hat{\theta} - \theta\|^2] = E((\hat{\theta} - \theta)^2) = \mathrm{bias}(\hat{\theta})^2 + \mathrm{var}(\hat{\theta})$$

where $E[\ ]$ denotes the expected value over the distribution $p(x|\theta)$, i.e. averaging over all possible observations x

An estimator is said to be unbiased if its bias is equal to zero for all values of parameter θ.

$\mathrm{bias}(\hat{\theta}) = 0$ then $E(\hat{\theta}) = \theta$. Hence, $\mathrm{MSE}(\hat{\theta}) = Var(\hat{\theta})$

Bias-Variance decomposition

Proof.

We here prove the bias-variance decomposition. By the definition of MSE,

$$
\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= \mathbb{E}[\hat{\theta} - \theta] \\
&= \mathrm{E}[\hat{\theta} - \mathrm{E}(\hat{\theta}) + \mathrm{E}(\hat{\theta}) - \theta] \\
&= \mathrm{E}[\hat{\theta} - \mathrm{E}(\hat{\theta})]^2 + \mathrm{E}[(\mathrm{E}(\hat{\theta}) - \theta)^2] + 2\mathrm{E}[(\hat{\theta} - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)] \\
&= \mathrm{var}(\hat{\theta}) + \mathrm{bias}(\hat{\theta})^2 + 2(\mathrm{E}(\hat{\theta} - \theta)\underbrace{\mathrm{E}(\hat{\theta} - \mathrm{E}(\hat{\theta}))}_{=0} \\
&= \mathrm{bias}(\hat{\theta})^2 + \mathrm{var}(\hat{\theta})
\end{aligned}
$$

which concludes the proof. □

Poisson distribution

Let $X_1, \ldots, X_n$ are random variables from a Poisson distribution with parameter λ. Find MSE of $\hat{\lambda} = \overline{x}$

$$
\begin{aligned}
\mathrm{MSE}(\hat{\lambda}) &= \mathrm{bias}(\hat{\lambda})^2 + \mathrm{var}(\hat{\lambda}) \\
&= \mathrm{var}(\hat{\lambda}) = \mathrm{var}(\overline{x}) \\
&= \mathrm{var}(\frac{1}{n}\sum_{i=1}^{n} x_i) \\
&= \frac{1}{n^2}\ \mathrm{var}(\sum_{i=1}^{n} x_i) \\
&= \frac{1}{n^2}\ n\ \lambda = \frac{\lambda}{n}
\end{aligned}
$$

Exercise

Let $Y_1, \ldots, Y_n$ Binomial random variables with parameters n and p.
We have two estimators $\widehat{p_1} = \frac{y}{n}$ and $\widehat{p_2} = \frac{y+1}{n+2}$
What value of p does $\widehat{p_2}$ achieve a lower MSE than $\widehat{p_1}$?

$\hat{P}_1$

$E(\hat{p}_1) = E\left(\frac{y}{n}\right) = \frac{1}{n} E(y) = \frac{1}{n} \cdot np = p$

$bias(\hat{p}_1) = E(\hat{p}_1) - p = p - p = 0$ Unbiased

$MSE(\hat{p}_1) = Var(\hat{p}_1) = Var\left(\frac{y}{n}\right) = \frac{1}{n^2} Var(y)$

$= \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

Exercise: cont.

$E(y) = np$
$Var(y) = np(1-p)$

Let $Y_1, \ldots, Y_n$ Binomial random variables with parameters n and p.
We have two estimators $\widehat{p_1} = \frac{y}{n}$ and $\widehat{p_2} = \frac{y+1}{n+2}$
What value of p does $\widehat{p_2}$ achieve a lower MSE than $\widehat{p_1}$?

