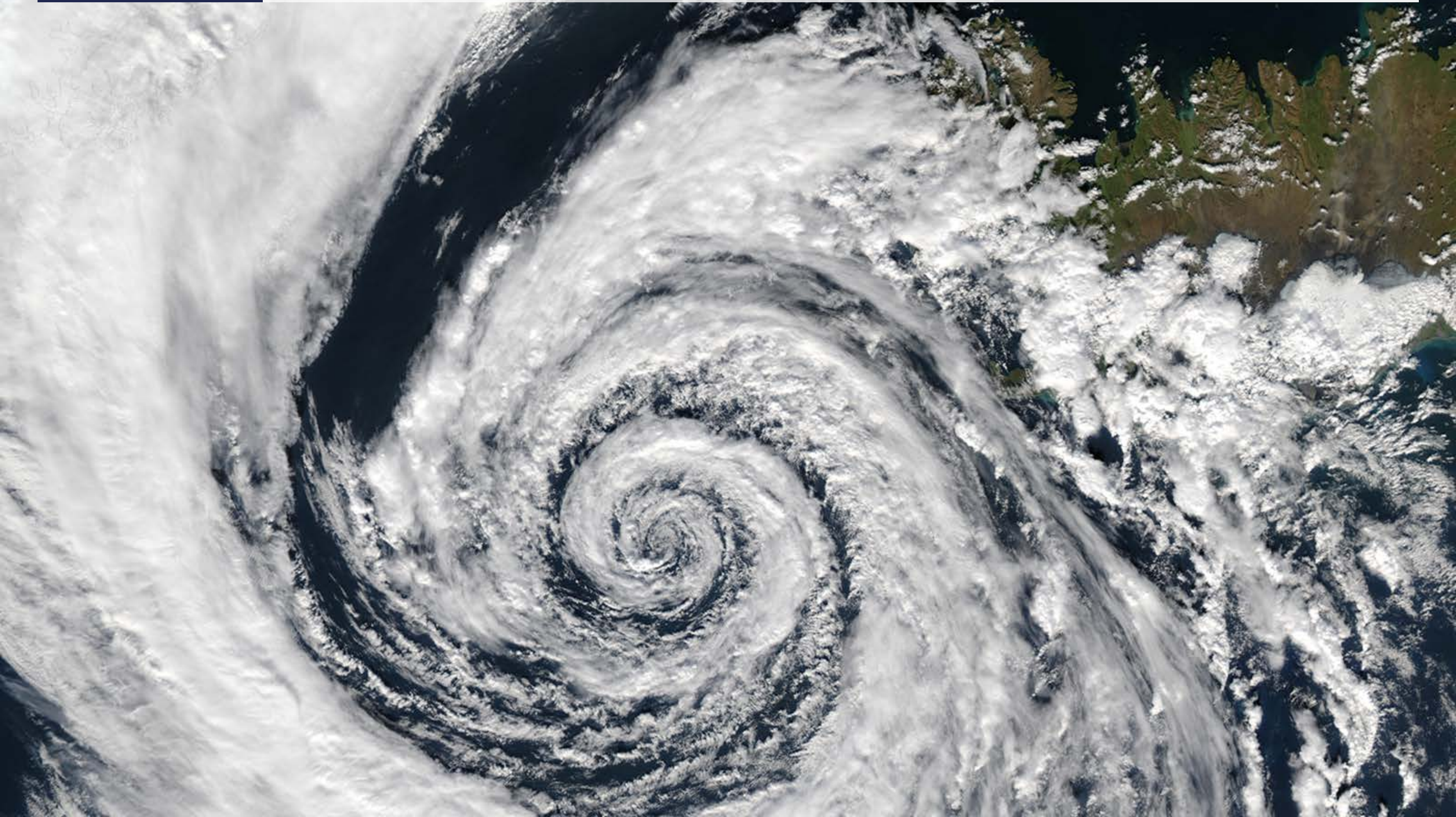
Statistics

What is a good estimator?

