Random variables III.

(Important discrete and continuous distributions)

Lecturer:
Francesco Dolce

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

Probability and Statistics

BIE-PST, WS 2023/24, Lecture 5

Content

- **Probability theory:**
  - Events, probability, conditional probability, Bayes' Theorem, independence of events.
  - Random variables, distribution function, functions of random variables, characteristics of random variables: expected value, variance, moments, generating function, quantiles, critical values, important discrete and continuous distributions, covariance and correlation.
  - Random vectors, joint and marginal distributions, functions of random vectors, independence of random variables, conditional distribution, conditional expected value.
  - Markov's and Chebyshev's inequality, weak law of large numbers, strong law of large numbers, Central limit theorem.
- **Mathematical statistics:**
  - Point estimators, sample mean, sample variance, properties of point estimators, Maximum likelihood method.
  - Interval estimators, hypothesis testing, one-sided vs. two-sided alternatives, linear regression, estimators of regression parameters, testing of linear model.

Recap

- A **random variable** X is a measurable function which assigns real values to the outcomes of a random experiment.
- The **distribution** of X gives the information of the probabilities of its values and is uniquely given by the **distribution function**:

$$F_X(x) = \mathrm{P}(X \leq x).$$

- There are two major types of random variables:
  - **Discrete**, taking only countably many possible values.
  - **Continuous**, taking values from an interval.
- The distribution can be given by:
  - for discrete distributions by the **probabilities** of possible values $\mathrm{P}(X = x_k)$.
  - for continuous distributions by the **density** f_X for which

$$F_X(x) = \int_{-\infty}^{x} f(t)dt.$$

Constant random variable

A constant random variable describes a non-random situation when we have only one possible result occurring with probability of 1.

Definition

A random variable X is called **constant**, if for some $c \in \mathbb{R}$ it holds that:

$$X(\omega) = c \text{ for all } \omega \in \Omega.$$

In other words it holds that:

$$\mathrm{P}(X = c) = 1, \qquad \mathrm{P}(X = x) = 0 \;\; \forall x \neq c.$$

We say that a constant random variable has a deterministic or degenerate distribution.

The **distribution function** of a constant random variable is

$$F_X(x) = \begin{cases} 0 & \text{for } \; x < c \\ 1 & \text{for } \; x \geq c. \end{cases}$$

Constant random variable – expectation, variance

$$\mathrm{P}(X = c) = 1, \qquad \mathrm{P}(X = x) = 0 \ \ \forall x \neq c$$

Expectation and variance:

$$\mathrm{E}(X) = \sum_{x_k} x_k \, \mathrm{P}(X = x_k) = c \, \mathrm{P}(x = c) = c$$

$$\mathrm{var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = c^2 - (c)^2 = 0.$$

In calculations we use:

$\mathrm{E}(c) = c$ – the center of mass of a constant c is c itself;

$\mathrm{var}(c) = 0$ – the width of the graph with only one number c is 0.

Bernoulli (Alternative) distribution

Suppose we perform a random experiment with two possible outcomes (alternatives). We assign values 0 (failure) and 1 (success) to these outcomes. We can use for example one toss with an unbalanced coin.

Suppose that a success occurs with the probability p.

Definition

A random variable X has the **Bernoulli** (alternative) **distribution** with parameter $p \in [0, 1]$, if it holds that:

$$\mathrm{P}(X = 1) = p, \qquad \mathrm{P}(X = 0) = 1 - p.$$

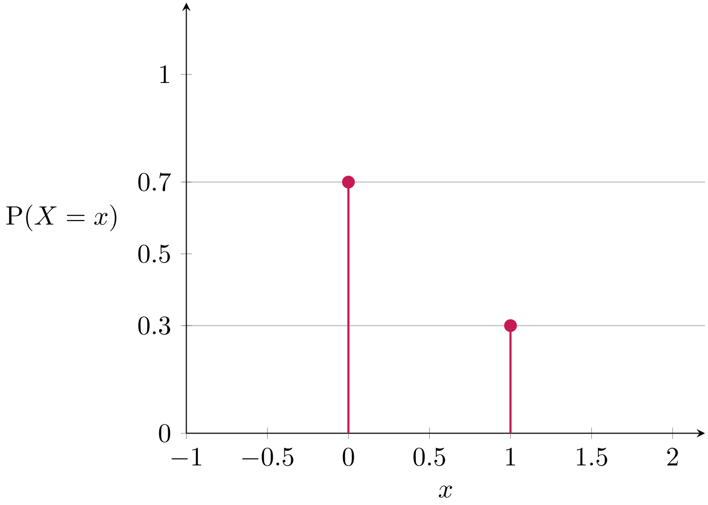
Notation: $X \sim \text{Be}(p)$ or $X \sim \text{Bernoulli}(p)$ or $X \sim \text{Alt}(p)$.

