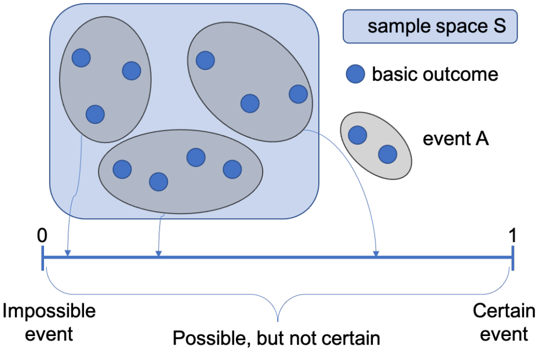
UNIVERSITÀ DEGLI STUDI DI TRIESTE

Statistics

Random Variables

Sara Geremia

April 5th, 2024



Probability model

A **probability model** is a mathematical description of a random experiment consisting of a sample space and a way of assigning probabilities to events

Random variable

A **random variable** (r.v.) X is a variable whose value represents a numerical outcome of a random phenomenon; that is, it is a well-defined but unknown number

- the number of tails on three coin tosses
- the number of defective items in a sample of 20 items from a large shipment
- the number of students attending the statistics class on Friday
- the delay time of the airplane
- the weight of a newborn
- the duration of a phone call with your mother

Random Variable

Probability distribution

The **probability distribution** of a random variable X tells us what values X can take and how to assign probabilities to those values

$$P(x) = P(X = x), \forall x$$

- the number of tails on tree coin tosses: $X : \{0, 1, 2, 3\}$ and each value x has probability $P(X = x)$

There are two main types of random variables: discrete if it has a finite list of possible outcomes, and continuous if it can take any value in an interval.

- D the number of tails on three coin tosses
- D the number of defective items in a sample of 20 items from a large shipment
- D the number of students attending the statistics class on Friday
- C the delay time of the airplane
- C the weight of a newborn
- C the duration of a phone call with your mother

There are two main types of random variables: discrete if it has a finite list of possible outcomes, and continuous if it can take any value in an interval.

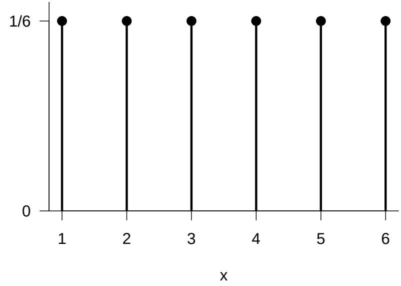
- D the number of tails on three coin tosses
- D the number of defective items in a sample of 20 items from a large shipment
- D the number of students attending the statistics class on Friday
- C the delay time of the airplane
- C the weight of a newborn
- C the duration of a phone call with your mother

For continuous random variables we can assign probabilities only to a range of values, using a mathematical function. This allows us to calculate the probability of events such as "today's high temperature will be between 25° and 26°."

Discrete Random Variables

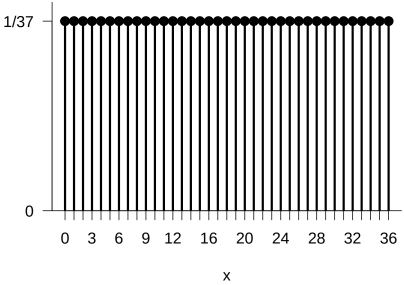
X = rolling a dice

x	P
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$



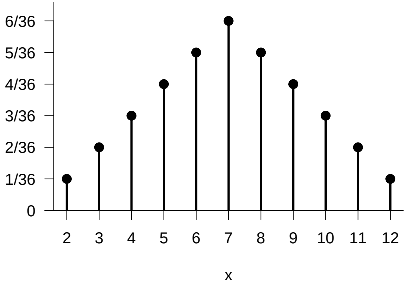
Y = roulette result

y	P
0	$1/37$
1	$1/37$
2	$1/37$
...	...
35	$1/37$
36	$1/37$



$Z =$ sum the results of rolling two dice

z	P
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$



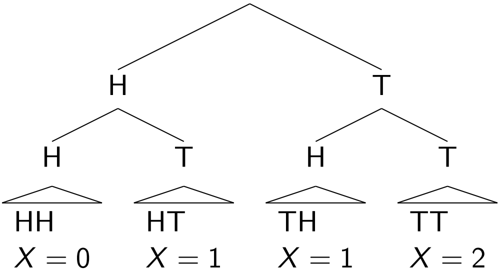
Number of tails on two flips of a coin

We toss a coin two times, then we sum the number of tails T

$X =$ number of tails in flipping a coin two times

X is a discrete random variable that can assume values: $\{0, 1, 2\}$

The random experiment is represented in the tree diagram:



4 possible outcomes $= 2^2$

Number of tails on two flips of a coin

We toss a coin two times, then we sum the number of tails T

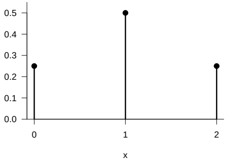
$X =$ number of tails in tossing a coin two times

X is a discrete random variable that can assume values: $\{0, 1, 2\}$

Given a fair coin, the probability distribution of X is

outcomes	P	X
HH	$1/4$	0
HT	$1/4$	1
TH	$1/4$	1
TT	$1/4$	2

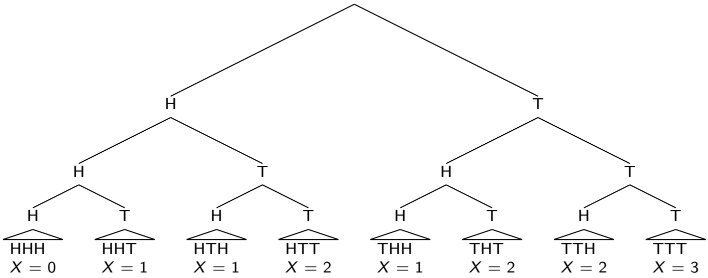
X	P
0	$P(HH) = 1/4$
1	$P(HT \cup TH) = 1/2$
2	$P(TT) = 1/4$



Number of tails on three flips of a coin

$X =$ number of tails in tossing a coin three times

X is a discrete random variable that can assume values: $\{0, 1, 2, 3\}$



8 possible outcomes $= 2^3$

Number of tails on three flips of a coin

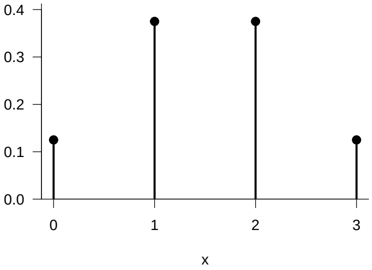
X = number of tails in tossing a coin three times

X is a discrete random variable that can assume values: $\{0, 1, 2, 3\}$

Assuming a fair coin, the probability distribution of X is

outcomes	P	X
HHH	1/8	0
HHT	1/8	1
HTH	1/8	1
THH	1/8	1
HTT	1/8	2
TTH	1/8	2
THT	1/8	2
TTT	1/8	3

X	P
0	1/8
1	3/8
2	3/8
3	1/8



$$P(X = 2) = P(HTT \cup TTH \cup THT) = P(HTT) + P(TTH) + P(THT)$$

Number of tails on n flips of a coin

$X =$ number of tails in tossing a coin n times

X is a discrete random variable that can assume values: $\{0, 1, 2, \ldots, n\}$

There are 2^n possible outcomes

a generic outcome	T	H	T	T	H	T	$\ldots$	T	H
flip	1	2	3	4	5	6	...	$n-1$	n

Given a fair coin, each outcome (sequence of n trials) has probability $\left(\frac{1}{2}\right)^n$

To compute $P(X = x)$ we have to count how many outcomes with x tails we can obtain in the random experiment:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Then, the probability distribution is:

$$P(X = x) = \binom{n}{x}\left(\frac{1}{2}\right)^n$$

Number of tails on n flips of a coin

$X =$ number of tails in tossing a coin n times

X is a discrete random variable that can assume values: $\{0, 1, 2, \ldots, n\}$

There are 2^n possible outcomes

a generic outcome	T	H	T	T	H	T	$\ldots$	T	H
flip	1	2	3	4	5	6	...	$n-1$	n
prob	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	...	$\frac{1}{2}$	$\frac{1}{2}$

Given a fair coin, each outcome (sequence of n trials) has probability $\left(\frac{1}{2}\right)^n$

To compute $P(X = x)$ we have to count how many outcomes with x tails we can obtain in the random experiment:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Then, the probability distribution is:

$$P(X = x) = \binom{n}{x}\left(\frac{1}{2}\right)^n$$

What if the coin is not fair?

We call $P(T) = p$, then $P(H) = 1 - p$
The number of possible outcomes is still 2^n

a generic outcome	T	H	T	T	H	T	$\dots$	T	H
flip	1	2	3	4	5	6	...	$n-1$	n

but the probability for a tail to occur is different from the probability for a head

It depends on the number of tails and the number of heads

$$P(THTT \dots TH) = p^{\#T}(1-p)^{n-\#T}$$

The $\binom{n}{x}$ sequences with x number of tails have all the same probability:

$$p^x(1-p)^{n-x}$$

Then, the probability distribution is:

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

What if the coin is not fair?

We call $P(T) = p$, then $P(H) = 1 - p$

The number of possible outcomes is still 2^n

a generic outcome	T	H	T	T	H	T	$\dots$	T	H
flip	1	2	3	4	5	6	...	$n-1$	n
prob	p	$1-p$	p	p	$1-p$	p	...	p	$1-p$

but the probability for a tail to occur is different from the probability for a head

It depends on the number of tails and the number of heads

$$P(THTT\dots TH) = p^{\#T}(1-p)^{n-\#T}$$

The $\binom{n}{x}$ sequences with x number of tails have all the same probability:

$$p^x(1-p)^{n-x}$$

Then, the probability distribution is:

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

Binomial distribution

Binomial distribution

A random variable X follows the binomial distribution with dimension $n \in \mathbb{N}$ and parameter $p \in [0, 1]$

$$X \sim \text{Binom}(n, p)$$

if $X \in \{0, 1, \ldots, n\}$ and

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

$X \sim \text{Binom}(n, p)$ is the number of successes in n independent trials with success probability p

- the number of observations/trials n is fixed
- the n observations are independent
- each observation can be a success or a failure

Blood Types

Genetic says that children receive genes from their parents independently

Each child of a particular pair of parents has a probability 0.25 of having type "0" blood

If these parents have 5 children, the number who have type "0" blood is the count X of successes in 5 independent observations with probability 0.25 of success in each observation

So X has the Binomial distribution with $n = 5$ and $p = 0.25$

$$X \sim \text{Binom}(5, 0.25)$$

$$P(X = x) = \binom{5}{x} 0.25^x (1 - 0.25)^{5-x}$$

Blood Types

X has the Binomial distribution with $n = 5$ and $p = 0.25$

$$X \sim \text{Binom}(5, 0.25)$$

$$P(X = x) = \binom{5}{x} 0.25^x (1 - 0.25)^{5-x}$$

What is the probability that two children have type "0" blood?

$$P(X = 2) = \binom{5}{2} 0.25^2 (1 - 0.25)^{5-2}$$

What is the probability that more than 4 children have type "0" blood?

$$P(X > 4) = P(X = 5) = 0.25^5$$

Overbooking

A small airline accepts reservations for a flight with 20 seats and knows that of the people who book a trip 10% do not show up

a) If there have been 20 reservations, what is the probability that the flight will be full?

