BIE-PST – Probability and Statistics

Lecture 5: Random variables III.

Winter semester 2023/2024

Lecturer:
Francesco Dolce

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

Table of contents

5 Important discrete distributions

5.1 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 5.1. 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

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

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.

5.2 Bernoulli 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. Experiments with repeated tossing a coin are a basic tool for understanding sequences of random variables (see example below). Similar like Bernoulli we usually choose $X(\text{Heads}) = 1$ and $X(\text{Tails}) = 0$. We denote the occurrence of Heads as a success.

Suppose that a success occurs with the probability p.

Definition 5.2. A random variable X has the *Bernoulli* (alternative) *distribution* with parameter $p \in [0, 1]$, if it holds that:

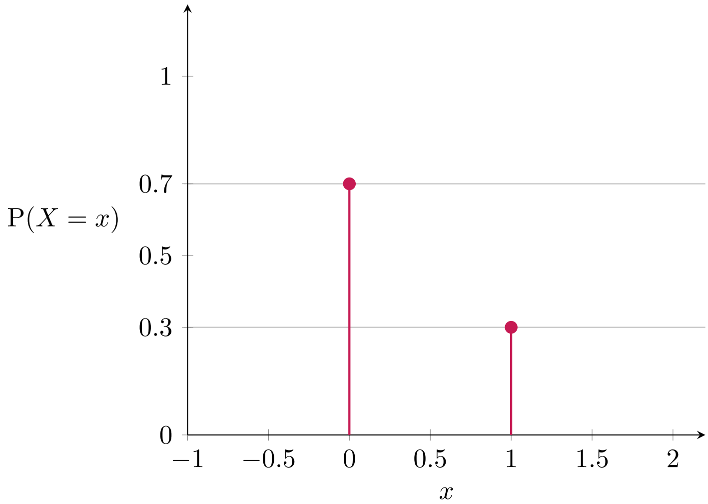
$$\mathrm{P}(X = 1) = p, \qquad \mathrm{P}(X = 0) = 1 - p.$$

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

Example 5.3 (– 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: $p = \mathrm{P}(\text{Heads})$.

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



Expectation and variance:

$$\mathrm{E}(X) = \sum_{x_k} x_k \, \mathrm{P}(X = x_k) = 1 \cdot p + 0 \cdot (1-p) = p$$

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

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

5.3 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

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

From combinatorics we know that k successes among n attempts can occur in $\binom{n}{k}$ different ways (we are choosing k-tuple of positions where success occur in sequence of the length n), which have the same probability p^k (k successes) times $(1-p)^{n-k}$ (and the rest $(n-k)$ failures).

Definition 5.4. A random variable X has the *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, if

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

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

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:

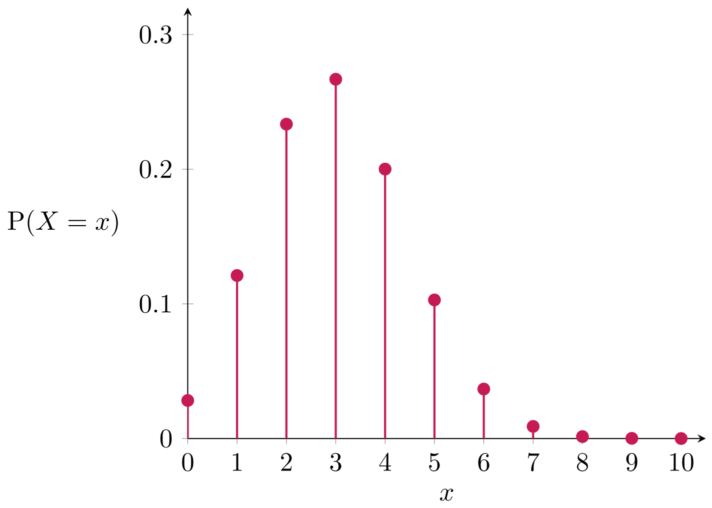
$$\sum_{k=0}^{n} P(X = k) = 1.$$

According to the *binomial theorem* it holds that

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

□

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



We repeat tossing a coin and we are interested in how many times in n tosses a Heads occur. A binomial random variable generally counts X = "number of successes" in n *identical and independent* repetitions of Bernoulli experiment (with $P(\textit{success}) = p$).