Name teacher



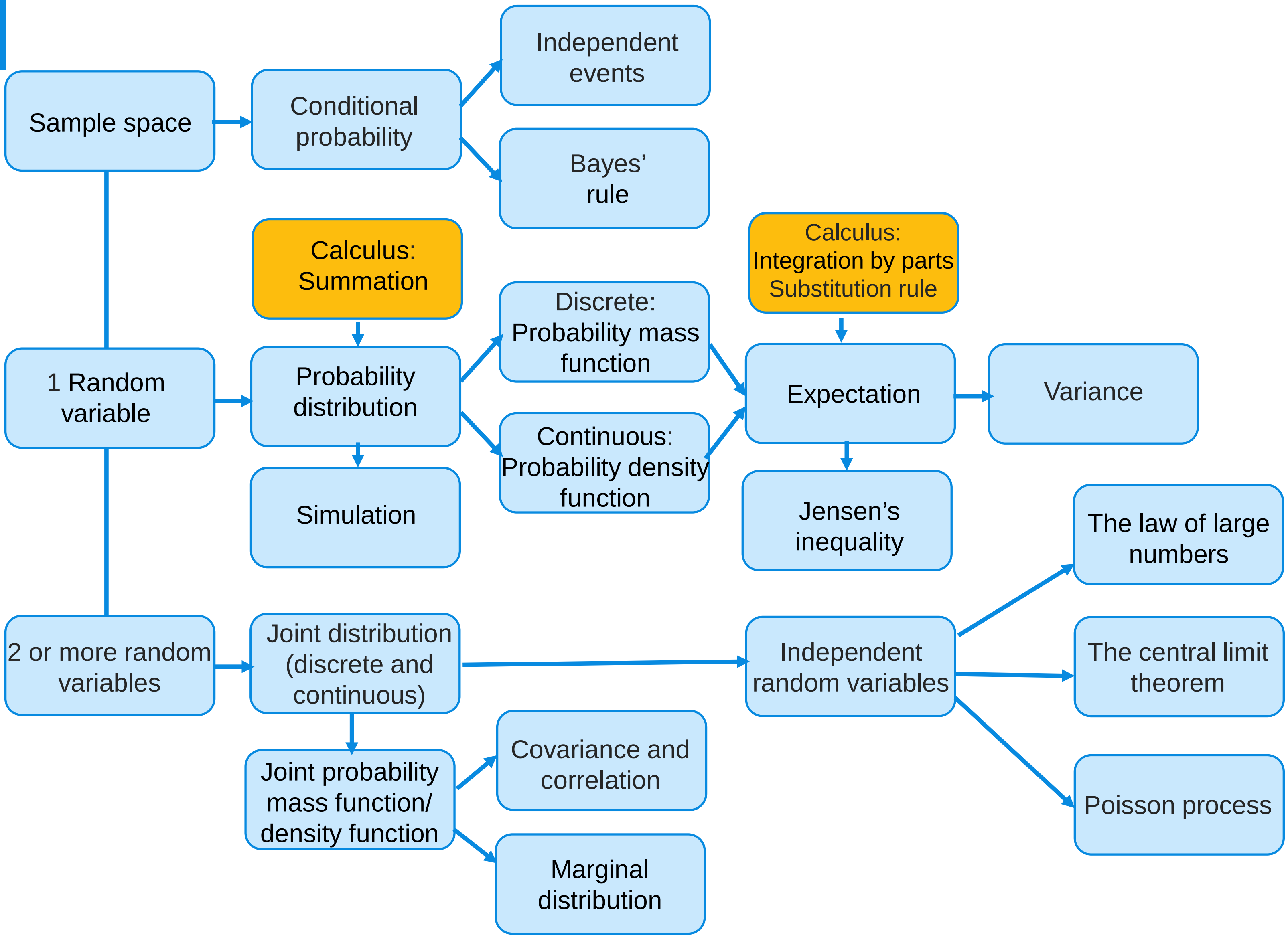
Learning Objective

After this class you can:

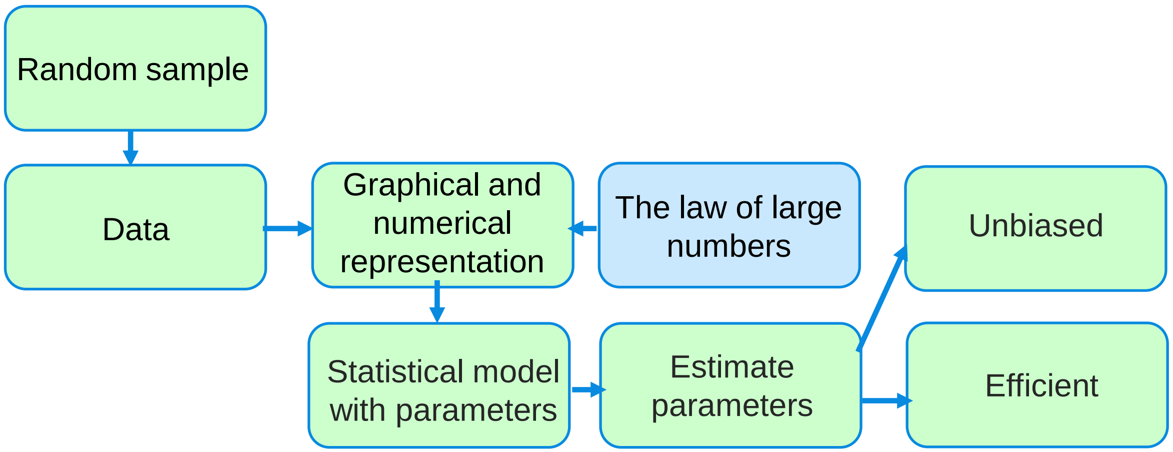
- Create estimators for parameters of a distribution, including the mean and the variance.
- Compare the quality of estimators using the concepts of unbiasedness, efficiency, and mean squared error.

Book: Chapters 19 and 20

Probability



Statistics



Random sample
Data
Graphical and numerical representation
The law of large numbers
Statistical model with parameters
Estimate parameters
Unbiased
Efficient

Before this class (week 3.6 lesson 2):

Watch prelecture *'Unbiased estimators'*

Book: Section 19.1, 19.2, 19.3

Programme

- Test statistics
- Tail probabilities
- Significance level
- Practice
- Relation to confidence intervals

You draw three balls from a bowl with balls numbered 1 through *N*. Every number has equal probability $p = 1/N$. What is the expectation of the average?

A) $\frac{1}{2}N$

B) $\frac{1}{2}(N+1)$

C) N

D) $\frac{3}{2}(N+1)$

You draw three balls from a bowl with balls numbered 1 through *N*. Every number has equal probability $p = 1/N$. What is an unbiased estimator for *N*?

A) $2\bar{X}_3$

B) $2\bar{X}_3 - 1$

C) $2\bar{X}_3 + 1$

D) $3\bar{X}_3$

Sampling Distribution

Definition:
Let $T = h(X_1, X_2, \ldots, X_n)$ be an estimator based on a random sample $X_1, X_2, \ldots, X_n$. The probability distribution of T is called the <u>sampling distribution</u> of T.

Exercises

Suppose $X_i \sim U(-\theta, \theta)$ are independently uniformly distributed according to some unknown parameter θ.

a. Show that $T = \frac{3}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)$ is an unbiased estimator for θ^2.

b. Is $\sqrt{T}$ an unbiased estimator for θ? If not, does it have positive or negative bias?