The company accepts more than 20 reservations (*overbooking*), hoping that no more than 20 people show up.

b) If 22 bookings have been accepted, what is the probability that some booked passengers will stay ashore?

c) How many bookings can you accept if you want the probability of a passenger to be ashore to be less than 15%?

a) If there have been 20 reservations, what is the probability that the flight will be full?

Let X be the number of passengers that might show up at the airport

The number of reservations is $n = 20$

The probability to show up for a passenger is $p = 0.9$

$$X \sim \text{Binom}(20, 0.9)$$

$$P(X = x) = \binom{20}{x} 0.9^x (1 - 0.9)^{20-x}$$

The plane will travel full if $X = 20$, that is

$$P(X = 20) = \binom{20}{20} 0.9^{20} (1 - 0.9)^{20-20} = 0.9^{20} = 0.1216$$

b) If 22 bookings have been accepted, what is the probability that some booked passengers will stay ashore?

X is still the number of passengers that might show up at the airport

The number of reservations is $n = 22$

The probability to show up for a passenger is $p = 0.9$

$$X \sim \text{Binom}(22, 0.9)$$

$$P(X = x) = \binom{22}{x} 0.9^x (1 - 0.9)^{22-x}$$

No passengers will remain ashore if $X \leq 20$

$$\begin{aligned} P(X > 20) &= P(X = 21) + P(X = 22) \\ &= \binom{22}{21} 0.9^{21} (1 - 0.9)^{22-21} + 0.9^{22} \\ &= 0.3392 \end{aligned}$$

c) How many bookings can you accept if you want the probability of a passenger to be ashore to be less than 15%?

We have computed the probability that at least one passenger remains ashore accepting 22 reservations is $0.3392 > 0.15$

If we increase the number of reservations, even the probability that at least one passenger remains ashore increases

$$X \sim \text{Binom}(21, 0.9)$$

$$P(X = 21) = \binom{21}{21} 0.9^{21} = 0.1095$$

If the company accept 21 reservations, the probability that a passenger remains ashore is about 11%

College admissions

Early in August, an undergraduate college discovers that it can accommodate a few extra students.
Enrolling those additional students would provide a substantial increase in revenue without increasing the operating costs of the college; that is, no new classes would have to be added.
From past experience the college knows that the frequency of enrollment given admission for all students is 40%

a) What is the probability that at most 6 students will enrol if the college offers admission to 10 more students?
b) What is the probability that more than 12 will enrol if admission is offered to 20 students?
c) If the frequency of enrollment given admission for all students was 70%, what is the probability that at least 12 out of 15 students will enrol?

Probability mass function with countably finite support

Probability mass function with countably finite support

Given a random variable X with finite support $\{x_1, x_2, \ldots, x_n\}$, we define the probability mass function of the rv X

$$P(X = x_i) = p(x_i), \forall i$$

such that

i. $p(x_i) \geq 0$
ii. $\sum_{i=1}^{n} p(x_i) = 1$

Probability mass function with countably infinite support

Probability mass function with countably infinite support

Given a random variable X that assumes a countably infinite set of values $\{x_1, x_2, \ldots, x_n, \ldots\}$, we define its probability mass function as

$$P(X = x_i) = p(x_i), \forall i$$

such that

i. $p(x_i) \geq 0$

ii. $\sum_{i=1}^{\infty} p(x_i) = 1$ (that is, the series must converge to 1)

Cumulative distribution function - discrete rv

Cumulative distribution function - discrete rv

Given a random variable X that assumes a countably infinite set of values $x_1, \ldots, x_n, \ldots$ and with probability mass function $p(x)$, we define the cumulative distribution function of X as

$$F(x) = P(X \leq x) = \sum_{i: x_i \leq x} p(x_i)$$

The cumulative distribution function represents the probability that X does not exceed the value x

i. $F(x) \geq 0, \quad \forall x \in \mathbb{R}$;
ii. $F(x)$ is non decreasing;
iii. $\lim_{x \to -\infty} F(x) = 0$;
iv. $\lim_{x \to +\infty} F(x) = 1$.

Assume that X is a discrete random variable that follows a Binomial distribution with $n = 4$ and $p = 0.4$, then

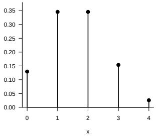
$$X \in \{0, 1, 2, 3, 4\}$$

and the probability mass function of X is

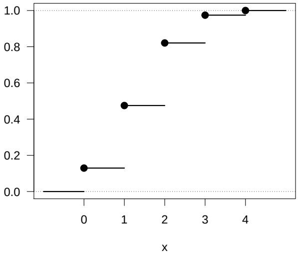
$$P(X = x_i) = \binom{4}{x_i} p^{x_i} (1 - p)^{4 - x_i}$$

x_i	p_i	F_i
0	0.12960	0.1296
1	0.34560	0.4752
2	0.34560	0.8208
3	0.15360	0.9744
4	0.02560	1.0000

Probability mass function



Cumulative distribution function



In order to obtain a measure of the center of a probability distribution, we introduce the notion of the expectation of a random variable

You know the sample mean as a measure of central location for sample data

The expected value is the corresponding measure of central location for a random variable

In order to obtain a measure of the center of a probability distribution, we introduce the notion of the expectation of a random variable

You know the sample mean as a measure of central location for sample data

The expected value is the corresponding measure of central location for a random variable

Let X be the number of errors on a page chosen at random from business area textbooks, from a review we found that 81% of all pages were error-free ($X = 0$), 17% of all pages contained one error ($X = 1$), and the remaining 2% contained two errors ($X = 2$).
Thus, the probability mass function of the variable X is

$$p(0) = 0.81, \quad p(1) = 0.17, \quad p(2) = 0.02$$

Let X be the number of errors on a page chosen at random from business area textbooks, from a review we found that 81% of all pages were error-free ($X = 0$), 17% of all pages contained one error ($X = 1$), and the remaining 2% contained two errors ($X = 2$).
Thus, the probability mass function of the variable X is

$$p(0) = 0.81, \quad p(1) = 0.17, \quad p(2) = 0.02$$

What is the expected value of X?
In computing the average number of possible values, $E(X) = (0 + 1 + 2)/3 = 1$, we are ignoring how each value is likely to occur (assuming the same probability on each value)

$$E(X) = 0 \cdot 0.81 + 1 \cdot 0.17 + 2 \cdot 0.02 = \sum_x xp(x) = 0.21$$

Expectation

The expected value E(X), of a discrete random variable X is defined as

$$E(X) = \mu = \sum_{i=1}^{\infty} x_i p(x_i)$$

Using the definition of relative frequency probability, we can view the expected value of a rv as the long-run weighted average value that it takes over a large number of trials

Variance

The variance V(X), of a discrete random variable X is defined as the expectation of the squared deviations about the mean, $(X - E(X))^2$

$$V(X) = \sigma^2 = E[(X - E(X))^2] = \sum_{i=1}^{\infty} (x_i - E(x))^2 p(x_i)$$

$$V(X) = E(X^2) - [E(X)]^2$$

The standard deviation σ is the positive square root of the variance

Binomial: expected value and variance

It can be shown that for a Binomial rv X with dimension n and probability p, that is

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

then

$$E(X) = np$$
$$V(X) = np(1-p)$$

Binomial: expected value and variance

It can be shown that for a Binomial rv X with dimension n and probability p, that is

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

then

$$E(X) = np$$
$$V(X) = np(1-p)$$

Overbooking example:

A small airline accepts reservations for a flight with 20 seats and knows that of the people who book a trip 10% do not show up

What is the expected number of people that show up at the airport?

Assuming $X \sim Binom(20, 0.9)$

$$E(X) = np = 20 \cdot 0.9 = 18$$

Linear transformations

We defined random variables as numbers, arithmetical operations are allowed
e.g. given a random variable X we can define a new rv Y applying a linear transformation

$$Y = aX + b$$

The values that the rv Y can assume and its probability distribution are derived from the ones of X
If X assumes values $\{x_i\}$, then $Y = aX + b$ assumes values $\{ax_i + b\}$, and the probability distribution of Y is

$$P(Y = ax_i + b) = P(X = x_i)$$

Also,

$$E(Y) = E(aX + b) = aE(X) + b$$

$$V(Y) = V(aX + b) = a^2 V(X)$$

Standardization

Given a rv X with mean $\mu = E(X)$ and variance $\sigma^2 = V(X)$, the standardization is the linear transformation

$$Z = \frac{X - \mu}{\sigma}$$

such that, the rv Z has a mean equal to 0 and variance (and standard deviation) equal to 1

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{E(X) - \mu}{\sigma} = 0$$

$$V(Z) = V\left(\frac{X - \mu}{\sigma}\right) = \frac{V(X)}{\sigma^2} = 1$$

the linear transformation $Z = a + Xb$, a and b are defined as: $a = -\frac{\mu}{\sigma}$ and $b = 1/\sigma$

Examples

Given the rv X with the probability mass function in the table below, obtain the probability mass functions for the random variables

$$Y = |X|, \quad W = X^3 \quad Z = (X - \mu)/\sigma$$

Compute the expected value and variance of X, Y, W, and Z

x_i	p_i
-3	0.1
-1	0.1
0	0.2
1	0.2
3	0.4

$E(X) = 1$
$V(X) = 3.8$

Examples

Given the rv X with the probability mass function in the table below, obtain the probability mass functions for the random variables

$$Y = |X|, \quad W = X^3 \quad Z = (X - \mu)/\sigma$$

Compute the expected value and variance of X, Y, W, and Z

x_i	p_i
-3	0.1
-1	0.1
0	0.2
1	0.2
3	0.4

y_i	p_i
0	0.2
1	0.3
3	0.5

w_i	p_i
-27	0.1
-1	0.1
0	0.2
1	0.2
27	0.4

z_i	p_i
-2.05	0.1
-1.03	0.1
-0.51	0.2
0	0.2
1.03	0.4

$E(X) = 1$
$V(X) = 3.8$

$E(Y) = 1.8$
$V(Y) = 1.56$

$E(W) = 8.2$
$V(W) = 297.6$

$E(Z) = 0$
$V(Z) = 1$

- The number of failures in a large computer system during a given day
- The number of replacement orders for a part received by a firm in a given month
- The number of ships arriving at a loading facility during a 6-hour loading period
- The number of delivery trucks to arrive at a central warehouse in an hour
- The number of customers to arrive at a checkout aisle in your local grocery store during a particular time interval

