Estimation Theory

Part I: What is an estimator?

Suwichaya Suwanwimolkul[1]

[1] Electrical Engineering
Chulalongkorn University

Lecture I
Semester I, 2023

Table of contents

1. Table of contents
2. Motivation
3. Notation
   - Sampling from distribution
   - What is an estimator?
   - Is there a hat ?
4. General Properties of Estimators
   - Overall
   - Bias
   - Efficiency
   - Consistency
   - Markov inequality & Chebyshev's inequality
   - Summary: Properties of Estimators
   - Asymptotic Properties of Estimators
5. Examples of estimators
   - Sample mean
   - Sample variance
   - Method of moments
6. Homework
   - Homework 1.1 (scores=7)
   - Homework 1.2 (scores=10)
   - Homework 1.3 (scores=3)
7. Index

Motivation

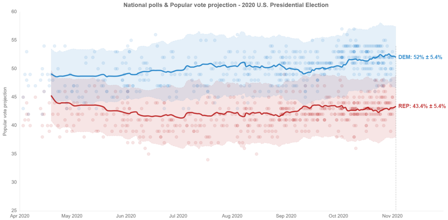
We have a random variable X, but its *probability density function (pdf)* $f_X(x)$ or *probability mass function (pmf)* $p_X(x)$ is not known.

Motivation

Why do we need to know the pdf/pmf of X ?
Answer. The characteristics of our data.

Political Polls

Images are taken from https://www.indystar.com



Images are taken from https://macleans.ca/

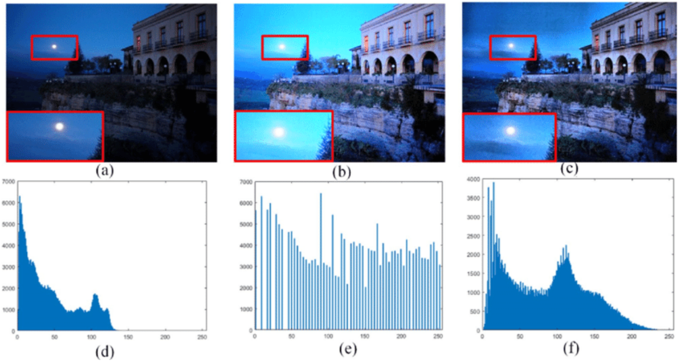
Can you compare the difference between the polls in May vs November?
What does it mean if there is a high overlapping?

Motivation

Why do we need to know the pdf/pmf of X ?

Answer. The characteristics of our data.

Image Histrogram



Images are taken from Yunliang Qi, Zhen Yang, Wenhao Sun, Meng Lou, "A Comprehensive Overview of Image Enhancement Techniques," Archives of Computational Methods in Engineering 2021.

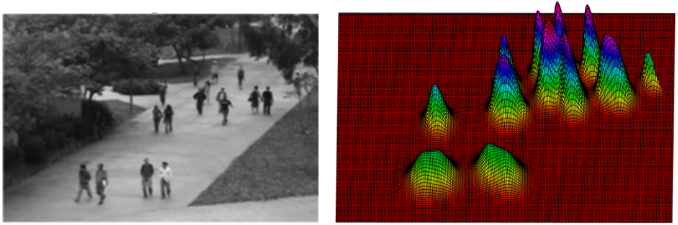
Same image but with different illumination (some content, different color)

Similar density function, but different parameters.

Motivation

Why do we need to know the pdf/pmf of X ?
Answer. The characteristics of our data.

Image Histrogram



Images are taken from https://teamkage.github.io

Content also depends on density of pixel value in spatial dimension
Similar density $==$ similar types of figures.

Motivation

We have a random variable X, but its *probability density function (pdf)* $f_X(x)$ or *probability mass function (pmf)* $p_X(x)$ is not known.

There are two possible scenarios:

1. $f_X(x)$ or $p_X(x)$ is completely unknown.
2. The formulae is known, but parameters θ are unknown

Why do we need an estimator?

We have a random variable X, but its *probability density function (pdf)* $f_X(x)$ or *probability mass function (pmf)* $p_X(x)$ is not known.

There are two possible scenarios:

1. $f_X(x)$ or $p_X(x)$ is completely unknown.
2. The formulae is known, but parameters θ are unknown.

This course *'Estimation Theory'* we consider only the second case...

For example, we know/assume that $X \sim \mathcal{N}(\mu, \sigma^2)$, but we may not know what is the value of μ and σ^2.

What if we still need to find the unknowns ?

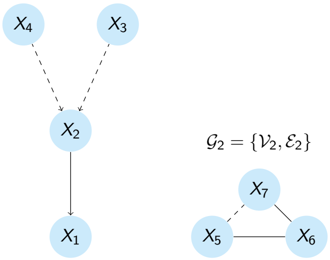
There are two possible scenarios:

1. $f_X(x)$ or $p_X(x)$ is unknown. $\Rightarrow$ Structure Learning
2. The parameters θ are unknown. $\Rightarrow$ Learning/estimating parameters

Example

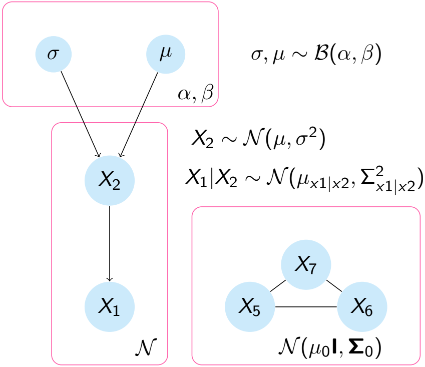
Structure Learning

$\mathcal{G}_1 = \{\mathcal{V}_1, \mathcal{E}_1\}$



$\mathcal{G}_2 = \{\mathcal{V}_2, \mathcal{E}_2\}$

Learning/estimating parameters



$\sigma, \mu \sim \mathcal{B}(\alpha, \beta)$

$X_2 \sim \mathcal{N}(\mu, \sigma^2)$

$X_1 | X_2 \sim \mathcal{N}(\mu_{x1|x2}, \Sigma^2_{x1|x2})$

$$p(x_1) = \int_{x_4}\int_{x_3}\int_{x_2} p(x_1|x_2)p(x_2|x_3, x_4)p(x_3)p(x_4)$$

$$p(x_5) = \int_{x_7}\int_{x_6} p(x_5, x_6, x_7)$$

$$p(x_1) = \int_{x_4}\int_{x_3}\int_{x_2} p(x_1|x_2)p(x_2|\sigma, \mu)p(\sigma; \alpha, \beta)p(\mu; \alpha, \beta)$$

$$p(x_5) = \int_{x_7}\int_{x_6} p(x_5, x_6, x_7; \mu_0 \mathbf{I}, \mathbf{\Sigma}_0)$$

What if we still need to find the unknowns ?

