Announcements

- No in-person Lectures 11, 12 (scheduled 22 May and 24 May)
- There will be recordings for Lecture 11, 12
- possibly an in-person Example Class in the week 29 May–2 June
Announcements

- No in-person Lectures 11, 12 (scheduled 22 May and 24 May)
- There will be recordings for Lecture 11, 12
- possibly an in-person Example Class in the week 29 May–2 June

- IA Examination Briefing on Wednesday 24 May 12:00-13:00 by Prof Robert Watson, Lecture Theatre A, Arts School (this venue!)
- for exam questions in this course, calculators are not required
A Distribution whose Average does not converge (Lecture 9)

The Cauchy distribution has “too heavy” tails (no expectation), in particular the average does not converge.
Outline

Introduction

Defining and Analysing Estimators

More Examples
Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbb{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\iff$ in principle we can reconstruct the unknown distribution
Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $P[X \leq a]$ for any $a \in \mathbb{R}$
  $\Rightarrow$ in principle we can reconstruct the unknown distribution

¿¿¿

- How can we estimate the variance or other parameters?
  $\Rightarrow$ estimator
- How can we measure the accuracy of an estimator?
  $\Rightarrow$ bias (this lecture) and mean-squared error (next lecture)
Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $P[X \leq a]$ for any $a \in \mathbb{R}$
  $\Rightarrow$ in principle we can reconstruct the unknown distribution

- How can we estimate the variance or other parameters? $\Rightarrow$ estimator
- How can we measure the accuracy of an estimator? $\Rightarrow$ bias (this lecture) and mean-squared error (next lecture)

Physical Experiments:
Measurement = Quantity of Interest + Measurement Error
Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $P[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the unknown distribution

- How can we estimate the variance or other parameters? $\rightsquigarrow$ estimator
- How can we measure the accuracy of an estimator? $\rightsquigarrow$ bias (this lecture) and mean-squared error (next lecture)

Physical Experiments:

Measurement = Quantity of Interest + Measurement Error
Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $P[X \leq a]$ for any $a \in \mathbb{R}$
  $\implies$ in principle we can reconstruct the unknown distribution

How can we estimate the variance or other parameters?
$\implies$ estimator

How can we measure the accuracy of an estimator?
$\implies$ bias (this lecture) and mean-squared error (next lecture)

Physical Experiments:
Measurement = Quantity of Interest + Measurement Error
Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $P[X \leq a]$ for any $a \in \mathbb{R}$
  \(\rightsquigarrow\) in principle we can reconstruct the unknown distribution

- How can we estimate the variance or other parameters?
  \(\rightsquigarrow\) estimator
- How can we measure the accuracy of an estimator?
  \(\rightsquigarrow\) bias (this lecture) and mean-squared error (next lecture)

Physical Experiments:
Measurement = Quantity of Interest + Measurement Error
Empirical Distribution Function

Definition of Empirical Distribution Function (Empirical CDF)

Let \( X_1, X_2, \ldots, X_n \) being i.i.d. samples, and \( F \) be the corresponding distribution function. For any \( a \in \mathbb{R} \), define

\[
F_n(a) := \frac{\text{number of } X_i \in (-\infty, a]}{n}.
\]
Empirical Distribution Function

Definition of Empirical Distribution Function (Empirical CDF)

Let \( X_1, X_2, \ldots, X_n \) being i.i.d. samples, and \( F \) be the corresponding distribution function. For any \( a \in \mathbb{R} \), define

\[
F_n(a) := \frac{\text{number of } X_i \in (-\infty, a]}{n}.
\]

Remark

The Weak Law of Large Numbers implies that for every \( \epsilon > 0 \) and \( a \in \mathbb{R} \),

\[
\lim_{n \to \infty} P \left[ \left| F_n(a) - F(a) \right| > \epsilon \right] = 0.
\]
Empirical Distribution Function

Definition of Empirical Distribution Function (Empirical CDF)

