Random variables II.

(Characteristics or random variables)

Lecturer:

Francesco Dolce

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

Probability and Statistics

BIE-PST, WS 2024/25, Lecture 4

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.
  - Random vectors, joint and marginal distributions, functions of random vectors, independence of random variables, conditional distribution, conditional expected value, covariance and correlation.
  - 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.

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.

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

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

Expected value

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

Definition

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

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

resp., as

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

if the sum or the integral converges absolutely.

Expected value

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

Definition

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

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

resp., as

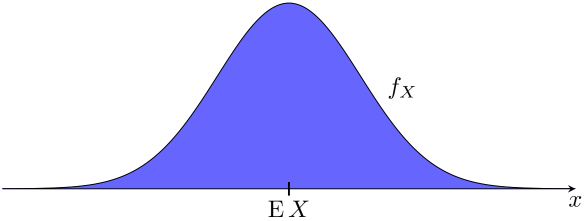
$$\mathrm{E}\, X = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x, \quad \text{(continuous)}$$

if the sum or the integral converges absolutely.

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.

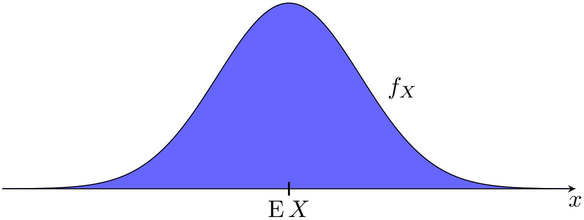
Visualization of the expectation

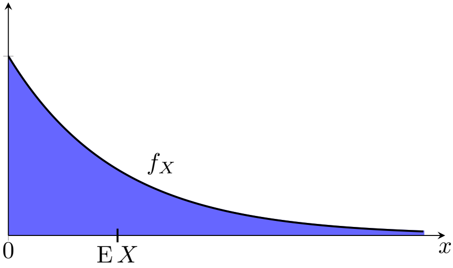
$\mathrm{E}\, X$ is taken as the expected value of the next experiment or as the weighted average (mean) or the center of mass of all possible values.



Visualization of the expectation

$\mathrm{E}\,X$ 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 of the computation of the expectation

Example – tossing two coins

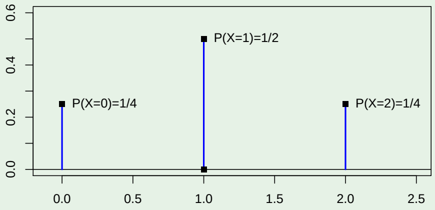
Suppose we throw two balanced coins. Let X denote the number of Heads appearing. Find the expectation of X.

Example of the computation of the expectation

Example – 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: $\Omega = \{\mathtt{TT}, \mathtt{HT}, \mathtt{TH}, \mathtt{HH}\}$. Therefore we can obtain 0, 1 or 2 Heads, with probabilities of $1/4$, $1/2$ and $1/4$, respectively.

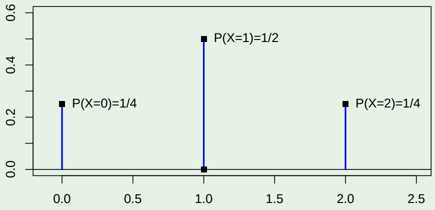


Example of the computation of the expectation

Example – 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: $\Omega = \{\mathtt{TT}, \mathtt{HT}, \mathtt{TH}, \mathtt{HH}\}$. 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 – discrete uniform disribution

Example – 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?

Example – discrete uniform disribution

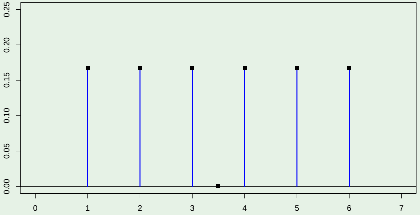
Example – rolling a six-sided die

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 – discrete non-uniform disribution

Example – 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?