There are two possible scenarios:

1. $f_X(x)$ or $p_X(x)$ is unknown. ⇒ Structure Learning
2. The parameters θ are unknown. ⇒ Learning/estimating parameters

Note that eventually we still need to assume something.
So, learning/estimating just the parameters are not that bad.

Notation

1 Table of contents

2 Motivation

3 Notation
    Sampling from distribution
    What is an estimator?
    Is there a hat ?

4 General Properties of Estimators

5 Examples of estimators

6 Homework

7 Index

Sampling from distribution

In a typical setting, we have a random variable X, but *its pdf* $f_X(x;\theta)$ or *pmf* $p_X(x;\theta)$ is not known.

Suppose that the unknown parameter θ is in real space, *i.e.*, $\theta \in \mathbb{R}$, we estimate the parameter θ by **randomly sampling** $X_1, X_2, ..., X_N$ from the same distribution as X.

What is ‘random sampling’?

Definition 1: ‘Randomly sampling/random samples’ implication

‘Randomly sampling/random samples’ implies

$$X_1, X_2, ..., X_N \overset{iid}{\sim} f_X(x),$$

where $f_X(x)$ denotes the pdf or pmf of X.

On the other words,

$X_1, X_2, ..., X_N$ is drawn independently from a distribution.

$X_1, X_2, ..., X_N$ is assumed to have the same distribution as X.

What is an estimator?

Our goal is to estimate the value of θ with an estimator $\hat{\Theta}$.

Definition 2: Estimator

An estimator $\hat{\Theta}$ can be defined as a function of random samples with a finite sample size $N < \infty$, that is,

$$X_1, X_2, ..., X_N \rightarrow \hat{\Theta} = f(X_1, X_2, ..., X_N) \tag{1}$$

Note. $\hat{\Theta}$ is a random variable because it is a function of the random variables $X_1, X_2, ..., X_N$.

Example of estimators

Example 1: Example of estimators

If $X_1, X_2, ..., X_N$ is a random sample from the same distribution as X, which is a Gaussian distribution with mean μ and variance σ^2, *i.e.*, $X_1, X_2, ..., X_N \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\hat{\Theta}_1 := \text{Mean}(X_1, X_2, ..., X_N);$$

$$\hat{\Theta}_2 := \min(X_1, X_2, ..., X_N)$$

$$\hat{\Theta}_3 := \frac{1}{2}(X_1 + X_N);$$

Is there a hat ?

θ: When we write θ as it is, *e.g.*, $\theta = \mu$, $\theta \in (\alpha, \beta)$, $\theta = \sigma^2$—without any hat nor a bar, it is the **true** value (aka **ground truth**).

$\hat{\Theta}$ & $\hat{\theta}$: But when we write it with a hat, *e.g.*, $\hat{\theta}$ or $\hat{\Theta}$, then it will be a **pseudo** which could be either an **estimate** or an **estimator** for θ.

Estimate vs. Estimator

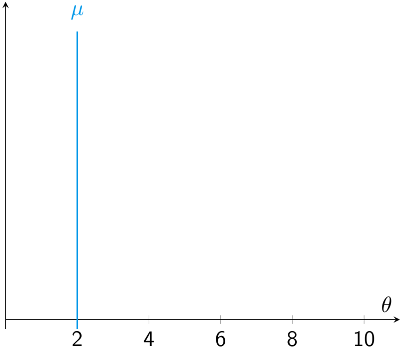
Definition 3: Estimate vs. Estimator

- **Estimate** $\hat{\theta}$:= observation or a number resulting from an estimation. For example, $\hat{\theta} = \bar{x} = 0.5$.
- **Estimator** $\hat{\Theta}$:= a random variable. For example, $\hat{\Theta} = \bar{X} \sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2)$.

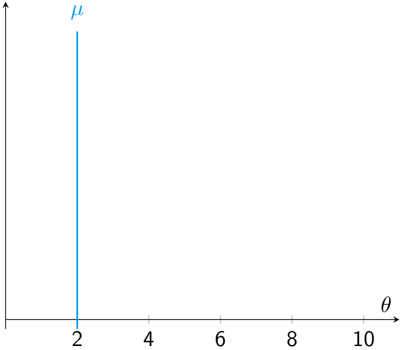
Example A — Estimate vs. Estimator

If the unknown θ is a **ground truth** parameter value, *e.g.*, $\theta = \mu = 2$.
We observe $X_1, X_2, ..., X_N$ and find the estimator $\hat{\Theta}$.

Estimate $\hat{\theta} =?$



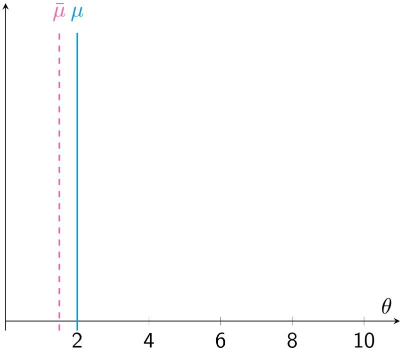
Estimator $\hat{\Theta} =?$



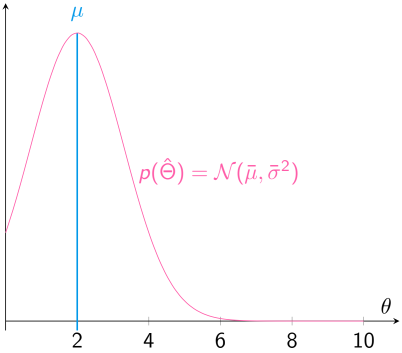
Example A — Estimate vs. Estimator

If the unknown θ is a **ground truth** parameter value, *e.g.*, $\theta = \mu = 2$.
We observe $X_1, X_2, ..., X_N$ and find the estimator $\hat{\Theta}$.

Estimate $\hat{\theta} = \bar{\mu}$



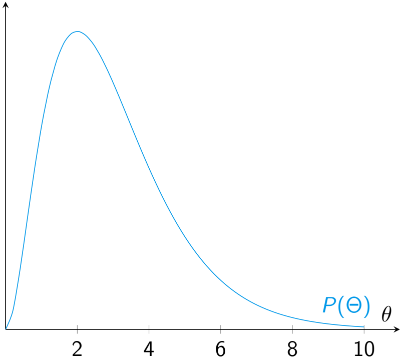
Estimator $\hat{\Theta} \sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2)$



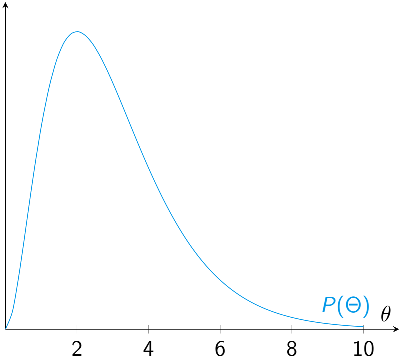
Examples B — Estimate vs. Estimator

If the unknown Θ is a random variable, *e.g.*, $\Theta \sim \Gamma(\alpha, \beta)$.
We observe $X_1, X_2, ..., X_N$ and find the estimator $\hat{\Theta}$.