Let $X_1, X_2, \ldots, X_n$ being i.i.d. samples, and $F$ be the corresponding distribution function. For any $a \in \mathbb{R}$, define

$$F_n(a) := \frac{\text{number of } X_i \in (-\infty, a]}{n}.$$ 

Remark

The Weak Law of Large Numbers implies that for every $\epsilon > 0$ and $a \in \mathbb{R}$,

$$\lim_{n \to \infty} P \left[ |F_n(a) - F(a)| > \epsilon \right] = 0.$$ 

Thus by taking enough samples, we can estimate the entire distribution (including its expectation and variance).
Empirical Distribution Functions (Example 1/2)

Example 1

Consider throwing an unbiased dice 8 times, and let the realisation be:

\[(x_1, x_2, \ldots, x_8) = (4, 1, 5, 3, 1, 6, 4, 1).\]

What is the Empirical Distribution Function \(F_8(a)\)?

\[
\begin{align*}
F_8(a) &= \frac{1}{8} & \text{if } a = 1 \\
&= \frac{2}{8} & \text{if } a = 2 \\
&= \frac{3}{8} & \text{if } a = 3 \\
&= \frac{4}{8} & \text{if } a = 4 \\
&= \frac{5}{8} & \text{if } a = 5 \\
&= \frac{6}{8} & \text{if } a = 6 \\
&= \frac{7}{8} & \text{if } a = 7 \\
&= 1 & \text{if } a \geq 8
\end{align*}
\]
Consider throwing an unbiased dice 8 times, and let the realisation be:

\[(x_1, x_2, \ldots, x_8) = (4, 1, 5, 3, 1, 6, 4, 1)\].

What is the Empirical Distribution Function \(F_8(a)\)?
Consider throwing an unbiased dice 8 times, and let the realisation be:

\((x_1, x_2, \ldots, x_8) = (4, 1, 5, 3, 1, 6, 4, 1)\).

What is the Empirical Distribution Function \(F_8(a)\)?

Answer:

\[
\begin{align*}
F_8(a) & : 0, 1/8, 1/6, 1/3, 2/6, 3/6, 4/6, 5/6, 7/8, 1 \\
F(a) & : 0, 1/8, 1/6, 1/3, 2/6, 3/6, 4/6, 5/6, 7/8, 1
\end{align*}
\]
Empirical Distribution Functions (Example 2/2)

Figure: Empirical Distribution Functions of samples from a Normal Distribution $\mathcal{N}(5, 4)$ ($n = 20$ left, $n = 200$ right)

Source: Modern Introduction to Statistics
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

An Example of an Estimation Problem

Consider the packages arriving at a network server.

- We might be interested in:

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
  1. number of packets that arrive within a “typical” minute

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
  1. number of packets that arrive within a “typical” minute

An Example of an Estimation Problem

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

Consider the packages arriving at a network server.
- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time \( \sim \) number of arrivals during one minute follows a Poisson distribution with unknown parameter \( \lambda \)

An Example of an Estimation Problem

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time \( \sim \) number of arrivals during one minute follows a Poisson distribution with unknown parameter \( \lambda \)

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time \( \sim \) number of arrivals during one minute follows a Poisson distribution with unknown parameter \( \lambda \)

\[ X \sim Poi(\lambda) \]

\[ P[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \]
An Example of an Estimation Problem

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time \( \rightsquigarrow \) number of arrivals during one minute follows a **Poisson distribution** with **unknown** parameter \( \lambda \)

\[
X \sim Poi(\lambda)
\]

\[
P[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}
\]
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive
- If arrivals occur at random time \( \sim \) number of arrivals during one minute follows a Poisson distribution with unknown parameter \( \lambda \)


\[
P[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}
\]
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time \( \sim \) number of arrivals during one minute follows a Poisson distribution with unknown parameter \( \lambda \)

Waiting Time (Lecture 5, Slide 22)

\[
\Delta \sim \text{Exp}(\lambda)
\]

\[
X \sim \text{Poi}(\lambda)
\]

\[
P[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}
\]

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time ⇝ number of arrivals during one minute follows a Poisson distribution with unknown parameter $\lambda$