Example – discrete non-uniform disribution

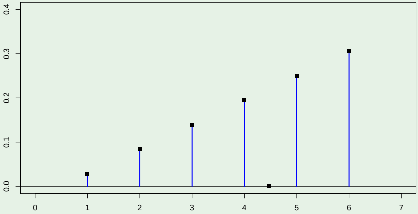
Example – rolling two six-sided dice

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 $Y = g(X)$.

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

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

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

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

under the assumption that the sum converges absolutely.

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

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

if the integral converges absolutely.

Expected value of the function of a random variable

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

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

$$\mathrm{E}\,X = \int_0^\infty [1 - F(x)]\,\mathrm{d}x = \int_0^\infty \mathrm{P}(X > x)\,\mathrm{d}x.$$

Expected value of the function of a random variable

Proof

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

$$\mathrm{E}(g(X)) = \mathrm{E}\,Y = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y$$

□

Expected value of the function of a random variable

Proof

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

$$\mathrm{E}(g(X)) = \mathrm{E}\,Y = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y = \int_0^\infty \mathrm{P}(g(X) > y)\,\mathrm{d}y$$

□

Expected value of the function of a random variable

Proof

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

$$\begin{aligned} \mathrm{E}(g(X)) = \mathrm{E}\,Y &= \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y = \int_0^\infty \mathrm{P}(g(X) > y)\,\mathrm{d}y \\ &= \int_0^\infty \left(\int_{\{x:\, g(x) > y\}} f_X(x)\,\mathrm{d}x \right) \mathrm{d}y \end{aligned}$$

□

Expected value of the function of a random variable

Proof

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

$$\mathrm{E}(g(X)) = \mathrm{E}\,Y = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y = \int_0^\infty \mathrm{P}(g(X) > y)\,\mathrm{d}y$$

see $(*)$ $$= \int_0^\infty \left(\int_{\{x:\, g(x)>y\}} f_X(x)\,\mathrm{d}x \right) \mathrm{d}y = \iint_{\{(x,y):\, 0<y<g(x)\}} f_X(x)\,\mathrm{d}(x,y)$$

$(*)$ We used $\mathrm{P}(X \in A) = \int_A f_X(x)\,\mathrm{d}x$ for $A = \{x \,:\, g(x) > y\}$.

□

Expected value of the function of a random variable

Proof

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

$$
\begin{aligned}
\mathrm{E}(g(X)) = \mathrm{E}\,Y &= \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y = \int_0^\infty \mathrm{P}(g(X) > y)\,\mathrm{d}y \\
&= \int_0^\infty \left(\int_{\{x:\, g(x) > y\}} f_X(x)\,\mathrm{d}x \right) \mathrm{d}y = \iint_{\{(x,y):\, 0<y<g(x)\}} f_X(x)\,\mathrm{d}(x,y) \\
&= \int_{\{x:\, 0<g(x)\}} \left(\int_0^{g(x)} f_X(x)\,\mathrm{d}y \right) \mathrm{d}x
\end{aligned}
$$

□

Expected value of the function of a random variable

Proof

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

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

□

Expected value of the function of a random variable

Proof

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

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

□

Expected value of the function of a random variable

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.

Theorem

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 $\mathrm{E}(X) = c$.*

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.

Theorem

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:

- These properties of expectation do not depend on the type of random variable – discrete, continuous or mixed.
- For discrete, continuous or mixed random variables X and Y with finite expectations it holds that (we will prove it later)

$$\mathrm{E}(aX + bY) = a\ \mathrm{E}\, X + b\ \mathrm{E}\, Y, \qquad \forall\, a, b \in \mathbb{R}.$$

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.

Theorem

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:

- These properties of expectation do not depend on the type of random variable – discrete, continuous or mixed.
- For discrete, continuous or mixed random variables X and Y with finite expectations it holds that (we will prove it later)

$$\mathrm{E}(aX + bY) = a\,\mathrm{E}\,X + b\,\mathrm{E}\,Y, \qquad \forall\, a, b \in \mathbb{R}.$$

These formulas can be used to simplify practical computing.

Properties of the expected value

Proof

i) For a discrete non-negative random variable X it holds that $x_k \operatorname{P}(X = x_k) \geq 0, \forall\, k$. Therefore $\operatorname{E}(X) = \sum_{\text{all } x_k} x_k \operatorname{P}(X = x_k) \geq 0$.