Example – toss with a coin

- Let us choose $X(\text{Heads}) = 1$ and $X(\text{Tails}) = 0$.
- We denote the occurrence of Heads as a success: $p = \mathrm{P}(\text{Heads})$.

Bernoulli distribution – graph of probabilities

Probabilities of values of the Bernoulli distribution with $p = 0.3$:



Bernoulli distribution – expectation, variance

Bernoulli random variable:

$$P(X = 1) = p \in [0, 1] \qquad \text{(Heads, success)}$$
$$P(X = 0) = 1 - p \qquad \text{(Tails, failure).}$$

Expectation and variance:

$$E(X) = \sum_{x_k} x_k \, P(X = x_k) = 1 \cdot p + 0 \cdot (1 - p) = p$$
$$E(X^2) = \sum_{x_k} x_k^2 \, P(X = x_k) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$$
$$\text{var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p).$$

Binomial distribution

If we repeat the coin tossing we can be interested in how many times from n tosses we have obtained Heads:

- Consider n independent experiments with two possible outcomes.
- Again suppose that we succeed in each experiment with probability p.
- The probability that **exactly k out of n attempts ended with a success** is

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Definition

A random variable X has the **binomial distribution** with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, if

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Notation: $X \sim \mathrm{Bin}(n,p)$, $X \sim \mathrm{Binom}(n,p)$.

Binomial distribution – normalization

To prove that the binomial distribution is correctly defined, we verify the **normalization condition**, i.e., that the sum of all probabilities is equal to 1:

$$\sum_{k=0}^{n} \mathrm{P}(X = k) = 1.$$

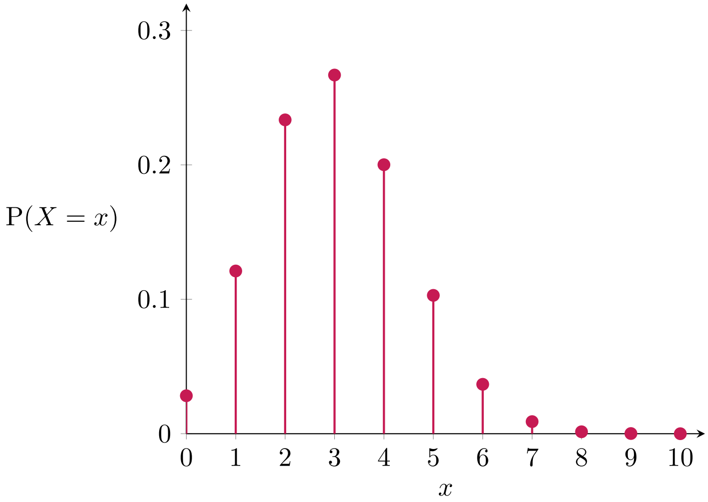
According to the binomial theorem it holds that

$$\sum_{k=0}^{n} \mathrm{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

□

Binomial distribution – graph of probabilities

Binomial distribution with parameters $n = 10$ and $p = 0.3$:



Binomial distribution – expectation

Binomial random variable $X \sim \text{Binom}(n,p)$:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n.$$

$$E(X) = \sum_{k=0}^{n} k \; P(X = k) = \sum_{k=0}^{n} \binom{n}{k} k \, p^k (1-p)^{n-k}.$$

The sum on the right hand side looks, except for a term $k \, p^k$, like

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

Notice that $(p^k)' = k \, p^{k-1}$ and thus $\qquad p \, (p^k)' = k \, p^k$.

After differentiating both sides with respect to p and multiplying by p we obtain the needed expression.