$\hat{P}_2$

$$E(\hat{P}_2) = E\left(\frac{y+1}{n+2}\right) = \frac{1}{n+2} E(y+1) = \frac{1}{n+2}\left(E(y) + E(1)\right) = \frac{np+1}{n+2}$$

$$bias(\hat{P}_2) = E(\hat{P}_2) - p = \frac{np+1}{n+2} - p$$

$$Var(\hat{P}_2) = Var\left(\frac{y+1}{n+2}\right) = \frac{1}{(n+2)^2}\left(Var(y) + Var(1)\right) = \frac{np(1-p)}{(n+2)^2}$$

$$MSE(\hat{P}_2) = var(\hat{P}_2) + bias(\hat{P}_2)^2$$

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

Exercise: cont.

Let $Y_1, \ldots, Y_n$ Binomial random variables with parameters n and p.
We have two estimators $\widehat{p_1} = \frac{y}{n}$ and $\widehat{p_2} = \frac{y+1}{n+2}$
What value of p does $\widehat{p_2}$ achieve a lower MSE than $\widehat{p_1}$?

required $MSE(\hat{p}_2) < MSE(\hat{p}_1)$

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

$$\frac{1}{2}\left(1 - \sqrt{\frac{n+1}{2n+1}}\right) < p < \frac{1}{2}\left(1 + \sqrt{\frac{n+1}{2n+1}}\right)$$

Standard Normal distribution

Your turn now :)

Let $X_1, \ldots, X_n$ be a random sample from a standard normal distribution with parameters θ and σ^2, and let

$$\widehat{\theta_1} = x_i, \;\; \widehat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Find the following:

MSE of $\hat{\theta}_1$
MSE of $\hat{\theta}_2$
Which is better?

Standard Normal distribution: cont.

$\mathrm{MSE}(\hat{\theta}_1)$

$$
\begin{aligned}
\mathrm{var}(\hat{\theta}_1) &= \mathrm{var}(x_i) = \sigma^2 \\
\mathrm{bias}(\hat{\theta}_1) &= \mathrm{E}(\hat{\theta}_1) - \theta \\
&= \mathrm{E}(x_i) - \theta \\
&= \theta - \theta = 0 \\
\mathrm{MSE}(\hat{\theta}_1) &= \mathrm{bias}(\hat{\theta})^2 + \mathrm{var}(\hat{\theta}) \\
&= \mathrm{var}(\hat{\theta}) \\
&= \sigma^2
\end{aligned}
$$

Standard Normal distribution: cont.

$\mathrm{MSE}(\hat{\theta}_2)$

$$\mathrm{var}(\hat{\theta}_2) = \mathrm{var}(\frac{1}{n}\sum_{i=1}^{n} x_i)$$

$$= \frac{1}{n^2}\, n\, \mathrm{var}(x_i) = \frac{\sigma^2}{n}$$

$$\mathrm{bias}(\hat{\theta}_2) = \mathrm{E}(\hat{\theta}_2) - \theta$$

$$= \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} x_i) - \theta$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(x_i) - \theta = \frac{n\,\theta}{n} - \theta = 0$$

$$\mathrm{MSE}(\hat{\theta}_1) = \mathrm{bias}(\hat{\theta})^2 + \mathrm{var}(\hat{\theta}) = \mathrm{var}(\hat{\theta}) = \frac{\sigma^2}{n}$$

Standard Normal distribution: cont.

Which estimator is better in terms of MSE

$$\mathrm{MSE}(\hat{\theta}_2) < \mathrm{MSE}(\hat{\theta}_1) \quad \text{for } n > 1$$

Hence, $\hat{\theta}_2$ is better estimator than $\hat{\theta}_1$

Consistent estimator

Definition (Consistent estimator)

A sequence of estimators $\{\theta_{(n)}\}$ is **consistent** if $\theta_{(n)} \xrightarrow{p} \theta$ as $n \to \infty$.

Corollary

A sequence of estimators $\{\theta_{(n)}\}$ *is consistent if* $\mathrm{MSE}(\theta_{(n)}) \to 0$ *as* $n \to \infty$.

The larger the sample size we can achieve the more accurate our estimation becomes.
We can reduce the mean squared error by either reducing the bias, the variance or both. It is easy to see that the mean squared error of an unbiased estimator is the variance itself. Therefore, when comparing among the class of unbiased estimators we only need to care about the variance of the estimators. This motivates the cram-Rao bound. (will be introduced later).

