BIE-PST – Probability and Statistics

Lecture 4: Random variables II.

Winter semester 2023/2024

Lecturer:
Francesco Dolce

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

Table of contents

4 Characteristics of random variables

4.1 Expected value

One of the important characteristics of a random variable is its *expected value*.

Definition 4.1. The *expected value* (or *expectation* or *mean value)* of a discrete random variable X with values $x_1, x_2, \ldots$, resp., of a continuous random variable X with density f_X, is given as

$$\mathrm{E}\,X = \sum_k x_k \,\mathrm{P}(X = x_k) \qquad \text{(discrete)}$$

resp., as

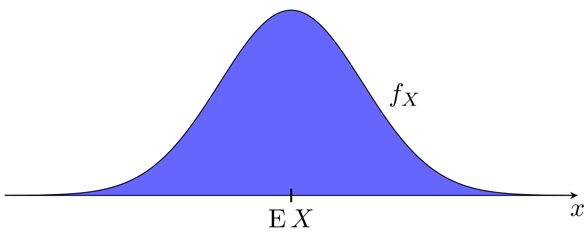
$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x, \qquad \text{(continuous)}$$

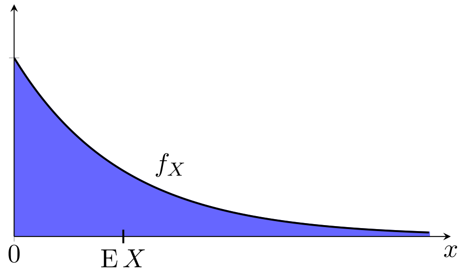
if the sum or the integral converges absolutely.

Note to the summation in the definition of the expected value of a discrete random variable: Thanks to the absolute convergence it does not depend on the order of summands in both series written above. (Generally for infinite series it does depend on the order of summands!) In the first series we sum over all possible values x_k of variable X without giving the order. It is often more explanatory than summing over all indexes k of values x_k ordered to some sequence. Similarly in the second series, instead of $\sum_k x_k \,\mathrm{P}(X = x_k)$ we write $\sum_{x:\mathrm{P}(X=x)>0} x\,\mathrm{P}(X = x)$. We know that $\mathrm{P}(X = x) > 0$ only for finite or countable many x and the order of summands is not important.

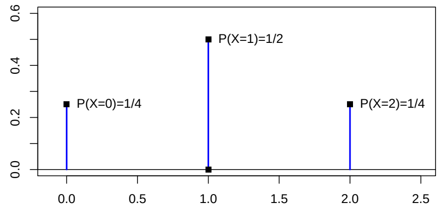
From the definition it follows that $\mathrm{E}\,X$ can be interpreted as the x coordinate of the *center of the mass* of the probability.

$\mathrm{E}\,X$ is taken as the expected value of the next experiment or as the weighted average (mean) or the center of mass of all possible values.





Example 4.2 (– tossing two coins)**.** Suppose we throw two balanced coins. Let X denote the number of Heads appearing. Find the expectation of X. There are four possible results, which are equally likely: $\Omega = \{TT, HT, TH, HH\}$. Therefore we can obtain 0, 1 or 2 Heads, with probabilities of 1/4, 1/2 and 1/4, respectively.



The expectation is then computed as the probability-weighted average of the possible values:

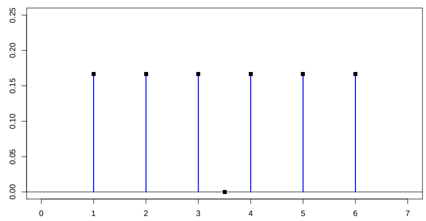
$$\mathrm{E}\,X = \sum_k x_k \,\mathrm{P}(X = x_k) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = \frac{1}{2} + \frac{2}{4} = 1.$$

Example 4.3 (– rolling a six-sided die)**.** Suppose we roll a balanced six-sided die one time. Let X denote the number of points rolled. What is the expectation of X?

k	1	2	3	4	5	6
$\mathrm{P}(X = k)$	1/6	1/6	1/6	1/6	1/6	1/6

The expectation is computed as the weighted average of possible results:

$$\mathrm{E}\,X = \sum_{k=1}^{6} k \cdot \mathrm{P}(X = k) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5.$$

