

Probability and random variables

Version 2.4 (28 July 2010)

Free University of Bolzano Bozen – Paolo Coletti

Introduction

This book contains a brief introduction on theory of probabilities and random variables. In order to be able to read it correctly a basic knowledge of calculus, in particular integral calculation, is necessary.

This book is for Decision Theory and Statistics B courses:

- section marked with DT is necessary for Decision Theory course A.Y. 2009/10;
- section marked with SB is absolutely necessary for the comprehension of Statistics B course, these concepts will be used during the course and will not be repeated;
- not marked sections are included for completeness but are not necessary for the courses.

Table of contents

Introduction	1	3.1.4 Cauchy	11
Table of contents	1	3.2 Cumulative distribution ^{DT SB}	12
1. Probability	2	3.2.1 Uniform ^{DT}	12
1.1 Outcomes and events.....	2	3.2.2 Normal ^{DT SB}	13
1.2 Complement	2	3.2.3 Exponential ^{DT}	13
1.3 Bayes theorem.....	3	3.3 Independence.....	13
2. Discrete random variables ^{DT}	3	4. Expected value ^{DT SB}	13
2.1 Discrete r.v. examples ^{DT}	3	4.1 Rules ^{DT SB}	14
2.1.1 Bernoulli ^{DT SB}	3	4.2 Calculations	14
2.1.2 Binomial ^{DT SB}	4	4.2.1 Bernoulli ^{DT SB}	14
2.1.3 Die roll ^{DT}	4	4.2.2 Die roll ^{DT}	14
2.1.4 Indicator	5	4.2.3 Binomial ^{DT}	14
2.1.5 Poisson	5	4.2.4 (die roll)-squared ^{DT}	14
2.2 Diagrams	5	4.2.5 Uniform ^{DT}	14
2.3 Independence ^{SB}	6	4.2.6 Exponential ^{DT}	14
2.4 Mathematical operations ^{DT}	6	4.2.7 Normal ^{DT SB}	15
2.4.1 Operations for two r.v.	7	4.2.8 Uniform-squared ^{DT}	15
2.5 Cumulative distribution ^{DT SB}	7	4.2.9 Cauchy	15
2.5.1 Bernoulli ^{DT}	8	5. Variance ^{DT SB}	16
2.5.2 Binomial ^{DT}	8	5.1 Rules ^{DT SB}	16
2.5.3 Die roll ^{DT}	8	5.2 Calculation	16
2.5.4 Poisson	8	5.2.1 Bernoulli ^{DT SB}	16
3. Continuous random variables ^{DT SB}	9	5.2.2 Die roll ^{DT}	17
3.1 Continuous r.v. examples.....	9	5.2.3 Binomial ^{DT}	17
3.1.1 Uniform or Rectangular ^{DT}	9	5.2.4 Uniform ^{DT}	17
3.1.2 Normal or Gaussian ^{DT SB}	10	5.2.5 Exponential ^{DT}	17
3.1.3 Exponential ^{DT}	10	5.2.6 Normal ^{DT}	17
		6. Momentum ^{SB}	18

6.1	Skewness ^{SB}	18	10. Mode	24
6.2	Kurtosis ^{SB}	18	10.1 Normal.....	24
7. Median ^{DT SB}	18	10.2 Exponential.....	24	
7.1 Exponential ^{DT}	19	10.3 Cauchy.....	25	
8. Lognormal r.v. ^{DT}	19	10.4 Lognormal	25	
8.1 Median ^{DT}	20	11. Covariance ^{SB}	25	
8.2 Expected value ^{DT}	20	11.1 Rules.....	26	
9. Standard normal ^{DT SB}	21	11.2 Correlation ^{SB}	26	
9.1 Calculating normal area ^{DT SB}	22	12. Central limit theorem ^{SB}	26	

1. Probability

1.1 Outcomes and events

Let Ω be the space of n (which can also be infinite) outcomes ϖ_i . These outcomes must satisfy these axioms:

1. $P(\varpi_i) \geq 0 \forall i$;
2. $\sum_{i=1}^n P(\varpi_i) = 1$.

Each outcome represents an individual fact which may happen. Note that one and only one outcome will happen at any time.

We call “event” a subset E of Ω , therefore an event is a set of possible outcomes.

The probability of an even is defined as the sum of the probabilities of its outcomes.

An event with zero probability is said to be impossible; clearly the empty set \emptyset is always impossible, meaning that an outcome must always happen.

Two events are said to be mutually exclusive when their intersection is empty, i.e. they do not share any outcome.

1.2 Complement

Given E subset of Ω , we define the complement of E in Ω , CE , as the subset of Ω containing all the outcomes which are not in E .

The complement has the following rules:

- $C(A \cup B) = CA \cap CB$;
- $C(A \cap B) = CA \cup CB$;
- $P(CA) = 1 - P(A)$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- $P(A \cap B) = P(A) \cdot P(B)$ only when A and B are independent. This is in fact used as official definition of event independence.

1.3 Bayes theorem

We define $P(A|B)$ as the probability that event A happens when event B happens. In other words using outcomes, we give for granted that an outcome from B will happen and we calculate the probability that the event belongs also to A .

Bayes theorem says that $P(A|B) = P(A \cap B)/P(B)$. Obviously if A and B are mutually exclusive, $P(A|B) = 0$.

2. Discrete random variables ^{DT}

Random variables are variables which do not take a fixed and predictable value but may take different values, called outcomes, every time they are looked at, according to probability rules. Random variables may be divided into discrete and continuous.

Let ξ be a discrete random variable. It takes values in \mathbb{R} according to a defined probability distribution.

$$\xi = \begin{cases} x_1 & p_1 \\ x_2 & p_2 \\ \vdots & \vdots \\ x_n & p_n \end{cases}; \quad x_i \in \mathbb{R}; \quad p_i \in [0,1]; \quad \sum_{i=1}^n p_i = 1; \quad P(x_i) = p_i$$

ξ takes as value any x_i of the $x_1 \dots x_n$ with the corresponding probability p_i , called marginal distribution. The number of values may also be infinite, but must be numerable.

2.1 Discrete r.v. examples ^{DT}

2.1.1 Bernoulli ^{DT SB}

$$\xi = \begin{cases} 0 & 1-p \\ 1 & p \end{cases} \quad \text{with } p \in [0,1]$$

Bernoulli random variable is the most famous discrete random variable. It is a very simple r.v. which may take only two values, 0 and 1, with probabilities $1-p$ and p respectively.

It is a one parameter (p , which must be between 0 and 1) random variable.

When p is 50% it corresponds, for example, to the flip of a fair coin (head corresponds to 1 and tail corresponds to 0).

The two probabilities are clearly always between 0 and 1 and the sum of them is clearly 1, so the assumptions are satisfied.

2.1.2 Binomial ^{DT SB}

$$\xi = \begin{cases} 0 & p_0 \\ 1 & p_1 \\ 2 & p_2 \\ \dots & \dots \\ n & p_n \end{cases} \quad p_i = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

Binomial random variable, often indicated as $B(n, p)$ is a discrete random variable which may take $n+1$ values according to a complex probability distribution.

It is a two parameters distribution: n , which must be integer and positive, and p , which must be between 0 and 1. When n is equal to 1, the Binomial r.v. corresponds to the Bernoulli r.v..

$B(n, p)$ can be seen as the sum of n independent¹ and identically distributed Bernoulli random variables. It can also be seen as the answer to the question “how many 1s will I get if I look for n times at a Bernoulli r.v. with parameter p ?”. Therefore $B(n, p) = \sum_{i=1}^n \xi_i$ where ξ_i are n independent Bernoulli r.v. with parameter p .

In order to check the r.v. assumptions, we note that probabilities are clearly between 0 and 1, while it can be shown that their sum is exactly 1.

Calculating effective Binomial probabilities is a difficult task whenever n is large, since $n!$ becomes a huge number. Therefore often tables are used, paying attention to the difference between tables reporting probability values $P(\xi = x)$ and tables reporting cumulative probability values $P(\xi \leq x)$. They are easily distinguishable from the fact that the former values are increasing and then decreasing while the latter are always increasing reaching 1 when x is n .