Estimate $\hat{\theta} =?$



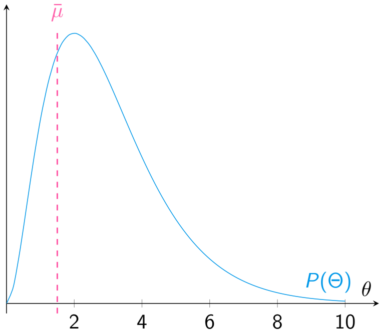
Estimator $\hat{\Theta} =?$



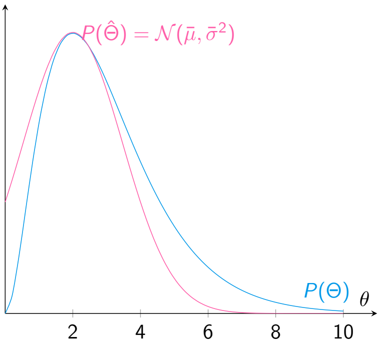
Examples B — Estimate vs. Estimator

If the unknown Θ is a random variable, *e.g.*, $\Theta \sim \Gamma(\alpha, \beta)$.
We observe $X_1, X_2, ..., X_N$ and find the estimator $\hat{\Theta}$.

Estimate $\hat{\theta} = \bar{\mu}$



Estimator $\hat{\Theta} \sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2)$



General Properties of Estimators

1. Table of contents
2. Motivation
3. Notation
4. General Properties of Estimators
    - Overall
    - Bias
    - Efficiency
    - Consistency
    - Markov inequality & Chebyshev's inequality
    - Summary: Properties of Estimators
    - Asymptotic Properties of Estimators
5. Examples of estimators
6. Homework
7. Index

Overall

What is the performance of the estimator $\hat{\Theta} = f(X_1, X_2, ..., X_N)$?
We can measure the estimator $\hat{\Theta}$ using the following (finite sample) properties...

- Bias $\Rightarrow$ On average how far $\hat{\Theta}$ is from the real value of θ.
- Efficiency $\Rightarrow$ A good $\hat{\Theta}$ should provide the lowest error.
- Consistency $\Rightarrow$ As N gets larger, $\hat{\Theta}$ converges to the real θ.

Next week: we will consider the asymptotic properties of the estimator as N becomes very large large, *i.e.*, $N \to \infty$.

Bias

Definition 4: Bias

Let $\hat{\Theta} := f(X_1, X_2, ..., X_N)$ be the point estimator for θ. The bias of the point estimator is defined as

$$\text{Bias}_\theta(\hat{\Theta}) = E(\hat{\Theta}) - \theta. \tag{2}$$

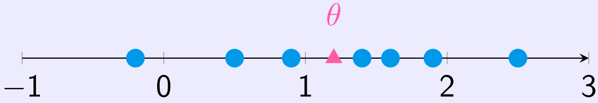
Unbiased estimator

Definition 5: Unbiased estimator

We say that $\hat{\Theta}$ is an unbiased estimator of θ, if

$$B(\hat{\Theta}) = 0 \text{ or } E[\hat{\Theta}] = \theta, \tag{3}$$

for all possible values of θ. Otherwise, $\hat{\Theta}$ is biased.



Examples

Example 2: Is the estimator $\hat{\Theta}$ is biased or unbiased ?

Our estimator is $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ where $X_1, X_2, ..., X_N$ is **randomly sampled** from a Bernoulli distribution with parameter p, *i.e.*, Bernoulli(p).

The goal is to show that $\mathrm{E}[\hat{\Theta}] = p$ ...

$$
\begin{aligned}
& \mathrm{E}[\hat{\Theta}] && = \mathrm{E}_X[\frac{1}{N}\sum_{i=1}^{N} X_i] \\
& X_1, X_2, ..., X_N \sim iid && = \frac{1}{N}\mathrm{E}_X[\sum_{i=1}^{N} X_i] = \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_X[X_i] \\
& \because \mathrm{E}_X[X_i] = p, && = \frac{1}{N}(Np) = p
\end{aligned}
$$

$\hat{\Theta}$ is an unbiased estimator.

Examples

Example 2 (Cont'): Is the estimator $\hat{\Theta}$ is biased or unbiased ?

Our estimator is $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ where $X_1, X_2, ..., X_N$ is **randomly sampled** from a Bernoulli distribution with parameter p, *i.e.*, Bernoulli(p).

The goal is to show that $\mathrm{E}[\hat{\Theta}] = p$ ...

$$\begin{aligned}
\mathrm{E}[\hat{\Theta}] \quad &= \mathrm{E}_X[\frac{1}{N}\sum_{i=1}^{N} X_i] \\
X_1, X_2, ..., X_N \sim iid \quad &= \frac{1}{N}\mathrm{E}_X[\sum_{i=1}^{N} X_i] = \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_X[X_i] \\
\because \mathrm{E}_X[X_i] = p, \quad &= \frac{1}{N}(Np) = p.
\end{aligned}$$

Bernoulli ⇒ plug-in the value

$\hat{\Theta}$ is an unbiased estimator.

Quiz

Example 3: Same estimator ($\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$), but what if $X_1, X_2, ..., X_N$ is from a Gaussian distribution?

What do you think? ...

From the previous example, did we use the fact that $X_1, X_2, ..., X_N$ came from a Bernoulli distribution?

Quiz

Example 3 (Cont'): Same estimator ($\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$), but what if $X_1, X_2, ..., X_N$ is from a Gaussian distribution?

What do you think? ...

From the above example, did we use the fact that $X_1, X_2, ..., X_N$ came from a Bernoulli distribution?

Answer: Only the last line.
There is a name for the estimator...
We call it *'sample mean.'*

Property 1: For $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$, one can show that $\mathrm{E}[\hat{\Theta}] = \mu$.

$$
\begin{aligned}
\mathrm{E}[\hat{\Theta}] &= \mathrm{E}_X[\frac{1}{N}\sum_{i=1}^{N} X_i] \\
&= \frac{1}{N}\mathrm{E}_X[\sum_{i=1}^{N} X_i] = \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_X[X_i] \\
&= \frac{1}{N}(N\mu) = \mu
\end{aligned}
$$

$\hat{\Theta}$ is (always) an unbiased estimator.

Efficiency

Efficiency of an estimator $\hat{\Theta}$ is measured by comparing their variance to the Cramer–Rao lower bound (LB).

Definition 6: Efficiency (formal definition)

If $\hat{\Theta}$ is an unbiased estimator of θ, the efficiency of $\hat{\Theta}$ is defined as:

$$\mathrm{Eff}_{\theta}(\hat{\Theta}) = \frac{LB}{\mathrm{Var}[\hat{\Theta}]} = \frac{1}{LB^{-1}\mathrm{Var}[\hat{\Theta}]}, \tag{4}$$

where $\mathrm{Var}[\hat{\Theta}] = \mathrm{E}[(\hat{\Theta} - \mu_{\hat{\Theta}})^2]$ and LB denotes Cramer–Rao lower bound.

