Probability

Law of Large Numbers and Central Limit Theorem

Name teacher

TU Delft

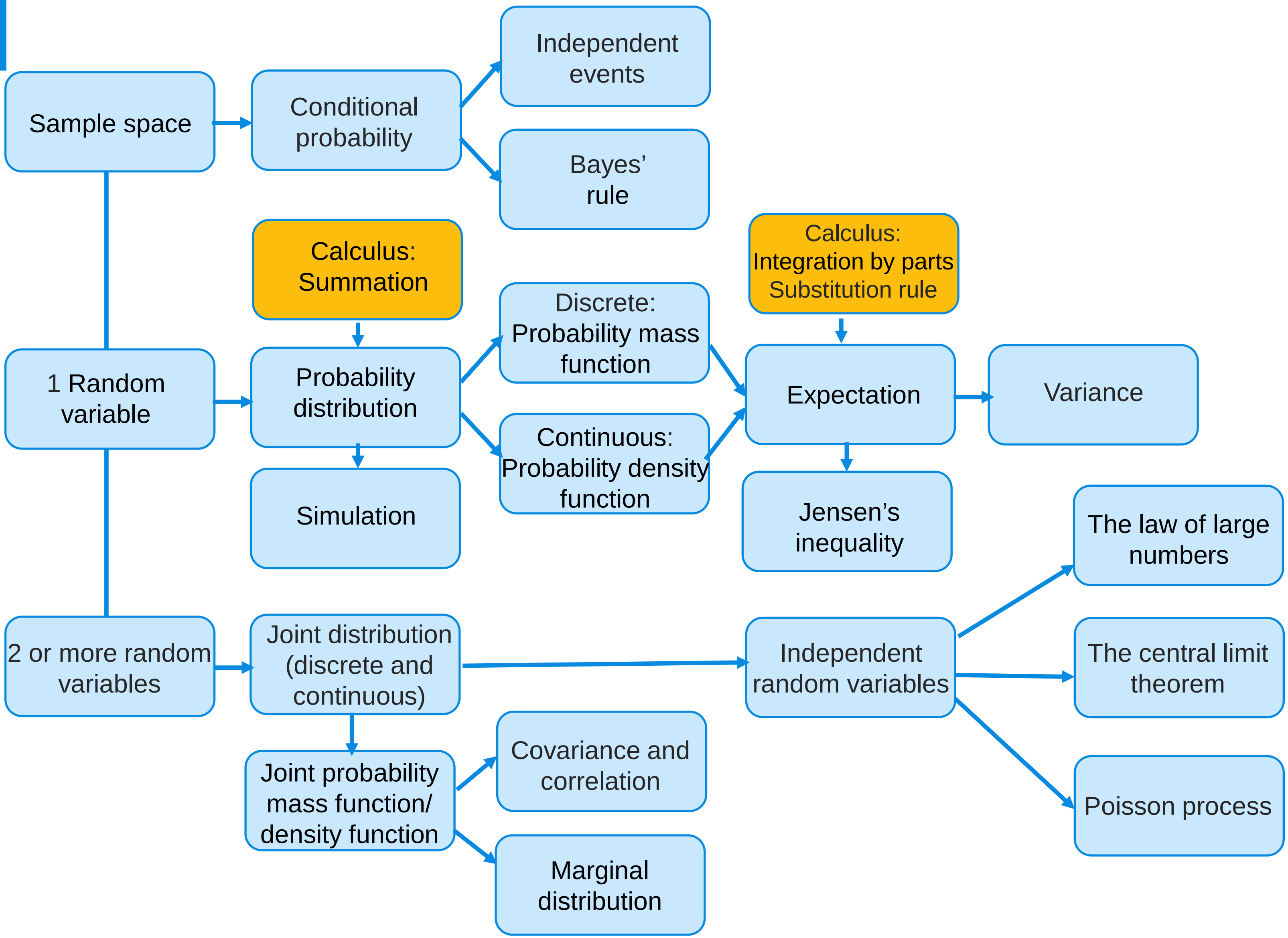
Learning objective

After this class you are able to

- **Compute the expectation and variance of sample mean**
- **Apply Chebyshev's inequality**
- **Apply Law of Large numbers**
- **Apply Central Limit Theorem**

Book: Chapters 13 and 14

Probability



Before this class (week 3.5 lesson 2):

Watch prelecture *'Averages and Chebyshev'*

Book: Sections 13.1 and 13.2

Programme

- Law of Large Numbers
- Exercises
- Central Limit Theorem
- Exercises

Suppose U has a uniform distribution on [-1, 1]. Note that $\mathrm{E}[U] = 0$. First find σ and then $\mathrm{P}(|U - 0| \geq 2\sigma)$.