Suppose you have a coin which has unknown probability *p* of turning up heads. You record how long it takes before you get the first heads. Thus we have independent random variables $X_i \sim Geo(p)$. Note that the expectation of $\bar{X}_n$ equals $1/p$.

a. Show that $T = 1/\bar{X}_n$ is a biased estimator for *p* and determine the sign of the bias.

b. Show that $S = \frac{\text{number of } X_i \leq 3}{n}$ is an unbiased estimator of the probability of getting a heads within 3 throws.

Unbiased estimators for mean and variance

Theorem:

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with finite expectation μ and variance σ^2.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

are unbiased estimators of μ and σ^2.

Exercise

Suppose that the random variables X_i have identical expectation μ, though not necessarily identical distribution.

a. Is $S = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$ an unbiased estimator for μ?

b. Under what conditions on the a_i is

$$T = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

an unbiased estimator for μ?

Efficiency and mean squared error

Definition:
Let T_1 and T_2 be two *unbiased* estimators for the same parameter. If

$$\mathrm{Var}(T_2) < \mathrm{Var}(T_1)$$

we say that T_2 is more efficient than T_1.

Definition:
Let T be an estimators for the parameter θ. The mean squared error of T equals

$$\mathrm{MSE}(T) = \mathrm{E}[(T-\theta)^2].$$

Number of German tanks in World War 2

The serial number X_i of the *i*'th captured tank is modeled as a random drawing from $\{1, 2, \ldots, N\}$ with uniform distribution.

Two unbiased estimators for *N* are

$$T_1 = 2\bar{X}_n - 1 \qquad\qquad T_2 = \tfrac{n+1}{n} \max X_i - 1$$

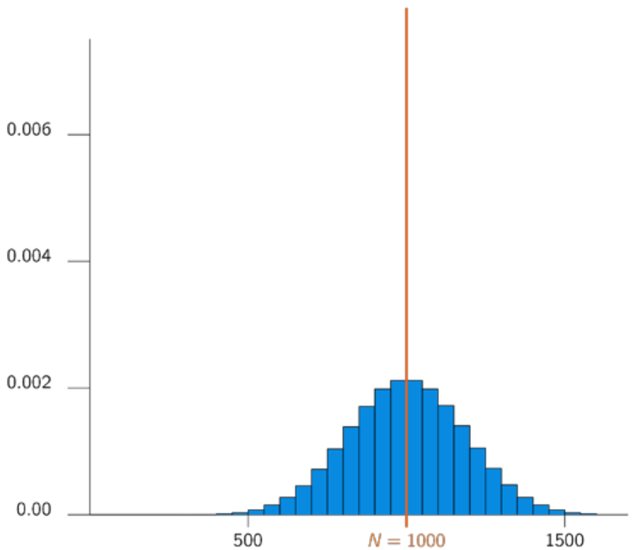
Number of German tanks in World War 2

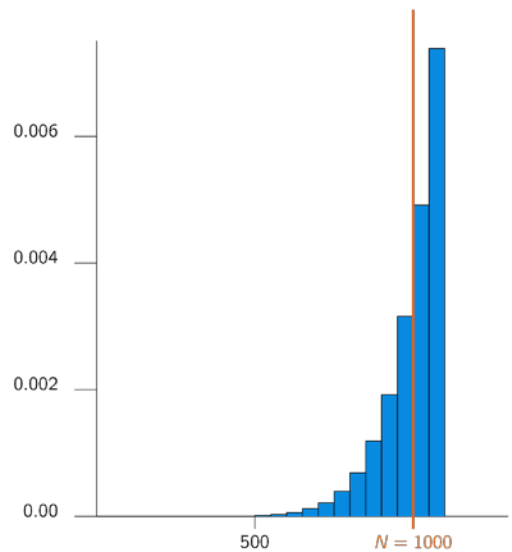
The serial number X_i of the i'th captured tank is modeled as a random drawing from $\{1, 2, \ldots, N\}$ with uniform distribution.