We will review the convergence in probability again in Lecture 6.
But we mention it here to show how $\mathrm{Eff}_{\theta}(\hat{\Theta})$ is the variance of the estimator $\mathrm{Var}[\hat{\Theta}]$ as well as the theoretical MSE (Slide 36).

Efficiency as Cramer–Rao lower bound (LB)

Previously,

$$\mathrm{Eff}_\theta(\hat{\Theta}) = \frac{LB}{\mathrm{Var}[\hat{\Theta}]} = \frac{1}{LB^{-1}\mathrm{Var}[\hat{\Theta}]}.$$

- Cramer–Rao lower bound (LB) is related to the PDF of X and θ.
- Specifically, Cramer–Rao lower bound is inversely proportional to Fisher information $I(\theta)$, *i.e.*, $LB^{-1} = NI(\theta)$.
- Fisher information $I(\theta)$ measures the amount of information that X carries about θ, related to pdf.
- An estimator $\hat{\Theta}$ is said to be efficient, if and only if $\mathrm{Eff}_\theta(\hat{\Theta}) = 1$.

Efficiency as mean-square error

If $\hat{\Theta}$ is an unbiased estimator of θ, *i.e.*, $\mu_{\hat{\Theta}} = \theta$, then $\mathrm{Eff}_\theta(\hat{\Theta})$ is **inverse proportional** to $\mathrm{Var}[\hat{\Theta}] = \mathrm{E}[(\hat{\Theta} - \mu_{\hat{\Theta}})^2]$.

Thus, **efficiency** can **also** be characterized by having **lower variances** to the **true value** θ, which can be measured by mean-square error (MSE) ...

Definition 7: Mean-square error (MSE)

The mean squared error of an estimator $\hat{\Theta}$ is defined as:

$$\mathrm{MSE}_\theta(\hat{\Theta}) = \mathrm{E}_{\hat{\Theta}}[(\hat{\Theta} - \theta)^2]. \tag{5}$$

The smaller MSE is generally indicative of the better estimator.

Efficiency as mean-square error—II

Definition 8: Mean-square error (MSE) — Alternative calculation

The mean squared error of an estimator $\hat{\Theta}$ can also be calculated as:

$$\mathrm{MSE}_{\theta}(\hat{\Theta}) = \mathrm{E}_{\hat{\Theta}}[(\hat{\Theta} - \theta)^2] = \mathrm{Var}_{\hat{\Theta}}[\hat{\Theta}] + \mathrm{Bias}_{\theta}(\hat{\Theta})^2. \tag{6}$$

Homework 1. . Can you show that Eq.(6) is true?

Note. This MSE is calculated using the theoretical expectation operation, so it is a theoretical MSE.

Example

Example 4: comparing the efficiency of $\hat{\Theta}_1$ vs $\hat{\Theta}_2$

Let X_1, X_2, ..., X_N be a random variable from a distribution with mean $E_X[X_i] = \theta$ and variance $\mathrm{Var}_X[X_i] = \sigma^2$. The following $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are the estimators for μ.

$$\hat{\Theta}_1 := \frac{1}{N}\sum_{i=1}^{N} X_i;$$

$$\hat{\Theta}_2 := \frac{1}{2}(X_1 + X_N);$$

Q1. Are the above estimators biased or unbiased?

Q2. Which estimator is better by showing $\mathrm{MSE}_\theta(\hat{\Theta}_1)$ and $\mathrm{MSE}_\theta(\hat{\Theta}_2)$? *e.g.*, if $\mathrm{MSE}_\theta(\hat{\Theta}_1) < \mathrm{MSE}_\theta(\hat{\Theta}_2)$, then $\hat{\Theta}_1$ is better than $\hat{\Theta}_2$.

Answering Q1—Are the estimators biased or unbiased?

Example 4 (Cont'): Answering Q1

Let X_1, X_2, ..., X_N be a random variable from a distribution with mean $E_X[X_i] = \theta$ and variance $\mathrm{Var}_X[X_i] = \sigma^2$. The following $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are the estimators for μ.

$$\hat{\Theta}_1 := \frac{1}{N}\sum_{i=1}^{N} X_i;$$

$$\hat{\Theta}_2 := \frac{1}{2}(X_1 + X_N);$$

Q1. Are the above estimators biased or unbiased?

$$\mathrm{E}[\hat{\Theta}_1] = ?$$

$$\mathrm{E}[\hat{\Theta}_2] = ?$$

<u>ANSWER:</u> Both estimators are unbiased.

Answering Q1—Are the estimators biased or unbiased?

Example 4 (Cont'): Answering Q1

Let X_1, X_2, ..., X_N be a random variable from a distribution with mean $E_X[X_i] = \theta$ and variance $\mathrm{Var}_X[X_i] = \sigma^2$. The following $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are the estimators for μ.

$$\hat{\Theta}_1 := \frac{1}{N}\sum_{i=1}^{N} X_i;$$

$$\hat{\Theta}_2 := \frac{1}{2}(X_1 + X_N);$$

Q1. Are the above estimators biased or unbiased?

$$\mathrm{E}[\hat{\Theta}_1] = \theta \text{ (see Page.31)}$$

$$\mathrm{E}[\hat{\Theta}_2] = \frac{1}{2}(\mathrm{E}[X_1] + \mathrm{E}[X_N]) = \theta;$$

ANSWER: Both estimators are unbiased.

Answering Q2 — the efficiency of $\hat{\Theta}_1$ vs $\hat{\Theta}_2$

Example 4 (Cont'): Answering Q2

Let X_1, X_2, ..., X_N be a random variable from a distribution with mean $E_X[X_i] = \theta$ and variance $\mathrm{Var}_X[X_i] = \sigma^2$. The following $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are the estimators for μ.

$\hat{\Theta}_1 := \frac{1}{N}\sum_{i=1}^{N} X_i$;

$\hat{\Theta}_2 := \frac{1}{2}(X_1 + X_N)$;

Q2. Which estimator is better by showing $\mathrm{MSE}_\theta(\hat{\Theta}_1)$ and $\mathrm{MSE}_\theta(\hat{\Theta}_2)$? *e.g.*, if $\mathrm{MSE}_\theta(\hat{\Theta}_1) < \mathrm{MSE}_\theta(\hat{\Theta}_2)$, then $\hat{\Theta}_1$ is better than $\hat{\Theta}_2$.