- The number of failures in a large computer system during a given day
- The number of replacement orders for a part received by a firm in a given month
- The number of ships arriving at a loading facility during a 6-hour loading period
- The number of delivery trucks to arrive at a central warehouse in an hour
- The number of customers to arrive at a checkout aisle in your local grocery store during a particular time interval

All the random phenomena above describe the number of independent occurrences (successes) on a given interval of time

Poisson distribution

Poisson distribution

A random variable $X \in \{0, 1, 2, \ldots, n, \ldots\}$ follows a Poisson distribution with parameter λ if and only if

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

$X \sim Poisson(\lambda)$ is the number of occurrences/successes of a certain event in a given continuous interval (such as time, surface area, or length)

- assume that the interval is divided into a large number of equal subintervals each with a very small probability of occurrence of an event
- the probability of the occurrence of an event is constant for all subintervals
- there can be no more than one occurrence in each subinterval
- occurrences are independent; that is, an occurrence in one interval does not influence the probability of an occurrence in another interval

Poisson: expected value and variance

It can be shown that for a Poisson rv X with parameter λ, that is

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

then

$$E(X) = \lambda, \qquad V(X) = \lambda$$

Thus λ represents the expected number of successes per space unit and it can assume only positive values

$\lambda = 1$

x_i	p_i
0	0.36788
1	0.36788
2	0.18394
3	0.06131
4	0.01533
5	0.00307
6	0.00051
7	0.00007
8	0.00001
9	0.00000
10	0.00000
> 10	0.00000

$E(X) = 1$

$V(X) = 1$

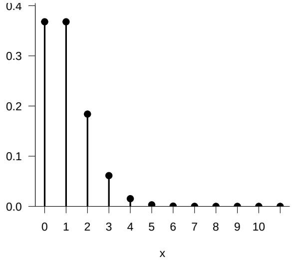
$\lambda = 0.2$

x_i	p_i
0	0.81873
1	0.16375
2	0.01637
3	0.00109
4	0.00005
5	0.00000
6	0.00000
7	0.00000
8	0.00000
9	0.00000
10	0.00000
> 10	0.00000

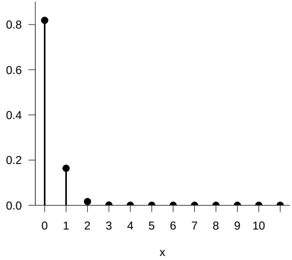
$E(X) = 0.2$

$V(X) = 0.2$

$\lambda = 1$



$\lambda = 0.2$



Football

A football team scores a number of goals per game that is assumed to be distributed as a Poisson distribution and on average, the team scores 1.5 goals per game

1. Compute the probability that in the next game, the number of goals by the football team is 0
2. Compute the probability that in the next game, the number of goals by the football team is greater than 4

Football

A football team scores a number of goals per game that is assumed to be distributed as a Poisson distribution and on average, the team scores 1.5 goals per game

1. Compute the probability that in the next game, the number of goals by the football team is 0
2. Compute the probability that in the next game, the number of goals by the football team is greater than 4

The number of goals per game follows a Poisson distribution with parameter $\lambda = 1.5$, thus

$$P(X = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-\lambda} = 0.2231$$

Football

A football team scores a number of goals per game that is assumed to be distributed as a Poisson distribution and on average, the team scores 1.5 goals per game

1. Compute the probability that in the next game, the number of goals by the football team is 0
2. Compute the probability that in the next game, the number of goals by the football team is greater than 4

$$
\begin{aligned}
P(X > 4) &= P\left(\bigcup_{i=5}^{+\infty}(X = i)\right) = \sum_{i=5}^{+\infty} \frac{\lambda^i}{i!} e^{-\lambda} \\
&= 1 - \sum_{i=0}^{4} \frac{\lambda^i}{i!} e^{-\lambda} = 1 - 0.9814 = 0.01858
\end{aligned}
$$

Uniform discrete distribution

Uniform discrete distribution

A random variable $X \in \{a, a+1, a+2, \ldots, b-2, b-1, b\}$ follows a discrete Uniform distribution on the interval $[a, b]$ if and only if its probability mass function is

$$P(X = x) = \frac{1}{n}$$

where a and b are integer numbers such that $a \leq b$ and $n = b - a + 1$

If $X \sim Unif\{a, b\}$ all the values of the support are equally likely to be observed

The expected value and variance of a rv X following a discrete Uniform distribution are respectively

$$E(X) = \frac{a+b}{2} \qquad V(X) = \frac{n^2 - 1}{12}$$

Toss a coin

If we flip a coin three times we can define several random variables such as
S = number of tails
M = number of tails before the first head

Outcome	P	S	M
HHH	1/8	0	0
HHT	1/8	1	0
HTH	1/8	1	0
THH	1/8	1	1
HTT	1/8	2	0
TTH	1/8	2	2
THT	1/8	2	1
TTT	1/8	3	3

S	P
0	1/8
1	3/8
2	3/8
3	1/8

M	P
0	4/8
1	2/8
2	1/8
3	1/8

We now want to look at them jointly, so we consider events as

$$(S = s) \cap (M = m)$$

Toss a coin

If we flip a coin three times we can define several random variables such as
$S =$ number of tails
$M =$ number of tails before the first head

Outcome	P	S	M
HHH	1/8	0	0
HHT	1/8	1	0
HTH	1/8	1	0
THH	1/8	1	1
HTT	1/8	2	0
TTH	1/8	2	2
THT	1/8	2	1
TTT	1/8	3	3

		S			
		0	1	2	3
M	0	HHH	HHT HTH	HTT	-
	1	-	THH	THT	-
	2	-	-	TTH	-
	3	-	-	-	TTT

We now want to look at them jointly, so we consider events as

$$(S = s) \cap (M = m)$$

Toss a coin

If we flip a coin three times we can define several random variables such as
S = number of tails
M = number of tails before the first head

Outcome	P	S	M
HHH	1/8	0	0
HHT	1/8	1	0
HTH	1/8	1	0
THH	1/8	1	1
HTT	1/8	2	0
TTH	1/8	2	2
THT	1/8	2	1
TTT	1/8	3	3

		S			
		0	1	2	3
M	0	1/8	2/8	1/8	0
	1	0	1/8	1/8	0
	2	0	0	1/8	0
	3	0	0	0	1/8

We now want to look at them jointly, so we consider events as

$$(S = s) \cap (M = m)$$

Joint probability distribution

Let X and Y be pair of discrete random variables with values $\{x_1, \ldots\}$ and $\{y_1, \ldots\}$. Their joint probability distribution expresses the probability that simultaneously X takes the specific value x_i, and Y takes the value y_j.

$$p(x_i, y_j) = P(X = x_i \cap Y = y_j)$$

It is easy to see the link with the probability of the intersection of bivariate events.
Properties:

- $p(x_i, y_j) \geq 0$
- $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(x_i, y_j) = 1$

In this framework, the probability distribution of the random variable X is called its marginal probability distribution and is obtained by summing the joint probabilities over all possible values—that is,
$p_X(x_i) = P(X = x_i) = \sum_{j=1}^{\infty} p(x_i, y_j)$

The table that contains all the joint probabilities is the joint probability function of the pair (M, S), i.e. the function

$$p(m, s) = P((M = m) \cap (S = s))$$

		S: 0	1	2	3	
M	0	1/8	2/8	1/8	0	4/8
	1	0	1/8	1/8	0	2/8
	2	0	0	1/8	0	1/8
	3	0	0	0	1/8	1/8
		1/8	3/8	3/8	1/8	1

$p_M(0) = P(M = 0) = \sum_{j=1}^{4} p(0, y_j) = \frac{4}{8}$
$p_M(1) = P(M = 1) = \sum_{j=1}^{4} p(1, y_j) = \frac{2}{8}$
$p_M(2) = P(M = 2) = \sum_{j=1}^{4} p(2, y_j) = \frac{1}{8}$
$p_M(3) = P(M = 3) = \sum_{j=1}^{4} p(3, y_j) = \frac{1}{8}$

Conditional probability distribution

Let X and Y be a pair of jointly distributed discrete random variables with values $\{x_1, \ldots\}$ and $\{y_1, \ldots\}$. The conditional probability distribution of the random variable Y, given that the random variable X takes the value x,

$$p_{Y|X=x_i}(y_j|x_i) = \frac{p(x_i, y_j)}{p_X(x_i)}$$

where $p(x_i, y_j)$ is the joint probability distribution of the two variables and $p_X(x_i)$ is the marginal probability distribution.

It expresses the probability that Y takes the value y, as a function of y, when the value x is fixed for X.

It is the well-known conditional probability

$$p_{Y|X=x_i}(y_j|x_i) = P(Y = y_j|X = x_i) = \frac{P((X = x_i) \cap (Y = y_j))}{P(X = x_i)} = \frac{p(x_i, y_j)}{p_X(x_i)}$$

Conditional probability distribution

Let X and Y be a pair of jointly distributed discrete random variables with values $\{x_1, \ldots\}$ and $\{y_1, \ldots\}$. The conditional probability distribution of the random variable Y, given that the random variable X takes the value x,

$$p_{Y|X=x_i}(y_j|x_i) = \frac{p(x_i, y_j)}{p_X(x_i)}$$

where $p(x_i, y_j)$ is the joint probability distribution of the two variables and $p_X(x_i)$ is the marginal probability distribution.

It expresses the probability that Y takes the value y, as a function of y, when the value x is fixed for X.

Similarly for $X|Y = y_j$

$$p_{X|Y=y_j}(x_i|y_j) = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

$$p_{M|S}(m|s) = P((M = m)|(S = s))$$

		S				
		0	1	2	3	
M	0	1	2/3	1/3	0	-
	1	0	1/3	1/3	0	-
	2	0	0	1/3	0	-
	3	0	0	0	1	-
		1	1	1	1	-

Each column is a conditional probability distribution.

$$p_{S|M}(s|m) = P((S = s)|(M = m))$$

		S				
		0	1	2	3	
M	0	1/4	1/2	1/4	0	1
	1	0	1/2	1/2	0	1
	2	0	0	1	0	1
	3	0	0	0	1	1
		-	-	-	-	-

Each row is a conditional probability distribution.

Joint expected value

Let X and Y be two random variables assuming values $\{x_1, \dots\}$ and $\{y_1, \dots\}$ with joint probability function $p(x, y)$, we define the joint expected value as

$$E(h(X, Y)) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} h(x_i, y_j) p(x_i, y_j)$$

where h is a known function.

In particular, we define the covariance as

$$\begin{aligned} \text{cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (x_i - E(X))(y_j - E(Y)) p(x_i, y_j) \end{aligned}$$

The covariance is a measure of linear association between two random variables and represents the joint variability of two random variables.

Covariance: properties

We define the covariance as

$$\begin{aligned}\text{cov}(X,Y) &= E((X-E(X))(Y-E(Y))) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i-E(X))(y_j-E(Y))p(x_i,y_j)\end{aligned}$$

The covariance is a measure of linear association between two random variables and represents the joint variability of two random variables. Given two random variables X and Y

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y)$$

Given two random variables X and Y and a, b, c, d real numbers

$$\text{cov}(aX+b, cY+d) = ac\,\text{cov}(X,Y)$$

Example

Roll two dice, each with 4 sides

$M =$ maximum value of the two dice

$S =$ sum of the values of the two dice

Find

1. the joint probability distribution
2. $P((M \leq 2))$
3. $P((M \leq 2) \cap (S \leq 4))$
4. the conditional probability of S given $M = 4$
5. the conditional probability of M given $S = 6$
6. the covariance between M and S

Example

Roll a red and green dice, each with 6 sides

$R =$ red dice result

$G =$ green dice result

Find

1. the joint probability distribution
2. the conditional probability of R given $G = 5$
3. the conditional probability of G given $R = 6$
4. the covariance between R and G

Independence of jointly distributed RV

The jointly distributed random variables X and Y are said to be independent if and only if their joint probability distribution is the product of their marginal probability distributions - that is, if and only if

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j)$$

for all possible pairs of values x_i and y_j.

Namely, we say that the two random variables X e Y are independent if all the pairs of events $(X = x_i)$ e $(Y = y_j)$ are independent:

$$p(x_i, y_j) = P((X = x_i) \cap (Y = y_j)) = P(X = x_i)P(Y = y_j) = p_X(x_i)p_Y(y_j)$$

Independence of jointly distributed RV

The jointly distributed random variables X and Y are said to be independent if and only if their joint probability distribution is the product of their marginal probability distributions - that is, if and only if

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j)$$

for all possible pairs of values x_i and y_j.

This means that

$$p_{X|Y}(x_i|y_j) = p_X(x_i) \quad [P(X = x_i|Y = y_j) = P(X = x_i)]$$

and

$$p_{Y|X}(y_j|x_i) = p_Y(y_j) \quad [P(Y = y_j|X = x_i) = P(Y = y_i)]$$

Functions of random variables

Starting from a single random variable X we defined a new random variable by applying a transformation as $Y = h(X)$. The same can be done with multiple random variables and transformations of the type

$$S = h(X, Y)$$

Two examples we have seen in the previous slides

$M =$ maximum value of the two dices

$S =$ sum of the values of the two dices

We concentrate on the case of the sum of multiple random variables

Linear combination of random variables

Given the random variables $X_1, \ldots, X_n$ and the real numbers $a_1, \ldots, a_n$, we define the random variable

$$Y = \sum_{i=1}^{n} a_i X_i$$

In general, it is not easy to obtain the probability distribution of Y, but the following results hold

(1) $E(Y) = \sum_{i=1}^{n} a_i E(X_i)$

(2) $V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i<j} a_i a_j \text{cov}(X_i, X_j)$

(3) if the random variables X_i are independent

$$V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i)$$

Binomial distribution as sum of binary variables

Let n be the number of independent and equally likely events with $P(E_i) = p$

$$Y = \#\{E_i \text{occurring events}\} \sim \text{Binom}(n, p)$$

We define

$$X_i = \begin{cases} 0 & \text{if } \bar{E}_i \\ 1 & \text{if } E_i \end{cases}$$

then,

$$Y = \sum_{i=1}^{n} X_i$$

From this result we can derive the mean and variance of the binomial distribution through points (1) and (3) of the previous slide.

Binomial distribution: mean and variance

Note that

$$E(X_i) = 0(1-p) + 1p = p$$
$$E(X_i^2) = 0(1-p) + 1p = p$$

thus,

$$V(X_i) = E(X_i^2) - (E(X_i))^2 = p - p^2 = p(1-p)$$

For $Y = \sum_{i=1}^{n} X_i$, this leads to

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = np$$

$$V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = np(1-p)$$

Example

Regarding a football match between two teams, we know that the number of goals scored by the team A, X_A, is distributed according to a Poisson with parameter $\lambda_A = 1.8$, while the number of goals of the team B, X_B, follows a Poisson with parameter $\lambda_B = 1.3$. The two results are considered independent.

Define the following events and, when possible, compute the probabilities

(1) the probability that the final match result is 0-0;

(2) the probability that the A team wins and that the B does not score any goal;

(3) the probability that the match ends in a draw;

(4) the probability that A wins.

Continuous Random Variables

Random variable

A **random variable** (r.v.) X is a variable whose value is a numerical outcome of a random phenomenon, that is, it is a well defined but unknown number

There are two main types of random variables: discrete, if it has a finite list of possible outcomes, and continuous, if it can take any value in an interval.

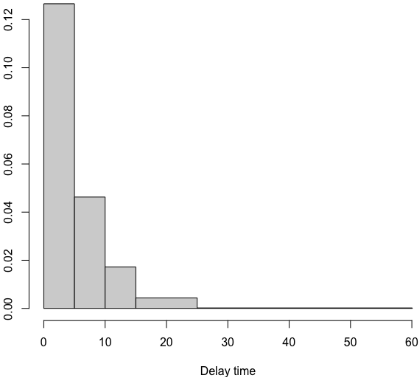
- D the number of tails on three coin tosses
- D the number of defective items in a sample of 20 items from a large shipment
- D the number of students attending the statistics class on a Friday
- C the delay time of the airplane
- C the weight of a newborn
- C the length of a phone call to your mother

For continuous random variables we can assign probabilities only to a range of values, using a mathematical function, we express the probability distribution in the continuous space

Delay time of a flight

About the delay time of a flight, we might observe the following data and plot them using a histogram

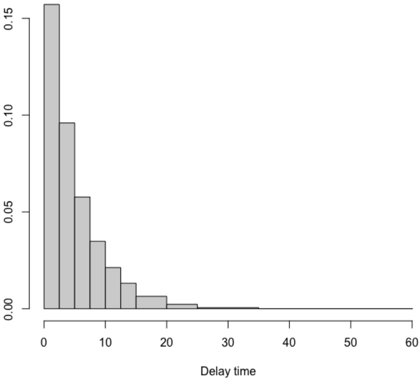
Delay	P
(0,5]	0.633
(5,10]	0.231
(10,15]	0.086
(15,25]	0.043
(25,Inf]	0.006



Delay time of a flight

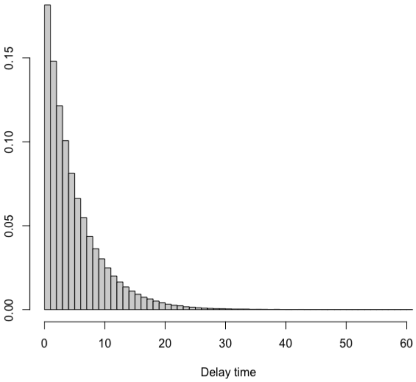
About the delay time of a flight, we might observe the following data and plot them using a histogram

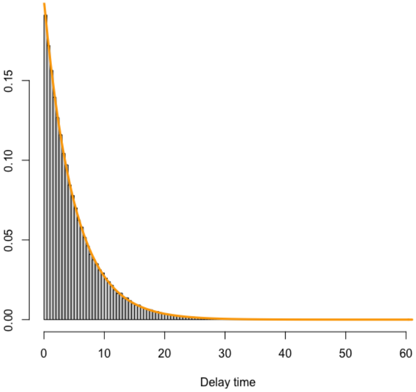
Delay	P
(0,2.5]	0.393
(2.5,5]	0.240
(5,7.5]	0.144
(7.5,10]	0.087
(10,12.5]	0.053
(12.5,15]	0.033
(15,20]	0.032
(20,25]	0.011
(25,35]	0.005
(35,Inf]	0.001



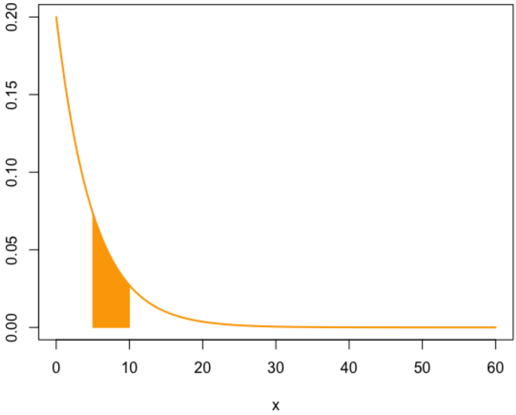
Delay time of a flight

Assuming that we have infinite observations we can increase the number of classes and approximate the histogram with a curve: the probability density function





The interpretation of the probability density function is analogous to the histogram: the areas represent probabilities rather than frequencies



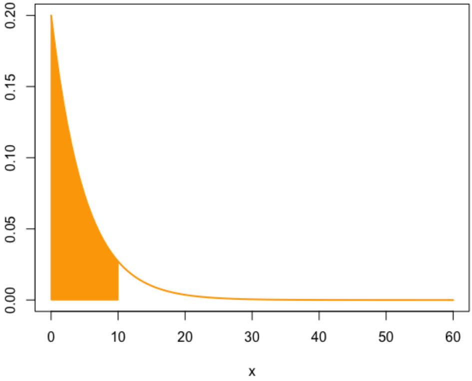
E.g. the probability of observing a value between 5 and 10 is:

$$P(5 \leq X \leq 10) = 0.234$$

Cumulative distribution function

A relevant quantity is

$$P(X \leq x)$$

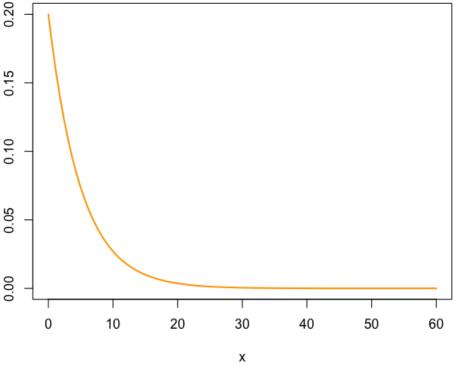


E.g. the probability of observing a value smaller than 10 is:

$$P(X \leq 10) = 0.865$$

Probability density function

The probability density function is a real-valued function assuming non-negative values



The function in the plot is

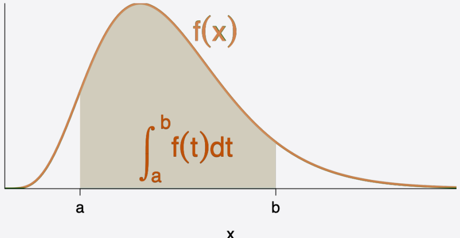
$$f(x) = \frac{1}{5}e^{-x/5}, \quad x \geq 0$$

Probability density function

Probability density function

The probability density function of a continuous random variable X is a non-negative function $f(x)$ whose area under the curve in a range of values is the probability of X in that range:

$$\int_a^b f(t)dt = P(a \leq X \leq b)$$



Probability density function

Probability density function

The probability density function of a continuous random variable X is a non-negative function $f(x)$ whose area under the curve in a range of values is the probability of X in that range:

$$\int_a^b f(t)dt = P(a \leq X \leq b)$$

The probability density function satisfies the following properties:

(i) $f(x) \geq 0$

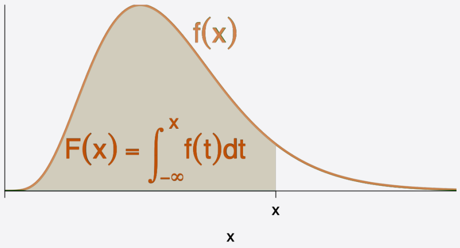
(ii) $\int_{-\infty}^{\infty} f(t)dt = 1$

Cumulative distribution function

Cumulative distribution function

The cumulative distribution function of a random variable X is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$



Cumulative distribution function

Cumulative distribution function

The cumulative distribution function of a random variable X is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

and satisfies the following properties:

i. $F(x) \geq 0, \qquad \forall x \in \mathbb{R};$

ii. $F(x)$ is not decreasing;

iii. $\lim_{x \to -\infty} F(x) = 0;$

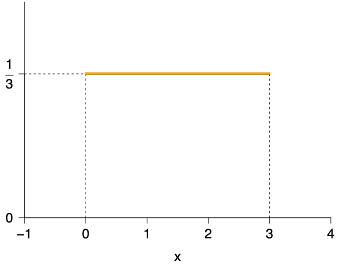
iv. $\lim_{x \to +\infty} F(x) = 1.$

and note that

$$P(a \leq X \leq b) = F(b) - F(a)$$

Example

There are several density functions, any non-negative function that integrates to 1 is a density function



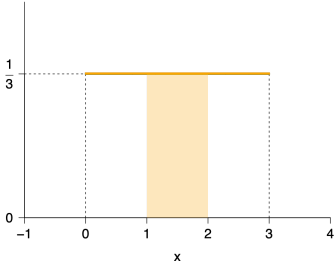
Uniform between 0 and 3

$$f(x) = \begin{cases} 1/3 & \text{for } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Find $P(1 \leq X \leq 2)$

Example

There are several density functions, any non-negative function that integrates to 1 is a density function



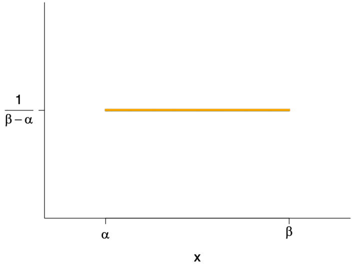
Uniform between 0 and 3

$$f(x) = \begin{cases} 1/3 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

$$P(1 \leq X \leq 2) = (2-1) \times \frac{1}{3} = \frac{1}{3}$$

Probability density functions and parameters

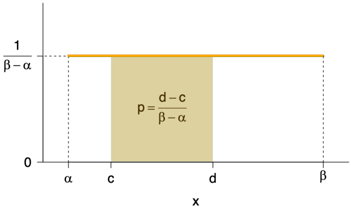
Often it is useful to define a probability density function up to one or more parameters, which is the equivalent of defining a set of density functions



$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

$f(x)$ is a probability density function for every α and β, with $\beta > \alpha$

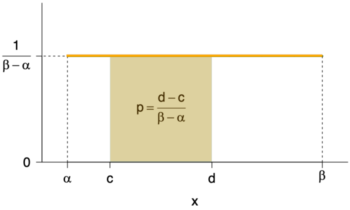
Uniform distribution on $[\alpha, \beta]$



$$P(c \leq X \leq d) = \frac{d - c}{\beta - \alpha}$$

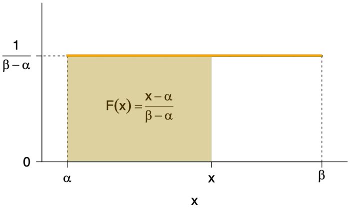
with $\alpha < c < d < \beta$

Uniform distribution on $[\alpha, \beta]$



$$P(c \leq X \leq d) = \frac{d-c}{\beta-\alpha}$$

with $\alpha < c < d < \beta$



$$F(x) = \begin{cases} 0 & \text{if} x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{if} \alpha \leq x \leq \beta \\ 1 & \text{if} x > \beta \end{cases}$$

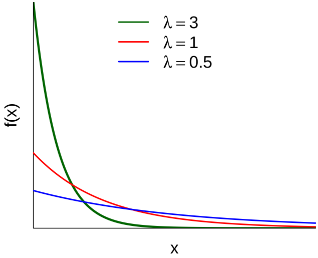
Exponential distribution

The Exponential distribution is a continuous probability distribution that describes the time between events in a Poisson process, where events occur continuously and independently at a constant average rate.

It has rate parameter λ and is defined as

$$f(x) = \lambda e^{-\lambda x}$$

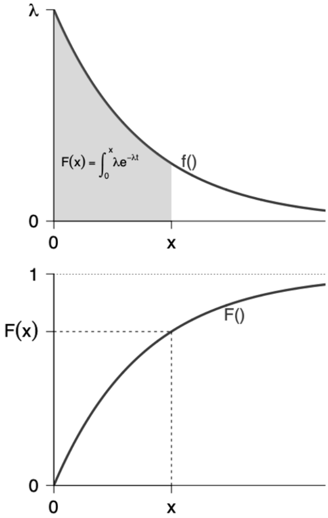
for $x \geq 0$ and $\lambda > 0$



Exponential distribution

The cumulative distribution function is

$$
\begin{aligned}
F(x) &= \int_0^x \lambda e^{-\lambda t} dt \\
&= \left[-e^{-\lambda t}\right]_0^x \\
&= 1 - e^{-\lambda x}
\end{aligned}
$$



Example

Knowing that X follows an exponential distribution with rate parameter λ

$$F(x) = 1 - e^{-\lambda x}$$

- If $\lambda = 2$, what is $P(X < 1) =$
- If $\lambda = 2$, $P(X < 0.5) =$
  - So, what is the probability that X is between 0.5 and 1?
- If $\lambda = 4$ what is the answer to the previous questions?

- The delay of a train (in minutes), is distributed according to an exponential of the parameter $\lambda = 0.1$, how likely is it to wait more than ten minutes?
- If the delay were distributed according to an exponential of parameter 0.2, would we wait longer or shorter (probably)?

Example

Knowing that X follows an exponential distribution with rate parameter λ

$$F(x) = 1 - e^{-\lambda x}$$

- If $\lambda = 2$, what is $P(X < 1) = 0.8646647$
- If $\lambda = 2$, $P(X < 0.5) = 0.6321206$
  - So, what is the probability that X is between 0.5 and 1?
- If $\lambda = 4$ what is the answer to the previous questions?

- The delay of a train (in minutes), is distributed according to an exponential of the parameter $\lambda = 0.1$, how likely is it to wait more than ten minutes?
- If the delay were distributed according to an exponential of parameter 0.2, would we wait longer or shorter (probably)?

Example

Knowing that X follows an exponential distribution with rate parameter λ

$$F(x) = 1 - e^{-\lambda x}$$

- If $\lambda = 2$, what is $P(X < 1) = 0.8646647$
- If $\lambda = 2$, $P(X < 0.5) = 0.6321206$
  - So, what is the probability that X is between 0.5 and 1?
- If $\lambda = 4$ what is the answer to the previous questions? (0.9816844, 0.8646647, 0.1170196)
- The delay of a train (in minutes), is distributed according to an exponential of the parameter $\lambda = 0.1$, how likely is it to wait more than ten minutes?
- If the delay were distributed according to an exponential of parameter 0.2, would we wait longer or shorter (probably)?

Expected value and variance of a continuous rv

Let $f(x)$ be the probability density function of a continuous random variable X

$$E(X) = \int x f(x) dx$$

$$E(h(X)) = \int h(x) f(x) dx$$

$$V(X) = \int (x - E(X))^2 f(x) dx$$

Moreover, the following properties hold (using the properties of the integrals)

$$E(aX + b) = aE(X) + b$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(aX + b) = a^2 V(X)$$

Uniform distribution: expected value and variance

Let X be a random variable following the Uniform distribution between α and β, thus

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

then

$$E(X) = \int_\alpha^\beta \frac{x}{\beta-\alpha} dx = \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_\alpha^\beta = \frac{1}{\beta-\alpha}\frac{\beta^2-\alpha^2}{2} = \frac{\alpha+\beta}{2}$$

$$E(X^2) = \int_\alpha^\beta \frac{x^2}{\beta-\alpha} dx = \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_\alpha^\beta = \frac{1}{\beta-\alpha}\frac{\beta^3-\alpha^3}{3} = \frac{\beta^2+\alpha\beta+\alpha^2}{3}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} = \frac{(\beta-\alpha)^2}{12}$$

Exponential distr: expected value and variance

Let X be a random variable following the Exponential distribution with rate parameter $\lambda > 0$, thus

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$, then, by integration by parts,

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \left[x e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} dx = -\left[\frac{e^{-\lambda x}}{\lambda} \right]_0^\infty = \frac{1}{\lambda}$$

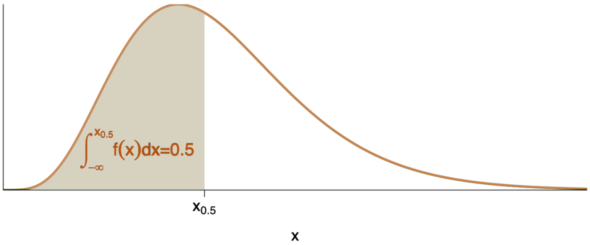
$$E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \left[x^2 e^{-\lambda x} \right]_0^\infty + \frac{2}{\lambda} \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$$

Median

We define the median of a continuous random variable X with pdf $f(x)$ the value $Me(X)$ or $x_{0.5}$ such that

$$F(Me(X)) = \int_{-\infty}^{Me(X)} f(x)dx = 0.5$$



Quantiles

We call q-quantile of a continuous random variable X with pdf $f(x)$ the value x_q such that

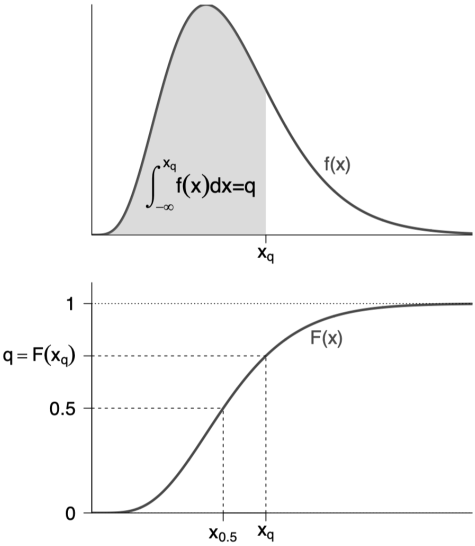
$$P(X \leq x_q) = q$$

that is

$$F(x_q) = \int_{-\infty}^{x_q} f(x)dx = q$$

therefore

$$x_q = F^{-1}(q)$$



Uniform distribution: quantiles

Let X be a continuous random variable following a Uniform distribution between α and β, then for $\alpha \leq x \leq \beta$

$$F(x) = \frac{x-\alpha}{\beta-\alpha}$$

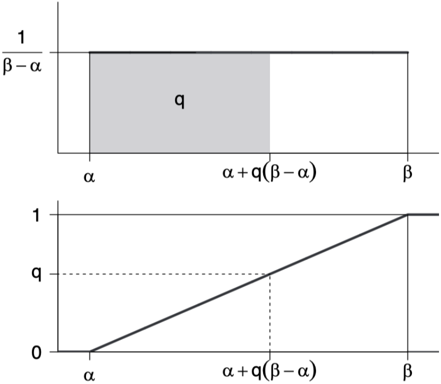
the q-quantile is then x_q such that

$$q = F(x_q) = \frac{x_q-\alpha}{\beta-\alpha}$$

computing the inverse we find

$$x_q = \alpha + q(\beta-\alpha)$$

Thus, the median is ...



Exponential distribution: quantiles

Let X be a continuous random variable following an exponential distribution with rate parameter $\lambda > 0$, then

$$F(x) = 1 - e^{-\lambda x}$$

the q-quantile id then x_q such that

$$q = F(x_q) = 1 - e^{-\lambda x_q}$$

computing the inverse we find

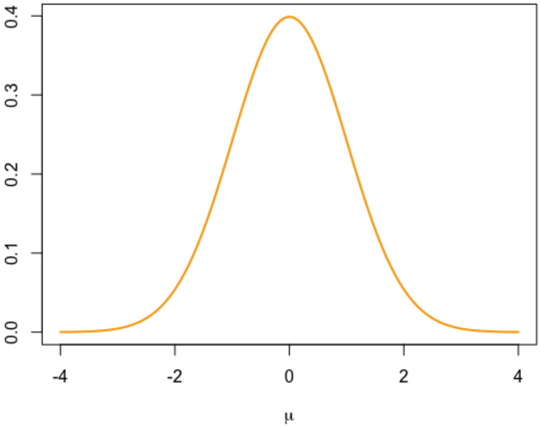
$$x_q = -\frac{\log(1-q)}{\lambda}$$

The median is then

$$\mathrm{Me} = x_{0.5} = \frac{\log 2}{\lambda}$$

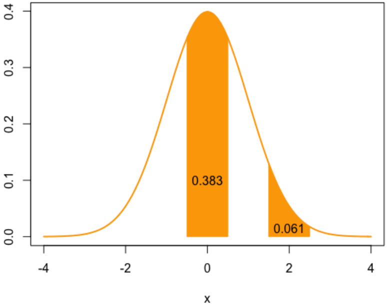
Normal distribution

The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



Normal distribution

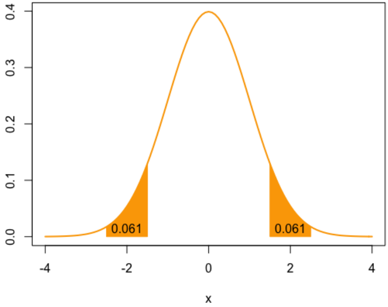
The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



- The most likely values to occur are the ones around the center (the mean)

Normal distribution

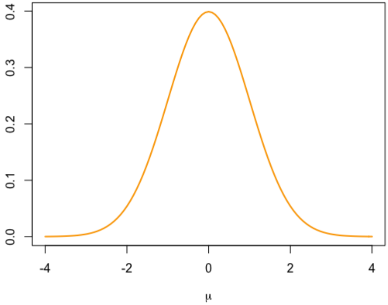
The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



- The most likely values to occur are the ones around the center (the mean)
- It is symmetric, then symmetric deviations on the right and left have the same probability

Normal distribution

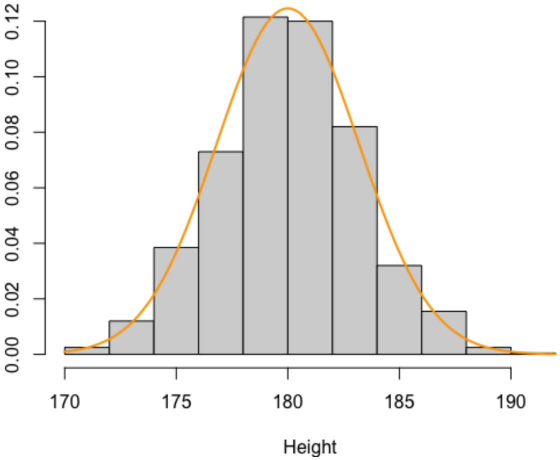
The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



- The most likely values to occur are the ones around the center (the mean)
- It is symmetric, then symmetric deviations on the right and left have the same probability
- The pdf is: $f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Example

The distribution of the male students' height

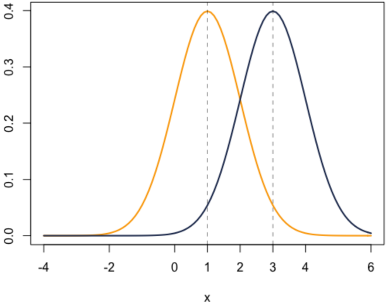


Normal distribution: parameters

The probability density function of the normal distribution is

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

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$



μ is the mean of the distribution

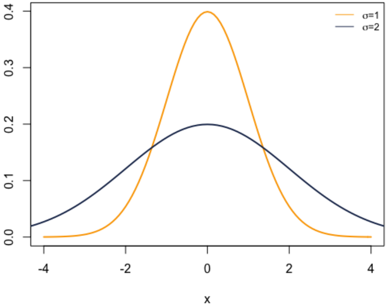
For different μ values, the distribution shifts on the x axis

Normal distribution: parameters

The probability density function of the normal distribution is

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

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$



σ^2 is the variance of the distribution

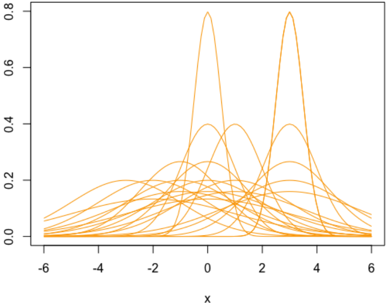
For different σ values, the distribution remains centered on the same value but becomes wider: larger values (in absolute values) of the r.v. are more likely

Normal distribution: parameters

The probability density function of the normal distribution is

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

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$



Varying μ e σ, we can obtain an infinite number of distributions

Normal distribution: parameters

The probability density function of the normal distribution is

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

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$

Normal distribution: probabilities

Assume that $X \sim N(\mu, \sigma^2)$, then its pdf is

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

we can use it to compute probabilities of a Normal rv

$$p = P(a \leq X \leq b)$$

where

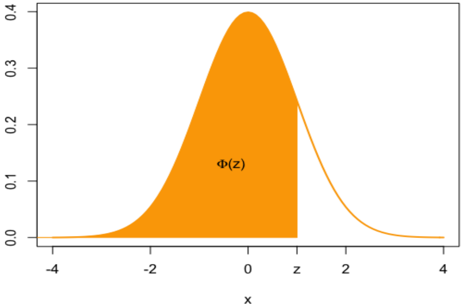
$$p = \int_a^b f(t)dt = F(b) - F(a)$$

Unfortunately, we do not have a closed form to compute this probability, but we can refer to a particular normal distribution...