Two unbiased estimators for N are

$$T_1 = 2\bar{X}_n - 1$$

$$T_2 = \frac{n+1}{n} \max X_i - 1$$





$$\mathrm{Var}(T_1) = \frac{(N+1)(N-n)}{3n}$$

$$\mathrm{Var}(T_2) = \frac{(N+1)(N-n)}{(n+2)n}$$

Downtime on a network server

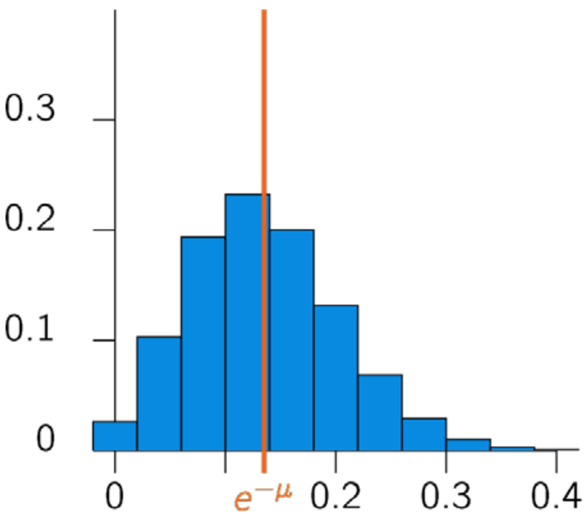
The number of packets arriving at a server in one minute is modeled as $X_i \sim \text{Pois}(\mu)$. Two estimators for $P(X_i = 0) = e^{-\mu}$ are

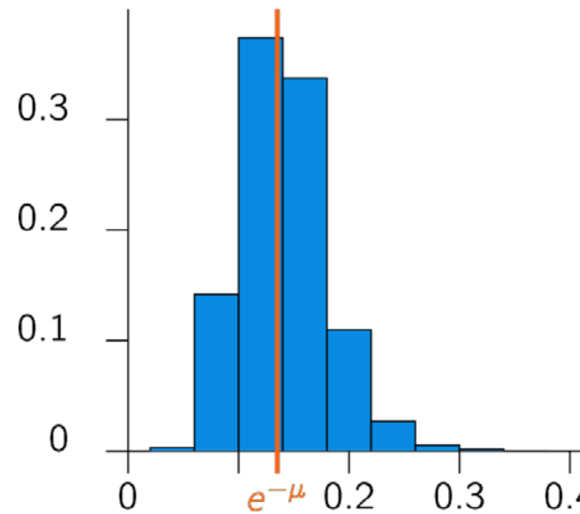
$S = \frac{\text{number of } X_i = 0}{n}$ and $T = e^{-\bar{X}_n}$.

Downtime on a network server

The number of packets arriving at a server in one minute is modeled as $X_i \sim \text{Pois}(\mu)$. Two estimators for $P(X_i = 0) = e^{-\mu}$ are

$$S = \frac{\text{number of } X_i = 0}{n} \quad \text{and} \quad T = e^{-\bar{X}_n}.$$