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

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.

$$\begin{aligned}
\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} &= 1 && \Big/\text{differentiate w.r.t. } p\\
\sum_{k=0}^{n}\binom{n}{k}\Big[k\,p^{k-1}(1-p)^{n-k} \;+\; p^k(1-p)^{n-k-1}\Big] &= 0 && \Big/\text{split the sum}\\
\sum_{k=0}^{n}\binom{n}{k}kp^{k-1}(1-p)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k-1} & && \Big/\text{multiply by } p\\
\sum_{k=0}^{n}\binom{n}{k}k\,p^k(1-p)^{n-k} = p\sum_{k=0}^{n}\binom{n}{k}p^{k-1}(1-p)^{n-k-1} & && \Big/k\binom{n}{k} = n\binom{n-1}{k-1}\\
E(X) = np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-1-(k-1)} & &&\\
= np\cdot(p+1-p)^{n-1} = np. & &&
\end{aligned}$$

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

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

From the previous argument we know that

$$\sum_{k=0}^{n}\binom{n}{k}kx^ky^{n-k} = nx(x+y)^{n-1},$$

if we differentiate the equality with respect to x and than multiply by x we get the needed sum. Thus:

$$\begin{aligned}
\sum_{k=0}^{n}\binom{n}{k}kx^ky^{n-k} &= nx(x+y)^{n-1} && \Big/\text{ differentiate with respect to } x\\
\sum_{k=0}^{n}\binom{n}{k}k^2x^{k-1}y^{n-k} &= n(x+y)^{n-1} + n(n-1)x(x+y)^{n-2} && \Big/\text{ multiply by } x\\
\sum_{k=0}^{n}\binom{n}{k}k^2x^ky^{n-k} &= nx(x+y)^{n-1} + n(n-1)x^2(x+y)^{n-2}.
\end{aligned}$$

Inserting $x = p$, $y = q = 1-p$ we have:

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

Therefore

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

The above mentioned computations directly from definition are quite demanding. Later after defining independence of random variables we can compute expectation an variance of binomial distribution easily as expectation and variance of sum of n independent Bernoulli random variables.

5.3.1 Indicator of an event

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

Definition 5.5. 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:

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

Examples 5.6 (– 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.

5.4 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 5.7. 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: $X \sim \mathrm{Geom}(p)$.

Again we verify the *normalization condition*:

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

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

$$\begin{aligned} F_X(k) = \mathrm{P}(X \le k) &= \sum_{i=1}^{k} p(1-p)^{i-1} = p \sum_{i=0}^{k-1} (1-p)^i \\ &= p \frac{1-(1-p)^k}{1-(1-p)} = 1-(1-p)^k. \end{aligned}$$

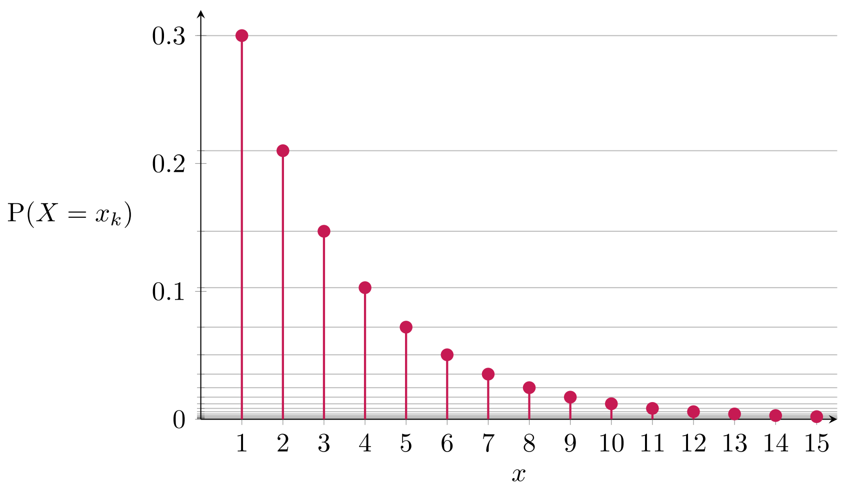
For non-integer points $x > 0$ the value of distribution function is equal to value at point $\lfloor x \rfloor$ (the lower integer part of x):

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

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 with parameter $p = 0.3$:



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

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

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

Switching order of sum and derivative is possible only for uniformly convergent series. Which is fulfilled in our case for all $|1-p|<1$.

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

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

5.5 Poisson distribution

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: $X \sim \mathrm{Binom}(n,p)$.
- Infinite population: we are interested in $X \sim \mathrm{Binom}(n,p)$ *for* $n \to \infty$.
  - Useful approximation for great populations (molecules of gas, internet users, etc.).

Example 5.8 (– 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: $X \sim \mathrm{Binom}(n,p)$;
- expected value: $\mathrm{E}\,X = np = nz/(\rho n) = z/\rho$ . . . constant.