2.1.3 Die roll ^{DT}

$$\xi = \begin{cases} 1 & \frac{1}{6} \\ 2 & \frac{1}{6} \\ 3 & \frac{1}{6} \\ 4 & \frac{1}{6} \\ 5 & \frac{1}{6} \\ 6 & \frac{1}{6} \end{cases} \quad \text{The die roll is a quite simple r.v. without any parameter. Also in this case it is easy to note that probabilities are between 0 and 1 and that their sum is exactly 1.}$$

¹ For a definition of independence, see section 2.3 on page 5.

2.1.4 Indicator

$$\chi_A(a) = \begin{cases} 0 & a \notin A \\ 1 & a \in A \end{cases}$$

Indicator random variable, traditionally indicated with the letter χ , is a random variable whose value depends on whether element a belongs to set A or not. It is theoretically equal to a Bernoulli r.v., with p equal to the probability that a belongs to A .

2.1.5 Poisson

$$\xi = \begin{cases} 0 & \frac{\lambda^0 e^{-\lambda}}{0!} \\ & \frac{\lambda^1 e^{-\lambda}}{1!} \\ & \vdots \\ & \frac{\lambda^x e^{-\lambda}}{x!} \\ & \vdots \end{cases} \quad \text{with } \lambda > 0$$

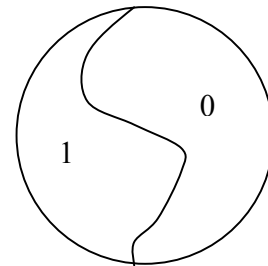
Poisson random variable is an example of a discrete r.v. with infinite values. It is a one parameter r.v..

It can be shown that probabilities are between 0 and 1 and it can be demonstrated that the sum of probabilities, which is infinite, is exactly one.

2.2 Diagrams

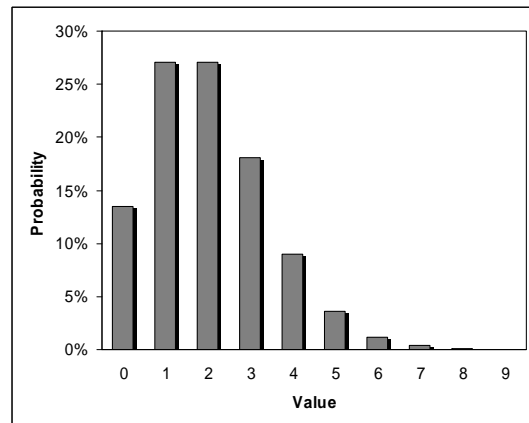
Venn diagram is a graphical representation of a discrete random variable. It consists on a circle divided into either slices or more generic areas with the value written inside and with the size of each area representing the probability of that value.

Here an example of a Venn diagram for a Bernoulli r.v. with generic parameter p .



A bar plot can also be used to represent discrete r.v.. It consists on a graph with values on the horizontal axis, probabilities on the vertical axis and vertical bars corresponding to every value's probability.

Here an example of a bar plot for a Poisson r.v. with parameter $\lambda = 2$.



2.3 Independence ^{SB}

For a definition of events independence, see section 1.2 on page 2.

Two discrete random variables ξ and ε are independent when, for every couple of values taken by the two r.v., the joint probability distribution $P(x_i, y_j)$, which is the probability that ξ takes value x_i and ε takes value y_j , is equal to product of the two marginal distributions. This means $P(x_i, y_j) = P(x_i) \cdot P(y_j)$ for every x_i and y_j . Note that in general this is not true for every r.v., since the joint probability distribution depends on the events which originate the values of the involved random variables.

An alternative equivalent definition uses the probability set theory and says that two discrete random variables ξ and ε are independent when the event set for which ξ takes value x_i and the event set for which ε takes value y_j are independent. Since in probability set theory the probability of the intersection of two independent sets is the product of probabilities, this means that

$$\begin{aligned} & P(\{\omega \in \Omega \text{ such that } \xi(\omega) = x_i\} \cap \{\omega \in \Omega \text{ such that } \varepsilon(\omega) = y_j\}) = \\ & = P(\{\omega \in \Omega \text{ such that } \xi(\omega) = x_i\}) \cdot P(\{\omega \in \Omega \text{ such that } \varepsilon(\omega) = y_j\}) \text{ for every } x_i \text{ and } y_j. \end{aligned}$$

2.4 Mathematical operations ^{DT}

Mathematical functions may be applied to random variables: in this case, the result is a new random variable with the same probabilities (which therefore automatically satisfies the r.v. assumption on probabilities) and the mathematical function applied only to its values. It is very important to note that probabilities are never affected by the mathematical function, only values are.

For example, applying the square function to the die roll r.v. leads to:

$$\xi^2 = \begin{cases} 1 & \frac{1}{6} \\ 4 & \frac{1}{6} \\ 9 & \frac{1}{6} \\ 16 & \frac{1}{6} \\ 25 & \frac{1}{6} \\ 36 & \frac{1}{6} \end{cases}$$

2.4.1 Operations for two r.v.

Mathematical operations may be done also using two or more r.v. at the same time, for example summing or multiplying them, obtaining a new r.v.. In this case, however, we need to use the joint probability distribution $p(x_i, y_j)$, which is the probability that the first r.v. takes value x_i and the second takes value y_j . For example, considering the sum of r.v. ξ , with n values x_i , and ε , with m values y_j , the resulting random variable is:

$$\xi + \varepsilon = \begin{cases} x_1 + y_1 & p(x_1, y_1) \\ \vdots & \vdots \\ x_n + y_1 & p(x_n, y_1) \\ \vdots & \vdots \\ x_1 + y_m & p(x_1, y_m) \\ \vdots & \vdots \\ x_n + y_m & p(x_n, y_m) \end{cases}$$

In general the joint probability distribution can not be derived directly from the marginal distributions but must be derived from the events which originate the values of the involved random variables. Only when the random variables are independent the joint probability distribution is simply the product of the marginal distributions.

2.5 Cumulative distribution ^{DT SB}

The cumulative distribution of a discrete r.v. with numerical outcomes is a function which assigns to every outcome the probability to get that outcome or any of the previous ones. Clearly it is not a r.v. since it does not respect conditions in section 2. However, cumulative distributions have their own features:

1. $F_\xi(x) \geq 0$;
2. $F_\xi(x)$ is (non strictly) increasing;
3. $F_\xi(x) = P(\xi \leq x)$;
4. $F_\xi(x) \leq 1$ and the cumulative distribution of the last outcome must be 1.

2.5.1 Bernoulli^{DT}

$$F_{\xi}(x) = \begin{cases} x < 0 & 0 \\ 0 \leq x < 1 & p \\ x \geq 1 & 1 \end{cases}$$

2.5.2 Binomial^{DT}

$$F_{\xi}(x) = \begin{cases} x < 0 & 0 \\ 0 \leq x < 1 & p_0 \\ 1 \leq x < 2 & p_0 + p_1 \\ \dots & \dots \\ n-1 \leq x < n & p_0 + \dots + p_{n-1} \\ x \geq n & 1 \end{cases}$$

$$p_i = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

2.5.3 Die roll^{DT}

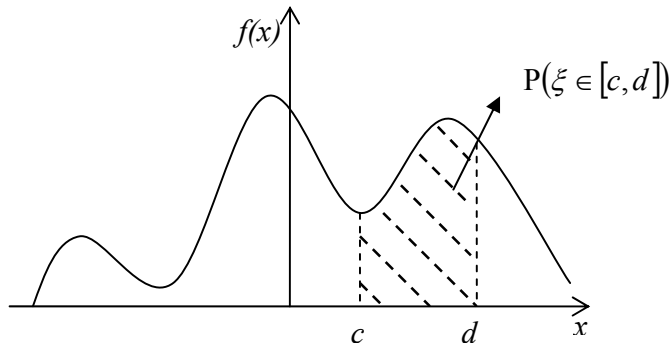
$$F_{\xi}(x) = \begin{cases} x < 1 & 0 \\ 1 \leq x < 2 & 1/6 \\ 2 \leq x < 3 & 2/6 \\ 3 \leq x < 4 & 3/6 \\ 4 \leq x < 5 & 4/6 \\ 5 \leq x < 6 & 5/6 \\ x \geq 6 & 1 \end{cases}$$