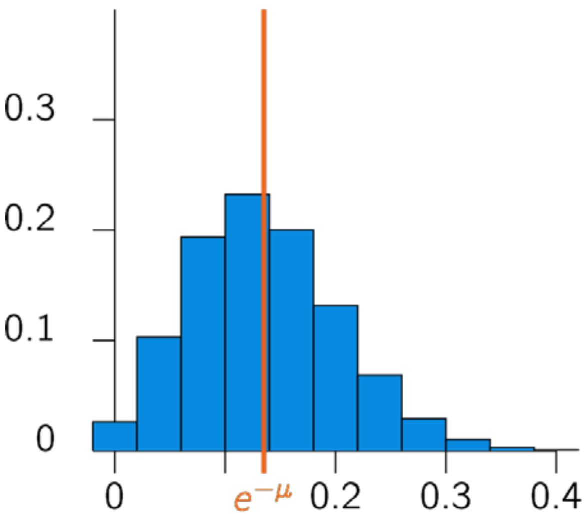


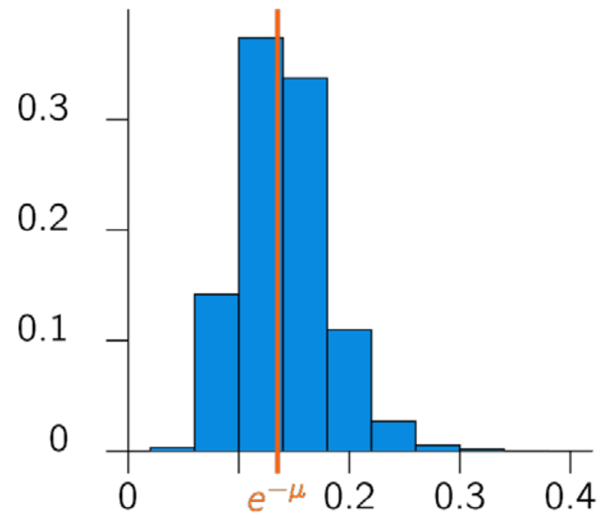


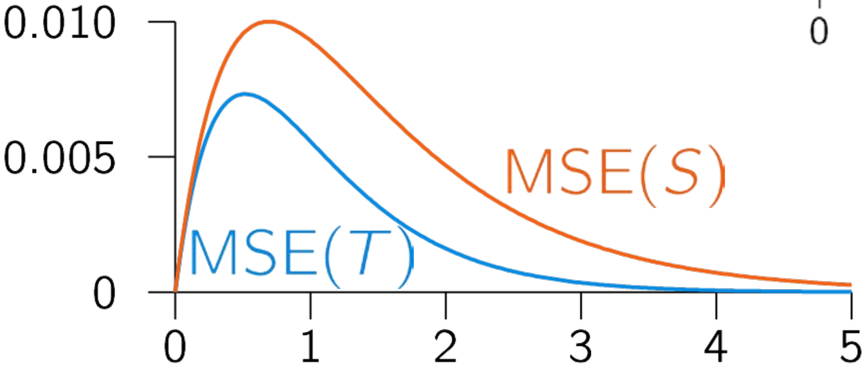
Downtime on a network server

The number of packets arriving at a server in one minute is modeled as $X_i \sim \text{Pois}(\mu)$. Two estimators for $P(X_i = 0) = e^{-\mu}$ are

$$S = \frac{\text{number of } X_i\!=\!0}{n} \quad \text{and} \quad T = e^{-\bar{X}_n}.$$







Exercises

Given a random sample $X_1, X_2, \ldots, X_{2n}$ from a distribution with finite variance σ^2. Let $\bar{X}_n$ denote the average of the first n measurements, and $\bar{X}_{2n}$ the average of all measurements.

a. What is the relative efficiency $\mathrm{Var}(\bar{X}_n)/\mathrm{Var}(\bar{X}_{2n})$?

b. Is $\bar{X}_n$ or $\bar{X}_{2n}$ more efficient?

Let X_1 and X_2 be two independent random variables with identical mean μ and variances σ_1^2 and σ_2^2. Recall $T = aX_1 + (1-a)X_2$ is an unbiased estimator for μ. For which value of a is this estimator most efficient?

Given estimators S and T for θ with

$$\mathrm{E}[S] = \theta, \quad \mathrm{Var}(S) = 30, \quad \mathrm{E}[T] = \theta + 5, \quad \mathrm{Var}(T) = 6.$$

Which estimator would you prefer and why?

For next class (week 3.7 lesson 1):

Complete MyStatlab assignments and

book exercises 19.3, 19.6, 20.3, 20.6, 20.5, 20.9

Watch prelecture *'Maximum likelihood principle'*

Book: Section: 21.1 and 21.2

After this class you can:

- Apply the Maximum Likelihood Principle in various settings
- Derive the Maximum Likelihood Estimate for model parameters
- Set up a linear regression model and estimates its parameters

Statistics

Good luck!