Binomial distribution – expectation

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1 \quad \Big/ \text{differentiate w.r.t. } p$$

$$\sum_{k=0}^{n} \binom{n}{k} \left[k\, p^{k-1}(1-p)^{n-k} \;+\; p^k(1-p)^{n-k-1}\right] = 0 \quad \Big/ \text{split the sum}$$

$$\sum_{k=0}^{n} \binom{n}{k} k p^{k-1}(1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k-1} \quad \Big/ \text{multiply by } p$$

$$\sum_{k=0}^{n} \binom{n}{k} k\, p^k (1-p)^{n-k} = p \sum_{k=0}^{n} \binom{n}{k} p^{k-1}(1-p)^{n-k-1} \quad \Big/ k\binom{n}{k} = n\binom{n-1}{k-1}$$

$$E(X) = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-1-(k-1)}$$

$$= np \cdot (p + 1 - p)^{n-1} = np.$$

Binomial distribution – variance

Similarly by means of differentiating we calculate $\mathrm{E}(X^2)$:

$$\mathrm{E}(X^2) = \sum_{k=0}^{n} \binom{n}{k} k^2 \, p^k (1-p)^{n-k} = np + n(n-1)p^2.$$

Therefore

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p)$$

Detailed computation of $\mathrm{E}(X^2)$ can be found in the lecture handout.

Indicator of an event

A special and important example of a Bernoulli random variable is the **indicator of an event**.

Definition

Let $A \in \mathcal{F}$ be an event. The random variable $\mathbb{1}_A : \Omega \to \{0, 1\}$ defined as

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

is called the **indicator** (or **characteristic function**) of the event A.

For the indicator of an event A it holds that:

$$p = \mathrm{P}(\mathbb{1}_A = 1) = \mathrm{P}(A),$$
$$1 - p = \mathrm{P}(\mathbb{1}_A = 0) = \mathrm{P}(A^c) = 1 - \mathrm{P}(A).$$

Indicator of event – examples

Examples – tossing a coin

- The Bernoulli random variable X from the previous example (tossing a coin) is nothing but an indicator of the event $\{\mathrm{H}\}$. Thus $X = \mathbb{1}_{\{\mathrm{H}\}} = \mathbb{1}_{\mathrm{H}}$.
- The Binomial random variable X corresponding to number of Heads in n tosses can be expressed as the sum

$$X = \sum_{i=1}^{n} \mathbb{1}_{\mathrm{H}_i},$$

where $\mathbb{1}_{\mathrm{H}_i}$ is the indicator of the event H_i = “Heads appears in the i-th toss”.

Remark:

Expressing a binomial variable as a sum of (Bernoulli) indicators often leads to a significant simplification of calculations.

Geometric distribution

Another important event is the first occurrence of Heads in a sequence of coin tosses:

- Consider a sequence of independent experiments with two possible outcomes.
- Suppose that each experiment ends with a success with probability p.
- Probability that the **first successful** attempt the is **k-th** in the sequence is

$$(1-p)^{k-1}p.$$

Definition

A random variable X has the **geometric distribution** with parameter $p \in (0,1)$, if

$$\mathrm{P}(X=k) = (1-p)^{k-1}p, \qquad k = 1, 2, \ldots .$$

Notation: $X \sim \text{Geom}(p)$.

Again we verify the **normalization condition**:

$$\sum_{k=1}^{\infty} \mathrm{P}(X=k) = \sum_{k=1}^{\infty} (1-p)^{k-1}p = p\sum_{k=0}^{\infty}(1-p)^k = \frac{p}{1-(1-p)} = 1.$$

Geometric distribution – distribution function

The **distribution function** of the geometric distribution can be expressed as

$$F_X(k) = \mathrm{P}(X \leq k) = \sum_{i=1}^{k} p(1-p)^{i-1} = p\sum_{i=0}^{k-1}(1-p)^i$$
$$= p\frac{1-(1-p)^k}{1-(1-p)} = 1-(1-p)^k.$$

For non-integer points $x > 0$ the value of distribution function is equal to value at point $\lfloor x \rfloor$ (the lower integer part of x):