Uniform distribution

Let $X_1, \ldots, X_n$ are random variables from Uniform distribution with parameter θ and the pdf is

$$f(x,\theta) = \begin{cases} nx^{n-1}/\theta^n & \text{if } 0 \leq x \leq \theta \\ 0 & otherwise \end{cases}$$

then

$$\widehat{\theta} \;=\; \max\left(x_1, x_2, \ldots, x_n\right)$$

is the maximum likelihood estimator (MLE) of θ.

Find the following:

bias of $\hat{\theta}$
MSE of $\hat{\theta}$
Is $\widehat{\theta}$ a consistent estimator of θ

Uniform distribution: cont.

$\text{bias}(\hat{\theta})$

$$\begin{aligned}
\mathrm{E}(\hat{\theta}) &= \int_0^{\theta} x\ f(x)\ dx \\
&= \int_0^{\theta} x\ \frac{n\ x^{n-1}}{\theta^n}\ dx \\
&= \int_0^{\theta} x^n\ \frac{n}{\theta^n}\ dx \\
&= \frac{n}{n+1}\ \theta \\
\text{bias}(\hat{\theta}) &= \frac{n}{n+1}\ \theta -\ \theta \\
&= \frac{-\ \theta}{n+1}
\end{aligned}$$

Uniform distribution: cont.

$\mathrm{MSE}(\hat{\theta})$

$$\begin{aligned}
\mathrm{E}(\hat{\theta}^2) &= \int_0^\theta x^2 \ f(x) \ dx \\
&= \int_0^\theta x^2 \ \frac{n \ x^{n-1}}{\theta^n} \ dx \\
&= \int_0^\theta x^{n+1} \ \frac{n}{\theta^n} \ dx \\
&= \frac{n}{n+2} \ \theta^2
\end{aligned}$$

Uniform distribution: cont.

$\mathrm{MSE}(\hat{\theta})$

$$
\begin{aligned}
\mathrm{var}(\hat{\theta}) &= \mathrm{E}(\hat{\theta}^2) - [\mathrm{E}(\hat{\theta})]^2 \\
&= \frac{n}{n+2}\,\theta^2 - \left(\frac{n}{n+1}\right)\,\theta^2 \\
&= \frac{n}{(n+2)(n+1)^2}\,\theta^2 \\
\mathrm{MSE}(\hat{\theta}) &= \mathrm{bias}(\hat{\theta})^2 + \mathrm{var}(\hat{\theta}) \\
&= (\frac{-\theta}{n+1})^2 + \frac{n}{(n+2)(n+1)^2}\,\theta^2 \\
&= \frac{2\,\theta^2}{(n+2)(n+1)}
\end{aligned}
$$

Uniform distribution: cont.

$\hat{\theta}$ is a consistent estimator of θ

$$\lim_{n\to\infty} \mathrm{MSE}(\hat{\theta}) = \lim_{n\to\infty} \frac{2\,\theta^2}{(n+2)(n+1)}$$
$$= 0$$

Hence, $\hat{\theta}$ is consistent estimator of θ

Exercise

Your turn now :)

Let $X_1, \ldots, X_n$ are random variables from a distribution with parameter θ and the pdf is

$$f(x,\theta) = \frac{1}{2}\,(1+\theta\,x), \quad -1 < x < 1,\ -1 < \theta < 1$$

Show that $3\,\overline{x}$ is a consistent estimation of θ.

$$E(X) = \int_{-1}^{1} x \cdot \frac{1}{2}(1+\theta x)\,dx$$

$$= \frac{1}{2}\int_{-1}^{1} x + \theta x^2 \, dx$$

$$= \frac{1}{2}\left(\frac{x^2}{2} + \theta\frac{x^3}{3}\right)\Big|_{-1}^{1}$$

$$= \frac{1}{2}\left[\left(\frac{1}{2}+\frac{\theta}{3}\right) - \left(\frac{1}{2}+\frac{\theta}{3}\right)\right]$$

$$= \frac{1}{2}\cdot\frac{2\theta}{3} = \frac{\theta}{3}$$

Exercise

Let $X_1, \ldots, X_n$ are random variables from a distribution with parameter θ and the pdf is

$$f(x,\theta) = \frac{1}{2}\ (1 + \theta\, x), \quad -1 < x < 1,\ -1 < \theta < 1$$

Show that $3\,\overline{x}$ is a consistent estimation of θ.

$$E(x^2) = \int_{-1}^{1} x^2 \cdot \frac{1}{2}(1+\theta x)\, dx$$

$$= \frac{1}{2} \int_{-1}^{1} x^2 + \theta x^3 \, dx$$

$$= \frac{1}{2} \left[\frac{x^3}{3} + \frac{\theta x^4}{4} \right] \Big|_{-1}^{1}$$

$$= \frac{1}{3}$$

Exercise

Let $X_1, \ldots, X_n$ are random variables from a distribution with parameter θ and the pdf is

$$f(x, \theta) = \frac{1}{2}\ (1 + \theta\, x), \quad -1 < x < 1,\ -1 < \theta < 1$$

Show that $3\,\overline{x}$ is a consistent estimation of θ.

$$Var(X) = E(X^2) - E(X)^2$$

$$= \frac{1}{3} - \left(\frac{\theta}{3}\right)^2$$

$$= \frac{3 - \theta^2}{9}$$

Exercise

Let $X_1, \ldots, X_n$ are random variables from a distribution with parameter θ and the pdf is

$$f(x, \theta) = \frac{1}{2}\ (1 + \theta\, x), \quad -1 < x < 1,\ -1 < \theta < 1$$

Show that $3\,\overline{x}$ is a consistent estimation of θ.

$$E(3\bar{X}) = 3E(\bar{X})$$
$$= 3E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$
$$= \frac{3}{n} E\left(\sum_{i=1}^{n} x_i\right)$$
$$= \frac{3}{n}\sum_{i=1}^{n} E(x_i)$$
$$= \frac{3}{n} \cdot n \cdot \frac{\theta}{3}$$
$$= \theta$$

Exercise

Let $X_1, \ldots, X_n$ are random variables from a distribution with parameter θ and the pdf is

$$f(x, \theta) = \frac{1}{2}\ (1 + \theta\, x), \quad -1 < x < 1,\ -1 < \theta < 1$$

Show that $3\,\overline{x}$ is a consistent estimation of θ.

$$bias(3\bar{x}) = E(3\bar{x}) - \theta = \theta - \theta = 0 \qquad \text{unbiased}$$

$$Var(3\bar{x}) = 9\, Var(\bar{x})$$

$$= 9\, Var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$$

$$= \frac{9}{n^2}\, Var\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{9}{n^2}\, n\, Var(x_i)$$

$$= \frac{3 - \theta^2}{n}.$$

Exercise

Let $X_1, \ldots, X_n$ are random variables from a distribution with parameter θ and the pdf is

$$f(x,\theta) = \frac{1}{2}\ (1+\theta\, x), \quad -1 < x < 1,\ -1 < \theta < 1$$

Show that $3\,\overline{x}$ is a consistent estimation of θ.

$$MSE(3\bar{x}) = Var(3\bar{x}) + bias(3\bar{x})$$

$$= Var(3\bar{x})$$

"Unbiased estimator"

$$= \frac{3-\theta^2}{n}$$

$$\lim_{n\to\infty} MSE(3\bar{x}) = \lim_{n\to\infty} Var(3\bar{x}) = \lim_{n\to\infty} \frac{3-\theta^2}{n} = 0$$

$\Rightarrow$ $3\bar{x}$ is consistent estimator

UMVUE and MVUE

Definition

$\hat{\theta}$ is an efficient unbiased estimator if the variance of $\hat{\theta}$ is smaller than the variance of any other unbiased estimators of θ based on the same sample size. It might also be referred to as a uniformly minimum variance unbiased estimator(UMVUE).

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be unbiased estimators of θ with equal sample sizes. Then, $\hat{\theta}_1$ is a more efficient estimator than $\hat{\theta}_2$ if $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$.