Standard Normal distribution

The standard normal distribution, that is $\mu = 0$ and $\sigma = 1$, play an important role

$$Z \sim N(0,1) \quad \rightarrow \quad f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$



We define

$$\Phi(z) = P(Z \leq z)$$

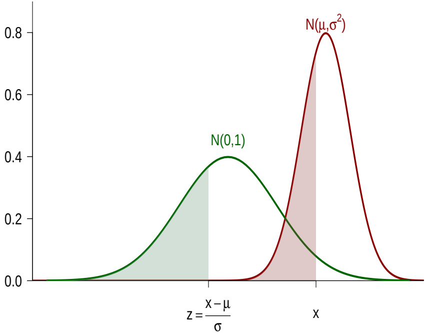
the area under the curve between $-\infty$ and z

Note that with Φ we call the cumulative distribution function of the standard normal distribution

Standard Normal vs Normal(μ, σ)

If $X \sim N(\mu, \sigma^2)$, then

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$



That is, the red area (equal to $P(X \leq x)$) is equal to the green one
Knowing $\Phi(z)$, we can compute any probability associated with a generic normal distribution $N(\mu, \sigma^2)$

Normal distribution table

It is not possible to express the function $\Phi(z)$ in an analytical form, but we find its values in a table

That is, the values of $\Phi(z)$ corresponding to different values of z are listed in a table like the following one

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
					.					
					.					
0.2					.					
0.3	.	.	.	.	$\Phi(0.34)$					
0.4										

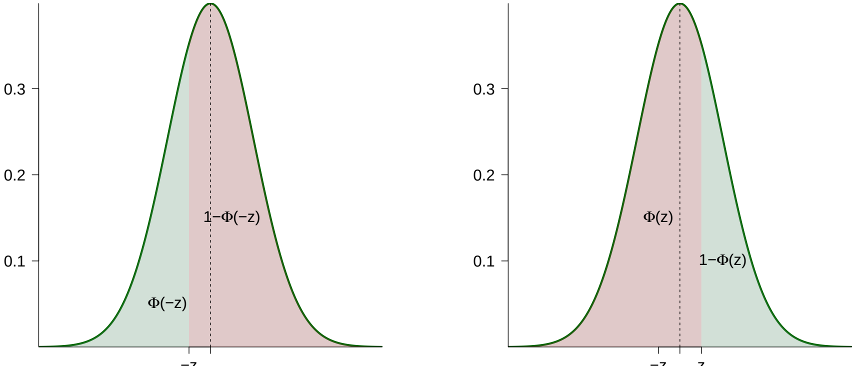
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Where are the negative values?

Knowing $\Phi(z)$ for positive values of z, when z is negative the following relation holds

$$\Phi(-z) = 1 - \Phi(z)$$

by the symmetry of the standard normal distribution around 0 (its mean) Graphically: same colors represent the same areas



Examples

Let Z follow a standard normal distribution

$$P(Z \leq 1.34) = ?$$

The relevant part of the table is

	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1							
1.1		0.8686	0.8708	0.8729	0.8749	0.8770	
1.2		0.8888	0.8907	0.8925	0.8944	0.8962	
1.3		0.9066	0.9082	0.9099	0.9115	0.9131	
1.4		0.9222	0.9236	0.9251	0.9265	0.9279	
1.5		0.9357	0.9370	0.9382	0.9394	0.9406	
1.6							

Examples

Let Z follow a standard normal distribution

$$P(Z > -1.23) = ?$$

The relevant part of the table is

	0	0.01	0.02	0.03	0.04	0.05	0.06
0.9							
1		0.8438	0.8461	0.8485	0.8508	0.8531	
1.1		0.8665	0.8686	0.8708	0.8729	0.8749	
1.2		0.8869	0.8888	0.8907	0.8925	0.8944	
1.3		0.9049	0.9066	0.9082	0.9099	0.9115	
1.4		0.9207	0.9222	0.9236	0.9251	0.9265	
1.5							

Examples

Let X follow a normal distribution with mean $\mu = 4$ and variance $\sigma^2 = 16$

$$P(X \leq 13) = ?$$

First, we compute $z = (13 - 4)/4 = 2.25$, then we have

$$P(X \leq 13) = \Phi(2.25) = ?$$

The relevant part of the table is

	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.9							
2		0.9783	0.9788	0.9793	0.9798	0.9803	
2.1		0.9830	0.9834	0.9838	0.9842	0.9846	
2.2		0.9868	0.9871	0.9875	0.9878	0.9881	
2.3		0.9898	0.9901	0.9904	0.9906	0.9909	
2.4		0.9922	0.9925	0.9927	0.9929	0.9931	
2.5							

Normal distribution: quantiles z_q s.t. $\Phi(z_q) = q$ I

What if we are interested in finding $z_{0.9}$, such that

$$\Phi(z_{0.9}) = 0.9$$

We have to follow the steps in the reverse order, looking for a value close to 0.9 on the table, and then we read the corresponding value of z

	0.06	0.07	0.08	0.09
0.9				
1		0.8577	0.8599	0.8621
1.1		0.8790	0.8810	0.8830
1.2		0.8980	0.8997	0.9015
1.3		0.9147	0.9162	0.9177
1.4		0.9292	0.9306	0.9319
1.5				

Normal distribution: quantiles z_q s.t. $\Phi(z_q) = q$ II

Knowing that

$$\Phi(1.28) = 0.9$$

We can easily find $z_{0.1}$, that is, the quantile z such that $\Phi(z_{0.1}) = 0.1$
Knowing that

$$\Phi(1.28) = 0.9$$

we can easily find $z_{0.1}$, that is the quantile z such that

$$\Phi(z_{0.1}) = 0.1$$

Indeed we can write

$$\Phi(-1.28) = 1 - \Phi(1.28) = 1 - 0.9 = 0.1$$

Since $\Phi(z) = 1 - \Phi(-z)$ we find

$$z_{1-q} = -z_q$$

Examples

We can also find $z_{0.95}$, that is, the value z such that

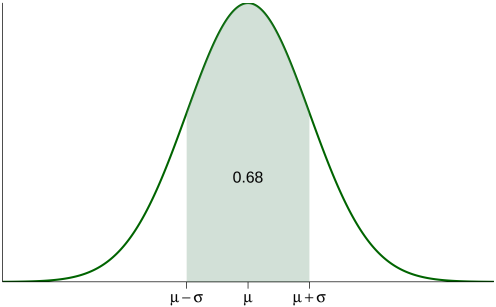
$$\Phi(z_{0.95}) = 0.95$$

On the table, we find a value close to 0.95 and we identify the corresponding value of z

	0.01	0.02	0.03	0.04	0.05	0.06	0.07
1.3							
1.4		0.9222	0.9236	0.9251	0.9265	0.9279	
1.5		0.9357	0.9370	0.9382	0.9394	0.9406	
1.6		0.9474	0.9484	0.9495	0.9505	0.9515	
1.7		0.9573	0.9582	0.9591	0.9599	0.9608	
1.8		0.9656	0.9664	0.9671	0.9678	0.9686	
1.9							

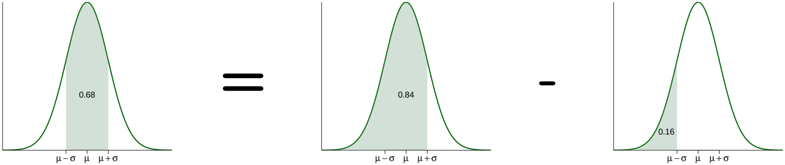
Intervals I

The interval defined from $\mu - \sigma$ to $\mu + \sigma$ includes a probability equal to 0.68



Let's prove it

Intervals II



$$0.68 = \Phi\left(\frac{(\mu+\sigma)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-\sigma)-\mu}{\sigma}\right)$$

$$= \Phi(1) - \Phi(-1)$$

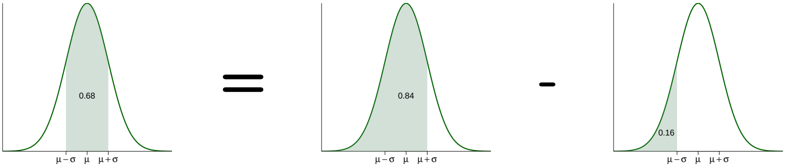
Intervals III

The interval defined from $\mu - \sigma$ to $\mu + \sigma$ includes a probability equal to 0.68

The relevant part of the table is

	0	0.01	0.02	0.03
0.7				
0.8	0.7881	0.7910	0.7939	
0.9	0.8159	0.8186	0.8212	
1	0.8413	0.8438	0.8461	
1.1	0.8643	0.8665	0.8686	
1.2	0.8849	0.8869	0.8888	
1.3				

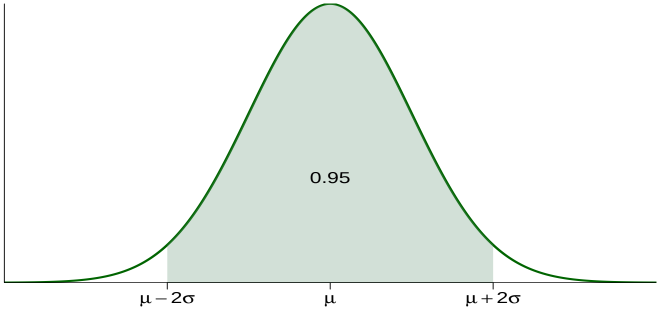
Intervals IV



$$
\begin{aligned}
0.68 &= \Phi\left(\frac{(\mu+\sigma)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-\sigma)-\mu}{\sigma}\right) \\
&= \Phi(1) - \Phi(-1) \\
&= 0.84 - 0.16
\end{aligned}
$$

Intervals V

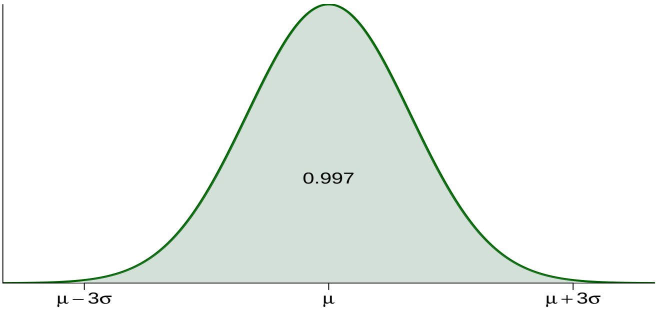
The interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ includes a probability equal to 0.95



Prove it by yourself!

Intervals VI

The interval from $\mu - 3\sigma$ to $\mu + 3\sigma$ includes a probability equal to 0.997



Almost all the observations should be included between $\mu - 3\sigma$ and $\mu + 3\sigma$

Exercise

The April daily average temperature in Rome is a random variable that can be assumed to follow a normal distribution with a mean of 13.5

Knowing that the probability that the temperature is lower than 15 is equal to 0.71, can we find the standard deviation?

Exercise

The April daily average temperature in Rome is a random variable that can be assumed to follow a normal distribution with a mean of 13.5

Knowing that the probability that the temperature is lower than 15 is equal to 0.71, can we find the standard deviation?