Example 4 (Cont'): Answering Q2 — I

Q2.1 Continue...We apply Eq.(6).

$$\mathrm{MSE}_\theta(\hat{\Theta}_1) = \mathrm{Var}[\hat{\Theta}_1] + \mathrm{Bias}_\theta(\hat{\Theta}_1)^2 = \mathrm{Var}[\hat{\Theta}_1] + 0, \quad \because \mathrm{Bias}_\theta(\hat{\Theta}_1) = 0 \text{ (Q1)}$$

$$\mathrm{Var}[\hat{\Theta}_1] = \mathrm{E}_{\hat{\Theta}_1}[(\hat{\Theta}_1 - \mathrm{E}[\hat{\Theta}_1])^2]$$

$$\text{Let } \mu_{\hat{\Theta}_1} := \mathrm{E}[\hat{\Theta}_1] \text{ and } \hat{\Theta}_1 = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$= \mathrm{E}_X[(\frac{1}{N}\sum_N X_i - \mu_{\hat{\Theta}_1})^2] = \mathrm{E}_X[\frac{1}{N^2}(\sum_N X_i - \sum_N \mu_{\hat{\Theta}_1})^2]$$

$$= \mathrm{E}_X[\frac{1}{N^2}((X_1 - \mu_{\hat{\Theta}_1}) + \ldots + (X_N - \mu_{\hat{\Theta}_1}))^2]$$

$$= \mathrm{E}_X[\frac{1}{N^2}\sum_N (X_i - \mu_{\hat{\Theta}_1})^2] = \frac{1}{N}(\frac{1}{N}\sum_N \mathrm{E}_X[(X_i - \mu_{\hat{\Theta}_1})^2]), \because \text{iid}$$

$$= \frac{1}{N}(\frac{N}{N}\mathrm{Var}_X[X_i]) = \frac{1}{N}\mathrm{Var}_X[X_i]^\dagger$$

† This is the variance of the sample mean estimator.

Example 4 (Cont'): Answering Q2 — II

Q2.2 Let examine $\mathrm{MSE}_\theta(\hat{\Theta}_2)$ ....

$$\begin{aligned}\mathrm{MSE}_\theta(\hat{\Theta}_2) &= \mathrm{E}[(\hat{\Theta}_2-\theta)^2]\\ &= \mathrm{Var}[\hat{\Theta}_2] + \mathrm{Bias}_\theta(\hat{\Theta}_2)^2 = \mathrm{Var}[\hat{\Theta}_2], \quad \because \text{see Q1.}\end{aligned}$$

Let $\mu_{\hat{\Theta}_2} := \mathrm{E}[\hat{\Theta}_2]$

$$\begin{aligned}\mathrm{Var}[\hat{\Theta}_2] &= \mathrm{E}_X[(\hat{\Theta}_2-\mu_{\hat{\Theta}_2})^2] = \mathrm{E}_X[(\frac{1}{2}\sum_{i\in\{1,N\}} X_i - \frac{1}{2}\sum_{i\in\{1,N\}}\mu_{\hat{\Theta}_2})^2]\\ &= \mathrm{E}_X[\frac{1}{2^2}((X_1-\mu_{\hat{\Theta}_2})+(X_N-\mu_{\hat{\Theta}_2}))^2]\\ &= \mathrm{E}_X[\frac{1}{2^2}((X_1-\mu_{\hat{\Theta}_2})^2+(X_N-\mu_{\hat{\Theta}_2})^2)], \quad \because X_1, X_2, \ldots \sim \text{iid}\\ &= \frac{1}{2^2}(\mathrm{Var}_X[X_1]+\mathrm{Var}_X[X_N]) = \frac{1}{2}\mathrm{Var}_X[X_i]\end{aligned}$$

Example 4 (Cont'): Answering Q2 — III

Let X_1, X_2, ..., X_N be a random variable from a distribution with mean $E_X[X_i] = \theta$ and variance $\mathrm{Var}_X[X_i] = \sigma^2$. The following $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are the estimators for μ.

$$\hat{\Theta}_1 := \frac{1}{N}\sum_{i=1}^{N} X_i;$$

$$\hat{\Theta}_2 := \frac{1}{2}(X_1 + X_N);$$

Q2. Which estimator is better by showing $\mathrm{MSE}_\theta(\hat{\Theta}_1)$ and $\mathrm{MSE}_\theta(\hat{\Theta}_2)$?

ANSWER: $\hat{\Theta}_1$ is better than $\hat{\Theta}_2$ because $\mathrm{MSE}_\theta(\hat{\Theta}_1) \leq \mathrm{MSE}_\theta(\hat{\Theta}_2)$
or $\frac{1}{N}\mathrm{Var}_X[X_i] \leq \frac{1}{2}\mathrm{Var}_X[X_i]$ for $N > 2$.

Note that both estimators are unbiased,
but the unbiasedness does not mean a lower error.

Consistency

We define an estimator $\hat{\Theta}$ as a function of $X_1, X_2, ..., X_n$ (Slide 17).

We say that the estimator is *consistent*, if as the sample size n increases, the estimator $\hat{\Theta}$ converges to the real value of θ.

Consistency (formal definition)

Definition 9: Consistency

Let $\hat{\Theta}_1, \hat{\Theta}_2, ..., \hat{\Theta}_n, ...,$ be a sequence of the point estimator $\hat{\Theta}$; each of which is sorted by the number of the random samples n.

Here, we write an estimator as a function of n random samples explicitly, *i.e.*, $\hat{\Theta} = \hat{\Theta}_n = f(X_1, X_2, ..., X_n)$.

For any $\epsilon > 0$, we say that $\hat{\Theta}$ is a consistent estimator of θ, if

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0 \quad \text{for all possible value of } \theta. \tag{7}$$

In other words, an estimator $\hat{\Theta}$ is a consistent estimator for θ, if $\hat{\Theta}$ converges to θ in probability, or $\hat{\Theta} \xrightarrow{p} \theta$.

We will review **convergence in probability** again in Lecture II.

Markov inequality & Chebyshev's inequality

Two useful tools for providing convergence in probability are:

Definition 10: Markov's inequality

Let X be a positive random variable, *i.e.*, $P(X > 0) = 1$.
For any $\epsilon > 0$,

$$P(X > \epsilon) \leq \frac{1}{\epsilon}\mathsf{E}_X[x]. \tag{8}$$

Definition 11: Chebyshev's inequality

Let X be a random variable with mean μ and variance σ^2.
For any $\epsilon > 0$,

$$P(|X - \mu| > \epsilon) \leq \frac{1}{\epsilon^2}\sigma^2. \tag{9}$$

Example: Convergence of sample mean estimators

Example 5: Convergence of sample mean estimators

Show that if the random sample $X_1, X_2, ..., X_N$ is the same normal distribution as X with mean μ and variance σ^2, then $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ converges in probability to μ.

In other words, an estimator X_n is a consistent estimator of μ.

Goal: show that

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0 \text{ for all possible value of } \theta.$$

There are 2 ways to show this:

1. Q-function.

2. Chebyshev's inequality.

Example 5 (Cont'): Q-function—I

Show that if the random sample $X_1, X_2, ..., X_N$ is from the same normal distribution as X with mean μ and variance σ^2, then $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ converges in probability to μ. In other words, an estimator X_n is a

consistent estimator of μ.