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 $n \to \infty$

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

Finally we have

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

This distribution is called Poisson. For a sequence of random variables the convergence

$$\mathrm{P}(X_n = x) \to \mathrm{P}(X = x), \quad n \to \infty, \quad \forall x$$

is called *convergence in distribution (according the law)* and is denoted by $X_n \overset{\mathcal{D}}{\to} X$. ($\mathcal{L}$ can be used instead of $\mathcal{D}$) We will define this convergence later.

Definition 5.9. A random variable X has the *Poisson distribution* with parameter $\lambda > 0$ if

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

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

Recalling the important formula:

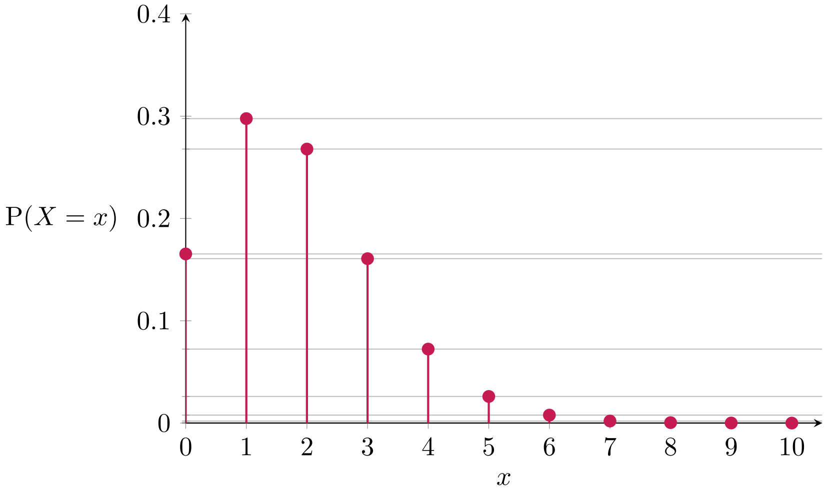
$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

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 with parameter $\lambda = 1.8$:



The expectation is

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

$\mathrm{E}(X^2)$ 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

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

6 Important continuous distributions

6.1 Uniform distribution

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

Definition 6.1. A continuous random variable X has the *uniform* distribution with parameters $a<b$, $a,b\in\mathbb{R}$, if its density has the form:

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

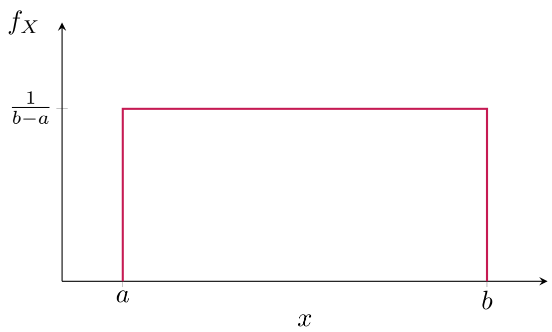
Notation: $X \sim \mathrm{Unif}(a,b)$, $X \sim \mathrm{U}(a,b)$.

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:

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



It is easy to compute that:

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

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

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

6.2 Exponential distribution

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

Definition 6.2. A random variable X has the *exponential* distribution with parameter $\lambda > 0$, if its density has the form:

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

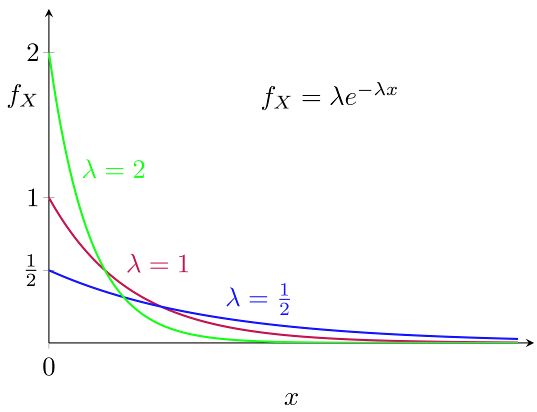
Notation: $X \sim \mathrm{Exp}(\lambda)$.

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:

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



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

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

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

✓ *Details during tutorials.*

6.3 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 6.3. A random variable X has the *normal* (Gaussian) distribution with parameters μ and $\sigma^2 > 0$, if the density has the form:

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

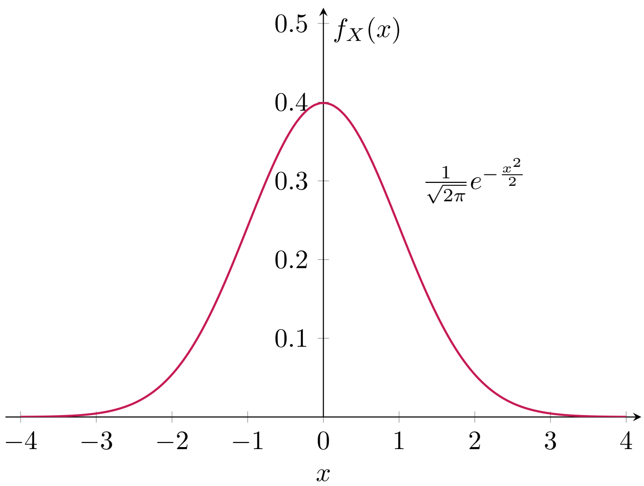
Notation: $X \sim \mathrm{N}(\mu, \sigma^2)$.