Waiting Time (Lecture 5, Slide 22)

$$X \sim Poi(\lambda)$$

$$\Delta \sim Exp(\lambda)$$

$$P[X=k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$


Intro to Probability

8
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
  1. number of packets that arrive within a “typical” minute
  2. percentage of minutes during which no packets arrive
- If arrivals occur at random time $\sim$ number of arrivals during one minute follows a Poisson distribution with unknown parameter $\lambda$


Waiting Time (Lecture 5, Slide 22)

$\Delta \sim Exp(\lambda)$

$X \sim Poi(\lambda)$

$P[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$
A random variable

\[ T = h(X_1, X_2, \ldots, X_n), \]

depending only on the samples is called estimator.
Definition of Estimator

A random variable

\[ T = h(X_1, X_2, \ldots, X_n), \]

depending only on the samples is called estimator. An estimate is a value that only depends on the dataset \( x_1, x_2, \ldots, x_n \), i.e.,

\[ t = h(x_1, x_2, \ldots, x_n). \]
A random variable

\[ T = h(X_1, X_2, \ldots, X_n), \]

depending only on the samples is called estimator. An estimate is a value that only depends on the dataset \( x_1, x_2, \ldots, x_n \), i.e.,

\[ t = h(x_1, x_2, \ldots, x_n). \]

Questions:
Estimator

A random variable

\[ T = h(X_1, X_2, \ldots, X_n), \]

depending only on the samples is called estimator. An estimate is a value that only depends on the dataset \( x_1, x_2, \ldots, x_n \), i.e.,

\[ t = h(x_1, x_2, \ldots, x_n). \]

Questions:
- What makes an estimator suitable? \( \rightsquigarrow \) unbiased (later: MSE)
- Does an unbiased estimator always exist? How to compute it?
- If there are several unbiased estimators, which one to choose?
Outline

Introduction

Defining and Analysing Estimators

More Examples
Example: Arrival of Packets (1/3)

- **Samples:** Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim \text{Pois}(\lambda)$
- **Meaning:** $X_i$ is the number of packets arriving in minute $i$

Example 2

Suppose we wish to estimate $\lambda$ by using the sample mean $\bar{X}_n$.

Answer
Example: Arrival of Packets (1/3)

- **Samples:** Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim \text{Pois}(\lambda)$
- **Meaning:** $X_i$ is the number of packets arriving in minute $i$

---

**Example 2**

Suppose we wish to estimate $\lambda$ by using the sample mean $\overline{X}_n$.

We have

$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and $\mathbb{E} \left[ \overline{X}_n \right] = \mathbb{E} \left[ X_1 \right] = \lambda$. 
Example: Arrival of Packets (1/3)

- **Samples:** Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim \text{Pois}(\lambda)$
- **Meaning:** $X_i$ is the number of packets arriving in minute $i$

Example 2

Suppose we wish to estimate $\lambda$ by using the sample mean $\bar{X}_n$.

We have

$$\bar{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1] = \lambda$. This suggests the estimator:

$$h(X_1, X_2, \ldots, X_n) := \bar{X}_n.$$
EXAMPLE: ARRIVAL OF PACKETS (1/3)

- **Samples:** Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim \text{Pois}(\lambda)$
- **Meaning:** $X_i$ is the number of packets arriving in minute $i$

---

**Example 2**

Suppose we wish to estimate $\lambda$ by using the sample mean $\overline{X}_n$.

\[
\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n},
\]

and $E[\overline{X}_n] = E[X_1] = \lambda$. This suggests the estimator:

\[
h(X_1, X_2, \ldots, X_n) := \overline{X}_n.
\]

Applying the **Weak Law of Large Numbers**:

\[
\lim_{n \to \infty} P \left[ \left| \overline{X}_n - \lambda \right| > \epsilon \right] = 0 \quad \text{for any } \epsilon > 0.
\]
Now suppose we wish to instead estimate the probability of zero arrivals $e^{-\lambda}$ by the relative frequency of samples which are zero.

Answer
Now suppose we wish to instead estimate the probability of zero arrivals $e^{-\lambda}$ by the relative frequency of samples which are zero.

Let $X_1, X_2, \ldots, X_n$ be the $n$ samples. Let

$$Y_i := 1_{X_i=0}.$$
Example: Arrival of Packets (2/3)

Now suppose we wish to instead estimate the probability of zero arrivals $e^{-\lambda}$ by the relative frequency of samples which are zero.

Let $X_1, X_2, \ldots, X_n$ be the $n$ samples. Let

$$Y_i := 1_{X_i = 0}.$$ 

Then

$$E[Y_i] = P[X_i = 0] = e^{-\lambda},$$
Now suppose we wish to instead estimate the probability of zero arrivals $e^{-\lambda}$ by the relative frequency of samples which are zero.

Let $X_1, X_2, \ldots, X_n$ be the $n$ samples. Let

$$Y_i := 1_{X_i=0}.$$  

Then

$$E[Y_i] = P[X_i = 0] = e^{-\lambda},$$

and thus we can define an estimator by

$$h_1(X_1, X_2, \ldots, X_n) := \frac{Y_1 + Y_2 + \cdots + Y_n}{n}.$$
Example 3b

Suppose we wish to estimate the probability of zero arrivals $e^{-\lambda}$ by using the sample mean $\overline{X}_n$.

Answer

---

Example: Arrival of Packets (3/3)
Example: Arrival of Packets (3/3)

Example 3b

Suppose we wish to estimate the probability of zero arrivals $e^{-\lambda}$ by using the sample mean $\bar{X}_n$.

We saw that $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ satisfies $E[\bar{X}_n] = E[X_1] = \lambda$.

Recall by the **Weak Law of Large Numbers**: 

\[
\lim_{n \to \infty} P\left[\left|\bar{X}_n - \lambda\right| > \epsilon\right] = 0 \quad \text{for any} \quad \epsilon > 0.
\]
Suppose we wish to estimate the probability of zero arrivals $e^{-\lambda}$ by using the sample mean $\overline{X}_n$.

We saw that $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ satisfies $E\left[\overline{X}_n\right] = E\left[X_1\right] = \lambda$.

Recall by the **Weak Law of Large Numbers**:

$$\lim_{n \to \infty} P\left[\left|\overline{X}_n - \lambda\right| > \epsilon\right] = 0 \quad \text{for any } \epsilon > 0.$$
Example 3b

Suppose we wish to estimate the probability of zero arrivals \( e^{-\lambda} \) by using the sample mean \( \bar{X}_n \).

We saw that \( \bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \) satisfies \( \mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1] = \lambda \).

Recall by the **Weak Law of Large Numbers**:

\[
\lim_{n \to \infty} \mathbb{P}\left( \left| \bar{X}_n - \lambda \right| > \epsilon \right) = 0 \quad \text{for any } \epsilon > 0.
\]

Then we estimate \( e^{-\lambda} \) by \( e^{-\bar{X}_n} \). Hence our estimator is

\[
h_2(X_1, X_2, \ldots, X_n) := e^{-\bar{X}_n}.
\]
Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$.

Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$. How good are these two estimators?

The first estimator can only attain values 0, 1/30, 2/30, ..., 1.

The second estimator can only attain values $1/30 e^{-1}$, $2/30 e^{-2}$, ...

For most values of $\lambda$, both estimators will never return the exact value of $e^{-\lambda}$ on the basis of 30 observations.
Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$.

Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

How good are these two estimators?
Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$

Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

How **good** are these two estimators?

⇒ The first estimator can only attain values $0, \frac{1}{30}, \frac{2}{30}, \ldots, 1$

⇒ The second estimator can only attain values $1, e^{-1/30}, e^{-2/30}, \ldots$
Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$.

Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

How good are these two estimators?