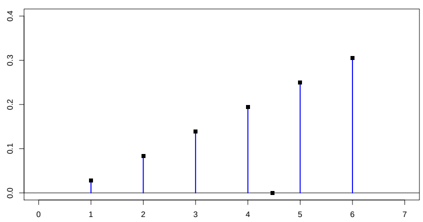


Example 4.4 (– rolling two six-sided dice)**.** Suppose we roll two balanced six-sided dice and keep the larger result of the two. Let X denote the number of points rolled, meaning $X = \max(\text{roll } 1, \text{roll } 2)$. What is the expectation of X?

k	1	2	3	4	5	6
$\mathrm{P}(X = k)$	1/36	3/36	5/36	7/36	9/36	11/36

The expectation is computed as the weighted average of possible results:

$$\mathrm{E}\,X = \sum_{k=1}^{6} k \cdot \mathrm{P}(X = k) = \frac{1 \cdot 1 + 2 \cdot 3 + 3 \cdot 5 + 4 \cdot 7 + 5 \cdot 9 + 6 \cdot 11}{36} = \frac{161}{36} \doteq 4.47.$$



Expected value of a function of a random variable

The expected value $\mathrm{E}(g(X))$ of a function of a random variable can be computed without determining the distribution of the random variable $Y = g(X)$.

Theorem 4.5. *Let X and $Y = g(X)$ for a given function g be random variables.*

i) If X has a discrete *distribution, then*

$$\mathrm{E}\,Y = \mathrm{E}\,g(X) = \sum_{\textit{all } x_k} g(x_k)\ \mathrm{P}(X = x_k),$$

under the assumption that the sum converges absolutely.

ii) If X has a continuous *distribution, then*

$$\mathrm{E}\,Y = \mathrm{E}\,g(X) = \int\limits_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x,$$

if the integral converges absolutely.

Proof. Suppose first that X is a discrete random variable. Denote the variable $Y = g(X)$ and its values $y_1, y_2, \ldots$. Then

$$\begin{aligned}
\mathrm{E}(g(X)) = \mathrm{E}\,Y &= \sum_{\text{all } y_j} y_j\ \mathrm{P}(Y = y_j) = \sum_{\text{all } y_j} y_j\ \mathrm{P}(g(X) = y_j) \\
&= \sum_{\text{all } y_j} \left(y_j \sum_{x_k : g(x_k) = y_j} \mathrm{P}(X = x_k) \right) = \sum_{\text{all } y_j} \sum_{x_k : g(x_k) = y_j} y_j\ \mathrm{P}(X = x_k) \\
&= \sum_{\text{all } y_j} \sum_{x_k : g(x_k) = y_j} g(x_k)\ \mathrm{P}(X = x_k) = \sum_{\text{all } x_k} g(x_k)\ \mathrm{P}(X = x_k).
\end{aligned}$$

The proof for continuous random variables is more difficult, we achieve it with the help of the following lemma only for function g taking non-negative values. □

Lemma 4.6. *If X is a non-negative random variable with the distribution function F, then*

$$\mathrm{E}\,X = \int\limits_0^{\infty} [1 - F(x)]\, \mathrm{d}x = \int\limits_0^{\infty} \mathrm{P}(X > x)\, \mathrm{d}x.$$

Proof . Suppose that X is a continuous random variable and the function g takes only non-negative values. Then

$$\begin{aligned}
\mathrm{E}(g(X)) = \mathrm{E}\,Y &= \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y = \int_0^\infty \mathrm{P}(g(X) > y)\,\mathrm{d}y \\
\text{see } (*) \qquad &= \int_0^\infty \left(\int_{\{x:\, g(x)>y\}} f_X(x)\,\mathrm{d}x \right) \mathrm{d}y = \iint_{\{(x,y):\, 0<y<g(x)\}} f_X(x)\,\mathrm{d}(x,y) \\
&= \int_{\{x:\, 0<g(x)\}} \left(\int_0^{g(x)} f_X(x)\,\mathrm{d}y \right) \mathrm{d}x \qquad (g(x) \text{ is non-negative}) \\
&= \int_{-\infty}^{\infty} f_X(x) \left(\int_0^{g(x)} \mathrm{d}y \right) \mathrm{d}x = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x.
\end{aligned}$$

$(*)$ We used $\mathrm{P}(X \in A) = \int_A f_X(x)\,\mathrm{d}x$ for $A = \{x\,!:\ g(x) > y\}$. If g is a general function we decompose it to its positive and negative parts which are both non-negative functions. Then we write $\mathrm{E}\,g(X) = \mathrm{E}\,Y = \mathrm{E}\,Y^+ - \mathrm{E}\,Y^- = \mathrm{E}\,g^+(X) - \mathrm{E}\,g^-(X)$ and use the above mentioned proof. □

Properties of the expected value

For *computation*, the following properties of the expected value are important. Notice that these properties hold for the expectation of both discrete and continuous random variables. More generally, these properties of expectation do not depend on the type of random variable – discrete, continuous or mixed.