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

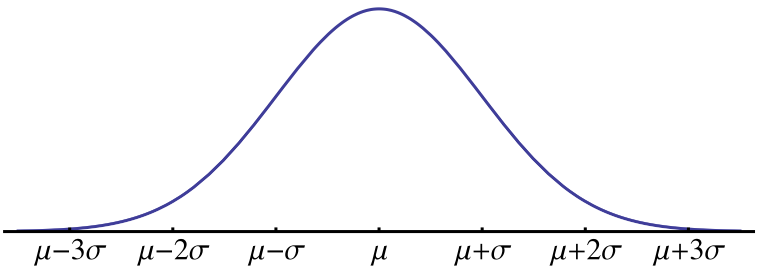
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

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



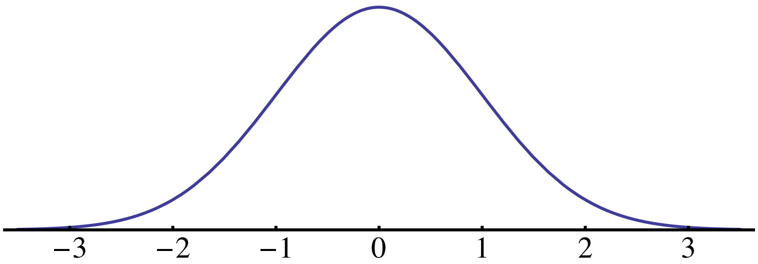
$$\Phi(-x) = 1 - \Phi(x)$$

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

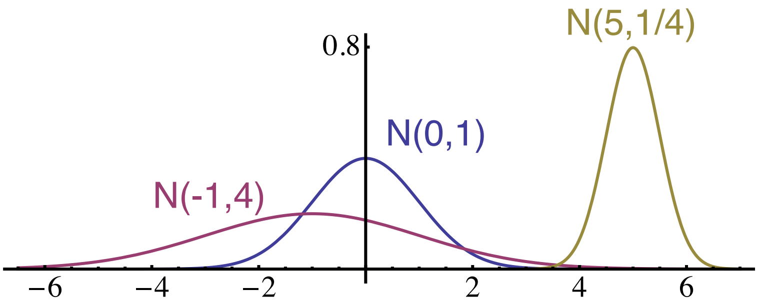


$$\mathrm{P}(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$$
$$\mathrm{P}(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$$
$$\mathrm{P}(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.997$$

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



Density of the normal distribution



For a normal distribution it is possible to compute the following quantities.

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

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

The computation uses substitution and is rather difficult.

6.3.1 Standardization of a random variable

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

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

- We subtract the expectation μ:

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

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

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

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.

6.3.2 Standardization of a normal random variable

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

Theorem 6.4. *Let a random variable X have the normal distribution $X \sim \mathrm{N}(\mu, \sigma^2)$. Then the random variable*

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

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

Proof.

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

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

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

□

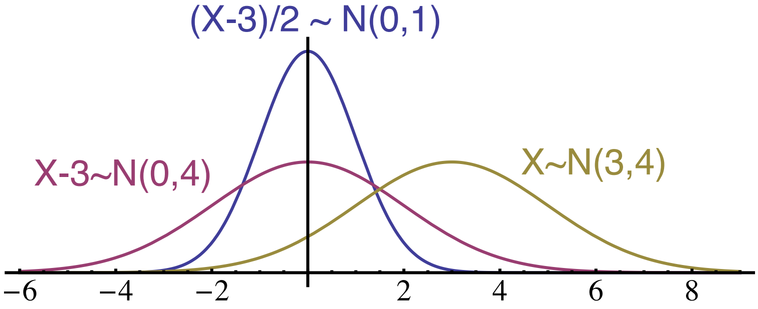
Remark 6.5. From the previous theorem it follows that:

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:

$$F_X(x) = \mathrm{P}(X \le x) = \mathrm{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right)$$

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


(X-3)/2 ~ N(0,1)
X-3~N(0,4)
X~N(3,4)
−6
−4
−2
2
4
6
8