---

1. Q-function.
For $Y \sim \mathcal{N}(\mu, \sigma^2)$, the Q-function of $|Y| \geq \epsilon$ is :

$$P(|Y| \geq \epsilon) = 2P(Y \geq \epsilon) = 2\left(1 - \Phi\left(\frac{\epsilon - \mu}{\sigma}\right)\right) \tag{10}$$

Let us consider $P(|\hat{\Theta} - \mu| \geq \varepsilon)$.

Example 5 (Cont'): Q-function—II

Show that if the random sample $X_1, X_2, ..., X_N$ is from the same normal distribution as X with mean μ and variance σ^2, then $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ converges in probability to μ. In other words, an estimator X_n is a consistent estimator of μ.

---

1. Q-function.

For $Y \sim \mathcal{N}(\mu, \sigma^2)$, the Q-function of $|Y| \geq \epsilon$ is :

$$P(|Y| \geq \epsilon) = 2P(Y \geq \epsilon) = 2\left(1 - \Phi\left(\frac{\epsilon - \mu}{\sigma}\right)\right) \tag{11}$$

Let us consider $P(|\hat{\Theta} - \mu| \geq \varepsilon)$...

$\because X_i \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \hat{\Theta} = \frac{1}{N}\sum_{i=1}^{N} X_i$ is also a Gaussian $\mathcal{N}(\mu, \left(\frac{\sigma}{\sqrt{N}}\right)^2)$.

Thus, $\hat{\Theta} - \mu \sim \mathcal{N}(0, \left(\frac{\sigma}{\sqrt{N}}\right)^2)$.

Example 5 (Cont'): Q-function—III

Show that if the random sample $X_1, X_2, ..., X_N$ is from the same normal distribution as X with mean μ and variance σ^2, then $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ converges in probability to μ. In other words, an estimator X_n is a

consistent estimator of μ.

---

1. Q-function. Cont'

Once we have $\hat{\Theta} - \mu \sim \mathcal{N}(0, \left(\frac{\sigma}{\sqrt{N}}\right)^2)$, we can apply Q-function Eq.(11) as follows:

$$P(|\hat{\Theta} - \mu| \geq \epsilon) = 2P(\hat{\Theta} - \mu \geq \epsilon) = 2\left(1 - \Phi\left(\frac{\epsilon}{\sigma/\sqrt{N}}\right)\right) \qquad (12)$$

$\because$ as $N \to \infty$, $\Phi\left(\frac{\epsilon}{\sigma/\sqrt{N}}\right) \to 1$. Then, $P(|\hat{\Theta} - \mu| \geq \epsilon) \to 0$.

Example 5 (Cont'): Chebyshev's inequality

Show that if the random sample $X_1, X_2, ..., X_N$ is from the same normal distribution as X with mean μ and variance σ^2, then $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ converges in probability to μ. In other words, an estimator X_n is a

consistent estimator of μ.

---

2. Chebyshev's inequality.

Because $\hat{\Theta} \sim \mathcal{N}(\mu, \left(\frac{\sigma}{\sqrt{N}}\right)^2)$, we can plug in the mean and variance into the Chebyshev's inequality: $P(|\hat{\Theta} - \mu| > \epsilon) \leq \frac{1}{\epsilon^2}\frac{\sigma^2}{N}$.

$$\text{So, as } N \to \infty,\ P(|\hat{\Theta} - \mu| > \epsilon) \leq 0.$$

Example 6: Chebyshev's inequality

What if this time, $X_i \sim$ Bernoulli with mean p and variance $p(1-p)$.
Will $\hat{\Theta} := \frac{1}{N}\sum_{i=1}^{N} X_i$ converges in probability to p ?

To apply **Chebyshev's inequality**, we need to know $\mathrm{E}_{\hat{\Theta}}[\hat{\Theta}]$ and $\mathrm{Var}_{\hat{\Theta}}[\hat{\Theta}]$.

1. $\mathrm{E}_{\hat{\Theta}}[\hat{\Theta}] = \mathrm{E}_{X_i}[\hat{\Theta}] = \mathrm{E}_{X_i}[\frac{1}{N}\sum_{i=1}^{N} X_i] = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_{X_i}[X_i] = p$
2. $\mathrm{Var}_{\hat{\Theta}}[\hat{\Theta}] = \mathrm{Var}_{X_i}[\hat{\Theta}] = \mathrm{E}_{\hat{\Theta}}[(\hat{\Theta} - \mathrm{E}_{\hat{\Theta}}[\hat{\Theta}])^2] = \frac{\sigma^2}{N}$

$$\begin{aligned} P(|\hat{\Theta} - \mu_{\hat{\Theta}}| > \epsilon) &= P(|\hat{\Theta} - p| > \epsilon) \\ &\leq \frac{\sigma^2_{\hat{\Theta}}}{\epsilon^2} \\ &= \frac{\sigma^2}{N\epsilon^2} \end{aligned}$$

Do we need to rely on the distribution of X_i? Answer: No.
Actually, this is known as the weak law of large numbers (WLLN).
The formal definition will be discussed in Lecture II.

Summary: Common Properties

Property 2: An estimators may satisfy the following properties

Let θ_o denote the **true** unknown parameter.

- $\hat{\Theta}$ is **unbiased**, if

$$\mathrm{Bias}_{\theta_o}(\hat{\Theta}) = \mathrm{E}[\hat{\Theta}] - \theta_o = 0. \tag{13}$$

- $\hat{\Theta}$ is **consistent**, if

$$\lim_{n\to\infty} P(|\hat{\Theta} - \theta_o| > \epsilon) = 0. \tag{14}$$

- $\hat{\Theta}$'s efficiency can be measured by

$$\mathrm{Eff}_{\theta}(\hat{\Theta}) = \frac{LB}{\mathrm{Var}[\hat{\Theta}]} = \frac{1}{LB^{-1}\mathrm{Var}[\hat{\Theta}]}, \tag{15}$$

  where $\mathrm{Var}[\hat{\Theta}] =\ \mathrm{E}[(\hat{\Theta} - \mu_{\hat{\Theta}})^2]$ and LB denotes Cramer–Rao lower bound.

Common Asymptotic Properties

Property 3: Some estimators may satisfy the following properties

Let θ_o denote the **true** unknown parameter.

- $\hat{\Theta}$ is *asymptotically* **unbiased**, if $\lim_{n\to\infty} \text{Bias}_{\theta_o}(\hat{\Theta}) = 0$.
- $\hat{\Theta}$ is *(asymptotically)* **consistent**, if $\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta_o| > \epsilon) = 0$.
- $\hat{\Theta}$ is $\sqrt{N}$*-asymptotically* normally distributed, if $\hat{\Theta}$ **converges in distribution** to $\mathcal{N}(\theta_o, N\sigma^2)$, where $\text{Avar}[\hat{\Theta}] = N\sigma^2$ denotes the asymptotical covariance of $\hat{\Theta}$.
- If $\hat{\Theta}$ is an unbiased and $\sqrt{N}$*-asymptotically* normally distributed. $\hat{\Theta}$ is said to be *asymptotically* **efficient**, if $\text{Avar}[\hat{\Theta}] \succeq \mathbb{C}$ where $\mathbb{C}$ is an important lower bound which could be the lower bound of $\text{Eff}_\theta(\hat{\Theta})$ defined in Eq.(15).

