BIE-PST – Probability and Statistics

Lecture 3: Random variables I.

Winter semester 2023/2024

Lecturer:
Francesco Dolce

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

Table of contents

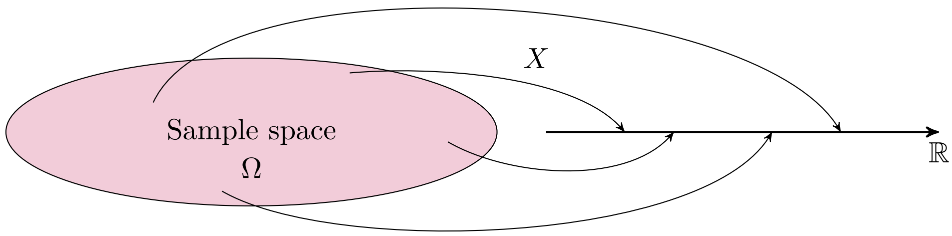
3 Random variables

3.1 Definition of a random variable, distribution function

Random variable

For a mathematical processing of a random experiment it is often useful to assign a number to each outcome ω. By this assignment we choose the part of information which is interesting from our point of view.

Such assignment can be established in many ways and will be called a *random variable*. For example, many gamblers are more concerned with their wins and losses than with the games which gives rise to them.

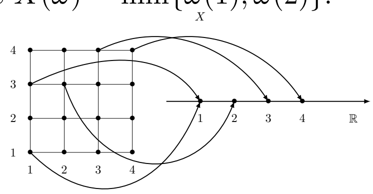


Examples 3.2.

- Number of Heads while tossing a coin: $X(\text{Heads}) = 1, X(\text{Tails}) = 0$.
- Number of winnings in the game with: $X(\text{Heads}) = 1, X(\text{Tails}) = -1$.
- How much a player won in a given game at a poker tournament.
- The highest rolled value or n rolls of a die.
- The height of a randomly chosen person.

Example 3.3 (– minimum of two rolls of a 4-sided die)**.** Two rolls of a 4-sided die. $\Omega = \{1, 2, 3, 4\}^2$.

Consider a random variable $X(\omega) = \min\{\omega(1), \omega(2)\}$:



$$
\begin{aligned}
\mathrm{P}(X = 1) &= \mathrm{P}(\{\omega | X(\omega) = 1\}) \\
&= \mathrm{P}(\{(1,1), (1,2), (1,3), (1,4), (2,1), (3,1), (4,1)\}) = \frac{7}{16}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{P}(X = 2) &= \mathrm{P}(\{(2,2), (2,3), (2,4), (3,2), (4,2)\}) = \frac{5}{16}, \\
\mathrm{P}(X = 3) &= \mathrm{P}(\{(3,3), (3,4), (4,3)\}) = \frac{3}{16}, \\
\mathrm{P}(X = 4) &= \mathrm{P}(\{(4,4)\}) = \frac{1}{16}.
\end{aligned}
$$

Random variable and its distribution function

Definition 3.4. A *random variable* X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a function $X\colon \Omega \to \mathbb{R}$, assigning to each outcome $\omega \in \Omega$ a number $X(\omega)$, with the property that:

$$\{X \leq x\} \in \mathcal{F}, \ \forall x \in \mathbb{R}.$$

Such a function is said to be *$\mathcal{F}$-measurable*.

Notes:

- In more details by means of pre-image of the set $X^{-1}(\cdot)$ we can write $\{X \leq x\} = \{X \in (-\infty, x]\} = X^{-1}((-\infty, x]) = \{\omega \in \Omega\colon X(\omega) \in (-\infty, x]\} = \{\omega \in \Omega\colon X(\omega) \leq x\}$.
- The measurability property in fact tells us that $\{X \leq x\}$ is **an event** and allows us to compute $\mathrm{P}(X \leq x)$, $\mathrm{P}(X = x)$, $\mathrm{P}(X \in (a, b))$, etc.
- This condition must be met, but in practice we never verify it.

The *probability distribution* of a random variable is given by its distribution function:

Definition 3.5. The *distribution function* of a random variable X is a function $F\colon \mathbb{R} \to [0, 1]$ defined as

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

There are various types of random variables.