Theorem 4.7. *The expected value of a random variable X has the following properties:*

i) If $X \geq 0$, then $\mathrm{E}(X) \geq 0$.

ii) If $a, b \in \mathbb{R}$, then $\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$ (if $\mathrm{E}\,X$ is finite).

iii) A constant random variable $X = c$ has expectation equal to the constant $\mathrm{E}(X) = c$.

Notes:

- Later we will prove that the expected value behaves as a linear operator (more precisely a linear functional) on a space of random variables. Now we are not familiar with random variables created as transformations of random vectors, thus we cannot handle variables $Z = aX + bY$.

These formulas can be used to simplify practical computing.

Proof.

i) For a discrete non-negative random variable X it holds that $x_k\,\mathrm{P}(X = x_k) \geq 0, \forall k$. Therefore $\mathrm{E}(X) = \sum_{\text{all } x_k} x_k\,\mathrm{P}(X = x_k) \geq 0$. For a continuous non-negative random variable X it holds that $f_X(x) = 0$ for $x < 0$. Therefore $\mathrm{E}(X) = \int_0^\infty x f_X(x)\,\mathrm{d}x \geq 0$.

ii) For a discrete random variable X it holds that

$$\begin{aligned} \mathrm{E}(aX+b) &= \sum_{\text{all } x_k} (ax_k+b)\ \mathrm{P}(X=x_k) \\ &= a \sum_{\text{all } x_k} x_k\ \mathrm{P}(X=x_k) + b \sum_{\text{all } x_k} \mathrm{P}(X=x_k) \\ &= a\ \mathrm{E}(X)+b. \end{aligned}$$

For a continuous random variable X the proof is similar.

iii) Consider $a=0$ in *ii)*.

□

4.2 Variance

Definition 4.8. The *variance* $\sigma^2 \equiv \operatorname{var} X$ of a random variable X is defined as

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

The *standard deviation* of a random variable X is defined as

$$\text{s.d. } X = \sqrt{\operatorname{var} X}.$$

The following properties of the variance are useful for *practical computations*:

Theorem 4.9. *For the variance it holds that:*

i) For all $a, b \in \mathbb{R}$ and a random variable X it holds that

$$\operatorname{var}(aX+b) = a^2 \operatorname{var} X.$$

ii) A constant random variable $X = c \in \mathbb{R}$ has zero variance ($\operatorname{var} c = 0$)*.*

Proof.

i) We just put $aX+b$ into the definition of var:

$$\begin{aligned} \operatorname{var}(aX+b) &= \mathrm{E}((aX+b) - \mathrm{E}(aX+b))^2 \\ &= \mathrm{E}(aX+b-a\,\mathrm{E}\,X-b)^2 \\ &= \mathrm{E}(aX-a\,\mathrm{E}\,X)^2 \\ &= \mathrm{E}(a^2(X-\mathrm{E}\,X)^2) \\ &= a^2\,\mathrm{E}(X-\mathrm{E}\,X)^2 \\ &= a^2 \operatorname{var} X. \end{aligned}$$

ii) $\operatorname{var} a = \mathrm{E}(a - \mathrm{E}\,a)^2 = \mathrm{E}(a-a)^2 = \mathrm{E}(0) = 0.$

□

While computing the variance it is often tedious to calculate the sum of values $(x_i - \mathrm{E}\,X)^2\,\mathrm{P}(X = x_i)$ or the integral of $(x - \mathrm{E}\,X)^2\; f_X(x)$.