2.5.4 Poisson

$$F_{\xi}(x) = \begin{cases} x < 0 & 0 \\ 0 \leq x < 1 & \frac{\lambda^0 e^{-\lambda}}{0!} \\ 1 \leq x < 2 & \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} \\ \vdots & \vdots \\ n-1 \leq x < n & \sum_{j=0}^{n-1} \frac{\lambda^j e^{-\lambda}}{j!} \\ \vdots & \vdots \end{cases}$$

3. Continuous random variables ^{DT SB}

Let ξ be a continuous random variable. Its possible values now belong to a dense interval, usually the whole \mathbb{R} . Therefore they may not be enumerated as before. The probability that a specified single value comes out is 0. Probability that a value inside an interval comes out is, in general, not zero and is defined as

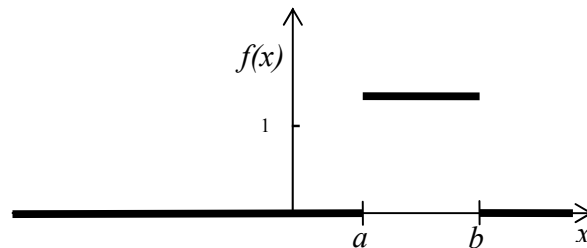


$$P(\xi \in [c, d]) = \int_c^d f_\xi(s) ds \quad \text{with} \quad \int_{-\infty}^{+\infty} f_\xi(s) ds = 1 \quad \text{and} \quad f_\xi \geq 0.$$

3.1 Continuous r.v. examples

3.1.1 Uniform or Rectangular ^{DT}

$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$

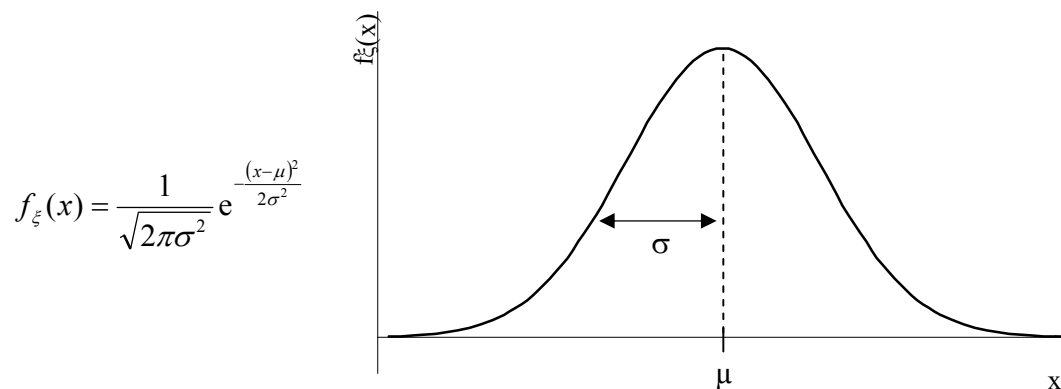


Uniform or Rectangular r.v. has a simple density function which is always 0 except in an interval where it takes a constant values. Therefore it may take values only in this interval. It is a two parameters r.v., with $a < b$, and it is indicated as $R(a,b)$ or $U(a,b)$.

The density function is clearly always zero or positive, while to check the second assumption we must calculate the integral:

$$\begin{aligned} \int_{-\infty}^{+\infty} f_\xi(s) ds &= \int_{-\infty}^a f_\xi(s) ds + \int_a^b f_\xi(s) ds + \int_b^{+\infty} f_\xi(s) ds = \int_{-\infty}^a 0 ds + \int_a^b \frac{1}{b-a} ds + \int_b^{+\infty} 0 ds = \\ &= 0 + \left[\frac{s}{b-a} \right]_a^b + 0 = \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1 \end{aligned}$$

3.1.2 Normal or Gaussian ^{DT SB}



The normal r.v. is the most famous continuous r.v.. It is a two parameters r.v., μ which may be any real number and defines the center of the distribution (which is symmetric around μ) and σ^2 , an always strictly positive parameter which defines how large is the bell-shape of the density function. It is usually indicated as $N(\mu, \sigma^2)$.

Its density function, being a multiplication of two positive factors, is always strictly positive. Its integral over \mathcal{R} is 1 (no demonstration is given in this handout).

An interesting feature of normal r.v. is that, when its values are multiplied by any constant c , the resulting r.v. is still normal with parameters $c\mu$ and $c^2\sigma^2$. In summary $c \cdot N(\mu; \sigma^2) = N(c\mu; c^2\sigma^2)$.

3.1.3 Exponential ^{DT}



Exponential density function is zero for negative values, thus implying that the r.v. may never take negative values. It is a one-parameter r.v. and its parameter λ must be strictly positive.

The density function is clearly always positive or zero, while to check the second assumption we need to calculate the integral

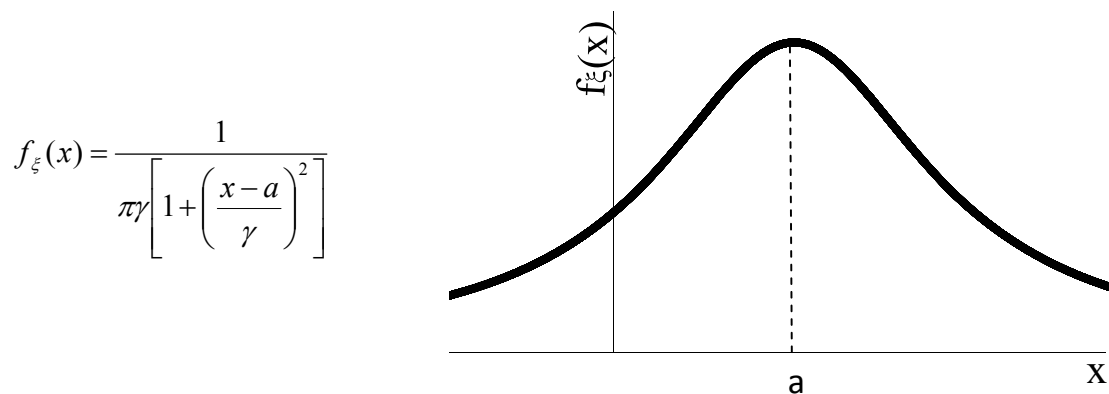
$$\int_{-\infty}^{+\infty} f_{\xi}(s) ds = \int_{-\infty}^0 f_{\xi}(s) ds + \int_0^{+\infty} f_{\xi}(s) ds = \int_{-\infty}^0 0 ds + \int_0^{+\infty} \lambda e^{-\lambda s} ds = 0 + \lambda \int_0^{+\infty} e^{-\lambda s} ds =$$

$$= \lambda \left[\frac{e^{-\lambda s}}{-\lambda} \right]_0^{+\infty} = \lambda \left[\lim_{s \rightarrow +\infty} \left(\frac{e^{-\lambda s}}{-\lambda} \right) - \frac{e^{-\lambda \cdot 0}}{-\lambda} \right] = \lambda \left[0 - \frac{1}{-\lambda} \right] = \frac{\lambda}{\lambda} = 1$$

This distribution is used in modeling time intervals between rare events, or waiting times, like:

1. Successive calls of your phone number.
2. Arrivals at a fast-food drive-through restaurant.
3. Emergency arrivals at a hospital.
4. The time until default of a money borrower.

3.1.4 Cauchy



Cauchy density function has two parameters: a determines the position of the distribution, while γ , which must be strictly positive, determines the width of the distribution.

The density function is clearly always positive, while to check the second assumption we need to calculate the integral (explicit calculation of the integral is beyond the scope of this book)

$$\int_{-\infty}^{+\infty} f_{\xi}(s) ds = \int_{-\infty}^{+\infty} \frac{1}{\pi \gamma \left[1 + \left(\frac{s-a}{\gamma} \right)^2 \right]} ds = \left[\frac{1}{\pi} \arctan \left(\frac{s-a}{\gamma} \right) + \frac{1}{2} \right]_{-\infty}^{+\infty} = 1$$

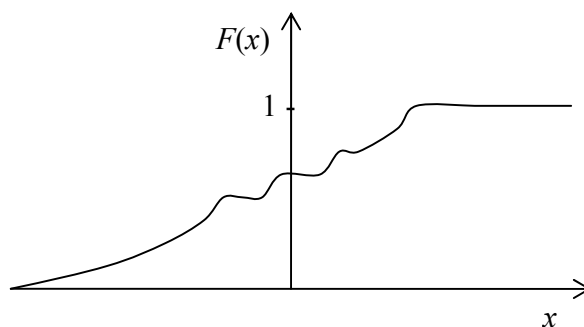
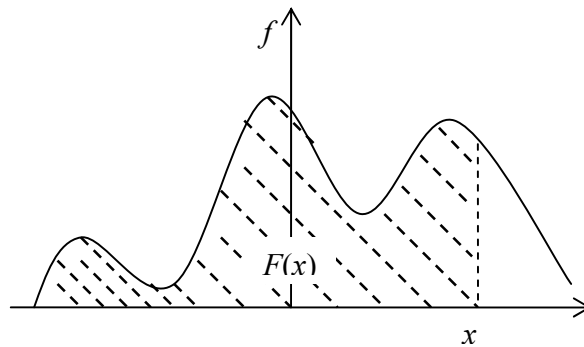
3.2 Cumulative distribution ^{DT SB}

Cumulative distribution is a function obtained as the integral between $-\infty$ and x of the density function.

$$F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds$$

Thanks to this definition, this function has always these features:

1. $F_{\xi}(x) \geq 0$;
2. $F_{\xi}(x)$ is (non strictly) increasing;
3. $F_{\xi}(x) = P(\xi \in (-\infty, x]) = P(\xi \leq x) = P(\xi < x)$, since the probability that $\xi = x$ is zero;
4. $F_{\xi}(x) \leq 1$ and its limit, as x tends to infinity, is 1.



3.2.1 Uniform ^{DT}

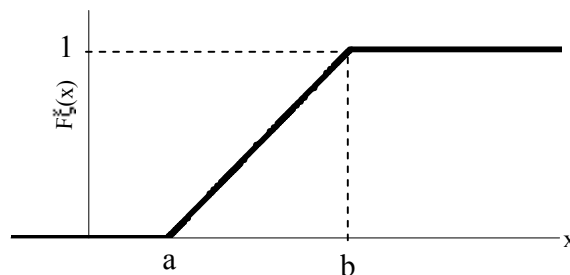
– for $x < a$: $F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds = \int_{-\infty}^x 0 ds = 0$;

– for $a \leq x \leq b$: $F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds = \int_{-\infty}^a 0 ds + \int_a^x \frac{1}{b-a} ds = 0 + \left[\frac{s}{b-a} \right]_a^x = \frac{x-a}{b-a}$;

– for $x > b$: $F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds = \int_{-\infty}^a 0 ds + \int_a^b \frac{1}{b-a} ds + \int_b^x 0 ds = \left[\frac{s}{b-a} \right]_a^b = \frac{b-a}{b-a} = 1$;

Therefore:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b; \\ 1 & x > b \end{cases}$$

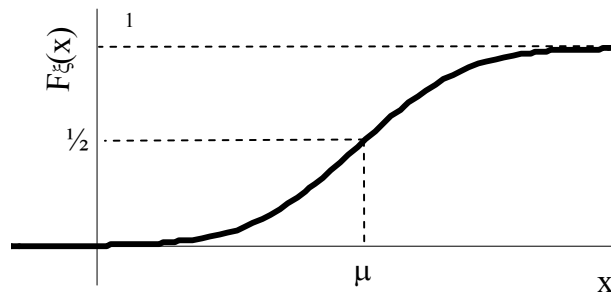


3.2.2 Normal ^{DT SB}

We are not able to calculate analytically the integral

$$\int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds, \text{ however}$$

we know, from numerical approximation computer programs, that the shape of the normal cumulative distribution function is like this picture.

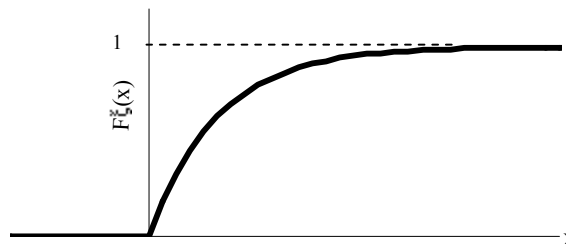


3.2.3 Exponential ^{DT}

– for $x < 0$: $F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds = \int_{-\infty}^x 0 ds = 0$;

– for $x \geq 0$: $F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds = \int_{-\infty}^0 0 ds + \int_0^x \lambda e^{-\lambda s} ds = 0 + \lambda \left[\frac{e^{-\lambda s}}{-\lambda} \right]_0^x = -e^{-\lambda x} + 1$;

$$F_{\xi}(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



3.3 Independence

Two continuous random variables ξ and ε are independent when, for every couple of values taken by the two r.v., the joint density distribution $f_{\xi\varepsilon}(x, y)$ is equal to product of the two densities $f_{\xi}(x) \cdot f_{\varepsilon}(y)$.

4. Expected value ^{DT SB}

The expected value is a number associated to every r.v.. It is calculated, for discrete

random variables as $E(\xi) = \sum_{i=1}^n x_i p_i$, while for continuous random variables

$$E(\xi) = \int_{-\infty}^{+\infty} s f(s) ds .$$

4.1 Rules ^{DT SB}

- $E(\xi + \zeta) = E(\xi) + E(\zeta)$, for any ξ and ζ random variables;
- $E(c\xi) = cE(\xi)$, for any ξ r.v. and c constant real number;
- $E(c) = c$, for any c constant real number;
- if ξ and ζ are independent (see sections 2.3 and 3.3 for definition), then $E(\xi \cdot \zeta) = E(\xi) \cdot E(\zeta)$.

4.2 Calculations

4.2.1 Bernoulli ^{DT SB}

$$E(\xi) = \sum_{i=1}^n x_i p_i = 1p + 0(1-p) = p;$$

4.2.2 Die roll ^{DT}

$$E(\xi) = \sum_{i=1}^n x_i p_i = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6} = 3.5;$$

4.2.3 Binomial ^{DT}

We use one of the expected value rules

$$E(B(n, p)) = E(\xi_1 + \xi_2 + \dots + \xi_n) = E(\xi_1) + E(\xi_2) + \dots + E(\xi_n) = np;$$

4.2.4 (die roll)-squared ^{DT}

$$E(\text{die roll}^2) = 1^2 \frac{1}{6} + 2^2 \frac{1}{6} + 3^2 \frac{1}{6} + 4^2 \frac{1}{6} + 5^2 \frac{1}{6} + 6^2 \frac{1}{6} = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6};$$

4.2.5 Uniform ^{DT}

$$E(U(a, b)) = \int_{-\infty}^{+\infty} sf(s)ds = \int_{-\infty}^a s \cdot 0 ds + \int_a^b s \frac{1}{b-a} ds + \int_b^{+\infty} s \cdot 0 ds =$$

$$= 0 + \left[\frac{s^2}{2(b-a)} \right]_a^b + 0 = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2};$$

4.2.6 Exponential ^{DT}

$$E(\xi) = \int_{-\infty}^{+\infty} sf(s)ds = \int_{-\infty}^0 s \cdot 0 ds + \int_0^{+\infty} s \lambda e^{-\lambda s} ds = 0 + \lambda \int_0^{+\infty} s e^{-\lambda s} ds;$$

In order to evaluate this integral we use the formula of integration by parts,

$\int_c^d hg' ds = [hg]_c^d - \int_c^d h' g ds$, where the role of h is taken by s and the role of g' is taken by $e^{-\lambda s}$:

$$\begin{aligned} E(\xi) &= \lambda \left[s \frac{e^{-\lambda s}}{-\lambda} \right]_0^{+\infty} - \lambda \int_0^{+\infty} \frac{e^{-\lambda s}}{-\lambda} ds = \lambda \left[\lim_{s \rightarrow +\infty} s \frac{e^{-\lambda s}}{-\lambda} - 0 \frac{e^{-\lambda 0}}{-\lambda} \right] + \int_0^{+\infty} e^{-\lambda s} ds = \\ &= \lambda [0 - 0] + \left[\frac{e^{-\lambda s}}{-\lambda} \right]_0^{+\infty} = 0 + \left[0 + \frac{1}{\lambda} \right] = \frac{1}{\lambda}. \end{aligned}$$

4.2.7 Normal ^{DT SB}

The expected value of the normal is $E(N(\mu, \sigma^2)) = \int_{-\infty}^{+\infty} s \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds =$

$$\begin{aligned} &= \int_{-\infty}^{+\infty} (s - \mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds + \int_{-\infty}^{+\infty} \mu \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \left[-\sigma^2 e^{-\frac{(s-\mu)^2}{2\sigma^2}} \right]_{-\infty}^{+\infty} + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = 0 - 0 + \mu \cdot 1 = \mu \end{aligned}$$

remembering that the integral from $-\infty$ to $+\infty$ of the normal density is 1.

4.2.8 Uniform-squared ^{DT}

We remember that when a r.v. is squared, the square is applied only to values and never to probabilities or density functions. Therefore, this calculation is:

$$\begin{aligned} E(U^2(a, b)) &= \int_{-\infty}^{+\infty} s^2 f(s) ds = \int_{-\infty}^a s^2 0 ds + \int_a^b s^2 \frac{1}{b-a} ds + \int_b^{+\infty} s^2 0 ds = \\ &= 0 + \left[\frac{s^3}{3(b-a)} \right]_a^b + 0 = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}. \end{aligned}$$

4.2.9 Cauchy

Cauchy distribution is a rare example of a r.v. for which the expected value does not exist, since the integral does not converge. In fact, using integration by substitution

$$t = \frac{s-a}{\gamma}$$

$$\begin{aligned}
E(\xi) &= \int_{-\infty}^{+\infty} s f_{\xi}(s) ds = \int_{-\infty}^{+\infty} \frac{s}{\pi\gamma \left[1 + \left(\frac{s-a}{\gamma}\right)^2\right]} ds = \int_{-\infty}^{+\infty} \frac{s-a}{\pi\gamma \left[1 + \left(\frac{s-a}{\gamma}\right)^2\right]} ds + \int_{-\infty}^{+\infty} \frac{a}{\pi\gamma \left[1 + \left(\frac{s-a}{\gamma}\right)^2\right]} ds = \\
&= \int_{-\infty}^{+\infty} \frac{s-a}{\pi\gamma \left[1 + \left(\frac{s-a}{\gamma}\right)^2\right]} ds + a \cdot 1 = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{t}{1+t^2} dt + a = \frac{1}{\pi} \left[\frac{1}{2} \ln(1+s^2) \right]_{-\infty}^{+\infty} \rightarrow +\infty
\end{aligned}$$

The Cauchy distribution is said to have "heavy" or "fat" tails. That is, as the argument unboundedly increases in absolute value, its density does not vanish fast enough (as it happens for exponential and normal r.v.) to ensure the existence of an expected value.

5. Variance ^{DT SB}

The variance is a number associated to every r.v.. It is calculated as

$\text{Var}(\xi) = E((\xi - E(\xi))^2)$, and this formula becomes for discrete random variables

$\text{Var}(\xi) = \sum_{i=1}^n (x_i - E(\xi))^2 p_i$ and for continuous random variables

$\text{Var}(\xi) = \int_{-\infty}^{+\infty} (s - E(\xi))^2 f(s) ds$.

5.1 Rules ^{DT SB}

- $\text{Var}(\xi) > 0$, for any r.v. ξ ;
- $\text{Var}(c) = 0$, for any c constant real number;
- if ξ and ζ are independent random variables, then $\text{Var}(\xi + \zeta) = \text{Var}(\xi) + \text{Var}(\zeta)$.
Otherwise $\text{Var}(\xi + \zeta) = \text{Var}(\xi) + \text{Var}(\zeta) + 2\text{Cov}(\xi, \zeta)$ (see section 11 on page 25 for a definition of covariance);
- $\text{Var}(c\xi) = c^2 \text{Var}(\xi)$, for any ξ r.v. and c constant real number;
- $\text{Var}(\xi) = E(\xi^2) - (E(\xi))^2$ and in particular $\text{Var}(\xi) = E(\xi^2)$ when ξ has zero expected value.

5.2 Calculation

5.2.1 Bernoulli ^{DT SB}

$$\text{Var}(\xi) = \sum_{i=1}^n (x_i - E(\xi))^2 p_i = (1-p)^2 p + (0-p)^2 (1-p) = (1-p)((1+p)p - p^2) = (1-p)p$$

5.2.2 Die roll ^{DT}

$$\text{Var}(\xi) = E(\xi^2) - (E(\xi))^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182 - 147}{12} = \frac{35}{12}$$

5.2.3 Binomial ^{DT}

Using the first rule of the variance and remembering that the Binomial is a sum of independent Bernoulli random variables, we easily get:

$$\begin{aligned} \text{Var}(B(n, p)) &= \text{Var}(\xi_1 + \xi_2 + \dots + \xi_n) \stackrel{\text{independence}}{=} \text{Var}(\xi_1) + \text{Var}(\xi_2) + \dots + \text{Var}(\xi_n) = \\ &= p(1-p) + p(1-p) + \dots + p(1-p) = np(1-p). \end{aligned}$$

5.2.4 Uniform ^{DT}

Having already calculated the variance of the square of the uniform r.v., we get

$$\begin{aligned} \text{Var}(U(a, b)) &= E(U^2(a, b)) - (E(U(a, b)))^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \\ &= \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 3b^2 - 6ab}{12} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

5.2.5 Exponential ^{DT}

$$\text{Var}(\xi) = \int_{-\infty}^{+\infty} \left(s - \frac{1}{\lambda}\right)^2 f(s) ds = \int_{-\infty}^0 \left(s - \frac{1}{\lambda}\right)^2 0 ds + \int_0^{+\infty} \left(s - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda s} ds = 0 + \lambda \int_0^{+\infty} \left(s - \frac{1}{\lambda}\right)^2 e^{-\lambda s} ds =$$

$$\stackrel{\text{by parts}}{=} \lambda \left[\left(s - \frac{1}{\lambda}\right)^2 \frac{e^{-\lambda s}}{-\lambda} \right]_0^{+\infty} - \lambda \int_0^{+\infty} 2\left(s - \frac{1}{\lambda}\right) \frac{e^{-\lambda s}}{-\lambda} ds = \lambda \left[\lim_{s \rightarrow +\infty} \left(s - \frac{1}{\lambda}\right)^2 \frac{e^{-\lambda s}}{-\lambda} - \frac{e^0}{-\lambda^3} \right] + 2 \int_0^{+\infty} \left(s - \frac{1}{\lambda}\right) e^{-\lambda s} ds =$$

$$\stackrel{\text{by parts}}{=} \frac{1}{\lambda^2} + 2 \left[\left(s - \frac{1}{\lambda}\right) \frac{e^{-\lambda s}}{-\lambda} \right]_0^{+\infty} - 2 \int_0^{+\infty} \frac{e^{-\lambda s}}{-\lambda} ds = \frac{1}{\lambda^2} + 2 \left[\lim_{s \rightarrow +\infty} \left(s - \frac{1}{\lambda}\right) \frac{e^{-\lambda s}}{-\lambda} - \frac{e^0}{-\lambda^2} \right] + \frac{2}{\lambda} \int_0^{+\infty} e^{-\lambda s} ds =$$

$$= \frac{1}{\lambda^2} + 2 \frac{1}{\lambda^2} + \frac{2}{\lambda} \left[\frac{e^{-\lambda s}}{-\lambda} \right]_0^{+\infty} = \frac{3}{\lambda^2} - \frac{2}{\lambda^2} = \frac{1}{\lambda^2}.$$

5.2.6 Normal ^{DT}

$$\text{Var}(N(\mu, \sigma^2)) = \int_{-\infty}^{+\infty} (s - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (s - \mu) \cdot (s - \mu) e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds \stackrel{\text{by parts}}{=} .$$

$$\stackrel{\text{by parts}}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \left[(s - \mu) \cdot -\sigma^2 e^{-\frac{(s-\mu)^2}{2\sigma^2}} \right]_{-\infty}^{+\infty} - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} 1 \cdot -\sigma^2 e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds =$$

$$= 0 - 0 + \sigma^2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = \sigma^2 \cdot 1 = \sigma^2$$

remembering that the integral from $-\infty$ to $+\infty$ of the normal density is 1.

6. Momentum ^{SB}

The n -th momentum of r.v. ξ is defined as $m_n = E(\xi^n)$.

From this definition it is clear that the first momentum is the expected value and that the first two momentums are related to the variance with the formula

$$\text{Var}(\xi) = E(\xi^2) - (E(\xi))^2 = m_2 - m_1^2.$$

6.1 Skewness ^{SB}

Skewness of a r.v. is

$$\frac{E\left(\left(\xi - E(\xi)\right)^3\right)}{\sqrt{[\text{Var}(\xi)]^3}} = E\left(\left(\frac{\xi - E(\xi)}{\sqrt{\text{Var}(\xi)}}\right)^3\right)$$

A negative skewness characterizes a distribution skewed to the left (with a long slim tail on the left and a short fat tail on the right), while a positive skewness corresponds to a distribution skewed to the right. A symmetric distribution, like a normal one, has skewness equal to 0.

6.2 Kurtosis ^{SB}

Kurtosis of a r.v. is

$$\frac{E\left(\left(\xi - E(\xi)\right)^4\right)}{[\text{Var}(\xi)]^2} - 3 = E\left(\left(\frac{\xi - E(\xi)}{\sqrt{\text{Var}(\xi)}}\right)^4\right) - 3$$

The name comes from the Greek word $\kappaυρτ\acute{o}\varsigma$ (kyrtos or kurtos), meaning bulging. It is a measure of the "peakedness" of the probability distribution. A high Kurtosis has a sharper peak and longer, fatter tails, while a low Kurtosis has a more rounded peak and shorter thinner tails. The "minus 3" at the end of this formula is often explained as a correction to make the Kurtosis of the normal distribution equal to 0.

7. Median ^{DT SB}

The median is another number associated to every r.v.. Its definition is that number η which divides the probability distribution in exactly two parts of 50% each, i.e.

$$P(\xi < \eta) = 0.5.$$

For discrete random variables it is defined as that value x_k for which $\sum_{i=1}^k p_i = 0.5$. This value clearly may not exist, for example in the case of a Bernoulli r.v. with a value of p which is not 0.5, or may be not unique for example in the case of a die roll where every number between 3 and 4 is a median.

For continuous random variables its definition is $0.5 = P(\xi < \eta) = F(\eta) = \int_{-\infty}^{\eta} f(s) ds$ and,

for all the continuous random variables analyzed here it exists, and coincides with the expected value whenever the density function is symmetric, as in the uniform, normal and Cauchy cases. Also in the continuous case we can have not unique median, while we are sure that at least a median exists whenever the density function is regular enough.

7.1 Exponential^{DT}

$$0.5 = \int_{-\infty}^{\eta} f(s) ds = \int_{-\infty}^0 0 ds + \int_0^{\eta} \lambda e^{-\lambda s} ds = 0 + \lambda \int_0^{\eta} e^{-\lambda s} ds = \lambda \left[\frac{e^{-\lambda s}}{-\lambda} \right]_0^{\eta} = -e^{-\lambda \eta} + 1$$

Therefore $e^{-\lambda \eta} = 0.5$ and $-\lambda \eta = \ln 0.5$ and $\eta = -\frac{\ln 0.5}{\lambda} = \frac{\ln 2}{\lambda}$.

8. Lognormal r.v.^{DT}

Consider a normal r.v. with values x and density function $f(x)$. Now apply the $\exp()$ function to it (which means applying the $\exp()$ function to the x values and not to the density function) and obtain $\text{LN}(\mu, \sigma^2) = \exp(\text{N}(\mu, \sigma^2))$. Its values now lie between 0 and $+\infty$.

Being a continuous r.v. itself, it has a density function which can be calculated in this way:

$$\begin{aligned} \int_0^x f_{\text{LN}}(s) ds &= P(\text{LN} < x) = P(\exp(\text{N}) < x) = P(\text{N} < \ln x) = \int_{-\infty}^{\ln x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = \\ &= \int_{-\infty}^{\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds + \int_{\mu}^{\ln x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = 0.5 + \int_{\mu}^{\ln x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds, \end{aligned}$$

where the last equality is due to the fact that μ is exactly the median of $\text{N}(\mu, \sigma^2)$. Deriving with respect to x the first and the last element of the equalities' chain, we get

$$\frac{d}{dx} \int_0^x f_{\text{LN}}(s) ds = 0 + \frac{d}{dx} \int_{\mu}^{\ln x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds$$

and, remembering that the derivative of an integral with a fixed lower limit is the integral argument calculated in the upper limit multiplied by the derivative of the upper limit, we get

$$f_{\text{LN}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \cdot \frac{1}{x},$$

which is the density function of the lognormal r.v..

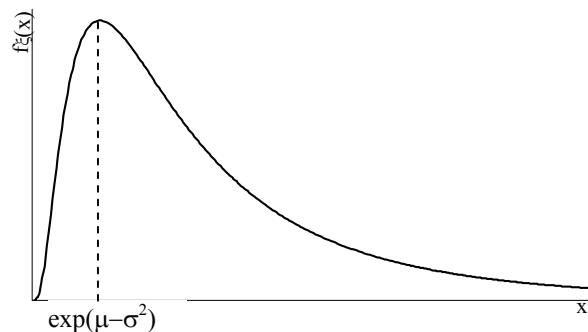
Therefore the lognormal r.v. can be seen in two perfectly equivalent ways:

- with values $\exp(x)$ and density

$$\text{function } \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}};$$

- with values x and density function

$$f_{\text{LN}}(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \frac{1}{x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$



Lognormal r.v. is used to model prices depending upon uncertain factors (like in the case of agricultural products, whose prices depend upon the weather conditions) and returns of risky assets (like the yield in one year time from 1 \$ invested today in stock).

8.1 Median^{DT}

Using the previous calculation it is very easy to get the value of the median of $\text{LN}(\mu, \sigma^2)$.

From $\int_0^x f_{\text{LN}}(s) ds = 0.5 + \int_{\mu}^{\ln x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds$ it is easy to see that if we fix x equal to $\exp(\mu)$,

we get exactly

$$\int_0^{\exp(\mu)} f_{\text{LN}}(s) ds = 0.5 + \int_{\mu}^{\ln \exp(\mu)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = 0.5 + \int_{\mu}^{\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = 0.5 + 0 = 0.5$$

and therefore the median is $\exp(\mu)$.

8.2 Expected value^{DT}

In order to calculate the expected value, we are going to calculate the n -th momentum using an integration technique which is common for every integration which involves the lognormal density function:

$$m_n(\text{LN}(\mu, \sigma^2)) = \int_0^{+\infty} s^n \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}} \frac{1}{s} ds = \int_0^{+\infty} e^{\ln s^n} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}} \frac{1}{s} ds =$$

$= \int_0^{+\infty} e^{n \ln s} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}} \frac{1}{s} ds = \int_0^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{-2\sigma^2 n \ln s + (\ln s - \mu)^2}{2\sigma^2}} \frac{1}{s} ds$. We now try to obtain at the nominator a perfect square:

$$\begin{aligned} -2\sigma^2 n \ln s + (\ln s - \mu)^2 &= -2\sigma^2 n \ln s + (\ln s)^2 + \mu^2 - 2\mu \ln s = \\ &= (\ln s)^2 - 2 \ln s (\sigma^2 n + \mu) + \mu^2 = [\ln s - (\sigma^2 n + \mu)]^2 - (\sigma^2 n + \mu)^2 + \mu^2 = \\ &= [\ln s - (\sigma^2 n + \mu)]^2 - \sigma^4 n^2 - 2\mu\sigma^2 n \text{ and we rewrite the integral as} \end{aligned}$$

$$m_n(\text{LN}(\mu, \sigma^2)) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[\ln s - (\sigma^2 n + \mu)]^2}{2\sigma^2}} e^{-\frac{-\sigma^4 n^2 - 2\mu\sigma^2 n}{2\sigma^2}} \frac{1}{s} ds = e^{\sigma^2 n^2 / 2 + \mu n} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[\ln s - (\sigma^2 n + \mu)]^2}{2\sigma^2}} \frac{1}{s} ds$$

and, remembering that $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[\ln s - (\sigma^2 n + \mu)]^2}{2\sigma^2}} \frac{1}{s}$ is the density for $\text{LN}(\sigma^2 n + \mu, \sigma^2)$ and therefore its integral is 1, we have $m_n(\text{LN}(\mu, \sigma^2)) = e^{\sigma^2 n^2 / 2 + \mu n}$ and $E(\text{LN}(\mu, \sigma^2)) = e^{\sigma^2 / 2 + \mu}$.

9. Standard normal ^{DT SB}

We call the normal r.v. $N(0,1)$, with $\mu = 0$ and $\sigma^2 = 1$, standard normal r.v.. Its

cumulative distribution function is indicated with $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$.

Let now calculate what is $\Phi\left(\frac{x - \mu}{\sigma}\right) = \int_{-\infty}^{\frac{x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$. Using the substitution $t = s\sigma + \mu$

(with $\frac{t - \mu}{\sigma} = s$, $dt = \sigma ds$ and considering that when s goes to $-\infty$ also t goes to $-\infty$ and when s is equal to $\frac{x - \mu}{\sigma}$ t is equal to x) we get

$$\Phi\left(\frac{x - \mu}{\sigma}\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} \frac{1}{\sigma} dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt,$$

which is exactly the cumulative distribution function for a generic normal random variable $N(\mu, \sigma^2)$. Therefore we obtained that:

$$F_{N(0,1)}\left(\frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) = F_{N(\mu, \sigma^2)}(x).$$

This result means to calculate the cumulative distribution of any normal r.v. from the cumulative distribution of the standard normal r.v. we simply need to subtract μ and divide by σ . This calculation procedure is called standardization.

Moreover, considering now the cumulative distribution of a lognormal

$$F_{LN(\mu, \sigma^2)}(x) = \int_{-\infty}^{\ln x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds \text{ we get immediately that } F_{LN(\mu, \sigma^2)}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),$$

which means that also the lognormal can be standardized (thus calculating its values from the cumulative distribution of the standard normal) applying the natural logarithm to x , subtracting μ and dividing by σ .

9.1 Calculating normal area ^{DT SB}

The analytical formula for the cumulative distribution function of normal r.v. is unfortunately still unknown but there are tables which print the values in the standard case calculated using computer approximation. These tables are similar to this one:

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
-3.0	0.0013	-2.0	0.0228	-1.0	0.1587	0.0	0.5000
-2.8	0.0026	-1.8	0.0359	-0.8	0.2119	0.2	0.5793
-2.6	0.0047	-1.6	0.0548	-0.6	0.2743	0.4	0.6554
-2.4	0.0082	-1.4	0.0808	-0.4	0.3446	0.6	0.7257
-2.2	0.0139	-1.2	0.1151	-0.2	0.4207	0.8	0.7881

As can be clearly seen from the table, since the normal distribution is symmetric with respect to 0, we have that $\Phi(-x) = 1 - \Phi(x)$. Therefore only half of the table is necessary since the other values can be calculated using this propriety.

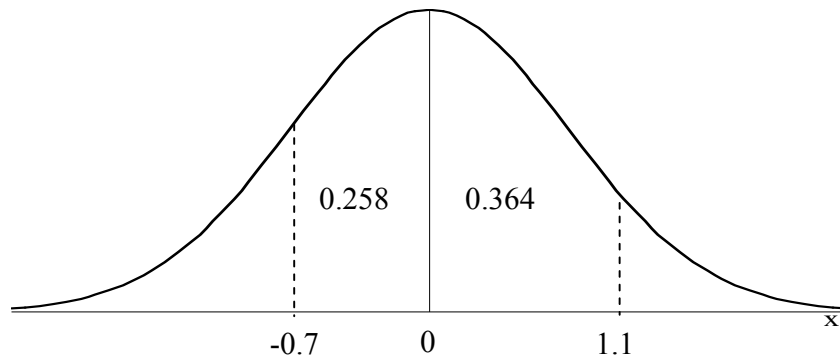
Another table more commonly found in statistics books displays the values of the standard normal distribution from 0 to x

x	$P_{N(0;1)}(0;x)$	x	$P_{N(0;1)}(0;x)$	x	$P_{N(0;1)}(0;x)$
0.0	0.000	1.0	0.341	2.0	0.477
0.1	0.040	1.1	0.364	2.1	0.482
0.2	0.079	1.2	0.385	2.2	0.486
0.3	0.118	1.3	0.403	2.3	0.489
0.4	0.155	1.4	0.419	2.4	0.492
0.5	0.191	1.5	0.433	2.5	0.494
0.6	0.226	1.6	0.445	2.6	0.495
0.7	0.258	1.7	0.455	2.7	0.497
0.8	0.288	1.8	0.464	2.8	0.497
0.9	0.316	1.9	0.471	2.9	0.498

For practical calculations it is easy to draw the standard normal density function and the available information on its areas.

For example, to get the value of $P_{N(0,1)}(-0.7 < x < 1.1)$ we can draw the normal density function and use the information that:

- area from 0 to +1.1 is 0.364 (taken from the table),
- area from -0.7 to 0 is the same as the area from 0 to +0.7 (symmetricity of the normal distribution), which is 0.258.



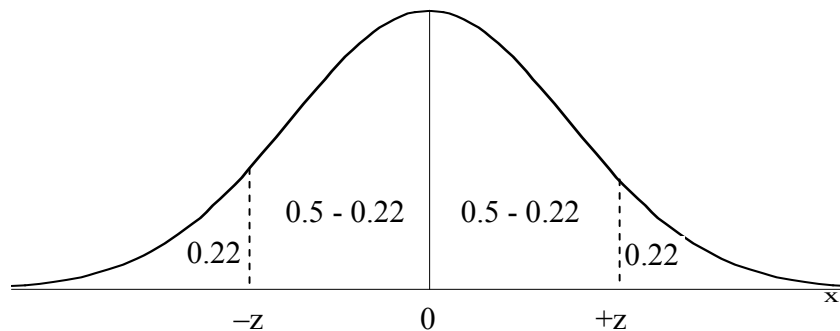
From the picture it is immediate to see that the considered area is 0.622.

For example, to get the value of $P_{N(0,1)}(x < -0.7)$ we can use the information that:

- area from minus infinity to 0 is 0.5 (since 0 is the median),
- area from -0.7 to 0 is 0.258 (symmetricity of the normal distribution),

and therefore the considered area is 0.242.

The same strategy can be applied for the reverse task, finding the x value corresponding to a given area. Suppose for example we know that the area of the left tail is 0.22; the picture in this case becomes



and we obtain the value $+z$ simply looking for the area corresponding to 0.28 in the table, which is approximately 0.8. Therefore the x value we were looking for is -0.8 .

If the normal is not standard, it is necessary to standardize the x values before using the table (or after using it in the case of the area is given and the x value is to be found) simply subtracting μ and dividing by σ . Apart from standardization, the rest of the calculation is exactly as for a standard normal.