- Some can take only isolated values (e.g., 0 or 1 for Heads and Tails of a coin toss, $1, \ldots, 6$ for a die roll).
- Some can take values from a continuous interval (e.g., weight of a newborn, time spent waiting for a bus, ...).

This divides the variables into *discrete* and *continuous*. For discrete random variables, we will be interested in probabilities of the singular values, whereas for continuous we will work with probabilities of intervals. Regardless of the type, the distribution function gives us a full description of the random variable. For any real number x, we can answer the question: "what is the probability that the random variable will be less than or equal to x"? This allows us to answer questions about any equalities and inequalities.

Properties of the distribution function

Theorem 3.6. *The distribution function F of a random variable X has following properties:*

i) F is non-decreasing: *if* $x < y$, *then* $F(x) \leq F(y)$

ii) F "starts at 0 and ends at 1": $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$

iii) F is right continuous: $\lim_{y \to x^+} F(y) = F(x)$

Proof. i) Recall the notation $\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$. Consider the disjoint partition

$$\{X \leq y\} = \{X \leq x\} \cup \{x < X \leq y\},$$

therefore $F(y) = \mathrm{P}(X \leq y) = \mathrm{P}(X \leq x) + \mathrm{P}(x < X \leq y) \quad \geq \quad \mathrm{P}(X \leq x) = F(x)$.

ii) For simplicity we only sketch the proof by means of a sequence of events $B_n = \{X \leq -n\}$. For $n \to \infty$ it is decreasing in the sense of inclusion with the intersection equal to $\emptyset$, i.e., $B_n \searrow \emptyset$. From the continuity of probability theorem we have $\mathrm{P}(B_n) \to \mathrm{P}(\emptyset) = 0$. For the proof of the second statement it is enough to consider a sequence $A_n = \{X \leq n\} \nearrow \Omega$ and from the same theorem we have $\mathrm{P}(A_n) \to \mathrm{P}(\Omega) = 1$.

iii) Similarly as *ii)* (see bibliography).

□

By means of the distribution function it is possible to express some important properties.

Lemma 3.7. *Let F be a distribution function of a random variable X, then it holds that:*

i) $\mathrm{P}(X > x) = 1 - F(x)$,

ii) $\mathrm{P}(X \in (x, y]) = \mathrm{P}(x < X \leq y) = F(y) - F(x)$,

iii) $\mathrm{P}(X < x) = \lim_{y \to x^-} F(y)$,

iv) $\mathrm{P}(X = x) = F(x) - \lim_{y \to x^-} F(y)$.

Proof. i) $\Omega = \{X > x\} \cup \{X \leq x\}$ is a disjoint partition. Therefore $\mathrm{P}(\{X > x\}) = \mathrm{P}(\{X \leq x\}^c)$.

ii) See proof of *i)* of the previous theorem.

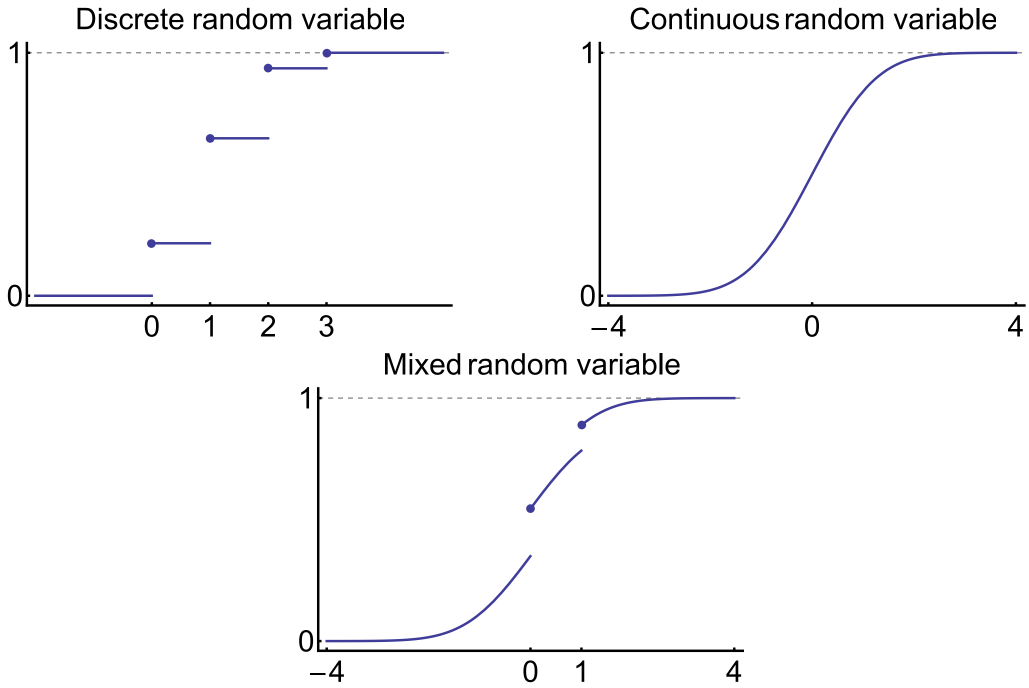
iii) See bibliography. Idea of the proof using a non-decreasing sequence and continuity of probability:

$$\{X \leq x - 1/n\} \nearrow \{X < x\} \qquad \Rightarrow \qquad F(x - 1/n) = \mathrm{P}(X \leq x - 1/n) \to \mathrm{P}(X < x).$$

iv) $\{X \leq x\} = \{X < x\} \cup \{X = x\}$ is a disjoint partition. Therefore $\mathrm{P}(X = x) = \mathrm{P}(X \leq x) - \mathrm{P}(X < x)$.

□

Types of random variables and their distribution functions



3.2 Discrete random variables

Definition 3.8. A random variable X is called *discrete* if it takes only values from some countable set $\{x_1, x_2, \dots\}$.

Probabilities of the values of a discrete random variable X are given by

$$\mathrm{P}(X = x_k), \qquad k = 1, 2, \dots$$

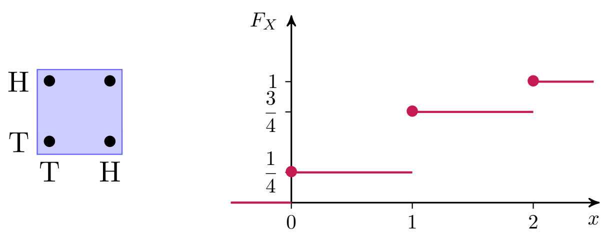
The probabilities $\mathrm{P}(X = x_k)$ can be viewed as a function of x and are sometimes called a *probability function*, or a *probability mass function* or a *discrete density* of the variable X.

The *distribution function* of a discrete random variable has the form

$$F_X(x) = \mathrm{P}(X \leq x) = \sum_{\text{all } x_k \leq x} \mathrm{P}(X = x_k).$$

From this it follows that $F_X(x)$ has jumps at points x_k and it is constant elsewhere. The size of the jump at point x_k is equal to $\mathrm{P}(X = x_k)$.

Example 3.9 (– toss with two coins)**.** The sample space is $\Omega = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\}$. Let the random variable X give the number of Heads. The distribution function is $F_X(x) = \mathrm{P}(X \leq x)$:

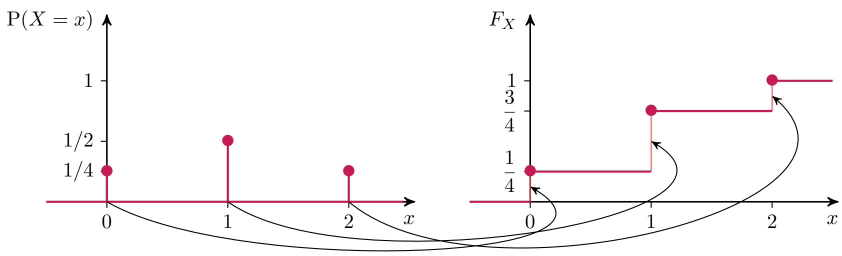


The distribution function $F_X = \mathrm{P}(X \le x)$ is given by

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 & \mathrm{P}(\emptyset) \\ 1/4 & \text{for } 0 \le x < 1 & \mathrm{P}(\{(\mathrm{T},\mathrm{T})\}) \\ 3/4 & \text{for } 1 \le x < 2 & \mathrm{P}(\{(\mathrm{T},\mathrm{T}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H})\}) \\ 1 & \text{for } 2 \le x & \mathrm{P}(\Omega). \end{cases}$$

Example 3.10 (– toss with two coins)**.** The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads.