We can use properties of the expectation to get a more useful formula:

$$\begin{aligned}\operatorname{var}(X) = \mathrm{E}((X - \mathrm{E}\,X)^2) &= \mathrm{E}\left(X^2 - 2X(\mathrm{E}\,X) + (\mathrm{E}\,X)^2\right)\\ &= \mathrm{E}(X^2) - \mathrm{E}(2X(\mathrm{E}\,X)) + \mathrm{E}((\mathrm{E}\,X)^2)\\ &= \mathrm{E}(X^2) - 2(\mathrm{E}\,X)(\mathrm{E}\,X) + (\mathrm{E}\,X)^2\\ &= \mathrm{E}(X^2) - (\mathrm{E}\,X)^2.\end{aligned}$$

Using only $\mathrm{E}\,X$ and $\mathrm{E}(X^2)$, which we often know or can be easily computed, we get the formula

$$\operatorname{var}(X) = \mathrm{E}((X - \mathrm{E}\,X)^2) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2$$

or simply

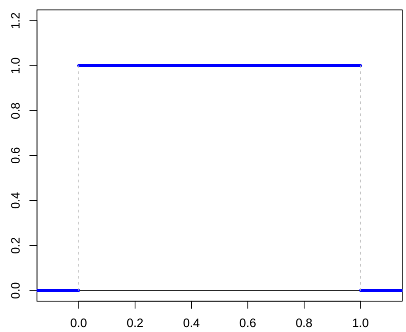
$$\operatorname{var} X = \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2.$$

Notice that $\operatorname{var}(X)$ is always non-negative (it is the expectation of a non-negative variable $(X - \mathrm{E}\,X)^2$). Therefore: $(\mathrm{E}\,X)^2 \leq \mathrm{E}(X^2)$.

Because the properties of the expectation are the same for discrete and continuous (even mixed) random variables, we can infer this way without specifying the type of the random variable.

Example 4.10 (– expectation and variance of the uniform distribution)**.** Suppose that Romeo arrives at the meeting point according to the uniform distribution with the density:

$$f_X(x) = \begin{cases} 1 & \text{for } x \in [0, 1]\\ 0 & \text{otherwise.}\end{cases}$$



What are the expectation and the variance of Romeo's arrival?

The expectation can be computed from the definition:

$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x 1 dx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}.$$

The expectation of the square is computed similarly:

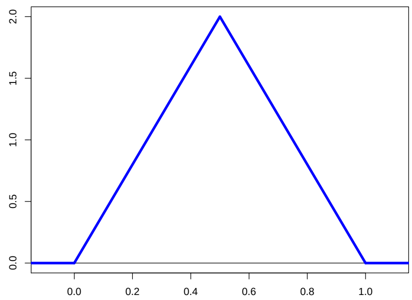
$$\mathrm{E}\,X^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 1 dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}.$$

The variance is obtained using the computational formula:

$$\operatorname{var} X = \operatorname{E} X^2 - (\operatorname{E} X)^2 = 1/3 - (1/2)^2 = 4/12 - 3/12 = 1/12.$$

Example 4.11 (– expectation and variance of a non-uniform distribution)**.** Suppose that Juliet arrives at the meeting point according to a non-uniform distribution with the density:

$$f_Y(y) = \begin{cases} 4y & \text{for } y \in [0, 1/2] \\ 4 - 4y & \text{for } y \in [1/2, 1] \\ 0 & \text{otherwise.} \end{cases}$$



What is the expectation and variance of Juliet's arrival? The expectation can be computed from the definition:

$$\operatorname{E} Y = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{0}^{1/2} y(4y) dy + \int_{1/2}^{1} y(4-4y) dy = \cdots = \frac{1}{2}.$$

The expectation of the square is computed similarly:

$$\operatorname{E} Y^2 = \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \int_{0}^{1/2} y^2(4y) dy + \int_{1/2}^{1} y^2(4-4y) dx = \cdots = \frac{7}{24}.$$

The variance is obtained using the computational formula:

$$\operatorname{var} Y = \operatorname{E} Y^2 - (\operatorname{E} Y)^2 = 7/24 - (1/2)^2 = 7/24 - 6/24 = 1/24.$$

The expectation is the same in both cases, but Romeo's arrivals have a twice larger variance than Juliet's.

Moments of random variables

Definition 4.12. For $k \in \mathbb{N}$ we define the *k-th moment* μ_k of a random variable X as

$$\mu_k = \operatorname{E}(X^k) = \begin{cases} \sum\limits_{\text{all } x_i} x_i^k \operatorname{P}(X = x_i) & \textit{discrete} \\ \int\limits_{-\infty}^{\infty} x^k f_X(x) \, \mathrm{d}x & \textit{continuous.} \end{cases}$$

Similarly, the *k-th central moment* σ_k is defined as

$$\sigma_k = \mathrm{E}((X-\mu_1)^k) = \begin{cases} \sum\limits_{\text{all } x_i} (x_i-\mu_1)^k \, \mathrm{P}(X=x_i) & \textit{discrete} \\ \int\limits_{-\infty}^{\infty} (x-\mu_1)^k \, f_X(x)\,\mathrm{d}x & \textit{continuous.} \end{cases}$$

Notation: usually we write $\mathrm{E}\,X^k$ instead of $\mathrm{E}(X^k)$ and $\mathrm{E}(X-\mu_1)^k$ instead of $\mathrm{E}((X-\mu_1)^k)$.