Examples of estimators

1 Table of contents

2 Motivation

3 Notation

4 General Properties of Estimators

5 Examples of estimators
   - Sample mean
   - Sample variance
   - Method of moments

6 Homework

7 Index

Sample mean

Let X_1, X_2, ..., X_n be a random sample sharing the same distribution as X, but the parameters (*e.g.*, mean and variance) could be unknown.

Definition 12: Sample mean

Sample mean, often, used as an estimator for the mean value of X, is defined as

$$\hat{\Theta}_\mu := \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{16}$$

We have shown that **sample mean** has the following properties:

- an unbiased estimator (see Property 1)
- efficiency measured by MSE, *i.e.*, $\text{MSE}_\theta(\hat{\Theta}_\mu) = \text{Var}[\hat{\Theta}_\mu] = \frac{\sigma^2}{N}$. see example in Slide 41.
- from above $\text{Var}[\hat{\Theta}_\mu]$,it is a consistent estimator for μ see Slide 52.

Sample variance

Let X_1, X_2, ..., X_n be a random sample sharing the same distribution as X, but the parameters (*e.g.*, mean and variance) could be unknown.

Definition 13: Sample variance

Sample variance, the estimator for variance of X, is defined as

$$\hat{\Theta}_{\sigma^2} := S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{17}$$

Homework 2. we will check if the sample variance is unbiased for estimating the variance.

The unbiased sample variance will be used as a reference in analyzing the MLE variance estimator in Lecture III.

Method of moments

In short, the method of moments estimates for the unknown parameters by

- equating sample moments with theoretical moments;
- then, solving for the parameters involved with the theoretical moments.

Definition 14: Theoretical moments vs. Sample moments

- $\mathrm{E}[X^k]$ is the k^{th} theoretical moment.
- $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ is the k^{th} sample moment.

If there is only one unknown, using the first-order moment is enough: 1 equation for 1 unknown.

If there are $k > 1$ unknowns, we will need k^{th} order for the theoretical moment and sample moment to be used in forming k equations.

Example for Method of Moments

Example 7: Method of Moment for exponential distribution

Let $X_1, X_2, ..., X_n$ be the random sample from an exponential distribution with parameter λ, *i.e.*, $e^{rate=\lambda}$.

- Q1. What is the estimator for λ using method of moments, *i.e.*, λ_{MME}?
- Q2. Is the estimator biased or unbiased?
- Q3. What is the efficiency (MSE) of λ_{MME}?

- Q4. Is λ_{MME} consistent for λ ? $\Rightarrow$ **Homework 3.**

Q1. Finding the estimator for λ using method of moments?

Example 7 (Cont'): Q1. What is λ_{MME}?

Let $X_1, X_2, ..., X_n$ be the random sample from an exponential distribution with parameter λ, *i.e.*, $e^{rate=\lambda}$.

Q1. What is λ_{MME}?

- The 1^{st} theoretical moment of exponential distribution is $E[X] := \frac{1}{\lambda}$.
- The 1^{st} sample moment is $M_1 := \frac{1}{n}\sum_{i=1}^{n} X_i$.

We equate the 1^{st} theoretical moment and the sample moment. Then, solve for λ.

$$\frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Rightarrow \lambda = \lambda_{MME} = \frac{n}{\sum_{i=1}^{n} X_i} \tag{18}$$

Thus, the estimator for λ via method of moments is $\lambda_{MME} = \frac{n}{\sum_{i=1}^{n} X_i}$.

Q2. Is the estimator biased or unbiased?

Example 7 (Cont'): Q2. Check if $\mathrm{E}[\lambda_{MME}] = \lambda$?

Q2. Previously we have $\lambda_{MME} = \frac{n}{\sum_{i=1}^n X_i} \Rightarrow \mathrm{E}[\lambda_{MME}] = \mathrm{E}_X[\frac{n}{\sum_{i=1}^n X_i}]$.

Let define $Y := \sum_{i=1}^n X_i$. Recall that if $X_i \sim \mathrm{e}^{rate=\lambda}$, then $Y \sim \Gamma(n, \lambda)$.

$$\begin{aligned}
\mathrm{E}[\lambda_{MME}] &= \mathrm{E}_X[\frac{n}{\sum_{i=1}^n X_i}] = \mathrm{E}_Y[\frac{n}{Y}] = n\mathrm{E}_Y[\frac{1}{Y}]. \\
&= n\int_{-\infty}^{\infty} \frac{1}{y} f_Y(y)\,dy \ = n\int_{-\infty}^{\infty} \frac{1}{y} \cdot \frac{1}{\Gamma(n)} \lambda^n y^{n-1} \mathrm{e}^{-\lambda y}\,dy \\
&= n\int_{-\infty}^{\infty} \frac{1}{\Gamma(n)} \lambda^n y^{n-2} \mathrm{e}^{-\lambda y}\,dy \\
&= n\lambda \frac{\Gamma(n-1)}{\Gamma(n)} \int_{-\infty}^{\infty} \frac{1}{\Gamma(n-1)} \lambda^{n-1} y^{n-2} \mathrm{e}^{-\lambda y}\,dy \\
&= n\lambda \frac{\Gamma(n-1)}{\Gamma(n)}, \quad \because \int_{-\infty}^{\infty} \frac{1}{\Gamma(n-1)} \lambda^{n-1} y^{n-2} \mathrm{e}^{-\lambda y}\,dy \ = 1.
\end{aligned}$$

Example 7 (Cont'): Q2. Check if $\mathrm{E}[\lambda_{MME}] = \lambda$?

Q2. Continue from the previous slide...

$$\begin{aligned}\mathrm{E}[\lambda_{MME}] &= n\lambda\frac{\Gamma(n-1)}{\Gamma(n)} \\ &= n\lambda\frac{\Gamma(n-1)}{(n-1)\Gamma(n-1)} = \frac{n}{n-1}\lambda\end{aligned}$$

Because $\mathrm{E}[\lambda_{MME}] = \frac{n}{n-1}\lambda \neq \lambda$, the estimator is biased.

Q3. What is the efficiency (MSE) for λ_{MME}?

Example 7 (Cont'): Q3. Compute $\mathrm{MSE}_\lambda(\lambda_{MME})$...