Draw the probabilities of the values and the distribution function.



When assigning probabilities to the values x_k, the *normalization condition* must hold:

$$\sum_{\text{all } x_k} \mathrm{P}(X = x_k) = 1.$$

Generally, for computing $\mathrm{P}(X \in B)$, with $B \subset \mathbb{R}$, it is enough to know the probabilities of the possible values X: $\mathrm{P}(\{X \in B\}) = \sum_{x_k \in B} \mathrm{P}(X = x_k)$.

A series with non-negative elements does not depend on the order of summands. Therefore the series can be summed over all possible values of random variable X without giving the exact ordering. Notice moreover, that $p_X(x) > 0$ only for a finite or countable number of points x.

The *distribution of* $\boldsymbol{X}$ can be equivalently given by F_X or by the probabilities. Considering that $\mathrm{P}(X = x_k) = F_X(x_k) - F_X(x_{k-1})$ (we are considering an increasing ordering $x_1 < x_2 < x_3 < \dots$), the knowledge of the distribution function is equivalent to the knowledge of the probabilities $\mathrm{P}(X = x_k)$.

Computation of the probabilities $\mathrm{P}(X = x_k)$:
Collect all ω for which $X(\omega) = x$ and sum their probabilities.

Computation of the distribution function $F_X(x) = \mathrm{P}(X \le x_k)$:
Collect all ω for which $X(\omega) \le x$ and sum their probabilities.

Remark 3.11. A random variable X can be discrete even if the sample space itself is not discrete.

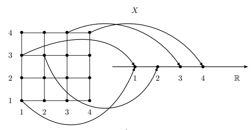
Example 3.12. Let us throw darts at a target $T \subset \mathbb{R}^2$.

The target can be divided into parts (often concentric annulus), denoted as T_1, T_2, T_3, T_4, T_5.

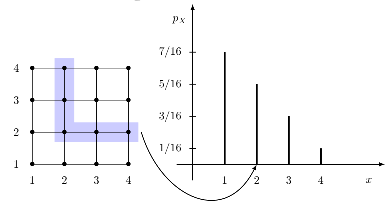
We can consider a discrete random variable X denoting the points obtained from one throw, for example

$$X(\omega) = \begin{cases} 10 & \text{for } \omega \in T_5 \\ 5 & \text{for } \omega \in T_4 \\ i & \text{for } \omega \in T_i,\ i = 1, 2, 3 \end{cases}$$

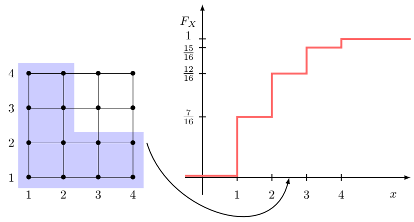
Example 3.13 (– minimum of two rolls of a 4-sided die (continuation))**.** $X = \min\{1^{\text{st}} \text{ roll}, 2^{\text{nd}} \text{ roll}\}$:



Probabilities:



Distribution function:



Example 3.14. [Important discrete probability distributions]

- *Bernoulli* (Alternating) distribution with a parameter $p \in [0, 1]$, $X \sim \text{Be}(p)$:
  (*One toss of an unbalanced coin.*)