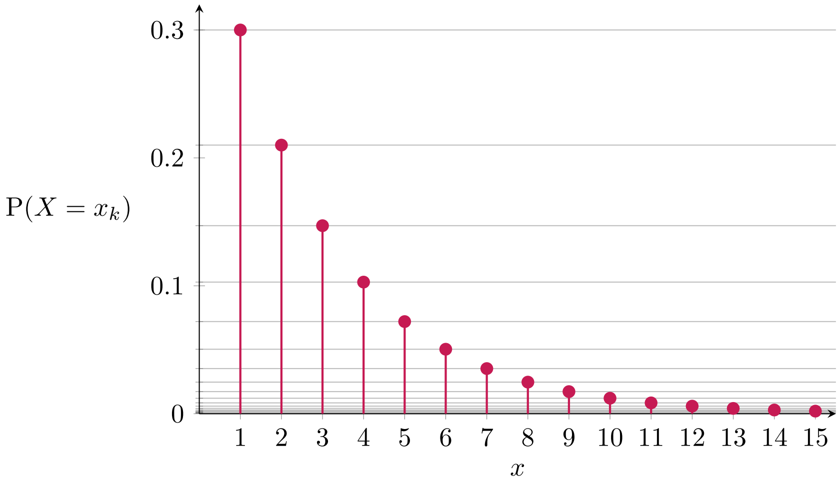
$$F_X(x) = F_X(\lfloor x \rfloor) = 1-(1-p)^{\lfloor x \rfloor}.$$

The probability that the success does not occur after k attempts can be computed as

$$\mathrm{P}(X > k) = (1-p)^k \qquad \text{and thus} \qquad F_X(k) = 1-\mathrm{P}(X>k) = 1-(1-p)^k.$$

Geometric distribution – graph of probabilities

Geometric distribution with parameter $p = 0.3$:



Geometric distribution – expectation

$$\mathrm{P}(X=k) = (1-p)^{k-1}p \qquad k = 1, 2, \ldots$$

$$\mathrm{E}(X) = \sum_{\text{all } x_k} x_k \, \mathrm{P}(X = x_k) = \sum_{k=1}^{\infty} k\,(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k\,(1-p)^{k-1}.$$

The sum on the right-hand side looks as the derivative of $-\sum_{k=0}^{\infty}(1-p)^k$:

$$\begin{aligned}\mathrm{E}\,X &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = -p\left(\sum_{k=1}^{\infty}(1-p)^k\right)' \\ &= -p\left(\frac{1}{1-(1-p)}\right)' = -p\left(\frac{-1}{p^2}\right) \\ &= \frac{1}{p}.\end{aligned}$$

Geometric distribution – variance

We can compute $\mathrm{E}(X^2)$ using the same procedure. From the above we know that

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k^2(1-p)^{k-1} \\
&= p\left(\sum_{k=1}^{\infty} -k(1-p)^k\right)' = p\left((1-p)\sum_{k=1}^{\infty} -k(1-p)^{k-1}\right)' \\
&= p\left((1-p)\left(\sum_{k=1}^{\infty}(1-p)^k\right)'\right)' = p\left((1-p)\left(\frac{1}{p}\right)'\right)' \\
&= p\left(\frac{p-1}{p^2}\right)' = p\,\frac{p^2-(p-1)2p}{p^4} = \frac{2-p}{p^2}.
\end{aligned}$$

Thus

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X)^2) = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

Poisson distribution – motivation

The number of random occurrences during a given time is often modeled by the Poisson distribution:

- For example $X =$ "number of server requests in 15 seconds".
- Or $X =$ "number of customers in a shop during lunch time".
- Finite population: n individuals independently decide whether to go to a shop or not.
  - Then X is a binomial random variable: $X \sim \text{Binom}(n, p)$.
- Infinite population: we are interested in $X \sim \text{Binom}(n, p)$ for $n \to \infty$.
  - Useful approximation for great populations (molecules of gas, internet users, etc.).

Example – number of customers in a shop during lunch time

- number of inhabitants in a city: n;
- number of shops proportional to the number of inhabitants: $n_o = \rho n$, where ρ is the density of shops (number of shops per one inhabitant);
- probability that an inhabitant decides to go shopping: z;
- probability that an inhabitant goes to a particular shop: $p = z/n_o = z/(\rho n)$;
- number of inhabitants going to the particular shop: $X \sim \text{Binom}(n, p)$;
- expected value: $\mathrm{E}\, X = np = nz/(\rho n) = z/\rho$ . . . constant.