Q3. Previously, $\lambda_{MME} \sim \Gamma(n, \lambda)$.

$$\begin{aligned}\mathrm{MSE}_\lambda(\lambda_{MME}) &= \mathrm{E}[(\lambda_{MME} - \lambda)^2] \\ &= \mathrm{Var}[\lambda_{MME}] + \mathrm{Bias}_\lambda(\lambda_{MME})^2 = XXX + (\frac{n\lambda}{n-1} - \lambda)^2 \\ &= XXX + (\frac{\lambda}{n-1})^2\end{aligned}$$

Need to derive $\mathrm{Var}[\lambda_{MME}]$ where $\mu_{\lambda_{MME}} = \frac{n\lambda}{n-1}$ & $\lambda_{MME} = \frac{n}{\sum_i X_i}$.
$\mathrm{Var}[\lambda_{MME}] = \mathrm{E}_Y[(\lambda_{MME} - \mu_{\lambda_{MME}})^2] = \mathrm{E}_Y[\lambda^2_{MME}] - \mu^2_{\lambda_{MME}} \ldots$
Next we have to derive $\mathrm{E}_Y[\lambda^2_{MME}]$

Example 7 (Cont'): Q3. Compute $\mathrm{MSE}_\lambda(\lambda_{MME})$...

Q3. Continue ...

$$
\begin{aligned}
\mathrm{E}[\lambda_{MME}^2] &= \mathrm{E}_X[\frac{n^2}{(\sum_i X_i)^2}] = \mathrm{E}_Y[\frac{n^2}{Y^2}] = n^2\mathrm{E}_Y[\frac{1}{Y^2}] = n^2\int_{-\infty}^{\infty}\frac{1}{y^2}f_Y(y)\,dy \\
&= n^2\int_{-\infty}^{\infty}\frac{1}{y^2}\cdot\frac{1}{\Gamma(n)}\lambda^n y^{n-1}\mathrm{e}^{-\lambda y}\,dy \\
&= n^2\lambda^2\frac{\Gamma(n-2)}{\Gamma(n)}\int_{-\infty}^{\infty}\frac{1}{\Gamma(n-2)}\lambda^{n-2}y^{n-3}\mathrm{e}^{-\lambda y}\,dy \\
&= n^2\lambda^2\frac{\Gamma(n-2)}{\Gamma(n)} = n^2\lambda^2\frac{\Gamma(n-2)}{(n-1)(n-2)\Gamma(n-2)} \\
&= n^2\lambda^2\frac{1}{(n-1)(n-2)} = \frac{n^2\lambda^2}{(n-1)(n-2)}
\end{aligned}
$$

Therefore, $MSE_\lambda(\lambda_{MME}) = \frac{n^2\lambda^2}{(n-1)(n-2)} + \frac{\lambda^2}{(n-1)^2}$.

Q4. Is the estimator λ_{MME} consistent?

Example 7 (Cont'): Q4. Is the estimator λ_{MME} consistent for λ?

Q4. $\Rightarrow$ As $n \to \infty$, check if

$$P(|\lambda_{MME} - \lambda| > \epsilon) \leq 0?$$

Homework 4. HINT! Use Chebyshev's inequality.

We can use Chebyshev's inequality, if we know the estimator's mean $\mathrm{E}[\lambda_{MME}]$ and variance $\mathrm{Var}[\lambda_{MME}]$.... What are they?

1. Table of contents

2. Motivation

3. Notation

4. General Properties of Estimators

5. Examples of estimators

6. Homework
   - Homework 1.1 (scores=7)
   - Homework 1.2 (scores=10)
   - Homework 1.3 (scores=3)

7. Index

Homework 1.1 (scores=7)

HW 1: Mean-square error (MSE) — Alternative calculation

Can you show that the mean squared error of an estimator $\hat{\Theta}$ can also be calculated as:

$$\mathrm{MSE}_{\theta}(\hat{\Theta}) = \mathrm{E}_{\hat{\Theta}}[(\hat{\Theta} - \theta)^2] = \mathrm{Var}_{\hat{\Theta}}[\hat{\Theta}] + \mathrm{Bias}_{\theta}(\hat{\Theta})^2. \tag{19}$$

Homework 1.2 (scores=10)

HW 2: Example for sample mean and variance

Let $X_1, X_2, \ldots$ be a random sample from a distribution with mean μ and variance σ^2. Note that the empirical variance is given by the following equation and is not the same as the sample variance.

$$S_\mu^2 = \frac{1}{n} \sum_i \left(X_i - \bar{X}\right)^2 \tag{20}$$

- Q1. Is the empirical variance, defined as Eq.(20), an unbiased estimator for variance σ^2? **[Check if $\mathbf{E}[S_\mu^2] - \sigma^2 = 0$?]**
- Q2. Is sample variance S^2 an unbiased estimator for variance σ^2?
- Q3. At $n = 10, 100, 1000, 10000$, and simulate the bias of both estimators, *i.e.*, $\mathrm{E}[S_\mu^2]$ and $\mathrm{E}[S^2]$.
  1. Write the description for each simulation at different n values.
  2. Provide **two sets of plots** for the estimators, *i.e.*, S^2 and S_μ^2, *i.e.*, the histogram of estimators, the mean of estimators, and the true variance.

Homework 1.3 (scores=3)

HW 3: Method of Moment for exponential distribution

Let $X_1, X_2, ..., X_n$ be the random sample from an exponential distribution with parameter λ, *i.e.*, $\mathrm{e}^{rate=\lambda}$.

- Q1. What is the estimator for λ using method of moments, *i.e.*, λ_{MME}?
- Q2. Is the estimator biased or unbiased?
- Q3. What is the efficiency (MSE) of λ_{MME}?

- Q4. Is λ_{MME} consistent for λ ? $\Rightarrow$ HOMEWORK

1. Table of contents
2. Motivation
3. Notation
   - Sampling from distribution
   - What is an estimator?
   - Is there a hat ?
4. General Properties of Estimators
   - Overall
   - Bias
   - Efficiency
   - Consistency
   - Markov inequality & Chebyshev's inequality
   - Summary: Properties of Estimators
   - Asymptotic Properties of Estimators
5. Examples of estimators
   - Sample mean
   - Sample variance
   - Method of moments
6. Homework
   - Homework 1.1 (scores=7)
   - Homework 1.2 (scores=10)
   - Homework 1.3 (scores=3)
7. Index

- H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*.
  Kappa Research, LLC, 2014.

- A. C. Robert Hogg, Joseph McKean, *Introduction to Mathematical Statistics*.
  Pearson; 7th edition, 2012.

- D. Romik, "Lecture Notes on Probability Theory, math 235a," Fall 2009.

- Z. Fan, "Lecture 3: Consistency and asymptotic normality of the MLE. stats 200," Autumn 2016.

- D. Panchenko, "Lecture 3: Properties of MLE: consistency, asymptotic normality, Fisher information. 18.650," Fall 2006.

- J. Songsiri, "Lecture Notes on Estimators, Statistics for Financial Engineering," 2022.

- J. Songsiri, "Lecture Notes on Statistical Estimation, system identification," 2022.

J. Songsiri, “Lecture Notes on Function of Random Variables,” 2023.