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

- *Binomial* distribution with parameter $p \in [0, 1]$, $X \sim \text{Binom}(n, p)$:
  (*Number of Heads in n tosses of an unbalanced coin.*)

$$\text{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- *Geometric* distribution with a parameter $p \in (0, 1)$, $X \sim \text{Geom}(p)$:
  (*Number of tosses of an unbalanced coin until the first Heads appear.*)

$$\text{P}(X = k) = (1 - p)^{k-1} p$$

- *Poisson* distribution with a parameter $\lambda > 0$, $X \sim \text{Poisson}(\lambda)$:
  (*Limit of the Binomial distribution for $n \to \infty$.*)

$$\text{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

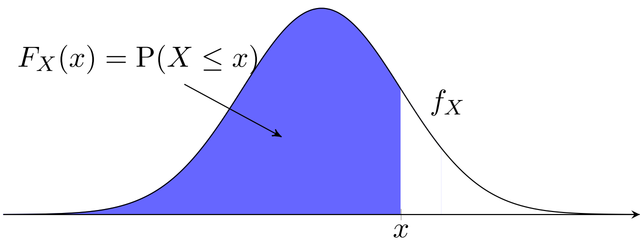
3.3 Continuous random variables

In some situations, a random variable can take *uncountably* many possible values. This arises when dealing with *continuous* models – measuring time, height, coordinates, etc. We cannot assign a positive probability $\mathrm{P}(X = x)$ to each value, because then the probabilities of the uncountable many values would sum up to infinity. Therefore we regard each singular value as having *zero probability* (intuitively, it is, e.g., infinitely improbable having to wait for the bus for exactly $3:00:00...$ minutes). Instead, we need a way to measure the probability of *intervals*. Recall the Romeo and Juliet problem, where each of them arrives at a random time point in an one-hour window, evenly chosen. Often we need to introduce an uneven distribution of values.

Definition 3.15. A random variable X is called (absolutely) *continuous*, if there exists a **non-negative** function $f_X : \mathbb{R} \to [0, +\infty)$ such that for all $x \in \mathbb{R}$ the distribution function F_X can be expressed as

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

The function f_X is called the *probability density* of the random variable X.



The *distribution function* of a continuous random variable is *continuous*.

Properties of continuous random variables

Theorem 3.16. *Let f_X be a density of a continuous random variable X. Then it holds that*

i) $\displaystyle\int_{-\infty}^{+\infty} f_X(t)\mathrm{d}t = 1$ *(*normalization condition*),*

ii) $\mathrm{P}(X = x) = 0$ *for all* $x \in \mathbb{R}$*,*

iii) $f_X(t) = \dfrac{\mathrm{d}F_X}{\mathrm{d}t}(t)$ *at points where the derivative exists,*

iv) $\mathrm{P}(a < X \leq b) = \displaystyle\int_{a}^{b} f_X(t)\,\mathrm{d}t = F_X(b) - F_X(a)$*,* *v)* $\mathrm{P}(X \in B) = \int\limits_B f_X(t)\mathrm{d}t$ *for all B in the Borel σ-algebra on $\mathbb{R}$, i.e., for all "common" sets.*

Consequences:

- $\mathrm{P}(X \le x) = \mathrm{P}(X < x)$ – from *ii)*
- $f_X(t)\mathrm{d}t \approx \mathrm{P}(t < X < t + \mathrm{d}t)$ for $\mathrm{d}t << 1$ – from *iv)*

Proof. i) $\int\limits_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = \lim\limits_{x\to+\infty} F_X(x) = 1.$

ii) Using the continuity of the distribution function and the previous theorem: $\mathrm{P}(X = x) = F(x) - \lim\limits_{y\to x-} F(y) = 0.$

iii) It follows from the properties of derivatives and integrals (first fundamental Theorem of calculus).

iv) $\mathrm{P}(a < X \le b) = F(b) - F(a) = \int\limits_{-\infty}^{b} f_X(t)\mathrm{d}t - \int\limits_{-\infty}^{a} f_X(t)\mathrm{d}t = \int\limits_{a}^{b} f_X(t)\mathrm{d}t.$ (second fundamental Theorem of calculus – Newton's formula)

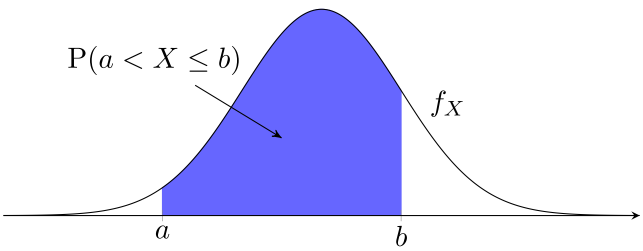
v) From the properties of the Lebesgue integral – advanced, see bibliography.

