

Introduction to Probability

Lecture 10: Estimators (Part I)

Mateja Jamnik, [Thomas Sauerwald](#)

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk

Easter 2023



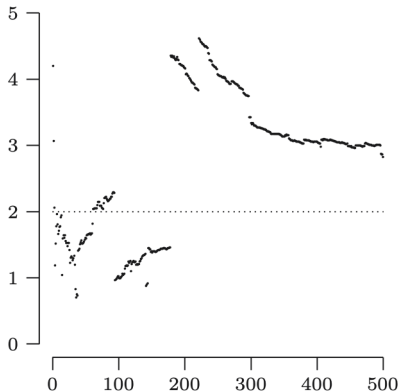
UNIVERSITY OF
CAMBRIDGE

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)



Cau(2, 1) distribution, Source: Modern Introduction to Statistics

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, \dots, X_n from an **unknown distribution**.

- Taking enough samples allows us to estimate the **mean** (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ 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:

$$\text{Measurement} = \text{Quantity of Interest} + \text{Measurement Error}$$

Empirical Distribution Function

Definition of Empirical Distribution Function (Empirical CDF)

Let X_1, X_2, \dots, 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 $a \in \mathbb{R}$,

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

Thus by taking enough samples, we can estimate the entire distribution (including its expectation and variance).

Empirical Distribution Functions (Example 1/2)

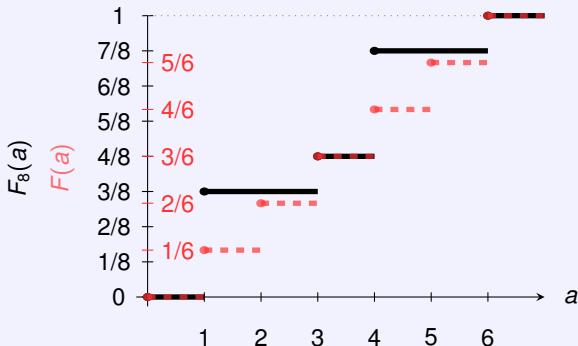
Example 1

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

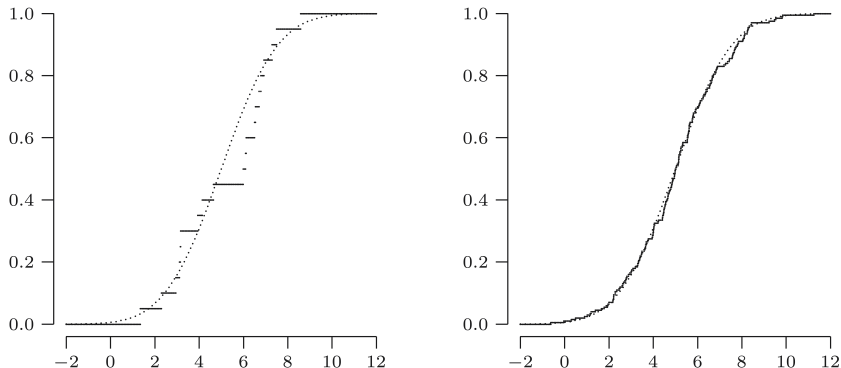
$$(x_1, x_2, \dots, x_8) = (4, 1, 5, 3, 1, 6, 4, 1).$$

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

Answer



Empirical Distribution Functions (Example 2/2)



Source: Modern Introduction to Statistics

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

An Example of an Estimation Problem

Scenario

Consider the **packages arriving at a network server**.

- We might be interested in:

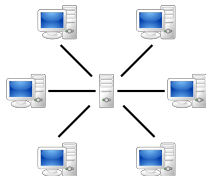
1. number of packets that arrive within a “typical” minute
2. percentage of minutes during which no packets arrive

Estimator for λ

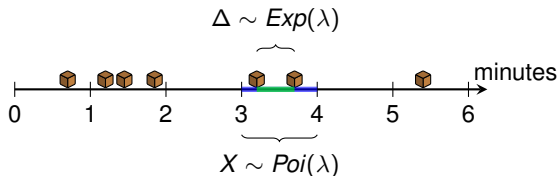
Estimator for $e^{-\lambda}$

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

Waiting Time (Lecture 5, Slide 22)



Source: Wikipedia



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

Definition of Estimator

A random variable

$$T = h(X_1, X_2, \dots, X_n),$$

depending only on the samples is called **estimator**.

An **estimate** is a value that only depends on the dataset x_1, x_2, \dots, x_n , i.e.,

$$t = h(x_1, x_2, \dots, x_n).$$

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, \dots, 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 λ by using the sample mean \bar{X}_n .