Poisson distribution – motivation

Binomial distribution with $n \to \infty$, $p \to 0$ and $np = \lambda$ is

$$\mathrm{P}(X = k) = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

We rearrange the product and take a limit $n \to \infty$

$$\begin{array}{ccccccccc} \mathrm{P}(X = k) = & \dfrac{n}{n} & \dfrac{(n-1)}{n} & \cdots & \dfrac{(n-k+1)}{n} & \dfrac{\lambda^k}{k!} & \left(1 - \dfrac{\lambda}{n}\right)^n & \left(1 - \dfrac{\lambda}{n}\right)^{-k} \\ & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 1 & \cdots & 1 & \dfrac{\lambda^k}{k!} & e^{-\lambda} & 1 \end{array}$$

Finally we have

$$\lim_{n \to \infty} \mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Poisson distribution

Definition

A random variable X has the **Poisson distribution** with parameter $\lambda > 0$ if

$$\mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0,\ 1, \dots .$$

Notation: $X \sim$ Poisson(λ)

Recalling the important formula:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

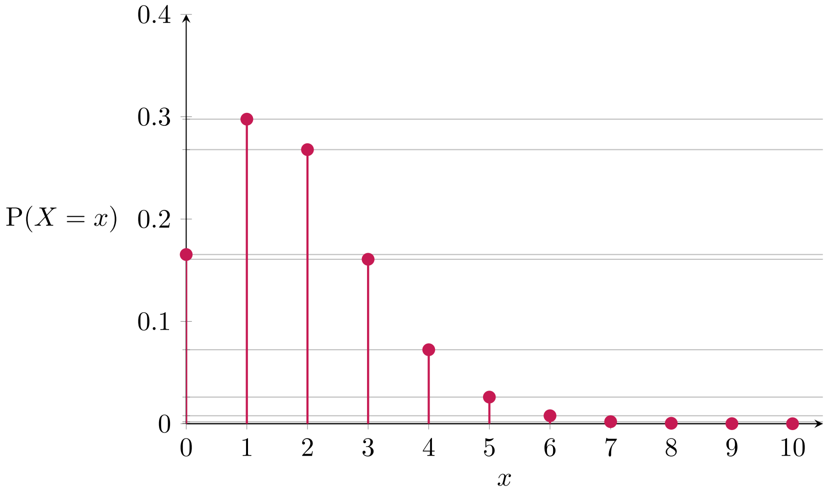
we can check that he normalization condition holds:

$$\sum_{k=0}^{\infty} \mathrm{P}(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

□

Poisson distribution – graph of probabilities

Poisson distribution with parameter $\lambda = 1.8$:



Poisson distribution – expectation

$$\mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$$

The expectation is

$$\begin{aligned}
\mathrm{E}(X) = \sum_{k=0}^{\infty} k\ \mathrm{P}(X = k) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda\, e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda\, e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!} \\
&= \lambda\, e^{-\lambda} e^{\lambda} = \lambda.
\end{aligned}$$

Poisson distribution – variance

$\mathrm{E}(X^2)$ is computed similarly:

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^{k-1}}{k(k-1)!} \\
&= \lambda e^{-\lambda} \left(\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) \\
&= \lambda e^{-\lambda} \left(\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) \\
&= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) = \lambda^2 + \lambda.
\end{aligned}$$

Thus

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\, X)^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

Recapitulation

- **Bernoulli** (Alternative) distribution with parameter p, $0 \le p \le 1$, $X \sim \text{Be}(p)$:
  (another notation $X \sim \text{Bernoulli}(p)$ $X \sim \text{Alt}(p)$)
  (One toss with an unbalanced coin.)
  $$\mathrm{P}(1) = p, \quad \mathrm{P}(0) = 1 - p, \qquad \mathrm{E}\, X = p, \ \operatorname{var} X = p(1-p).$$
- **Binomial** distribution with parameters n and p, $0 \le p \le 1$, $X \sim \text{Binom}(n, p)$:
  (Number of Heads in n tosses with an unbalanced coin.)
  $$\mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \mathrm{E}\, X = np, \ \operatorname{var} X = np(1-p).$$
- **Geometric** distribution with parameter p, $0 < p < 1$, $X \sim \text{Geom}(p)$:
  (Number of tosses with an unbalanced coin until first Heads appears.)
  $$\mathrm{P}(X = k) = (1-p)^{k-1} p, \ k = 1, 2, \ldots, \qquad \mathrm{E}\, X = \frac{1}{p}, \ \operatorname{var} X = \frac{1-p}{p^2}.$$
- **Poisson** distribution with parameter $\lambda > 0$, $X \sim \text{Poisson}(\lambda)$:
  (Limit of the binomial distribution for $n \to \infty$.)
  $$\mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, 2, \ldots, \qquad \mathrm{E}\, X = \operatorname{var} X = \lambda.$$

Uniform distribution

All values in some interval (a, b) can occur with “equal” probability.

Definition

A continuous random variable X has the **uniform** distribution with parameters $a < b$, $a, b \in \mathbb{R}$, if its density has the form:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } x \in (a, b), \\ 0 & \text{elsewhere.} \end{cases}$$

Notation: $X \sim \mathsf{Unif}(a, b)$, $\quad X \sim \mathsf{U}(a, b)$.

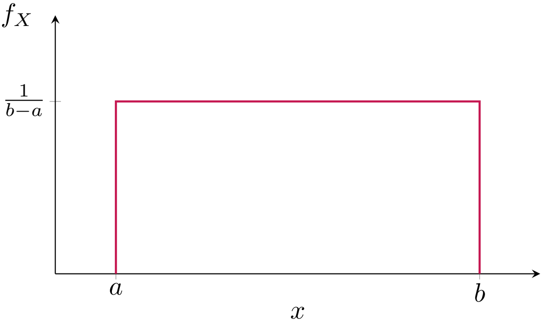
Normalization condition:

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = \int_a^b \frac{1}{b-a}\mathrm{d}x = \frac{b-a}{b-a} = 1.$$

Distribution function:

$$F_X(x) = \int_a^x \frac{1}{b-a}\mathrm{d}t = \left[\frac{t}{b-a}\right]_a^x = \frac{x-a}{b-a} \qquad \text{for} \quad x \in [a, b].$$

Uniform distribution – graph of density



Uniform distribution – expectation, variance

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } x \in (a,b), \\ 0 & \text{elsewhere.} \end{cases}$$

$$\mathrm{E}(X) = \int_a^b x\, f_X(x)\,\mathrm{d}x = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{a+b}{2},$$

$$\mathrm{E}(X^2) = \int_a^b x^2 f_X(x)\,\mathrm{d}x = \int_a^b \frac{x^2}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{a^2+ab+b^2}{3},$$

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Exponential distribution

Very often used in queuing theory and theory of random processes.

Definition

A random variable X has the **exponential** distribution with parameter $\lambda > 0$, if its density has the form:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \in [0, +\infty), \\ 0 & \text{elsewhere.} \end{cases}$$

Notation: $X \sim \text{Exp}(\lambda)$.

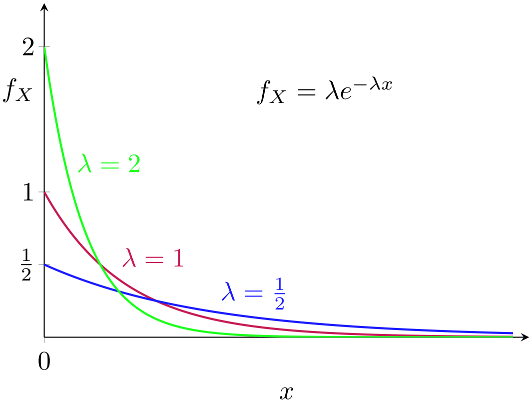
Normalization:

$$\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = \int_0^{\infty} \lambda e^{-\lambda x}\mathrm{d}x = \left[-e^{-\lambda x}\right]_0^{+\infty} = 0 - (-1) = 1.$$

Distribution function:

$$F_X(x) = \int_0^{x} \lambda e^{-\lambda t}\mathrm{d}t = \left[-e^{-\lambda t}\right]_0^{x} = 1 - e^{-\lambda x}.$$

Exponential distribution – graph of density



Exponential distribution – expectation, variance

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{elsewhere.} \end{cases}$$

$$\mathrm{E}(X) = \int_0^\infty x\, f_X(x)\,\mathrm{d}x = \int_0^\infty x\lambda e^{-\lambda x}\mathrm{d}x \overset{\text{by parts}}{=} \frac{1}{\lambda}$$

$$\mathrm{E}(X^2) = \int_0^\infty x^2\, f_X(x)\,\mathrm{d}x = \int_0^\infty x^2\lambda e^{-\lambda x}\mathrm{d}x \overset{\text{2x by parts}}{=} \frac{2}{\lambda^2}$$

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

✓ Details during tutorials.