□

Relation between density and probability

Now we recall and illustrate the important property of the probability density:

$$\mathrm{P}(a < X \le b) = \int\limits_a^b f_X(x)\,\mathrm{d}x = \Big[F(x)\Big]_a^b = F(b) - F(a).$$



Note that when dealing with *continuous* random variables, it does not matter whether the inequalities are strict or non-strict.

$$\mathrm{P}(a < X \le b) = \mathrm{P}(a \le X \le b) = \mathrm{P}(a \le X < b) = \mathrm{P}(a < X < b).$$

Example 3.17 (– uniform distribution of Romeo's arrival)**.** Denote the time when Romeo arrives at the meeting point as a random variable X. Suppose that X has the *uniform distribution* on the interval $[0, 1]$, meaning that its density is constant on this interval and zero elsewhere.

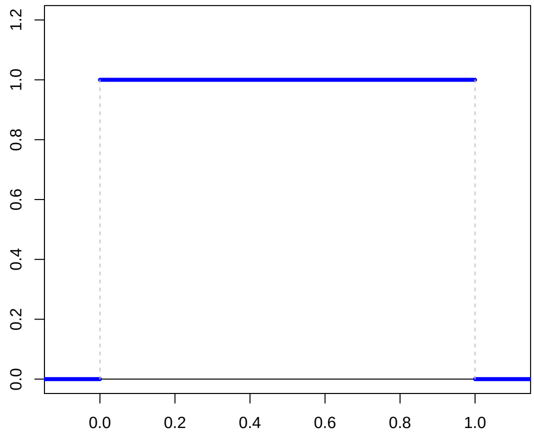
$$f_X(x) = \begin{cases} c & \text{for } x \in [0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

Determine the value of c, so that f truly forms a density of a random variable. From the normalization condition we know that the area under the graph of the density needs to be equal to one. Therefore the density needs to integrate to one:

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 c \cdot dx = [c \cdot x]_0^1 = c \cdot 1 - c \cdot 0 = c = 1.$$

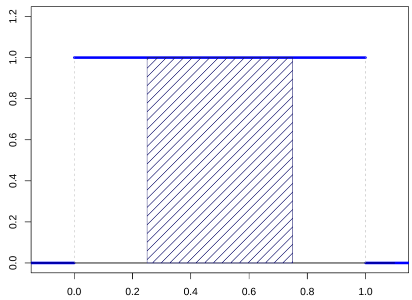
The constant c has to be equal to one.

Example 3.17 (– uniform distribution of Romeo's arrival (continued))**.** Density of the continuous uniform distribution on the interval $[0, 1]$:



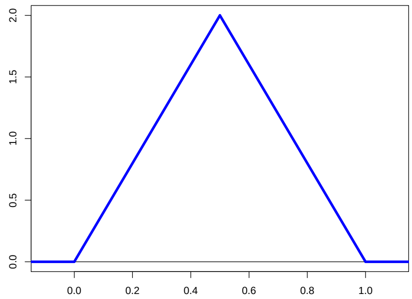
Example 3.17 (– uniform distribution of Romeo's arrival (continued))**.** What is the probability that Romeo arrives between 12:15 and 12:45? Probabilities concerning intervals are obtained as the corresponding area under the density:

$$\int_{1/4}^{3/4} 1dx = [x]_{1/4}^{3/4} = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}.$$



Example 3.18 (– non-uniform distribution of Juliet's arrival)**.** Suppose that Juliet arrives at the meeting point according to a non-uniform distribution with density:

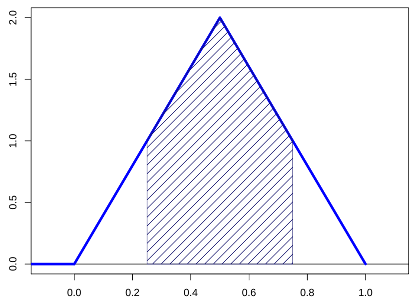
$$f_X(x) = \begin{cases} 4x & \text{for } x \in [0, 1/2] \\ 4 - 4x & \text{for } x \in [1/2, 1] \\ 0 & \text{otherwise.} \end{cases}$$



What is the probability that Juliet arrives between 12:15 and 12:45?