For example, $P_{N(1;4)}(-1 < x < 2.3) = P_{N(0;1)}\left(\frac{-1-1}{2} < x < \frac{2.3-1}{2}\right) = P_{N(0;1)}(-1 < x < 0.65) \approx 0.341 + 0.23 = 0.571$.

Same procedure if we have to calculate a probability for the lognormal: we apply the logarithm to the values and we subtract μ and divide by σ . The rest of the calculation is exactly as for a standard normal.

For example, $P_{LN(1;9)}(-1 < x < 2.3) = P_{N(0;1)}\left(\frac{\ln 1 - 1}{3} < x < \frac{\ln 2.3 - 1}{3}\right) \approx P_{N(0;1)}(-0.33 < x < -0.056) \approx 0.12 - 0.02 = 0.1$ since the area from 0 to 0.33 is approximately 0.12 and the area from 0 to 0.056 is approximately 0.02.

10. Mode

The mode is the outcome with the highest probability, for discrete random variables, or the global maximum point of the density function for continuous random variables. While for all the random variables considered here it exists, there are many cases when it is not unique, for example, the die roll r.v. or the uniform r.v..

10.1 Normal

To calculate the mode of continuous random variables whose density function is differentiable we can study the first derivative which gives us information local maxima.

The first derivative of normal density function is

$$f'_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left(-\frac{2(x-\mu)}{2\sigma^2} \right)$$

The first two factors are always positive, so we study the sign of only $-\frac{x-\mu}{\sigma^2}$. This factor is positive for $x < \mu$ and negative for $x > \mu$. Therefore $x = \mu$ is a local maximum point. Since the function goes towards zero at $-\infty$ and $+\infty$, the local maximum is a global maximum.

10.2 Exponential

The exponential density function is discontinuous and therefore its first derivative is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ \lambda & x = 0 \\ 0 & x < 0 \end{cases} \quad f'(x) = \begin{cases} -\lambda^2 e^{-\lambda x} & x > 0 \\ \text{does not exist} & x = 0 \\ 0 & x < 0 \end{cases}$$

which means that for $x < 0$ the function is constantly 0 and for $x > 0$ the density is positive but decreasing, and therefore its global maximum value is λ taken in $x = 0$.

10.3 Cauchy

The first derivative of Cauchy density function is

$$f'(x) = \frac{-\pi\gamma^2 \frac{x-a}{\gamma}}{\pi^2 \gamma^2 \left[1 + \left(\frac{x-a}{\gamma}\right)^2\right]^2} = \frac{-2(x-a)}{\pi\gamma^2 \left[1 + \left(\frac{x-a}{\gamma}\right)^2\right]^2}$$

The denominator is always positive, so we study the sign of only $-2(x-a)$. This factor is positive for $x < a$ and negative for $x > a$. Therefore $x = a$ is a local maximum point. Since the function goes towards zero at $-\infty$ and $+\infty$, the local maximum is a global maximum.

10.4 Lognormal

The first derivative of lognormal density function, when $x > 0$, is

$$\begin{aligned} f'_{\text{LN}}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \left[e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \left(-\frac{2(\ln x - \mu)}{2\sigma^2} \right) \frac{1}{x} \cdot \frac{1}{x} + e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \cdot \left(-\frac{1}{x^2} \right) \right] = \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \frac{1}{x^2} \left[\left(-\frac{(\ln x - \mu)}{\sigma^2} \right) - 1 \right] \end{aligned}$$

The first three factors are always positive, so we study the sign of only $\left[\frac{-\ln x + \mu}{\sigma^2} - 1 \right]$. This factor is positive for $x < e^{\mu - \sigma^2}$ and negative for $x > e^{\mu - \sigma^2}$. Therefore $x = e^{\mu - \sigma^2}$ is a local maximum point. Since the function goes towards zero at 0 and $+\infty$, the local maximum is a global maximum.

11. Covariance ^{SB}

The covariance of two r.v. ξ and ε is defined as $Cov(\xi, \varepsilon) = E((\xi - E(\xi)) \cdot (\varepsilon - E(\varepsilon)))$.

In order to practically calculate the covariance, there are two equivalent ways:

1. when the joint probability distribution $P(\xi = x, \varepsilon = y)$ is known for every couple of (x, y) values of the two random variables, we can use a formula which considers every couple of values: $cov(\xi, \varepsilon) = \sum_{\substack{i=1, n \\ j=1, m}} (x_i - E(\xi))(y_j - E(\varepsilon))P(x_i, y_j)$;
2. when the space of outcomes Ω is known, we can use a formula which considers every outcome: $cov(\xi, \varepsilon) = \sum_{k=1, n} (\xi(\omega_k) - E(\xi))(\varepsilon(\omega_k) - E(\varepsilon))P(\omega_k)$.

For continuous r.v. which have the same density function² $f(s)$, for example because they are one a function of the other, the covariance formula is

$$\text{cov}(\xi, \varepsilon) = \int_{-\infty}^{+\infty} (\text{value of } \xi - E(\xi))(\text{value of } \varepsilon - E(\varepsilon))f(s) ds .$$

11.1 Rules

- $\text{Cov}(\xi, \xi) = \text{Var}(\xi)$, for any ξ random variable;
- $\text{Cov}(\xi, \zeta) = E(\xi \cdot \zeta) + E(\xi) \cdot E(\zeta)$, for any ξ and ζ random variables;
- $\text{Var}(\xi + \varepsilon) = \text{Var}(\xi) + \text{Var}(\varepsilon) + 2\text{Cov}(\xi, \varepsilon)$, for any ξ and ζ random variables;
- if ξ and ζ are independent (see sections 2.3 and 3.3 for definition) then $\text{Cov}(\xi, \zeta) = 0$. Note that the reverse is not generally true.

11.2 Correlation^{SB}

Correlation is defined as $\text{corr}(\xi, \varepsilon) = \frac{\text{Cov}(\xi, \varepsilon)}{\sqrt{\text{Var}(\xi) \cdot \text{Var}(\varepsilon)}}$. Correlation varies from -1

(completely negatively correlated) to 0 (uncorrelated) and up to $+1$ (completely correlated).

Note that independent r.v. are uncorrelated, but the reverse is not generally true.

12. Central limit theorem^{SB}

If we have a series of n equal independent random variables ξ_j , the sum of these random variables is, when n is large enough, approximately normally distributed with $\mu = nE(\xi_j)$ and $\sigma^2 = n\text{Var}(\xi_j)$. A similar result clearly holds true for the average of these r.v. (which is the sum divided by n), with $\mu = E(\xi_j)$ and $\sigma^2 = \text{Var}(\xi_j)/n$.

$$\sum_{j=1}^n \xi_j \approx N(nE(\xi_j); n\text{Var}(\xi_j)) \quad \text{Average}(\xi_j) \approx N(E(\xi_j); \text{Var}(\xi_j)/n)$$

Using a standardization procedure, we can derive more practically useful approximations:

$$\frac{\sum_{j=1}^n \xi_j - nE(\xi_j)}{\sqrt{n\text{Var}(\xi_j)}} \approx N(0;1) \quad \text{and} \quad \frac{\text{Average}(\xi_j) - E(\xi_j)}{\sqrt{\frac{\text{Var}(\xi_j)}{n}}} \approx N(0;1).$$

² In case they do not have the same density function, the formula needs the joint distribution density function and becomes more complicated: $\text{Cov}(\xi, \varepsilon) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (s - E(\xi))(t - E(\varepsilon))f_{\xi, \varepsilon}(s, t) dt ds$.

These formulas replace the distribution of the sum and the distribution of the average in practical calculations whenever n is large, when it is very long to calculate the effective distributions.

Since the Binomial is a sum of equal independent Bernoulli r.v., it also can be approximated with the formula for the sum $\frac{B(n, p) - np}{\sqrt{np(1-p)}} \approx N(0;1)$.

How large n should be for the approximation to be good depends on the distribution of the ξ_j . If they are already normally distributed, the approximation is good for any value of n , since the sum of normal r.v. is a normal r.v. Instead, to approximate the Binomial distribution we need that these four conditions be respected $0 < p \pm 3\sqrt{\frac{p(1-p)}{n}} < 1$.