- The first estimator can only attain values $0, \frac{1}{30}, \frac{2}{30}, \ldots, 1$
- The second estimator can only attain values $1, e^{-1/30}, e^{-2/30}, \ldots$

For most values of $\lambda$, both estimators will never return the exact value of $e^{-\lambda}$ on the basis of 30 observations.
Simulation of the two Estimators

- The unknown parameter is \( p = e^{-\lambda} = 0.1 \) (i.e., \( \lambda = \ln 10 \approx 2.30 \ldots \)
Simulation of the two Estimators

- The unknown parameter is \( p = e^{-\lambda} = 0.1 \) (i.e., \( \lambda = \ln 10 \approx 2.30 \ldots \))
- We consider \( n = 30 \) minutes and compute \( h_1 \) and \( h_2 \)
- We repeat this 500 times and draw a frequency histogram \( (h_1 = \overline{Y}_n \text{ left}, \ h_2 = e^{-\overline{X}_n} \text{ right}) \)
Simulation of the two Estimators

- The unknown parameter is \( p = e^{-\lambda} = 0.1 \) (i.e., \( \lambda = \ln 10 \approx 2.30 \ldots \))
- We consider \( n = 30 \) minutes and compute \( h_1 \) and \( h_2 \)
- We repeat this 500 times and draw a frequency histogram (\( h_1 = \bar{Y}_n \) left, \( h_2 = e^{-\bar{X}_n} \) right)

Source: Modern Introduction to Statistics
Simulation of the two Estimators

- The unknown parameter is $p = e^{-\lambda} = 0.1$ (i.e., $\lambda = \ln 10 \approx 2.30 \ldots$)
- We consider $n = 30$ minutes and compute $h_1$ and $h_2$
- We repeat this 500 times and draw a frequency histogram ($h_1 = \bar{Y}_n$ left, $h_2 = e^{-X_n}$ right)

Both estimators concentrate around the true value 0.1, but the second estimator appears to be more concentrated.
An estimator $T$ is called an unbiased estimator for the parameter $\theta$ if

\[ \mathbb{E}[T] = \theta, \]

irrespective of the value $\theta$. 

Source: Edwin Leuven (Point Estimation)
An estimator $T$ is called an unbiased estimator for the parameter $\theta$ if

$$E[T] = \theta,$$

irrespective of the value $\theta$. The bias is defined as

$$E[T] - \theta = E[T - \theta].$$
An estimator $T$ is called an unbiased estimator for the parameter $\theta$ if

$$E[T] = \theta,$$

irrespective of the value $\theta$. The bias is defined as

$$E[T] - \theta = E[T - \theta].$$
Unbiased Estimators and Bias

A **definition**

An estimator $T$ is called an **unbiased estimator** for the parameter $\theta$ if

$$\mathbb{E}[T] = \theta,$$

irrespective of the value $\theta$. The **bias** is defined as

$$\mathbb{E}[T] - \theta = \mathbb{E}[T - \theta].$$

Which of the two estimators $h_1, h_2$ are unbiased?

Source: Edwin Leuven (Point Estimation)
Analysis of the Bias of the First Estimator

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

Recall we defined $Y_i := \mathbb{1}_{X_i = 0}$. Yes, because:

$\mathbb{E}[h_1(X_1, X_2, \ldots, X_n)] = n \cdot \mathbb{E}[Y_1] = P[X_1 = 0] = e^{-\lambda}$. 
Example 4a

Is \( h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n} \) an unbiased estimator for \( e^{-\lambda} \)?

Recall we defined \( Y_i := \mathbf{1}_{X_i=0} \).
Analysis of the Bias of the First Estimator

Example 4a

Is \( h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n} \) an unbiased estimator for \( e^{-\lambda} \)?

Recall we defined \( Y_i := 1_{X_i=0} \). Yes, because:

\[
E \left[ h_1(X_1, X_2, \ldots, X_n) \right]
\]
Is \( h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n} \) an unbiased estimator for \( e^{-\lambda} \)?

Recall we defined \( Y_i := 1_{X_i=0} \). Yes, because:

\[
\mathbb{E} \left[ h_1(X_1, X_2, \ldots, X_n) \right] = \frac{n \cdot \mathbb{E}[Y_1]}{n}
\]
Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

Recall we defined $Y_i := 1_{X_i=0}$. Yes, because:

\[
E[h_1(X_1, X_2, \ldots, X_n)] = \frac{n \cdot E[Y_1]}{n} = E[Y_1] = P[X_1 = 0]
\]
Is \( h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n} \) an unbiased estimator for \( e^{-\lambda} \)?