Normal distribution

The normal distribution occurs in nature (population lengths, weights, etc.) and is used as an approximation for sums and means of random variables.

Definition

A random variable X has the **normal** (Gaussian) distribution with parameters μ and $\sigma^2 > 0$, if the density has the form:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } x \in (-\infty, +\infty).$$

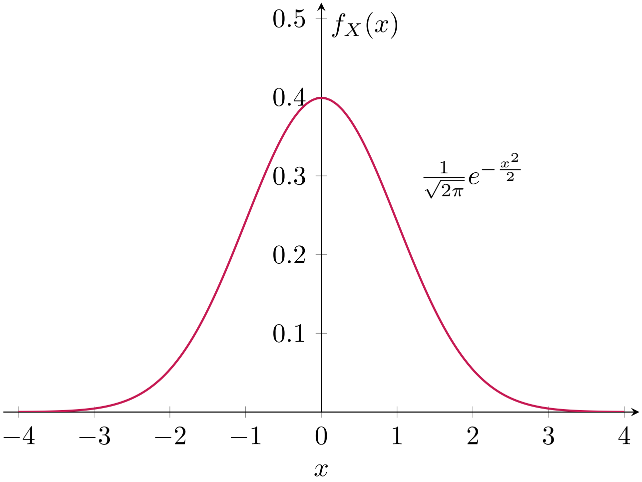
Notation: $X \sim \mathsf{N}(\mu, \sigma^2)$.

- Attention: Some literature and software uses $X \sim \mathsf{N}(\mu, \sigma)$.
- We will further use the symbol σ for $\sqrt{\sigma^2}$.
- $\mathsf{N}(0, 1)$ is called the **standard normal** distribution.

Distribution function: cannot be given explicitly, only numerically. The standard normal distribution function is tabulated and denoted as Φ.

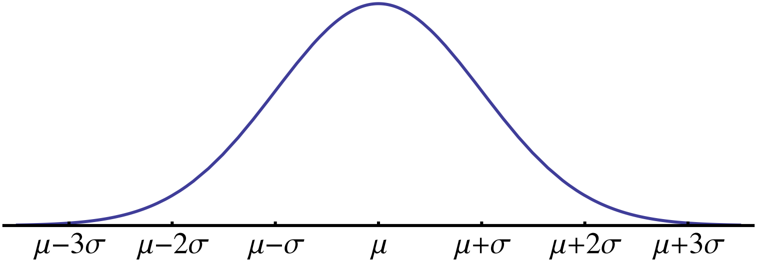
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Standard normal distribution $\mathsf{N}(0, 1)$



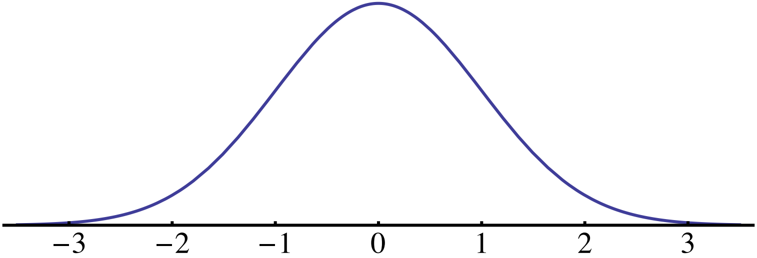
$$\Phi(-x) = 1 - \Phi(x)$$

Density of the normal distribution: $X \sim \mathrm{N}(\mu, \sigma^2)$

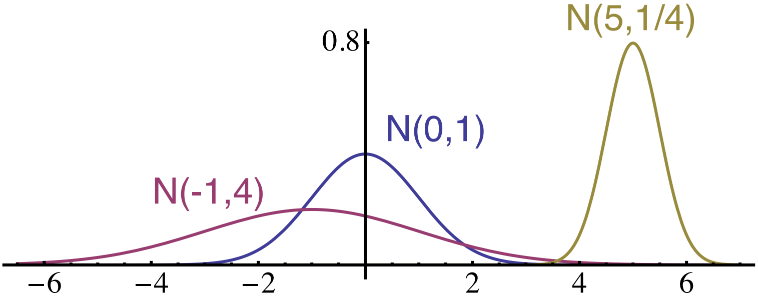


$$\mathrm{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$$
$$\mathrm{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$$
$$\mathrm{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$$

Density of the normal distribution: $Z \sim \mathrm{N}(0, 1)$



Density of the normal distribution



Normal distribution – expectation, variance

Normal random variable $X \sim \mathrm{N}(\mu, \sigma^2)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } x \in (-\infty, +\infty).$$

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x \, \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x \overset{\text{substitution}}{=} \mu.$$

$$\mathrm{var}(X) = \sigma^2.$$

Standardization of random variable

Consider a random variable X with expected value $\operatorname{E} X = \mu$ and variance $\operatorname{var} X = \sigma^2$.

