Introductory Statistics

Tatiana Tyukina
tt51@leicester.ac.uk

Topic 2 - Estimators
2.2 - Properties of Estimators
Efficiency

- **Topic 0: Introduction**
- **Topic 1: Descriptive Statistics**
- **Topic 2: Estimators: Point estimation, Interval estimation**
  - 2.1 **Point estimation**,
  - 2.2 Properties of estimators,
  - 2.3 Interval estimation,
  - 2.4 Inference based on Normal Distribution
- Topic 3: Hypothesis Testing
- Topic 4: Goodness of Fit: The χ^2 test
- Topic 5: Bayesian Estimation

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number.
1) Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = (2X_1 + X_2 + 5X_3)/8$ are unbiased estimators for μ.
2) Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number.
1) Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = (2X_1 + X_2 + 5X_3)/8$ are unbiased estimators for μ.
2) Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Solution.
1)For $\hat{\theta}_1$: $E(\hat{\theta}_1) = E(\bar{X}) = \frac{1}{3} \cdot 3\mu = \mu$.

Properties of Point Estimators: Efficiency

Example

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number.
1) Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = (2X_1 + X_2 + 5X_3)/8$ are unbiased estimators for μ.
2) Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Solution.
1)For $\hat{\theta}_1$: $E(\hat{\theta}_1) = E(\bar{X}) = \frac{1}{3} \cdot 3\mu = \mu$. $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\, n\, E(X)$
For $\hat{\theta}_2$: $E(\hat{\theta}_2) = \frac{1}{8} \cdot (2E(X_1) + E(X_2) + 5E(X_3)) = \frac{2\mu+\mu+5\mu}{8} = \mu$.
Hence, both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number.
1) Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = (2X_1 + X_2 + 5X_3)/8$ are unbiased estimators for μ.
2) Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Solution.

1)For $\hat{\theta}_1$: $E(\hat{\theta}_1) = E(\bar{X}) = \frac{1}{3} \cdot 3\mu = \mu$.
For $\hat{\theta}_2$: $E(\hat{\theta}_2) = \frac{1}{8} \cdot (2E(X_1) + E(X_2) + 5E(X_3)) = \frac{2\mu+\mu+5\mu}{8} = \mu$.
Hence, both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators.
2) The variance of $\hat{\theta}_1$: $\mathrm{Var}(\hat{\theta}_1) = \frac{\sigma^2}{3}$,
whereas the variance of $\hat{\theta}_2$ is

$$Var(\hat{\theta}_1) = Var(\bar{X}) = Var\left(\frac{X_1}{3} + \frac{X_2}{3} + \frac{X_3}{3}\right) = \frac{1}{9} Var(X_1) + \frac{1}{9} Var(X_2) + \frac{1}{9} Var(X_3) =$$

$$= \frac{1}{9}\left(\sigma^2 + \sigma^2 + \sigma^2\right) = \frac{\sigma^2}{3}$$

$$Var(\hat{\theta}_1) = \frac{\sigma^2}{n}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number.
1) Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = (2X_1 + X_2 + 5X_3)/8$ are unbiased estimators for μ.
2) Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Solution.

1)For $\hat{\theta}_1$: $E(\hat{\theta}_1) = E(\bar{X}) = \frac{1}{3} \cdot 3\mu = \mu$.
For $\hat{\theta}_2$: $E(\hat{\theta}_2) = \frac{1}{8} \cdot (2E(X_1) + E(X_2) + 5E(X_3)) = \frac{2\mu+\mu+5\mu}{8} = \mu$.
Hence, both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators.
2) The variance of $\hat{\theta}_1$: $\mathrm{Var}(\hat{\theta}_1) = \frac{\sigma^2}{3}$,
whereas the variance of $\hat{\theta}_2$ is

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}\left(\frac{2X_1+X_2+5X_3}{8}\right) = \frac{4}{64}\sigma^2 + \frac{1}{64}\sigma^2 + \frac{25}{64}\sigma^2 = \frac{30}{64}\sigma^2$$