Answer

We have

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

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

$$h(X_1, X_2, \dots, X_n) := \bar{X}_n.$$

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

$$\lim_{n \rightarrow \infty} \mathbf{P} \left[\left| \bar{X}_n - \lambda \right| > \epsilon \right] = 0 \quad \text{for any } \epsilon > 0.$$

Example: Arrival of Packets (2/3)

Example 3a

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

Answer

Let X_1, X_2, \dots, X_n be the n samples. Let

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

Then

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

and thus we can define an estimator by

$$h_1(X_1, X_2, \dots, X_n) := \frac{Y_1 + Y_2 + \dots + Y_n}{n}.$$

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 .

Answer

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

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

$$\lim_{n \rightarrow \infty} \mathbf{P} \left[\left| \bar{X}_n - \lambda \right| > \epsilon \right] = 0 \quad \text{for any } \epsilon > 0.$$

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

$$h_2(X_1, X_2, \dots, X_n) := e^{-\bar{X}_n}.$$

Behaviour of the Estimators

- Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$
- Consider the **two estimators** $h_1(X_1, \dots, X_n)$ and $h_2(X_1, \dots, X_n)$.

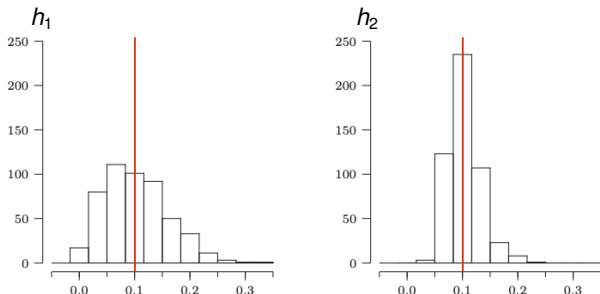
How **good** are these two estimators?

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

For most values of λ , 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., $\lambda = \ln 10 \approx 2.30 \dots$)
- 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

Both estimators concentrate around the true value 0.1, but the second estimator appears to be more concentrated.

Unbiased Estimators and Bias

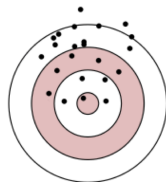
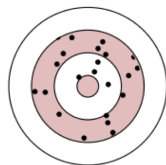
Definition

An **estimator** T is called an **unbiased estimator** for the parameter θ if

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

irrespective of the value θ . The **bias** is defined as

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



Source: Edwin Leuven (Point Estimation)

Which of the two estimators h_1, h_2 are unbiased?



Example 4a

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

Answer

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

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

Answer

$$\lambda g(a) + (1 - \lambda)g(b) \geq g(\lambda a + (1 - \lambda)b)$$

Jensen's Inequality

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

$$\mathbf{E}[g(X)] \geq g(\mathbf{E}[X]).$$

If g is **strictly convex** and X is not constant, then the inequality is strict.

Asymptotic Bias of the Second Estimator (non-examinable)

Example 4c

$\mathbf{E}[h_2(X_1, \dots, X_n)] \xrightarrow{n \rightarrow \infty} e^{-\lambda}$ (hence it is **asymptotically unbiased**).

Answer

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

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

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

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

$$\begin{aligned} \Rightarrow \mathbf{E}[h_2(X_1, \dots, X_n)] &= \sum_{k=0}^{\infty} e^{-n\lambda} \cdot \frac{(n\lambda)^k}{k!} \cdot e^{-k/n} \\ &\stackrel{\text{By LOTUS}}{=} e^{-n\lambda} \cdot e^{n\lambda e^{-1/n}} \sum_{k=0}^{\infty} e^{-n\lambda e^{-1/n}} \cdot \frac{(n\lambda e^{-1/n})^k}{k!} \\ &= e^{-n\lambda \cdot (1 - e^{-1/n})} \cdot 1 \end{aligned}$$

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

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

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

Outline

Introduction

Defining and Analysing Estimators

More Examples

Unbiased Estimators for Expectation and Variance

Let X_1, X_2, \dots, X_n be **identically distributed** samples from a distribution with finite expectation μ and finite variance σ^2 . Then

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

is an **unbiased** estimator for μ .

Furthermore,

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

is an **unbiased** estimator for σ^2 .

Example 5

We need to prove: $\mathbf{E}[S_n] = \sigma^2$.

Answer

Example 5 (cntd.)

$$\mathbf{E}[S_n] = \mathbf{E}\left[\frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X}_n)^2\right] = \sigma^2. \text{ Why is it } \frac{1}{n-1} \text{ and not } \frac{1}{n}?$$

Answer

Warning: An Unbiased Estimator may not always exist

Example 6

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

Answer

Warning: An Unbiased Estimator may not always exist (cntd.)

Example 6 (cntd.)

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

Answer

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

$$\begin{aligned} 1 &= p \cdot \mathbf{E}[T(X)] \\ &= p \cdot \sum_{k=0}^n \mathbf{P}[X = k] \cdot T(k) \\ &= p \cdot \sum_{k=0}^n \binom{n}{k} p^k \cdot (1-p)^{n-k} \cdot T(k) \end{aligned}$$

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