UMVUE and MVUE

Minimum Variance Unbiased Estimator(MVUE)

When you take multiple samples from a population, each of those samples will (probably) have different statistics: a slightly different mean or standard deviation/variance. The MVUE is the statistic with the lowest variance.

There isn't a simple formula to find the MVUE and UMVUE, and it may not actually exist for your samples. There are two main ways you can find or verify a MVUE:
Use the Cramer-Rao Lower Bound. This sets a lower bound for the variance. If you can find an estimator that meet this condition, you've found the MVUE. Find a sufficient statistic and then use the Rao-Blackwell theorem.

Statistical model: Poisson distribution

Let $X_1, \ldots, X_n$ be a random sample (that is, independent and identically distributed) from a Poisson distribution with parameter λ

$$p(x_i|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, 3, ... \ \lambda > 0$$

Derive a formula for $\widehat{\lambda}$, the maximum likelihood estimate of λ. Show that $\widehat{\lambda}$ is an efficient estimator.

Steps to find the MLE estimator: Step 1

Define the likelihood function $L(\lambda)$

The likelihood function of a Poisson distribution is:

$$L(\lambda) = \prod_{i=1}^{n} p(x_i|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Steps to find the MLE estimator: Step 2

Define the log likelihood function $\ell(\lambda)$

The natural log of the likelihood function of a Poisson distribution is:

$$
\begin{aligned}
\ell(\lambda) &= \log L(\lambda) = \log \prod_{i=1}^{n} p(x_i|\lambda) \\
&= \log \prod_{i=1}^{n} \frac{e^{-\lambda}\, \lambda^{x_i}}{x_i!} \\
&= \log \frac{\left(e^{-n\lambda}\right)\, \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \log(e^{-n\lambda}) + \log(\lambda^{\sum_{i=1}^{n} x_i}) - \log(\prod_{i=1}^{n} x_i!) \\
&= -n\lambda + \sum_{i=1}^{n} x_i \log(\lambda) - \sum_{i=1}^{n} \log(x_i!)
\end{aligned}
$$

Steps to find the MLE estimator: Step 3

Take the derivative of the log likelihood function

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter

$$\begin{aligned}
\frac{\partial \ell(\lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \log \left(\prod_{i=1}^{n} p(x_i|\lambda) \right) \\
&= \frac{\partial}{\partial \lambda} \left(-n\lambda + \sum_{i=1}^{n} x_i \log(\lambda) - \sum_{i=1}^{n} \log(x_i!) \right) \\
&= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}
\end{aligned}$$

Steps to find the MLE estimator: Step 4

Set the derivative of the log likelihood function $\stackrel{!}{=} 0$

Maximize the log likelihood by taking the derivative w.r.t the distribution parameter

$$\frac{\partial \ell(\lambda)}{\partial \lambda} \stackrel{!}{=} 0$$

$$\widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Steps to find the MLE estimator: Checking step

Take the second derivative of the log likelihood function

Check that the estimator is the maximum of the likelihood by taking the second derivative w.r.t the distribution parameter, It SHOULD be < 0

$$\begin{aligned} \frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} &= \frac{\partial^2 \ell(\lambda)}{\partial \lambda^2}\left(-n + \frac{\sum_{i=1}^{n} x_i}{\lambda}\right) \\ &= \frac{-\sum_{i=1}^{n} x_i}{\lambda^2} < 0 \end{aligned}$$

Steps to find the MLE estimator: Step 5

Take the expectation of the second derivative of the log likelihood function

Check that the estimator is the maximum of the likelihood by taking the second derivative w.r.t the distribution parameter, It SHOULD be < 0

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2}\right) &= \frac{-\mathrm{E}[\sum_{i=1}^{n} x_i]}{\lambda^2} \\
&= -\frac{n\,\lambda}{\lambda^2} = -\frac{n}{\lambda} \\
\mathrm{CRLB} &= \frac{\lambda}{n} = \mathrm{var}\,(\hat{\lambda})
\end{aligned}
$$

Hence, $\hat{\lambda}$ is an efficient estimator, it's also unbiased $\Longrightarrow$ *is UMVUE*.