For a continuous non-negative random variable X it holds that $f_X(x) = 0$ for $x < 0$. Therefore $\operatorname{E}(X) = \int_0^{\infty} x f_X(x)\,\mathrm{d}x \geq 0$.

ii) For a discrete random variable X it holds that

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

For a continuous random variable X the proof is similar.

iii) Consider $a = 0$ in **ii)**. □

Variance

Definition

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

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

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

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

Variance

Definition

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

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

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

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 ($\operatorname{var} c = 0$).*

Variance

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:

Variance

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:

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

We get the formula: $\mathrm{var}(X) = \mathrm{E}((X - \mathrm{E}\,X)^2) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2$
or simply: $\mathrm{var}\,X = \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2$.

Variance

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:

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

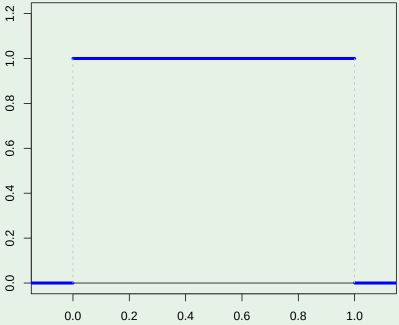
We get the formula: $\mathrm{var}(X) = \mathrm{E}((X - \mathrm{E}\,X)^2) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2$
or simply: $\mathrm{var}\,X = \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2$.
Notice that $\mathrm{var}(X)$ is always non-negative (it is the expectation of a non-negative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

Romeo, Juliet and the expectation and variance

Example – 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?

Romeo, Juliet and the expectation and variance

Example – expectation and variance of the uniform distribution

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

Romeo, Juliet and the expectation and variance

Example – expectation and variance of the uniform distribution

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

The expectation can be computed from the definition:

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

Romeo, Juliet and the expectation and variance

Example – expectation and variance of the uniform distribution

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:

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

Romeo, Juliet and the expectation and variance

Example – expectation and variance of the uniform distribution

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:

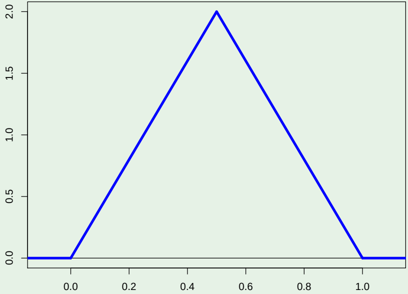
$$\mathrm{var}\,X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = 1/3 - (1/2)^2 = 4/12 - 3/12 = 1/12.$$

Romeo, Juliet and the expectation and variance

Example – 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 are the expectation and the variance of Juliet's arrival?

Romeo, Juliet and the expectation and variance

Example – expectation and variance of a non-uniform distribution

What is the expectation and variance of Juliet's arrival?

Romeo, Juliet and the expectation and variance

Example – expectation and variance of a non-uniform distribution

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

$$\mathrm{E}\,Y = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^{1/2} y(4y)dy + \int_{1/2}^{1} y(4-4y)dy = \cdots = \frac{1}{2}.$$

Romeo, Juliet and the expectation and variance

Example – expectation and variance of a non-uniform distribution

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

$$\mathrm{E}\,Y = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^{1/2} y(4y)dy + \int_{1/2}^{1} y(4-4y)dy = \cdots = \frac{1}{2}.$$

The expectation of the square is computed similarly:

$$\mathrm{E}\,Y^2 = \int_{-\infty}^{\infty} y^2 f_Y(y)dy = \int_0^{1/2} y^2(4y)dy + \int_{1/2}^{1} y^2(4-4y)dx = \cdots = \frac{7}{24}.$$

The variance is obtained using the computational formula:

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

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

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

$$\mu_k = \mathrm{E}(X^k) = \begin{cases} \displaystyle\sum_{\text{all } x_i} x_i^k \; \mathrm{P}(X = x_i) & \text{discrete} \\ \displaystyle\int_{-\infty}^{\infty} x^k \, f_X(x) \, \mathrm{d}x & \text{continuous.} \end{cases}$$

Moments of random variables

Definition

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

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

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

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

Moments of random variables

Definition

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

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

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

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

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, expectation, variance, standard deviation

- Moments of a given random variable X do not always exist (if the corresponding sum or integral does not converge).