Example 3.18 (– non-uniform distribution of Juliet's arrival (continued))**.** What is the probability that Juliet arrives between 12:15 and 12:45? Probabilities concerning intervals are obtained as the corresponding area under the density:

$$\int_{1/4}^{3/4} f(x)dx = \cdots = \frac{3}{4}.$$



Note that when the distribution of the arrivals is not uniform, the probability that they will meet cannot be obtained using the geometric approach as before.

3.4 Functions of random variables

For a random variable X with a *known distribution*, we are often interested in the *distribution* of values somehow *calculated* from the values of X, say $Y = g(X)$.

Example 3.19 (– linear transformation)**.** Let X Be a random temperature in degrees Celsius. Then $Y = 1.8X + 32$ corresponds to the temperature in degrees Fahrenheit.

In the case of a *discrete* random variables the situation is relatively easy.

- $g(X)$ is always a random variable.
- The distribution of the random variable $g(X)$ is always discrete.

If X is a *continuous* random variable, the following complications arise:

- It can happen that $g(X)$ is not a random variable. (Therefore the assumption of *measurability* of g is needed.)
- The distribution of a random variable $g(X)$ can be discrete, continuous or mixed.

Function of a discrete random variable

Lemma 3.20 (– function of discrete random variable)**.** *Consider a function* $g : \mathbb{R} \to \mathbb{R}$ *and a* discrete *random variable* X*, and define the function of the random variable* $g(X)$ *by* $g(X)(\omega) = g(X(\omega))$ *for all* $\omega \in \Omega$*.*

Then $g(X)$ *is a discrete random variable with probabilities of the values*

$$\mathrm{P}(g(X) = y) = \sum_{x_k : g(x_k) = y} \mathrm{P}(X = x_k).$$

Proof. The probabilities of the values of $g(X)$ can be obtained from the (countable) disjoint partition

$$\{g(X) = y\} = \bigcup_{x_k : g(x_k) = y} \{X = x_k\}.$$

□

Functions of random variables

Lemma 3.21 (– function of a general random variable)**.** *Consider a* measurable *function* $g\colon \mathbb{R} \to \mathbb{R}$ *and an* arbitrary *random variable* X *and define the function of random the variable* $g(X)$ *as* $g(X)(\omega) = g(X(\omega))$ *for all* $\omega \in \Omega$*.*

Then the function $g(X)$ *of the random variable* X *is a random variable.*

Note: g is measurable if the set $\{x \in \mathbb{R}\colon g(x) \le y\}$ belongs to the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ for all $y \in \mathbb{R}$.

Proof. The proof that $g(X)$ is a random variable consists in verifying the measurability of $Y = g(X)$, i.e., that $\{Y \le y\}$ is an event for all y:

$$\{g(X) \le y\} = \{\omega \in \Omega\colon g(X(\omega)) \le y\} \in \mathcal{F}, \ \forall y \in \mathbb{R}.$$

A detailed proof can be found in the bibliography. □

Remark 3.22. Generally for a *distribution* function $F_Y(y)$ of a random variable $Y = g(X)$ it holds that

$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(g(X) \le y) = \mathrm{P}\left(\{\omega \in \Omega\colon g(X(\omega)) \le y\}\right).$$

If Y is continuous we obtain f_Y as the derivative of $F_Y(y)$ with respect to y.
Possible simplification:

- If the inverse g^{-1} of g exists and is increasing, then it holds that
$$F_Y(y) = \mathrm{P}(g(X) \le y) = \mathrm{P}\left(X \le g^{-1}(y)\right) = F_X(g^{-1}(y)).$$
- If g is strictly monotone, then g^{-1} is differentiable and $Y = g(X)$ is continuous with
$$f_Y(y) = f_X(g^{-1}(y))\, \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}.$$

✓ *Proofs and more information can be found in bibliography.*

3.5 Quantile function and simulations

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 $\mathrm{P}(X \le x) = \alpha$.

Definition 3.23. 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 and only if

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

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

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

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

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

Theorem 3.24. *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*

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

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. 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 $X_{n+1} = (aX_n + b) \bmod m$;
- 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: