Cumulative distribution function (Lecture 9)

Definition (Cumulative distribution function)

Let $X : \Omega \to \mathbb{R}$ *be a random variable. Then the cumulative distribution function of* X *is the following function* $F : \mathbb{R} \to [0, 1]$*:*

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}) \qquad \text{for every real number } t \in \mathbb{R}.$$

To every real number t, function F maps the probability that the value of the random variable is at most t. For example, if X is the number of heads out of three fair coin tosses:

$$F(1) = \mathbb{P}(X \leq 1) = \mathbb{P}(\text{at most 1 heads}) = 1/2;$$

$$F(2) = \mathbb{P}(X \leq 2) = \mathbb{P}(\text{at most 2 heads}) = 7/8;$$

$$F(2.3) = \mathbb{P}(X \leq 2.3) = \mathbb{P}(\text{at most 2.3 heads}) = 7/8.$$

Cumulative distribution function (Lecture 9)

Definition (Cumulative distribution function)

Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the cumulative distribution function of X is the following function $F : \mathbb{R} \to [0,1]$:

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}) \qquad \textit{for every real number } t \in \mathbb{R}.$$

To every real number t, function F maps the probability that the value of the random variable is at most t. For example, if X is the number of heads out of three fair coin tosses:

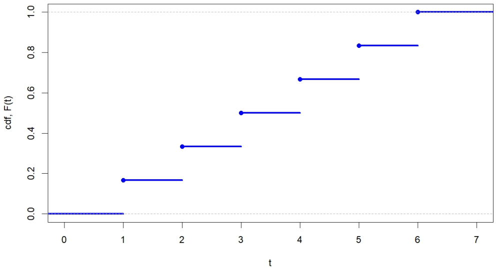
$$F(1) = \mathbb{P}(X \leq 1) = \mathbb{P}(\text{at most 1 heads}) = 1/2;$$

$$F(2) = \mathbb{P}(X \leq 2) = \mathbb{P}(\text{at most 2 heads}) = 7/8;$$

$$F(2.3) = \mathbb{P}(X \leq 2.3) = \mathbb{P}(\text{at most 2.3 heads}) = 7/8.$$

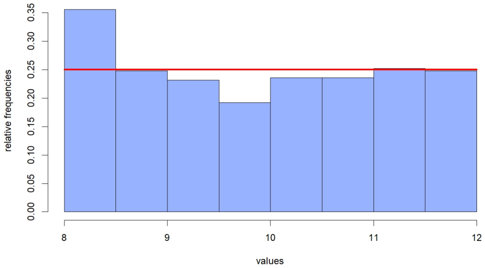
If X has a finite range, then its cumulative distribution function is a monotone step function (its range is finite), and the size of the "steps" are given by the probabilities of the possible values.

Cumulative distribution function: example



Cumulative distribution function of the value of a roll with a fair die
horizontal: t, vertical: $F(t) = \mathbb{P}(X \leq t)$.

Uniform distribution



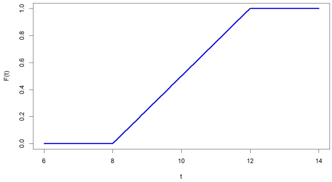
Histogram of a sample of size 500 from uniform distribution on the interval $[8, 12]$

Uniform distribution

Definition (Uniform distribution)

The random variable X has **uniform distribution** *on the interval $[a, b]$, if its distribution function is:*

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0, & \text{if } t \leq a; \\ \frac{t-a}{b-a}, & \text{if } a < t < b; \\ 1, & \text{if } t \geq b. \end{cases}$$



Properties of cumulative distribution functions

If $a, b \in \mathbb{R}$ are real numbers, and F is the cumulative distribution function of X, then we have

$$\mathbb{P}(a < X \leq b) = F(b) - F(a).$$

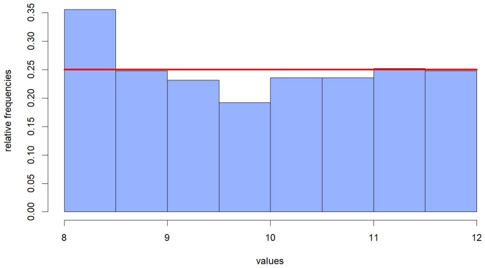
Indeed, the probability that X is larger than a, but smaller than b, can be obtained by taking $\mathbb{P}(X \leq b)$, and subtracting $\mathbb{P}(X \leq a)$.

Let F be the cumulative distribution function of an arbitrary random variable. Then the following properties hold:

1. F is monotone increasing: if $a < b$, then $F(a) \leq F(b)$.
2. $\lim_{t \to -\infty} F(t) = 0$; $\qquad \lim_{t \to \infty} F(t) = 1$.
3. F is continuous from the right, that is, for every real number $t \in \mathbb{R}$ we have $\lim_{s \to t+} F(s) = F(t)$.

Characterization: if F satisfies the properties above, then there exists X, whose cumulative distribution function is F.

Uniform distribution



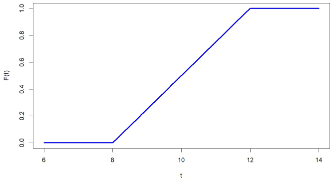
Histogram of a sample of size 500 from uniform distribution on the interval $[8, 12]$

Uniform distribution

Definition (Uniform distribution)

The random variable X has **uniform distribution** *on the interval $[a, b]$, if its distribution function is:*

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0, & \text{if } t \leq a; \\ \frac{t-a}{b-a}, & \text{if } a < t < b; \\ 1, & \text{if } t \geq b. \end{cases}$$



Uniform distribution

A package is delivered at a random time, which has uniform distribution between 10 and 12 (hours). We suppose that the time of delivery is uniformly distributed on the interval $[10, 12]$. Then the following hold ($a = 10, b = 12$):

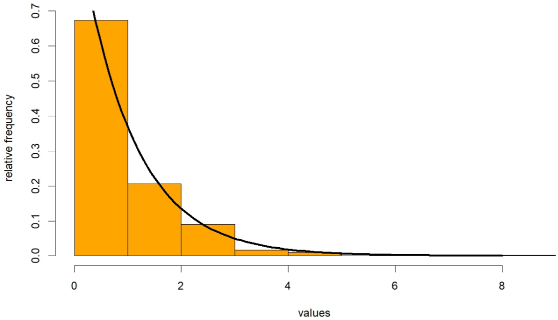
- The probability that the delivery is between 10 and 11: $(11 - 10)/(12 - 10) = 1/2$.
- The probability that the delivery is between $10:15$ and $10:30$: $1/8 = 12.5\%$.
- The probability that the delivery is after $10:30$: $3/4 = 75\%$.

Exponential distribution

Exponential distribution is often used to model random time intervals, for example,

- the time needed for a certain operation: service time of a customer, or an operation on a computer
- reaction time of a person
- time between the two occurences of an event, e.g. arrivals of two customers in a shop
- epidemic spread: time until recovery or infection
- radioactive decay: time until the decay of a particle

Exponential distribution



Density function of the exponential distribution with parameter $\lambda = 1$ and histogram of a sample of size 500 from the same distribution

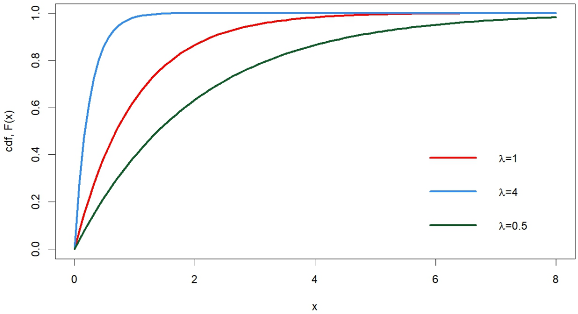
Exponential distribution

Definition

Let $\lambda > 0$ be real number. Random variable X has **exponential distribution** *with parameter λ, if its cumulative distribution function is:*

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 1 - e^{-\lambda t}, & \text{if } t > 0; \\ 0 & \text{otherwise.} \end{cases}$$

Exponential distribution



Cumulative distribution functions of random variables of exponential distribution of different parameters $\left(\lambda = \frac{1}{2}, 1, \text{ and } 4\right)$

Exponential distribution

Example. Suppose that the service time of a customer in a shop has **exponential distribution with parameter** $1/3$ (in minutes).

What is the probability that the service time is **at least** 5 **minutes**? What is the probability that the service time is **at least** 2**, but at most** 4 **minutes**?

Exponential distribution

Example. Suppose that the service time of a customer in a shop has **exponential distribution with parameter** $1/3$ (in minutes).

What is the probability that the service time is **at least** 5 **minutes**? What is the probability that the service time is **at least** 2**, but at most** 4 **minutes**?

Based on the definition, we have

$$\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X < 5) = 1 - F(5) = 1 - (1 - e^{-\lambda \cdot 5}) = e^{-\lambda \cdot 5} =$$
$$= e^{-5/3} = 18.9\%.$$

Exponential distribution

Example. Suppose that the service time of a customer in a shop has **exponential distribution with parameter** $1/3$ (in minutes).

What is the probability that the service time is **at least** 5 **minutes**? What is the probability that the service time is **at least** 2**, but at most** 4 **minutes**?

Based on the definition, we have

$$\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X < 5) = 1 - F(5) = 1 - (1 - e^{-\lambda \cdot 5}) = e^{-\lambda \cdot 5} = \\ = e^{-5/3} = 18.9\%.$$

Similarly, we have

$$\mathbb{P}(2 \leq X \leq 4) = \mathbb{P}(X \leq 4) - \mathbb{P}(X \leq 2) = F(4) - F(2) = \\ = (1 - e^{-4/3}) - (1 - e^{-2/3}) = e^{-2/3} - e^{-4/3} = 25\%.$$

Exponential distribution

Proposition (Memoryless property of exponential distribution)

Let X be a random variable with exponential distribution, and s, t be positive numbers. Then we have

$$\mathbb{P}(X \geq s+t | X \geq s) = \mathbb{P}(X \geq t).$$

Proof. We use the definition of conditional probability and the definition of exponential distribution as follows:

$$\begin{aligned}\mathbb{P}(X \geq s+t | X \geq s) &= \frac{\mathbb{P}(\{X \geq s+t\} \cap \{X \geq s\})}{\mathbb{P}(X \geq s)} = \frac{1-\mathbb{P}(X < s+t)}{1-\mathbb{P}(X < s)} = \\ &= \frac{1-F(s+t)}{1-F(s)} = \frac{1-(1-e^{-\lambda(s+t)})}{1-(1-e^{-\lambda s})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \\ &= e^{-\lambda t} = 1-(1-e^{-\lambda t}) = 1-F(t) = \mathbb{P}(X \geq t).\end{aligned}$$

□

Pareto distribution

In actuarial areas, or for modelling incomes, the following distribution is often used.

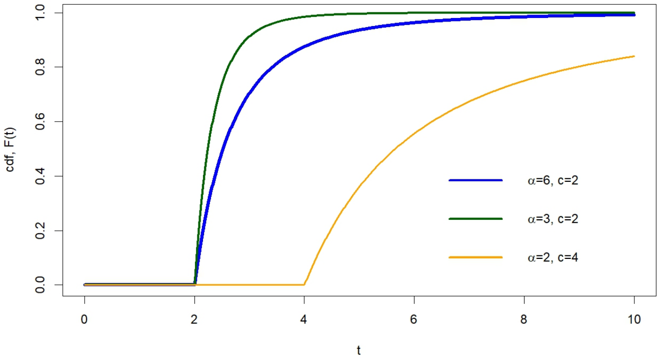
Definition (Pareto distribution)

Random variable X has Pareto distribution with parameters $\alpha > 0$ and $c > 0$, if its cumulative distribution function is

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0; & \text{if } t \leq c; \\ 1 - \left(\frac{c}{t}\right)^{\alpha}; & \text{if } t > c. \end{cases}.$$

The definition implies that the minimal value of the Pareto distribution is c. We will see later on that there exists α for which the expected value of this distribution is not well defined, and another α for which the expected value is finite, but the variance is not well defined.

Pareto distribution



Cumulative distribution functions of Pareto distributions of different parameters

Multivariate random variables

Examples:

- a random process (stock exchange index, population of a country, unemployment rate) at different time points;
- different characteristics of a person (or country, company etc.), e.g. the age, the income and the expenses of a person;
- values of a measurement repeated several times (e.g. we ask a hundred people about their time spent on sports).

A family of random variables is called **a multivariate random variable**. This can consist of **correlated** (first two cases) or **uncorrelated** (third case) random variables.

Multivariate random variable

A function

$$\underline{X} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$$

is a **multivariate random variable**, if $X_1, X_2, \ldots, X_n$ are all random variables.

- 1000 are asked about their time spent on sports in a week. Let X_i be the answer of the ith participant. Then $(X_1, X_2, \ldots, X_{1000})$ is a multivariate random variable.
- $(X_1, X_2, \ldots, X_{31})$ is a multivariate random variable, where X_j is the amount of rain on jth day of December $(j = 1, 2, \ldots, 31)$.

If $\underline{X}$ is a multivariate random variable, then the distribution of X_i is called the ***i*th marginal distribution** of $\underline{X}$.

A multivariate random variable $\underline{X}$ is **discrete**, if its range is finite or countably infinite.

Joint distribution function

A multivariate random variable $\underline{X} = (X_1, \ldots, X_n)$ has **joint distribution function** $F : \mathbb{R}^n \to [0,1]$ defined by

$$F(\underline{t}) = F(t_1, \ldots, t_n) = \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n),$$

if $(t_1, \ldots, t_n) \in \mathbb{R}^n$ are real numbers. For example:

- a randomly chosen person is asked about her time spent on sports in a week (X_1), her income (X_2), and age (X_3);
- then (X_1, X_2, X_3) is a multivariate random variable, and
- if its joint distribution function is F, then, for example,

$$F(3, 500000, 40) = \mathbb{P}(X_1 \leq 3, X_2 \leq 500000, X_3 \leq 40)$$

  is the probability that a randomly chosen person spends at most 3 hours on sports, has income at most 500000 (forints), and age at most 40 (age).