- Moments of a given random variable X do not always exist (if the corresponding sum or integral does not converge).
- $\mu_1 = \mathrm{E}\,X$ is the *expected value* of the variable X (often denoted as μ or μ_X).
- $\sigma_2 = \mathrm{E}(X-\mathrm{E}\,X)^2$ is the *variance* of the variable X denoted by $\mathrm{var}(X)$, $\mathrm{var}\,X$, σ^2 or σ_X^2.
- $\sigma = \sqrt{\mathrm{var}(X)}$ is the *standard deviation* of the variable X (possible notation σ_X).

Remark 4.13. Note that the variance is quadratic and therefore is measured in the units of X *squared*. The standard deviation is the square root of the variance and is therefore measured in the same units as X. This will be useful later.

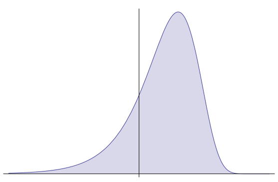
4.3 Skewness and Kurtosis

Skewness

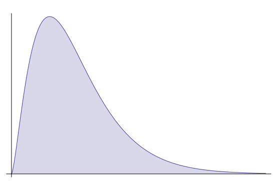
The *measure of asymmetry* around the mean is called *skewness*:

$$\gamma_1 = \frac{\sigma_3}{\sigma^3} = \frac{\mathrm{E}((X-\mathrm{E}(X))^3)}{(\mathrm{E}(X^2)-(\mathrm{E}\,X)^2)^{3/2}}.$$

Measure of asymmetry: for a unimodal density the coefficient γ_1 is *negative* if the *left* tail is longer and *positive* if the *right* tail is longer. It tells us to which side from the expected value is the bulk skewed:



$\gamma_1 = -1.14$



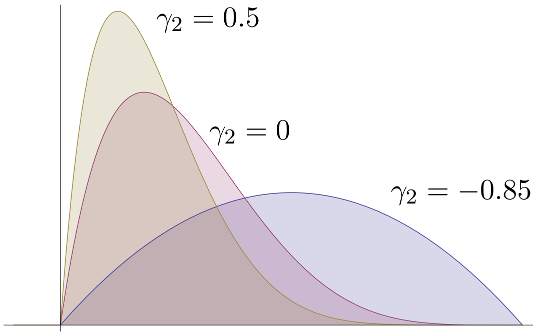
$\gamma_1 = 1.26$

Kurtosis

The *measure of "peakedness"* is called *(excess) kurtosis*:

$$\gamma_2 = \frac{\sigma_4}{\sigma^4} - 3 = \frac{\mathrm{E}((X - \mathrm{E}(X))^4)}{(\mathrm{E}(X^2) - (\mathrm{E}\,X)^2)^2} - 3.$$

This characteristics compares the shape ("peakedness") of the density with the normal distribution:



4.4 Quantiles

Quantiles

Important characteristics of random variables used especially in statistics are the *quantiles* and the *critical values*:

Definition 4.14. Let X be a random variable with distribution function F_X and let $\alpha \in (0, 1)$. The point q_α is called the α*-quantile* of the variable X if

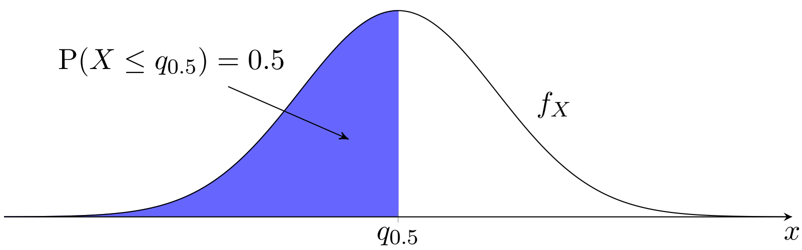
$$q_\alpha = \inf\{x | F_X(x) \geq \alpha\}.$$