Recall we defined \( Y_i := 1_{X_i=0} \). Yes, because:

\[
E \left[ h_1(X_1, X_2, \ldots, X_n) \right] = \frac{n \cdot E \left[ Y_1 \right]}{n} = \frac{1}{n} \sum_{i=1}^{n} E \left[ Y_i \right] = \frac{1}{n} \sum_{i=1}^{n} P(X_i = 0) = e^{-\lambda}.
\]
Is $h_2(X_1, X_2, \ldots, X_n) = e^{-X_n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

No! (recall: $E[X^2] \geq E[X]^2$)

We have

$$E[e^{-X_n}] > e^{-E[X_n]} = e^{-\lambda}$$

This follows by Jensen’s inequality, and the inequality is strict since $z \mapsto e^{-z}$ is strictly convex.

Thus $h_2(X_1, X_2, \ldots, X_n)$ is not unbiased – it has positive bias.

Example 4b

For any random variable $X$, and any convex function $g : \mathbb{R} \rightarrow \mathbb{R}$, we have

$$E[g(X)] \geq g(E[X]).$$

If $g$ is strictly convex and $X$ is not constant, then the inequality is strict.
Bias of the Second Estimator (and Jensen’s Inequality)

Example 4b

Is \( h_2(X_1, X_2, \ldots, X_n) = e^{-X_n} \) an unbiased estimator for \( e^{-\lambda} \)?

Answer

No! (recall: \( E[X^2] \geq E[X]^2 \))
Bias of the Second Estimator (and Jensen’s Inequality)

Example 4b

Is \( h_2(X_1, X_2, \ldots, X_n) = e^{-X_n} \) an unbiased estimator for \( e^{-\lambda} \)?

No! (recall: \( E[X^2] \geq E[X]^2 \))

Jensen’s Inequality

For any random variable \( X \), and any convex function \( g : \mathbb{R} \to \mathbb{R} \), we have

\[
E[g(X)] \geq g(E[X]).
\]

If \( g \) is strictly convex and \( X \) is not constant, then the inequality is strict.
Bias of the Second Estimator (and Jensen’s Inequality)

Is \( h_2(X_1, X_2, \ldots, X_n) = e^{-X_n} \) an unbiased estimator for \( e^{-\lambda} \)?

**Answer**

No! (recall: \( E[X^2] \geq E[X]^2 \))

Jensen’s Inequality

For any random variable \( X \), and any convex function \( g : \mathbb{R} \to \mathbb{R} \), we have

\[
E[g(X)] \geq g(E[X]).
\]

If \( g \) is strictly convex and \( X \) is not constant, then the inequality is strict.
Bias of the Second Estimator (and Jensen’s Inequality)

Example 4b

Is \( h_2(X_1, X_2, \ldots, X_n) = e^{-\bar{X}_n} \) an unbiased estimator for \( e^{-\lambda} \)?

**Answer**

No! (recall: \( E[X^2] \geq E[X]^2 \))

- We have

\[
E\left[ e^{-\bar{X}_n} \right] > e^{-E[\bar{X}_n]} = e^{-\lambda}
\]

Jensen’s Inequality

For any random variable \( X \), and any convex function \( g : \mathbb{R} \to \mathbb{R} \), we have

\[
E\left[ g(X) \right] \geq g(E\left[ X \right]).
\]

If \( g \) is strictly convex and \( X \) is not constant, then the inequality is strict.
Bias of the Second Estimator (and Jensen’s Inequality)

Example 4b

Is \( h_2(X_1, X_2, \ldots, X_n) = e^{-\bar{X}_n} \) an unbiased estimator for \( e^{-\lambda} \)?

Answer

No! (recall: \( E[ X^2] \geq E[ X]^2 \))

- We have

\[
E\left[ e^{-\bar{X}_n} \right] > e^{-E[ \bar{X}_n ]} = e^{-\lambda}
\]

- This follows by Jensen’s inequality, and the inequality is strict since \( z \mapsto e^{-z} \) is strictly convex.