Moments, expectation, variance, standard deviation

- 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).

Moments, expectation, variance, standard deviation

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

Moments, expectation, variance, standard deviation

- 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^2_X.
- $\sigma = \sqrt{\mathrm{var}(X)}$ is the standard deviation of the variable X (possible notation σ_X).

Moments, expectation, variance, standard deviation

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

Remark

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.

Skewness

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

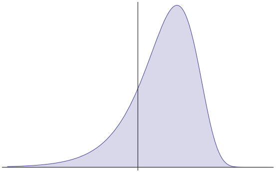
$$\gamma_1 = \frac{\sigma_3}{\sigma^3} = \frac{\mathrm{E}((X - \mathrm{E}(X))^3)}{(\mathrm{E}(X^2) - (\mathrm{E}\,X)^2)^{3/2}}.$$

Skewness

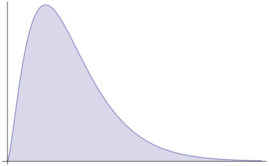
The measure of asymmetry around the mean is called **skewness**:

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

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**:

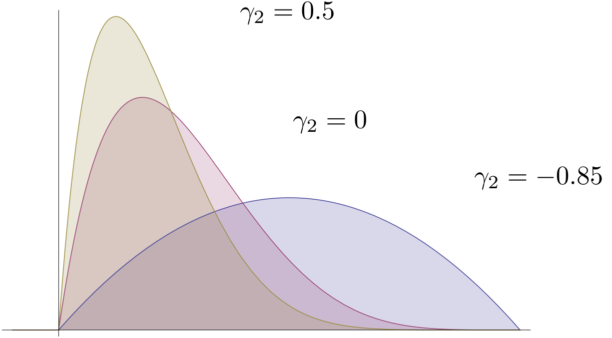
$$\gamma_2 = \frac{\sigma_4}{\sigma^4} - 3 = \frac{\mathrm{E}((X - \mathrm{E}(X))^4)}{(\mathrm{E}(X^2) - (\mathrm{E}\, X)^2)^2} - 3.$$

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:



Moment generating function

Definition

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

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

Moment generating function

Definition

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

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

Moment generating function

Definition

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.

Moment generating function

Definition

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

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

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

Examples of generating functions

Example – Poisson random variable

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

Examples of generating functions

Example – Poisson random variable

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

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

Examples of generating functions

Example – Poisson random variable

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

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

We have:

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

Examples of generating functions

Example – Poisson random variable

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

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

We have:

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

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

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

Examples of generating functions

Example – Exponential random variable

$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0.$

Examples of generating functions

Example – Exponential random variable

$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0.$

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

Examples of generating functions

Example – 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.

Examples of generating functions

Example – 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

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

Examples of generating functions

Example – 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

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

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

Quantile function

The distribution function gives us the probability that the random variable in question will be less than or equal to x. Sometimes we are interested in a reverse approach – for a given probability α, find such x, so that $\mathrm{P}(X \leq x) = \alpha$.

Quantile function

The distribution function gives us the probability that the random variable in question will be less than or equal to x. Sometimes we are interested in a reverse approach – for a given probability α, find such x, so that $\mathrm{P}(X \leq x) = \alpha$.

Definition

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 : $\quad c_\alpha = q_{1-\alpha}$.

Quantile function

The distribution function gives us the probability that the random variable in question will be less than or equal to x. Sometimes we are interested in a reverse approach – for a given probability α, find such x, so that $\mathrm{P}(X \leq x) = \alpha$.

Definition

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 : $c_\alpha = q_{1-\alpha}$.

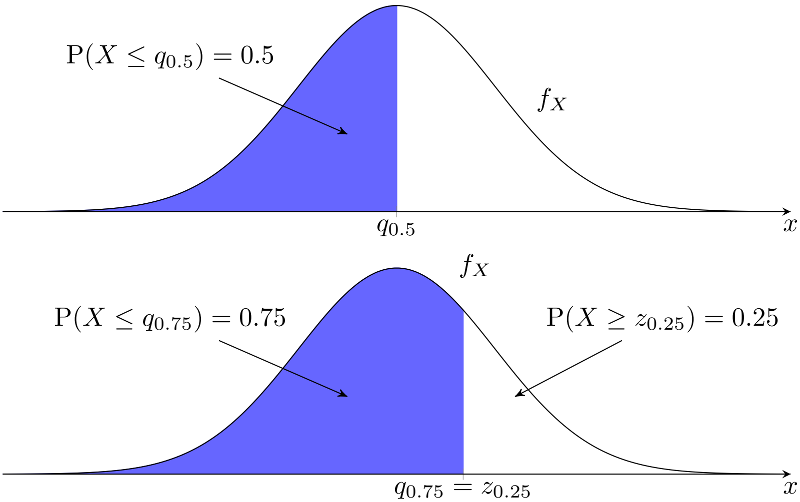
For F_X strictly increasing and continuous, q_α is the point for which it holds that

$$F_X(q_\alpha) = \mathrm{P}(X \leq q_\alpha) = \alpha,$$

thus the notation F_X^{-1} denotes the actual inverse of F_X.

Quantiles of the standard normal distribution

For some particular distributions, special notation is used, e.g., the quantiles of the Gaussian distribution (see later) are denoted as u_α and the critical values as z_α.



Romeo, Juliet and quantiles

Example – 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.

Romeo, Juliet and quantiles

Example – 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 $[0, 1]$:

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

Romeo, Juliet and quantiles

Example – 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:

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

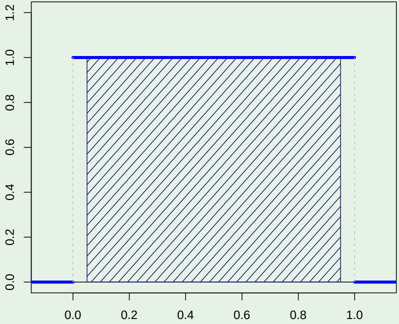
Therefore the quantiles are:

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

Romeo, Juliet and quantiles

Example – quantiles of the uniform distribution

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



Romeo, Juliet and quantiles

Example – 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.

Romeo, Juliet and quantiles

Example – 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 $y \in [1/2, 1]$:

$$F_Y(y) = \int_0^{1/2} 4tdt + \int_{1/2}^{y} (4-4t)dt = 1/2 + \left[4t - 2t^2\right]_{1/2}^{y} = 4y - 2y^2 - 1 = 1 - 2(y-1)^2.$$

Romeo, Juliet and quantiles

Example – 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:

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

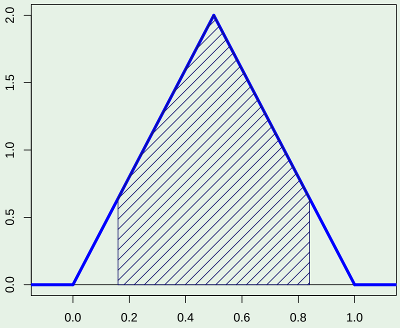
Similarly:

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

Romeo, Juliet and quantiles

Example – quantiles of a non-uniform distribution

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**.

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:

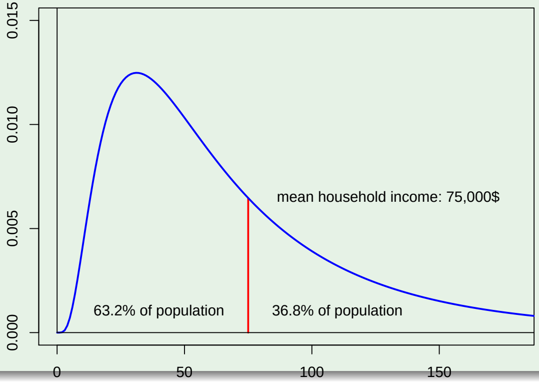
- The **median** provides a measure of **location** as an alternative to the expectation.
- The **interquartile range** $q_{0.75} - q_{0.25}$ provides a measure of **dispersion** as an alternative to the variance.

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

Expectation vs. median

Example – 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 $\$56,000$.



Quantile function – random number generation

Theorem

Suppose that X has a distribution with a distribution function F_X. Suppose that U has a uniform distribution on the interval $[0,1]$, meaning that

$$f_U(u) = \begin{cases} 1 & \text{for } u \in (0,1) \\ 0 & \text{elsewhere.} \end{cases}$$

Then the random variable $F_X^{-1}(U)$ has the same distribution as X.

Proof

For a continuous F_X:

$$\mathrm{P}(F_X^{-1}(U) \le x) = \mathrm{P}(U \le F_X(x)) = \int_0^{F_X(x)} 1 \cdot du = F_X(x).$$

□

This way, we can generate values from any distribution by generating values from the uniform distribution $U(0,1)$ and finding the corresponding quantiles.

Generating uniform random numbers

Truly random numbers can be generated by measuring physical phenomena, such as using oscillators or thermal devices.

Generating uniform random numbers

Truly random numbers can be generated by measuring physical phenomena, such as using oscillators or thermal devices.

Computer algorithms can only produce **pseudo-random numbers**, which try to appear as truly random. There are many ways to generate pseudo-random numbers.

Generating uniform random numbers

Truly random numbers can be generated by measuring physical phenomena, such as using oscillators or thermal devices.

Computer algorithms can only produce **pseudo-random numbers**, which try to appear as truly random. There are many ways to generate pseudo-random numbers.

Congruent generators (fast and easy to implement):

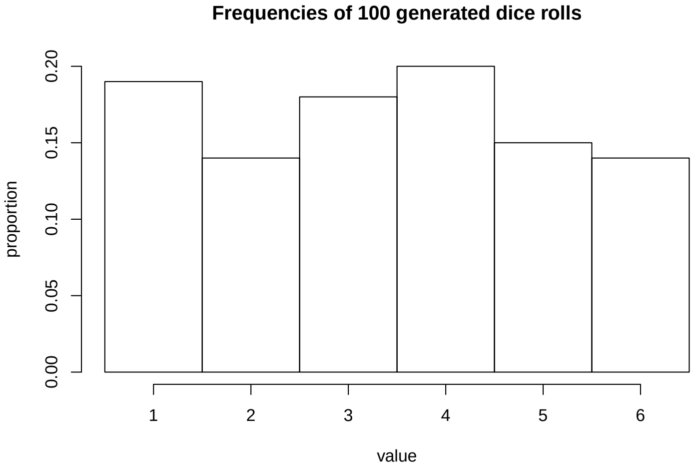
- select large integers a, b and m;
- choose a starting value X_0;
- generate a sequence $X_{n+1} = (aX_n + b)$ mod m;
- divide all results by m.

More sophisticated generators (used in R, Matlab, etc):

- Mersenne Twister
- Wichmann-Hill
- many others (see literature).

Generating dice rolls

When rolling a six-sided dice, we easily find out that $F_X^{-1}(U) = \lceil 6 \cdot U \rceil$. We generated 100 random dice rolls and counted the percentage of each outcome:



Recap

The **expectation** of a random variable X gives us its **center of mass** or the **expected average outcome**.

- For discrete random variables it is the average of its possible values weighted by their probabilities:
$$\mathrm{E}\,X = \sum_k x_k\,\mathrm{P}(X = x_k).$$
- For continuous random variables it is the integral average of its possible values weighted by the density:
$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x.$$

Recap

The **expectation** of a random variable X gives us its **center of mass** or the **expected average outcome**.

- For discrete random variables it is the average of its possible values weighted by their probabilities:

$$\mathrm{E}\,X = \sum_k x_k \,\mathrm{P}(X = x_k).$$

- For continuous random variables it is the integral average of its possible values weighted by the density:

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

The **variance** of a random variable X gives us the expected quadratic distance of the random variable from its expectation. It is defined as

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

and can be computed as:

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