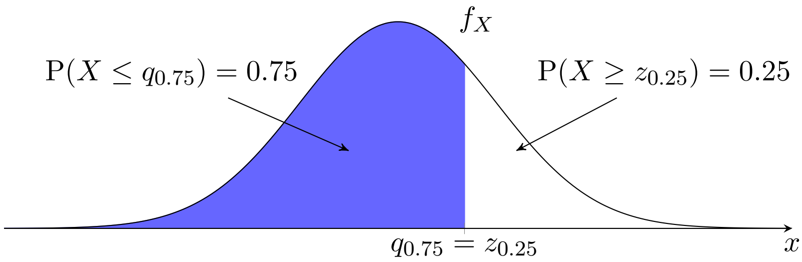
Quantiles treated as a function of α are called the *quantile function* and are denoted as $F_X^{-1}(\alpha)$.

The $(1 - \alpha)$-quantile is called the α*-critical value* of the variable X : $\quad c_\alpha = q_{1-\alpha}$.

- For F_X strictly increasing and continuous, the quantile function $F_X^{-1}(\alpha)$ is the inverse of F_X in the classic sense.
- For some particular distributions, special notation is used, e.g., the quantiles of the Gaussian distribution are denoted as u_α and the critical values as z_α.

Quantiles of the standard normal distribution





Example 4.15 (– quantiles of the uniform distribution)**.** Suppose that Romeo arrives at the meeting point according to the uniform distribution on the interval $[0, 1]$. Find the 5% and 95% quantiles of his arrival. The distribution function is found by integrating the density.

We are interested in the region, where the density is positive – the interval $[0, 1]$:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} 1dt = [t]_0^x = x.$$

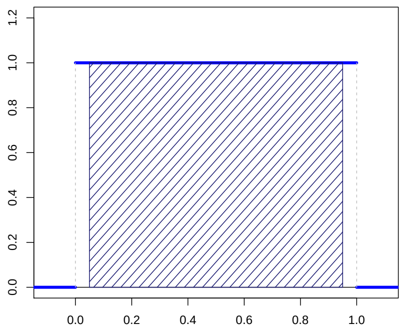
The distribution function is monotone, thus we can easily find the quantile function as its inverse:

$$F_X(q_\alpha) = \alpha \quad \Rightarrow \quad q_\alpha = \alpha \quad \Rightarrow \quad F_X^{-1}(\alpha) = \alpha.$$

Therefore the quantiles are:

$$q_{0.05} = 0.05 = 3 \text{ min.} \quad \text{and} \quad q_{0.95} = 0.95 = 57 \text{ min.}$$

With a 90% probability, Romeo arrives between the 3rd minute and the 57th minute.



Example 4.16 (– quantiles of a non-uniform distribution)**.** Suppose that Juliet arrives at the meeting point according to the non-uniform distribution with the triangular density from above. Find the 5% and 95% quantiles of her arrival. The distribution function is found by integrating the density. The observed interval has to be separated into two parts, because the function term is different. For $y \in [0, 1/2]$:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = \int_{0}^{y} 4tdt = \left[2t^2\right]_0^y = 2y^2.$$

For $y \in [1/2, 1]$:

$$F_Y(y) = \int_{0}^{1/2} 4tdt + \int_{1/2}^{y} (4-4t)dt = 1/2 + \left[4t - 2t^2\right]_{1/2}^y = 4y - 2y^2 - 1 = 1 - 2(y-1)^2.$$

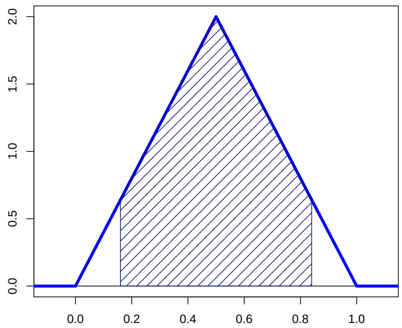
The quantile function is found as the inverse of the distribution function:

$$F_Y(q_{0.05}) = 0.05 \Leftrightarrow 2q_{0.05}^2 = 0.05 \Leftrightarrow q_{0.05} = \sqrt{0.05/2} \doteq 0.16 = 9.5 \text{ min.}$$

Similarly:

$$F_Y(q_{0.95}) = 0.95 \Leftrightarrow 1 - 2(q_{0.95} - 1)^2 = 0.95 \Leftrightarrow q_{0.95} = 1 - \sqrt{0.05/2} \doteq 0.84 = 50.5 \text{ min.}$$

With a 90% probability, Juliet arrives between the 9.5th minute and the 50.5th minute.



The central interval denoting the time, between which the person arrives with a 90% probability, is considerably shorter for Juliet than for Romeo. This is in accordance with Juliet's arrival having a smaller variance.

Important quantiles

Quantiles divide the population into groups according to probabilities. The important dividing points are called:

- $q_{0.5}$ – *median*,
- $q_{0.25}$ – *lower quartile*,
- $q_{0.75}$ – *upper quartile*.

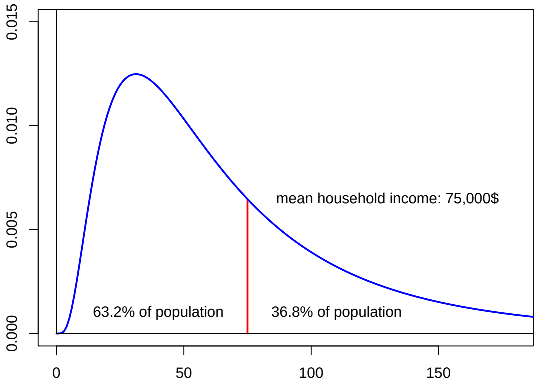
This quantiles can give us an overview of the variable in question:

- The *median* provides a measure of *location* as an alternative to the expectation.
- The *interquartile range* $q_{0.75} - q_{0.25}$ provides a measure of *dispersion* as an alternative to the variance.

The expectation can sometimes differ from the median significantly. Especially for one-sided heavy-tailed distributions.

Expectation vs. median

Example 4.17 (– U.S. household incomes)**.** According to the U.S Census Bureau, the mean yearly household income in 2014 was \$75, 000. But 63.2% of population had lower incomes. The median income was \$56, 000.



4.5 Moment generating function

Definition 4.18. The *moment generating function* of a random variable X is a function $M(s) = M_X(s)$ defined as

$$M(s) = \mathrm{E}(e^{sX}).$$

i.e., for a discrete or a continuous random variable X we have

$$M(s) = \sum_k e^{sk}\,\mathrm{P}(X = k), \qquad M(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)dx.$$

The generating function unambiguously determines the density f_X of the variable X (or the probabilities of its values). In fact the generating function is the *Laplace transformation* of the density. In particular, it allows us to easily compute the moments of the variable X.

Theorem 4.19. *For a random variable X with a generating function $M(s)$ it holds that:*

$$\mathrm{E}(X^n) = \frac{\mathrm{d}^n}{\mathrm{d}s^n} M(s)\big|_{s=0}.$$

Example 4.20 (– Poisson random variable)**.** $\mathrm{P}(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots$

$$M(s) = \sum_{k=0}^{\infty} e^{sk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{\lambda(e^s - 1)}.$$

We have:

$$\frac{d}{ds} e^{\lambda(e^s-1)} = \lambda e^s e^{\lambda(e^s-1)} \implies \mathrm{E}(X) = \lambda,$$

$$\frac{d^2}{ds^2} e^{\lambda(e^s-1)} = ((\lambda e^s)^2 + \lambda e^s) e^{\lambda(e^s-1)} \implies \mathrm{E}(X^2) = \lambda + \lambda^2.$$

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

Example 4.21 (– Exponential random variable)**.** $f_X(x) = \lambda e^{-\lambda x},\ x \geq 0$.

$$M(s) = \lambda \int_0^{\infty} e^{sx} e^{-\lambda x} \mathrm{d}x = \left[\lambda \frac{e^{(s-\lambda)x}}{s - \lambda}\right]_0^{\infty} = \frac{\lambda}{\lambda - s}.$$

Notice that $M(s)$ is well defined only for $s \in [0, \lambda)$. For $s \geq \lambda$ the integral diverges. Hence

$$\frac{d}{ds}\frac{\lambda}{\lambda - s} = \frac{\lambda}{(\lambda - s)^2} \quad \Longrightarrow \quad \mathrm{E}(X) = \frac{1}{\lambda},$$

$$\frac{d^2}{ds^2}\frac{\lambda}{\lambda - s} = \frac{2\lambda}{(\lambda - s)^3} \quad \Longrightarrow \quad \mathrm{E}(X^2) = \frac{2}{\lambda^2} \quad \text{and} \quad \mathrm{var}(X) = \frac{1}{\lambda^2}.$$