Jensen’s Inequality

For any random variable \( X \), and any convex function \( g : \mathbb{R} \to \mathbb{R} \), we have

\[
E[ g(X) ] \geq g(E[ X ]).
\]

If \( g \) is strictly convex and \( X \) is not constant, then the inequality is strict.
Bias of the Second Estimator (and Jensen’s Inequality)

Example 4b

Is \( h_2(X_1, X_2, \ldots, X_n) = e^{-\bar{X}_n} \) an unbiased estimator for \( e^{-\lambda} \)?

\[ \text{Answer} \]

No! (recall: \( E[X^2] \geq E[X]^2 \))

- We have

\[
E \left[ e^{-\bar{X}_n} \right] > e^{-E[\bar{X}_n]} = e^{-\lambda}
\]

- This follows by Jensen’s inequality, and the inequality is strict since \( z \mapsto e^{-z} \) is strictly convex.
- Thus \( h_2(X_1, X_2, \ldots, X_n) \) is not unbiased – it has positive bias.

Jensen’s Inequality

\[
\lambda g(a) + (1 - \lambda)g(b) \geq g(\lambda a + (1 - \lambda)b)
\]

For any random variable \( X \), and any convex function \( g : \mathbb{R} \rightarrow \mathbb{R} \), we have

\[
E \left[ g(X) \right] \geq g(E[X]).
\]

If \( g \) is strictly convex and \( X \) is not constant, then the inequality is strict.
Asymptotic Bias of the Second Estimator (non-examinable)

$E \left[ h_2(X_1, \ldots, X_n) \right] \xrightarrow{n \to \infty} e^{-\lambda}$ (hence it is asymptotically unbiased).

- Recall $h_2(X_1, \ldots, X_n) = e^{-\bar{X}_n}$. For any $0 \leq k \leq n$,

$$P \left[ h_2(X_1, \ldots, X_n) = e^{-k/n} \right] = P \left[ \sum_{i=1}^{n} X_i = k \right] = P \left[ Z = k \right],$$

where $Z \sim \text{Pois}(n \cdot \lambda)$ (since $\text{Pois}(\lambda_1) + \text{Pois}(\lambda_2) = \text{Pois}(\lambda_1 + \lambda_2)$)

$$\Rightarrow P \left[ h_2(X_1, \ldots, X_n) = e^{-k/n} \right] = \frac{e^{-n\lambda} \cdot (n\lambda)^k}{k!}$$

$$\Rightarrow E \left[ h_2(X_1, \ldots, X_n) \right] = \sum_{k=0}^{\infty} e^{-n\lambda} \cdot \frac{(n\lambda)^k}{k!} \cdot e^{-k/n}$$

By LOTUS

$$= e^{-n\lambda} \cdot e^{n\lambda e^{-1/n}} \sum_{k=0}^{\infty} e^{-n\lambda e^{-1/n}} \cdot \frac{(n\lambda e^{-1/n})^k}{k!}$$

$$= e^{-n\lambda} \cdot (1 - e^{-1/n}) \cdot 1$$

since $e^x = 1 + x + O(x^2)$ for small $x$

$$\approx n \to \infty e^{-n\lambda} \cdot (1 - 1 + 1/n + O(1/n^2)) = e^{-\lambda + O(\lambda/n)}.$$

Hence in the limit, the positive bias of $h_2$ diminishes.
Outline

Introduction

Defining and Analysing Estimators

More Examples
Unbiased Estimator for Expectation and Variance

Let $X_1, X_2, \ldots, X_n$ be identically distributed samples from a distribution with finite expectation $\mu$ and finite variance $\sigma^2$. Then

$$\bar{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is an unbiased estimator for $\mu$.

Furthermore,

$$S_n = S_n(X_1, \ldots, X_n) := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

is an unbiased estimator for $\sigma^2$. 


We need to prove: \( \mathbb{E} [ S_n ] = \sigma^2. \)

Multiplying by \( n - 1 \) yields:
\[
(n - 1) \cdot S_n = \sum_{i=1}^{n} \left( X_i - \overline{X}_n \right)^2
\]
\[
= \sum_{i=1}^{n} \left( X_i - \mu + \mu - \overline{X}_n \right)^2
\]
\[
= \sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} (\overline{X}_n - \mu)^2 - 2 \sum_{i=1}^{n} (X_i - \mu) (\overline{X}_n - \mu)
\]
\[
= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\overline{X}_n - \mu)^2 - 2 (\overline{X}_n - \mu) \cdot n \cdot (\overline{X}_n - \mu)
\]
\[
= \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X}_n - \mu)^2.
\]

Let us now take expectations:
\[
(n - 1) \cdot \mathbb{E} [ S_n ] = \sum_{i=1}^{n} \mathbb{E} \left[ (X_i - \mu)^2 \right] - n \cdot \mathbb{E} \left[ (\overline{X}_n - \mu)^2 \right]
\]
\[
= n \cdot \sigma^2 - n \cdot \sigma^2 / n
\]
\[
= (n - 1) \cdot \sigma^2.
\]

Recall: \( \mathbb{E} \left[ (\overline{X}_n - \mu)^2 \right] = \textup{Var} \left[ \overline{X}_n \right] = \sigma^2 / n \)
\[ E[S_n] = E \left[ \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right] = \sigma^2. \]

Why is it \( \frac{1}{n-1} \) and not \( \frac{1}{n} \)?

**Answer**

- **First Explanation.** Consider \( n = 1 \). Having just one estimate should not tell us anything about the variance (it could be infinite!).

- **Second Explanation.** Assume \( \mu \) is known, but \( \sigma^2 \) unknown. Define

\[ \sum_{i=1}^{n} (X_i - \mu)^2 =: A. \]

Additionally, define

\[ \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 =: B. \]

- \( B \leq A \), as \( \bar{X}_n \) solves a quadratic minimisation problem.
- It is easy to verify that \( \frac{1}{n} \cdot A \) is an unbiased estimator for \( \sigma^2 \).
- The factor \( \frac{1}{n-1} \) (instead of \( \frac{1}{n} \)) corrects the fact that \( \bar{X}_n \) is a more “favourable” average than the true mean \( \lambda \).
Example 6

Suppose that we have one sample $X \sim Bin(n, p)$, where $0 < p < 1$ is unknown but $n$ is known. Prove there is no unbiased estimator for $1/p$.

Answer

- First a simpler proof which exploits that $p$ might be arbitrarily small
- **Intuition:** For very small $p$, one $T(k)$, $k \in \{0, 1, \ldots, n\}$ must be very large, but then $E[T(X)]$ is too large for, e.g., $p = 1/2$
- **Formal Argument:**
  - Assume $T(X)$ is an unbiased estimator for $1/p$ for all $0 < p < 1$
  - Define $M := \max_{0 \leq k \leq n} T(k)$. Then,

$$
E[T(X)] = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \cdot T(k)
\leq M \cdot \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = M.
$$

- Hence this estimator does not work for $p < \frac{1}{M}$, since then $E[T(X)] \leq M < \frac{1}{p}$ (negative bias!)
- The next proof will work even if $p \in [a, b]$ for $0 < a < b \leq 1$.  

Intro to Probability More Examples 24
Warning: An Unbiased Estimator may not always exist (cntd.)

Example 6 (cntd.)

Suppose that we have one sample $X \sim Bin(n, p)$, where $0 < p < 1$ is unknown but $n$ is known. Prove there is no unbiased estimator for $1/p$.

- Suppose there exists an unbiased estimator with $E[T(X)] = 1/p$.
- Then

$$1 = p \cdot E[T(X)]$$

$$= p \cdot \sum_{k=0}^{n} P[X = k] \cdot T(k)$$

$$= p \cdot \sum_{k=0}^{n} \binom{n}{k} p^k \cdot (1 - p)^{n-k} \cdot T(k)$$

- Last term is a polynomial of degree $n + 1$ with constant term zero
  $\Rightarrow p \cdot E[T(X)] - 1$ is a (non-zero) polynomial of degree $\leq n + 1$
  $\Rightarrow$ this polynomial has at most $n + 1$ roots
  $\Rightarrow E[T(X)]$ can be equal to $1/p$ for at most $n + 1$ values of $p$, and thus cannot be an unbiased.