Because $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, $\bar{X}$ is the better unbiased estimator.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

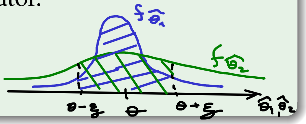
EXAMPLE

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number.
1) Show that both $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = (2X_1 + X_2 + 5X_3)/8$ are unbiased estimators for μ.
2) Compare the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Solution.
1)For $\hat{\theta}_1$: $E(\hat{\theta}_1) = E(\bar{X}) = \frac{1}{3} \cdot 3\mu = \mu$.
For $\hat{\theta}_2$: $E(\hat{\theta}_2) = \frac{1}{8} \cdot (2E(X_1) + E(X_2) + 5E(X_3)) = \frac{2\mu+\mu+5\mu}{8} = \mu$.
Hence, both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators.
2) The variance of $\hat{\theta}_1$: $\mathrm{Var}(\hat{\theta}_1) = \frac{\sigma^2}{3}$,
whereas the variance of $\hat{\theta}_2$ is

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}\left(\frac{2X_1+X_2+5X_3}{8}\right) = \frac{4}{64}\sigma^2 + \frac{1}{64}\sigma^2 + \frac{25}{64}\sigma^2 = \frac{30}{64}\sigma^2$$

Because $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, $\bar{X}$ is the better unbiased estimator.
For any value of $\xi > 0$:
$P(\theta - \xi \leq \hat{\theta}_1 \leq \theta + \xi) > P(\theta - \xi \leq \hat{\theta}_2 \leq \theta + \xi)$



PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

DEFINITION

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for a parameter θ. If

$$Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$$

*we say that $\hat{\theta}_1$ is **more efficient** than $\hat{\theta}_2$.*
*The **relative efficiency of** $\hat{\theta}_1$ **with respect to** $\hat{\theta}_2$ is the ratio $Var(\hat{\theta}_1)/Var(\hat{\theta}_2)$.*

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

DEFINITION

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for a parameter θ. If

$$Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$$

*we say that $\hat{\theta}_1$ is **more efficient** than $\hat{\theta}_2$.*
*The **relative efficiency of** $\hat{\theta}_1$ **with respect to** $\hat{\theta}_2$ is the ratio $Var(\hat{\theta}_1)/Var(\hat{\theta}_2)$.*

DEFINITION

Let Θ denote a set of all estimators $\hat{\theta} = h(X_1, X_2, ...X_n)$ that are unbiased for the parameter θ in the continuous pdf $f_X(x,\theta)$ (or discrete pmf $p_X(x,\theta)$).
*The estimator $\hat{\theta}^*$ is the **best** (or **unbiased minimum variance**) **estimator** if $\hat{\theta}^* \in \Theta$ and*

$$Var(\hat{\theta}^*) \leq Var(\hat{\theta}) \text{ for all } \hat{\theta} \in \Theta.$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

THEOREM (THE CRAMÉR-RAO LOWER BOUND)

Let $f_X(X,\theta)$ *be a continuous pdf with continuous first-order and second-order derivatives.*
Let $X_1, X_2, ..., X_n$ *be a random sample from* $f_X(x,\theta)$*, and suppose that the set of* x *values, where* $f_X(x,\theta) \neq 0$ *does not depend on* θ*. Let* $\hat{\theta} = g(X_1, X_2, ...X_n)$ *be any* ***unbiased*** *estimator of* θ*. Then*

$$Var\left(\hat{\theta}\right) \geq \left\{nE\left[\left(\frac{\partial \ln f_X(X,\theta)}{\partial\theta}\right)^2\right]\right\}^{-1} = \left\{\underbrace{-nE\left[\frac{\partial^2 \ln f_X(X,\theta)}{\partial\theta^2}\right]}_{\text{Fisher information}}\right\}^{-1}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

THEOREM (THE CRAMÉR-RAO LOWER BOUND)

Let $f_X(X,\theta)$ be a continuous pdf with continuous first-order and second-order derivatives.
Let $X_1, X_2, ..., X_n$ be a random sample from $f_X(x,\theta)$, and suppose that the set of x values, where $f_X(x,\theta) \neq 0$ does not depend on θ. Let $\hat{\theta} = g(X_1, X_2, ...X_n)$ be any ***unbiased*** *estimator of θ. Then*

$$Var\left(\hat{\theta}\right) \geq \left\{nE\left[\left(\frac{\partial \ln f_X(X,\theta)}{\partial\theta}\right)^2\right]\right\}^{-1} = \left\{-nE\left[\frac{\partial^2 \ln f_X(X,\theta)}{\partial\theta^2}\right]\right\}^{-1}$$

If the variance of a given $\hat{\theta}$ is equal to the Cramér-Rao lower bound we say that the estimator is *optimal* in a sense that no unbiased $\hat{\theta}$ can estimate θ with greater precision.

$\hat{\theta}^*$ - optimal $\Leftrightarrow Var(\hat{\theta}^*) = CRLB \Rightarrow \forall \hat{\theta} \in \Theta \quad Var(\hat{\theta}) \geq Var(\hat{\theta}^*)$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

THEOREM (THE CRAMÉR-RAO LOWER BOUND)

Let $f_X(X, \theta)$ be a continuous pdf with continuous first-order and second-order derivatives.
Let $X_1, X_2, ..., X_n$ be a random sample from $f_X(x, \theta)$, and suppose that the set of x values, where $f_X(x, \theta) \neq 0$ does not depend on θ. Let $\hat{\theta} = g(X_1, X_2, ...X_n)$ be any ***unbiased*** *estimator of θ. Then*

$$Var\left(\hat{\theta}\right) \geq \left\{nE\left[\left(\frac{\partial \ln f_X(X,\theta)}{\partial \theta}\right)^2\right]\right\}^{-1} = \left\{\overbrace{-nE\left[\frac{\partial^2 \ln f_X(X,\theta)}{\partial \theta^2}\right]}^{\text{Fisher Information}}\right\}^{-1}$$

If the variance of a given $\hat{\theta}$ is equal to the Cramér-Rao lower bound we say that the estimator is *optimal* in a sense that no unbiased $\hat{\theta}$ can estimate θ with greater precision.

The unbiased estimator $\hat{\theta}$ is said to be *efficient* if the variance of $\hat{\theta}$ equals to the Cramér-Rao lower bound associated with $f_X(x, \theta)$.
The *efficiency* of an unbiased estimator $\hat{\theta}$ is the ratio of the Cramér-Rao lower bound for $f_X(x, \theta)$ to the variance of $\hat{\theta}$.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution $p_X(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$
Compare the Cramér-Rao lower bound for $p_X(x, \lambda)$ to the variance of the maximum likelihood estimator for λ.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution $p_X(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$

Compare the Cramér-Rao lower bound for $p_X(x, \lambda)$ to the variance of the maximum likelihood estimator for λ.

Solution.

1) Obtain the ML estimator:

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

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution $p_X(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, ...$
Compare the Cramér-Rao lower bound for $p_X(x,\lambda)$ to the variance of the maximum likelihood estimator for λ.

Solution.

1) Obtain the ML estimator:

$L(\lambda) = \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-n\lambda}\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \Rightarrow$ $L(\lambda) = e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i} \cdot \prod_{i=1}^n (x_i!)^{-1}$

$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln\lambda - \sum_{i=1}^n \ln(x_i!)$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution
$p_X(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$
Compare the Cramér-Rao lower bound for $p_X(x,\lambda)$ to the variance of the maximum likelihood estimator for λ.

Solution.

1) Obtain the ML estimator:

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

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln\lambda - \sum_{i=1}^n \ln(x_i!) \quad \Rightarrow$$

$$\frac{d\ln L(\lambda)}{d\lambda} = -n + \sum_{i=1}^n x_i \underbrace{\lambda^{-1}} = 0$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution $p_X(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$
Compare the Cramér-Rao lower bound for $p_X(x,\lambda)$ to the variance of the maximum likelihood estimator for λ.

Solution.

1) Obtain the ML estimator:

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

$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln\lambda - \sum_{i=1}^n \ln(x_i!) \quad \Rightarrow$

$\frac{d\ln L(\lambda)}{d\lambda} = -n + \sum_{i=1}^n x_i\lambda^{-1} = 0 \quad \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{X}.$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution $p_X(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$

Compare the Cramér-Rao lower bound for $p_X(x, \lambda)$ to the variance of the maximum likelihood estimator for λ.

Solution.

1) Obtain the ML estimator:

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

$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln \lambda - \sum_{i=1}^n \ln(x_i!) \quad \Rightarrow$

$\frac{d \ln L(\lambda)}{d\lambda} = -n + \sum_{i=1}^n x_i \lambda^{-1} = 0 \quad \Rightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}.$

2) Let's find the Cramér-Rao Lower Boundary (CRLB):

$$CRLB = -\frac{1}{nE\left(\frac{d^2 \ln(p_X(x,\lambda))}{d\lambda^2}\right)}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a random sample from the Poisson distribution $p_X(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$
Compare the Cramér-Rao lower bound for $p_X(x,\lambda)$ to the variance of the maximum likelihood estimator for λ.

Solution.

1) Obtain the ML estimator:

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

$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln\lambda - \sum_{i=1}^n \ln(x_i!) \quad \Rightarrow$

$\frac{d\ln L(\lambda)}{d\lambda} = -n + \sum_{i=1}^n x_i\lambda^{-1} = 0 \quad \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{X}.$

2) Let's find the Cramér-Rao Lower Boundary (CRLB):

$$\frac{d\ln p_X(X,\lambda)}{d\lambda} = \frac{d[(\ln e^{-\lambda}) + \ln(\lambda^X) - \ln(X!)]}{d\lambda} = \frac{d(-\lambda + X\ln(\lambda) - \ln(X!))}{d\lambda}$$

$$\frac{d\ln p_X(X,\lambda)}{d\lambda} = -1 + \frac{X}{\lambda} \qquad \Rightarrow \qquad \frac{d^2\ln p_X(X,\lambda)}{d\lambda^2} = -\frac{X}{\lambda^2}$$

EXAMPLE

Then the Fisher information $-\frac{x}{\lambda}$ $E(x)=\lambda$

$$E\left(\frac{d^2 \ln p_X(X,\lambda)}{d\lambda^2}\right) = -\frac{1}{\lambda^2}E(X) = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

EXAMPLE

Then the Fisher information

$$E\left(\frac{d^2 \ln p_X(X,\lambda)}{d\lambda^2}\right) = -\frac{1}{\lambda^2}E\left(X\right) = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

$$\text{CRLB} = -\frac{1}{n(-\frac{1}{\lambda})} = \frac{\lambda}{n}$$

EXAMPLE

Then the Fisher information

$$E\left(\frac{d^2 \ln p_X(X,\lambda)}{d\lambda^2}\right) = -\frac{1}{\lambda^2}E\left(X\right) = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

$$\text{CRLB} = -\frac{1}{n(-\frac{1}{\lambda})} = \frac{\lambda}{n}$$

$$\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n \overset{\lambda}{\text{Var}(X_i)}}{n^2} = \frac{\sum_{i=1}^n \lambda}{n^2} = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

EXAMPLE

Then the Fisher information

$$E\left(\frac{d^2 \ln p_X(X,\lambda)}{d\lambda^2}\right) = -\frac{1}{\lambda^2}E(X) = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

$$\text{CRLB} = -\frac{1}{n(-\frac{1}{\lambda})} = \frac{\lambda}{n}$$

$$\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2} = \frac{\sum_{i=1}^n \lambda}{n^2} = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

Conclusion: The maximum-likelihood estimator $\hat{\lambda} = \bar{X}$ for the parameter λ of the Poisson distribution is **efficient**.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

The unbiased estimator $\hat{\theta}$ that minimizes the mean square error is called the **Minimum-Variance Unbiased Estimator** (**MVUE**) of θ.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Suppose that $X_1, X_2, ..., X_n$ is a random sample with a fixed sample size n, obtained from a population with the pdf

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

i.e $X_i \sim N(\mu, \sigma^2)$. We know that $\bar{X}$ is the ML estimator for μ.
Is it the best estimator for μ in terms of unbiasedness and minimum variance?

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Suppose that $X_1, X_2, ..., X_n$ is a random sample with a fixed sample size n, obtained from a population with the pdf

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

i.e $X_i \sim N(\mu, \sigma^2)$. We know that $\bar{X}$ is the ML estimator for μ.
Is it the best estimator for μ in terms of unbiasedness and minimum variance?
Solution:
1) By the theorem from Topic 2.2 - Unbiasedness (Slide 7), we know that $\bar{X}$ is an unbiased estimator for μ.
2) Let's find the CRLB for $f_X(x)$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Suppose that $X_1, X_2, ..., X_n$ is a random sample with a fixed sample size n, obtained from a population with the pdf

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

i.e $X_i \sim N(\mu, \sigma^2)$. We know that $\bar{X}$ is the ML estimator for μ.
Is it the best estimator for μ in terms of unbiasedness and minimum variance?
Solution:
1) By the theorem from Topic 2.2 - Unbiasedness (Slide 7), we know that $\bar{X}$ is an unbiased estimator for μ.
2) Let's find the CRLB for $f_X(x)$

$$\ln f_X(x, \mu) = -\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial \ln f_X(x, \mu)}{\partial \mu} = \frac{1}{2\sigma^2} 2(x-\mu) \quad \Rightarrow \quad \frac{\partial^2 \ln f_X(x, \mu)}{\partial \mu^2} = -\frac{1}{\sigma^2}$$

Properties of Point Estimators: Efficiency

Example

Hence, the CRLB:

$$CRLB = \frac{1}{-nE(\frac{\partial^2 \ln f_X(x,\mu)}{\partial \mu^2})} = \frac{1}{-nE(-\frac{1}{\sigma^2})} = \frac{\sigma^2}{n}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Hence, the CRLB:

$$CRLB = \frac{1}{-nE(\frac{\partial^2 \ln f_X(x,\mu)}{\partial \mu^2})} = \frac{1}{-nE(-\frac{1}{\sigma^2})} = \frac{\sigma^2}{n}$$

Now we can either derive the variance for $\bar{X}$, or we can recall the theorem from Topic 2.1 about the sampling distribution of $\bar{X}$ and note that the variance of the estimator is:

$$Var(\bar{X}) = \frac{\sigma^2}{n},$$

which coincides with the CRLB.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

Hence, the CRLB:

$$CRLB = \frac{1}{-nE(\frac{\partial^2 \ln f_X(x,\mu)}{\partial \mu^2})} = \frac{1}{-nE(-\frac{1}{\sigma^2})} = \frac{\sigma^2}{n}$$

Now we can either derive the variance for $\bar{X}$, or we can recall the theorem from Topic 2.1 about the sampling distribution of $\bar{X}$ and note that the variance of the estimator is:

$$Var(\bar{X}) = \frac{\sigma^2}{n},$$

which coincides with the CRLB.
Thus, we have shown that the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the ”best” estimator (MVUE) for the population mean.

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

DEFINITION

*The **mean square error** of the estimator $\hat{\theta}$, denoted by $MSE\left(\hat{\theta}\right)$, is defined as*

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

Properties of Point Estimators: Efficiency

Definition

*The **mean square error** of the estimator $\hat{\theta}$, denoted by $MSE\left(\hat{\theta}\right)$, is defined as*

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

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

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

DEFINITION

*The **mean square error** of the estimator $\hat{\theta}$, denoted by $MSE\left(\hat{\theta}\right)$, is defined as*

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

$$\begin{aligned}
\mathrm{MSE}\left(\hat{\theta}\right) &= E\left(\hat{\theta} - \theta\right)^2 = E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right) + \left(E\left(\hat{\theta}\right) - \theta\right)\right]^2 \\
&= E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + \left(E\left(\hat{\theta}\right) - \theta\right)^2 + 2\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right) - \theta\right)\right] \\
&= \underbrace{E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2}_{Var(\hat{\theta})} + E\Big(\underbrace{E\left(\hat{\theta}\right) - \theta}_{Bias(\hat{\theta})}\Big)^2 + 2\underbrace{E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)}_{=0}\left(E\left(\hat{\theta}\right) - \theta\right)
\end{aligned}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

DEFINITION

*The **mean square error** of the estimator $\hat{\theta}$, denoted by MSE $\left(\hat{\theta}\right)$, is defined as*

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

$$\begin{aligned}
\text{MSE}\left(\hat{\theta}\right) &= E\left(\hat{\theta} - \theta\right)^2 = E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right) + \left(E\left(\hat{\theta}\right) - \theta\right)\right]^2 \\
&= E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + \left(E\left(\hat{\theta}\right) - \theta\right)^2 + 2\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right) - \theta\right)\right] \\
&= E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + E\left(E\left(\hat{\theta}\right) - \theta\right)^2 + 2E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right) - \theta\right) \\
&= \text{Var}\left(\hat{\theta}\right) + \underbrace{\left(E\left(\hat{\theta}\right) - \theta\right)^2}_{\left(Bias\left(\hat{\theta}\right)\right)^2} = \text{Var}\left(\hat{\theta}\right) + \left(Bias\left(\hat{\theta}, \theta\right)\right)^2
\end{aligned}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

If X has a binomial distribution with parameters n and p, then $\hat{p}_1 = X/n$ is an unbiased estimator of p. Another estimator of p is $\hat{p}_2 = (X+1)/(n+2)$.
1) Derive the bias of $\hat{p}_2$.
2) Derive MSE$(\hat{p}_1)$ and MSE$(\hat{p}_2)$.
3) Show that for $p \approx 0.5$ MSE$(\hat{p}_2) <$ MSE$(\hat{p}_1)$.

Solution:

1) $$\text{bias}(\hat{p}_2) = E(\hat{p}_2) - p = E\left(\frac{X+1}{n+2}\right) - p = \frac{E(X)+1}{n+2} - p =$$

$$= \frac{np+1-np-2p}{n+2} = \frac{1-2p}{n+2}$$

PROPERTIES OF POINT ESTIMATORS: EFFICIENCY

EXAMPLE

2) $\underline{MSE(\hat{p}_1)} = Var(\hat{p}_1) + Bias^2(\hat{p}_1) = Var(\hat{p}_1) = Var\left(\frac{X}{n}\right) =$

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

$\underline{MSE(\hat{p}_2)} = Var(\hat{p}_2) + Bias^2(\hat{p}_2) = \frac{Var(X+1)}{(n+2)^2} + \frac{(1-2p)^2}{(n+2)^2} =$

$$= \frac{np(1-p) + (1-2p)^2}{(n+2)^2} = \frac{np - np^2 + 1 - 4p + 4p^2}{(n+2)^2} =$$

$$= \underline{\frac{1 + (n-4)p - (n-4)p^2}{(n+2)^2}}$$

3) $p \approx \frac{1}{2}$

$$MSE(\hat{p}_1) = \frac{1}{4n}$$

$$MSE(\hat{p}_2) = \frac{1 + (n-4)\frac{1}{2} - (n-4)\frac{1}{4}}{(n+2)^2} = \frac{4 + 2n - 8 - n + 4}{4(n+2)^2} = \frac{n}{4(n+2)^2}$$

$$MSE(\hat{p}_2) - MSE(\hat{p}_1) = \frac{n}{4(n+2)^2} - \frac{1}{4n} = \frac{n^2 - n^2 - 4n - 4}{4n(n+2)^2} =$$

$$= -\frac{n+1}{(n+2)^2} < 0 \quad n \geq 1$$

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

sometimes a biased estimator can be more preferable than a unbiased in terms of smaller MSE.