In the easiest possible way, try to convert the variable X to the variable Z with parameters $\operatorname{E} Z = 0$ and $\operatorname{var} Z = 1$ (**standardization**):

- We subtract the expectation μ:

$$\operatorname{E}(X-\mu) = \operatorname{E} X - \mu = 0 \qquad \text{and} \qquad \operatorname{var}(X-\mu) = \operatorname{var} X = \sigma^2.$$

- We rescale with the value $\sigma = \sqrt{\operatorname{var} X}$:

$$\operatorname{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{\operatorname{E}(X-\mu)}{\sigma} = 0 \text{ and } \operatorname{var}\left(\frac{X-\mu}{\sigma}\right) = \frac{\operatorname{var}(X-\mu)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1.$$

The required transformation is thus linear and the random variable

$$Z = \frac{X-\mu}{\sigma}$$

indeed has a zero mean and a variance of 1.

Standardization of a normal random variable

For practical uses we are interested in the standardization of the normal random variable.

Theorem

Let a random variable X have the normal distribution $X \sim \mathsf{N}(\mu, \sigma^2)$. Then the random variable

$$Z = \frac{X - \mu}{\sigma}$$

has the standard normal distribution, $Z \sim \mathsf{N}(0, 1)$.

Proof

$$F_Z(z) = \mathrm{P}(Z \leq z) = \mathrm{P}\left(\frac{X - \mu}{\sigma} \leq z\right) = \mathrm{P}\left(X \leq \sigma z + \mu\right) = F_X(\sigma z + \mu)$$

$$f_Z(z) = \frac{\partial F_Z}{\partial z}(z) = \frac{\partial F_X}{\partial z}(\sigma z + \mu) = \sigma\, f_X(\sigma z + \mu)$$

$$= \sigma \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

□

Standardization of a normal random variable

Remark

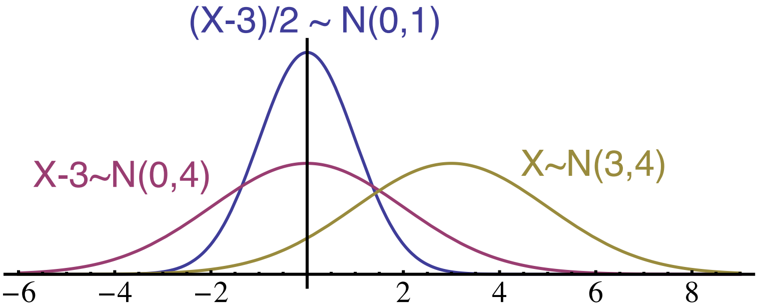
From the previous theorem it follows that:

If $X \sim \mathrm{N}(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$.

This is used for obtaining the values of the distribution function of the variable X from the tables of the standard normal distribution Z:

$$F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right)$$
$$= \mathrm{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Standardization of a normal random variable



Recapitulation

- **Uniform** distribution on the interval $[a, b]$, $X \sim \text{Unif}(a, b)$ or $X \sim \text{U}(a, b)$:

$$f_X(x) = \frac{1}{b-a},\ x \in [a, b] \qquad \mathrm{E}\,X = \frac{a+b}{2}, \qquad \operatorname{var} X = \frac{(b-a)^2}{12}.$$

- **Exponential** distribution with parameter $\lambda > 0$, $X \sim \text{Exp}(\lambda)$:

$$f_X(x) = \lambda e^{-\lambda x},\ x \in [0, +\infty) \qquad \mathrm{E}\,X = \frac{1}{\lambda}, \qquad \operatorname{var} X = \frac{1}{\lambda^2}.$$

- **Normal (Gaussian)** distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, $X \sim \text{N}(\mu, \sigma^2)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty, +\infty) \qquad \mathrm{E}\,X = \mu, \quad \operatorname{var} X = \sigma^2.$$