A) $\sigma = 1/\sqrt{3}$

B) $\sigma = 1/3$

C) $\sigma = 1/\sqrt{6}$

D) $\sigma = 1/6$

Suppose U has a uniform distribution on [-1, 1]. Note that $\mathrm{E}[U] = 0$. First find σ and then $\mathrm{P}(|U - 0| \geq 2\sigma)$.

A) $2\sqrt{3} - 2$

B) $\sqrt{3} - 1$

C) $1/9$

D) 0

Suppose $Y_1, \ldots, Y_{100}$ are independent random variables with $\mathrm{E}[Y_i] = m$ and $\mathrm{Var}(Y_i) = 2^2$ (i.e. $\sigma = 2$). Let $\bar{Y}_{100} = \frac{1}{100}\sum_{i=1}^{100} Y_i$, the average.

Find the smallest p for which $\mathrm{P}\left(|\bar{Y}_{100} - m| > 1\right) \leq p$.

A) 1/100

B) 1/25

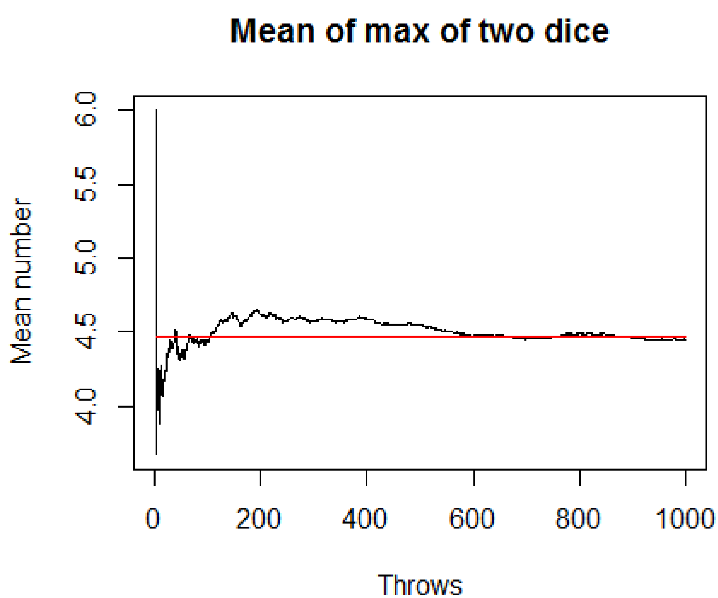
C) 1/5

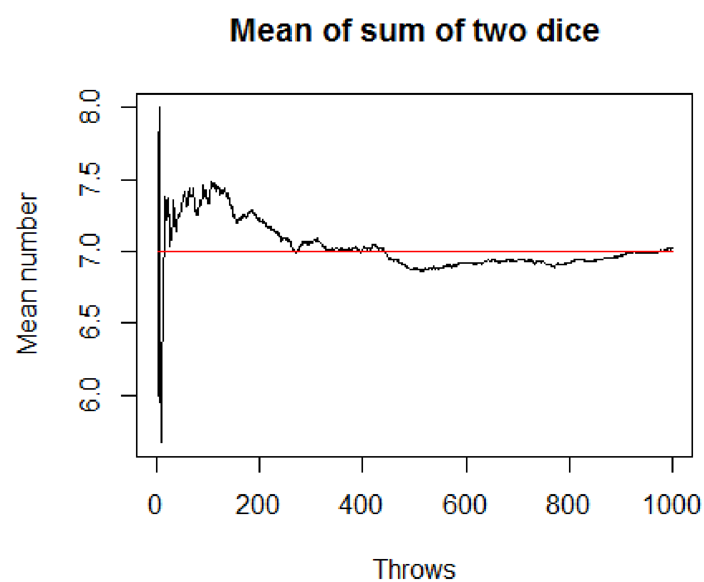
D) 1/40

Law of Large Numbers

Theorem:
Let $X_1, \ldots, X_n$ be an i.i.d. sequence with
$E[X] = \mu$ and $Var(X) = \sigma^2$. Then, for any $\epsilon > 0$:

$$P\left(|\bar{X}_n - \mu| > \epsilon\right) \to 0, \text{ as } n \to \infty.$$





Application of Law of Large Numbers

Let A be event of interest. How to estimate probability $p = \mathrm{P}(A)$?

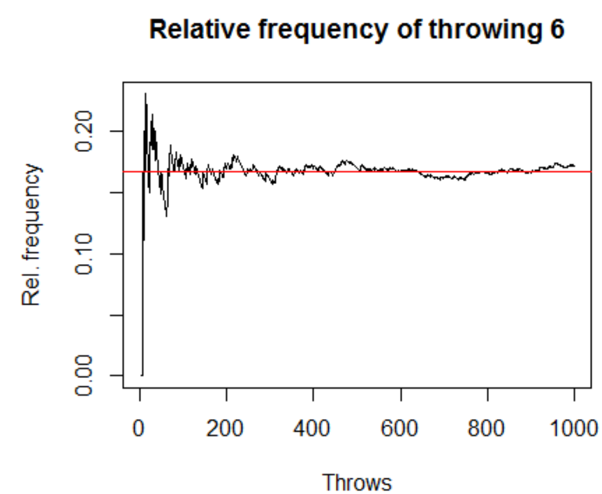
Define the random variables $X_i = \begin{cases} 1, & A \text{ occurs,} \\ 0, & \text{other.} \end{cases}$

$\mathrm{E}[X_i] = p$, thus Law of Large Numbers yields for any $\epsilon > 0$

$$\mathrm{P}\left(|\bar{X}_n - p| > \epsilon\right) \to 0, \text{ as } n \to \infty.$$

In other words: $\bar{X}_n \approx p$ for n large.

NB: $$\bar{X}_n = \frac{\text{number of times } A \text{ occurs}}{n} = \text{relative frequency}$$



Exercise

A casino is for sale and you might want to buy it, but you want to know how much money you are going to make.
You know that the roulette game Red or Black is played about 1000 times a night, 365 days a year.
Each time it is played you have probability 19/37 of winning the player's bet of 1 euro and probability 18/37 of having to pay the player 1 euro.

Explain in detail why the law of large numbers can be used to determine the income of the casino, and determine how much it is.

Exercise

Let $X_1, X_2, \ldots$ be an independent sequence of $U(-1, 1)$ RVs.

Let $T_n = \frac{1}{n} \sum_{i=1}^{n} X_i^2$.

Determine and explain the value a such that for any $\epsilon > 0$

$$\mathrm{P}(|T_n - a| > \epsilon) \to 0, \text{ as } n \to \infty.$$

Central Limit Theorem

Theorem:

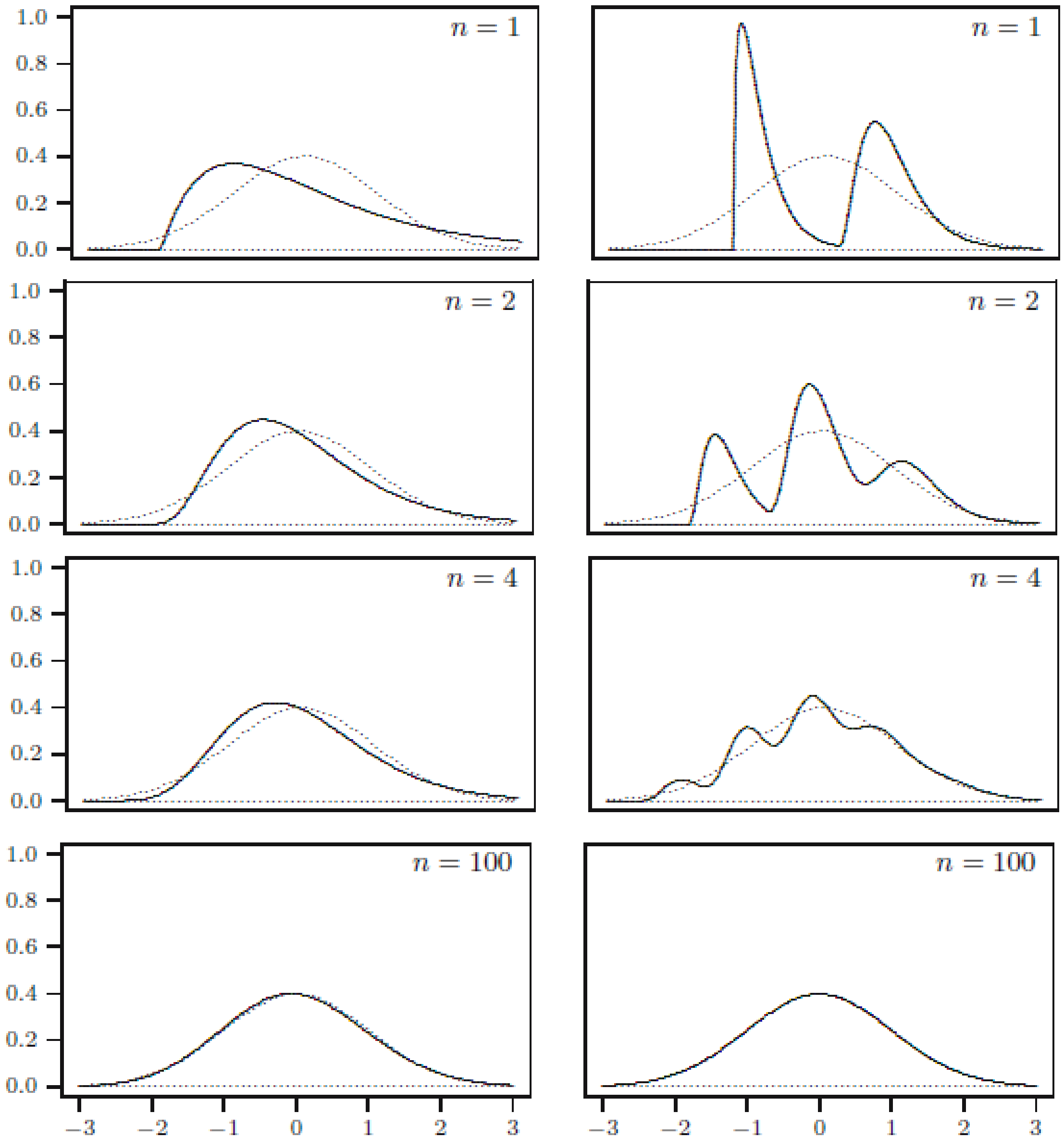
Let $X_1, \ldots, X_n$ be an i.i.d. sequence with $\mathrm{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. For $n \geq 1$, let Z_n be defined by

$$Z_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}.$$

Then for any a it holds that $\mathrm{P}(Z_n \leq a) \to \mathrm{P}(Z \leq a)$, as $n \to \infty$, where Z has a $N(0, 1)$ distribution.

In other words: for large n, Z_n has approximately a standard normal distribution.

Central Limit Theorem in pictures



Left: skewed unimodal distr.

Right: bimodal distr.

Central Limit Theorem: sum instead of mean

Theorem:

Same setting as before. But we can rewrite Z_n:

$$Z_n = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma} = \frac{\sqrt{n}}{n}\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma}$$

$$= \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}.$$

Then for any a it holds that $\mathrm{P}(Z_n \leq a) \to \mathrm{P}(Z \leq a)$, as $n \to \infty$, where Z has a $N(0, 1)$ distribution.

In other words: for large n, Z_n has approximately a standard normal distribution.

Let $X_1, X_2, \ldots, X_{169}$ be i.i.d. RVs with $E[X_i]=2$ and $E[X_i^2] = 13$. Find a such that $P(X_1 + \ldots + X_{169} > 350) \approx P(Z > a)$, where $Z \sim N(0, 1)$.

A) 0.103

B) 0.024

C) 0.308

D) 0.256

Application of Central Limit Theorem

Simplify bookkeeping by rounding amounts to the nearest integer, for example, rounding 99.53 euro and 100.46 euro both to 100 euro.

What is the cumulative effect of this if there are 100 amounts?

Model the rounding errors by 100 independent $U(-0.5, 0.5)$ random variables $X_1, X_2, \ldots, X_{100}$

What is the probability that the cumulative rounding error exceeds 10 euro?

Two exercises

Let $X_1, X_2, \ldots, X_{144}$ be i.i.d. RVs, each with expectation $\mu = E[X_i] = 2$ and variance $\sigma^2 = \mathrm{Var}(X_i) = 4$.

Approximate $P(X_1 + X_2 + \ldots + X_{144} > 144)$, using the central limit theorem.

Proof the following theorem.

Let X be a $Bin(n, p)$ distributed random variable and n large. Then the random variable $\dfrac{X - np}{\sqrt{np(1-p)}}$ has a distribution that is approximately standard normal.

For next class (week 3.6 lesson 1):

Complete MyStatlab assignments and book exercises

Watch prelectures

'Data Analysis: Graphical Representation' and

'Data Analysis: Numerical Representation'

Book: Section 15.1, 15.2, 15.3, 15.4, 16.1, 16.2, 16.3

After this class you can:

- **Create and interpret graphical summaries**
- **Create and interpret numerical summaries**
- **Build statistical models**
- **Estimate features of the model distribution**

Probability

Good luck!