- Yes, we can

Exercise

Let σ be the unknown standard deviation

Knowing σ we could compute the probability that the temperature is lower than 15 degrees

$$\Phi\left(\frac{15-13.5}{\sigma}\right)$$

But we know that this probability is equal to 0.71

Let's have a look at the table:

Exercise

Let σ be the unknown standard deviation
Knowing σ we could compute the probability that the temperature is lower than 15 degrees

$$\Phi\left(\frac{15-13.5}{\sigma}\right)$$

But we know that this probability is equal to 0.71
Let's have a look at the table:

	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.2							
0.3		0.6293	0.6331	0.6368	0.6406	0.6443	
0.4		0.6664	0.6700	0.6736	0.6772	0.6808	
0.5		0.7019	0.7054	0.7088	0.7123	0.7157	
0.6		0.7357	0.7389	0.7422	0.7454	0.7486	
0.7		0.7673	0.7704	0.7734	0.7764	0.7794	
0.8							

Exercise

Let σ be the unknown standard deviation

Knowing σ we could compute the probability that the temperature is lower than 15 degrees

$$\Phi\left(\frac{15-13.5}{\sigma}\right)$$

But we know that this probability is equal to 0.71

Let's have a look at the table: $\Phi(0.55) = 0.71$ (about)

Exercise

Let σ be the unknown standard deviation

Knowing σ we could compute the probability that the temperature is lower than 15 degrees

$$\Phi\left(\frac{15-13.5}{\sigma}\right)$$

But we know that this probability is equal to 0.71

Let's have a look at the table: $\Phi(0.55) = 0.71$ (about)

So we know

$$\frac{15-13.5}{\sigma} = 0.55$$

Exercise

Let σ be the unknown standard deviation

Knowing σ we could compute the probability that the temperature is lower than 15 degrees

$$\Phi\left(\frac{15-13.5}{\sigma}\right)$$

But we know that this probability is equal to 0.71

Let's have a look at the table: $\Phi(0.55) = 0.71$ (about)

So we know

$$\frac{15-13.5}{\sigma} = 0.55$$

then

$$\sigma = \frac{15-13.5}{0.55} = 2.7$$

In April, the daily average temperature in Rome follows a normal distribution with mean 13.5 and standard deviation 2.7

Exercise

Knowing that the average daily temperature in April in Rome follows a normal distribution with mean 13.5 and standard deviation 2.7

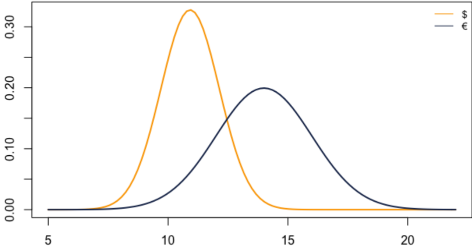
1. Identify an interval which contains the average temperature with probability 0.99.
2. What is the probability that the daily average temperature is 3.0 degrees?
3. What is the probability that the daily average temperature on a given day will be less than 3 degrees?
4. What is the probability that the daily average temperature on a given day is less than or equal to 3 degrees?

Normal distribution: transformations

Let X be distributed as a normal distribution $N(\mu, \sigma^2)$, and let a, and b be two real numbers, then

$$Y = aX + b$$

follows a normal distribution $Y \sim N(a\mu + b, a^2\sigma^2)$



As an example, if the stock price on a given day expressed in euros follows a Normal distribution $N(14, \sigma^2 = 2)$, then the value in dollars (1 \$ $= 0.78$ €) is still distributed as a Normal with mean 0.78×14 and variance $2 \times (0.78)^2$

Example

The average daily temperature in a locality expressed in Fahrenheit scale follows a $N(68, \sigma^2 = 9)$, what is the probability that the temperature will exceed 20 degrees in Celsius scale?

Note that the relationship between Celsius and Fahrenheit is

$$C = (5/9) \times (F - 32)$$

that is, a linear transformation $aX + b$, with $a = 5/9$ and $b = -(5/9) \times 32$

From the result in the previous slide, we know that the temperature in Celsius scale follows a Normal distribution $N(20, \sigma^2 = 2.78)$, since $(5/9)(68 - 32) = 20$ and $(5/9)^2 \times 9 = 2.78$

Thus, the requested probability is ...

Sum of two Normal random variables

Let $X_1 \sim N(\mu_1, \sigma_1^2)$, and $X_2 \sim N(\mu_2, \sigma_2^2)$, and assume X_1, X_2 independent, then

$$Y = a_1 X_1 + a_2 X_2$$

follows a normal distribution

$$Y \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$$

Example

Suppose that the weight of individuals of a population follows a $N(70, \sigma^2 = 50)$

Two subjects, randomly selected, go on a goods lift that bears a maximum weight of $160kg$

What is the probability that the weight of the two subjects is smaller than $160kg$?

For a randomly selected subject (assuming independence), the weight follows a $N(70, \sigma^2 = 50)$ thus the weight of the two subjects, $X_1 + X_2$, follows a Normal distribution

$$X_1 + X_2 \sim N(70 + 70 = 140, 50 + 50 = 100)$$

then, the probability that their weight is smaller than 160 kg is

$$\Phi\left(\frac{160 - 140}{10}\right) = \Phi(2) = 0.98$$

Example

Suppose that the weight of individuals of a population follows a $N(70, \sigma^2 = 50)$

What if we choose three subjects at random?

The total weight, $X_1 + X_2 + X_3$, follows a normal distribution

$$X_1 + X_2 + X_3 \sim N(70 + 70 + 70 = 210, 50 + 50 + 50 = 150)$$

then, the probability is

$$\Phi\left(\frac{160 - 210}{12.25}\right) = \Phi(-4.1) \approx 0$$

Example

Suppose that the weight of individuals of a population follows a $N(70, \sigma^2 = 50)$

What if we choose n subjects at random?

The total weight, $X_1 + X_2 + X_3 + \ldots + X_n$ follows a normal distribution

$$\sum_{i=1}^{n} X_i \sim N(70n, 50n)$$

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \sim N\left(70, \frac{50}{n}\right)$$

Since,

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} 70n$$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} 50n$$

Example

Suppose that the weight of individuals of a population follows a $N(70, \sigma^2 = 50)$

What if we choose n subjects at random?

The total weight, $X_1 + X_2 + X_3 + \ldots + X_n$ follows a normal distribution

$$\sum_{i=1}^{n} X_i \sim N(70n, 50n)$$

Note that, the sample mean is the sum multiplied by $1/n$: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, therefore it follows a normal distribution

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \sim N\left(70, \frac{50}{n}\right)$$

Since,

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}70n$$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}50n$$

Sum of Normal random variables

Let $X_1, \ldots, X_n$ be independent normal random variables with mean μ_i and variance σ_i^2, then

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

follows a normal distribution

$$N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Let $X_1, \ldots, X_n$ be normal random variables with mean μ_i, variance σ_i^2, and covariances $Cov(X_i, X_j) = \sigma_{ij}$, then

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

follows a normal distribution

$$N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{i>j} a_ia_j\sigma_{ij}\right)$$

Sum of iid Normal random variables

It is worth noting the case in which the variables X_i are independent and identically distributed (iid)

Let $X_1, \dots, X_n$ be independent normal variables such that

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, \dots, n$$

then

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Limits and probability

Let's define a sequence, i.e. a countably infinite number, of random variables

$$X_1, \ldots, X_n, \ldots$$

defined on the same probability space and with their distributions defined as

$$F_1, \ldots, F_n, \ldots$$

What is the behavior of the sequence as n approaches infinity?

In the probability theory there are several definitions for the limit of a sequence of random variables (we study two of them)

Convergence in probability

A sequence of random variables X_n converges in probability to X if and only if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

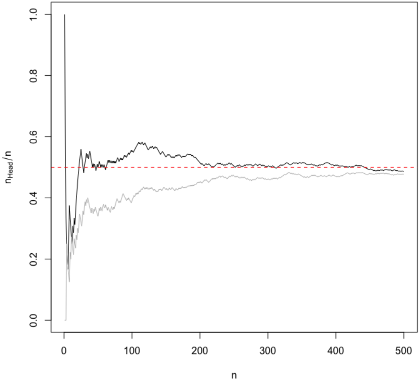
for all $\varepsilon > 0$, and we write

$$X_n \underset{p}{\to} X$$

In particular, if X is a constant, i.e. c, then $X_n \underset{p}{\to} c$ means that for n large enough, the probability that X_n is in a small interval around c is close to 1.

A flashback

Each trajectory is a realization of a sequence of random variables



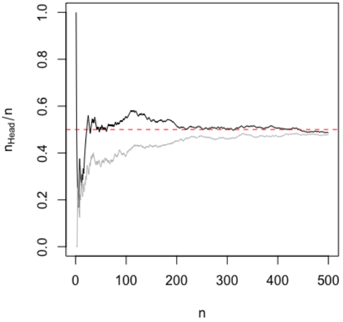
The idea is that as n increases, the results concentrate around the limit of the sequence

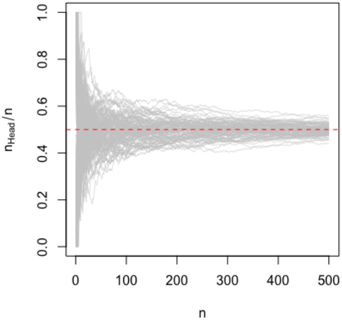
$$X_i = \begin{cases} 1 & \text{if head} \\ 0 & \text{otherwise} \end{cases}$$

The mean $\bar{X}_n$ is the proportion of heads: $E(X_i) = P(\text{head})$

A flashback

Each trajectory is a realization of a sequence of random variables





The mean $\bar{X}_n$ is the proportion of heads: $E(X_i) = P(\text{head})$

The weak law of large numbers [LLN]

Given a sequence of independent random variables $X_1, \ldots, X_n, \ldots$ with the same mean $E(X_i) = \mu$ and the same (finite) variance $V(X_i) = \sigma^2$, then the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \underset{p}{\to} \mu$$

So, as n increases, the probability that $\bar{X}_n$ is close to μ approaches 1

$$P\left(\mu - \varepsilon \leq \bar{X}_n \leq \mu + \varepsilon\right) \underset{n \to \infty}{\to} 1$$

for any $\varepsilon > 0$

Note that we specified the term weak law since it also exists a strong law that involves a different type of convergence, stronger than the one in probability

LLN in practice

From an urn half filled with white balls, that is the proportion of white balls is $p = 0.5$, we draw $n_1 = 10$ balls with replacement

The expected result is to extract $pn_1 = 5$ white balls, that is the Binomial expected value, i.e. that the expected percentage of extracted white balls is $p = 0.5$

Note that we cannot be sure that this will be exactly the result, we can only compute the probability to observe this outcome

We define S_1 as the proportion of white balls

$$P(S_1 = 0.5) = 0.246$$

the proportion could be also equal to 0.4 or 0.6 with probability, respectively

$$P(S_1 = 0.4) = 0.205, \quad P(S_1 = 0.6) = 0.205$$

So, the probability of observing proportions of white balls with a distance lower or equal to 0.1 from the expected one ($p = 0.5$) is 0.656

LLN in practice

If we perform 10 draws, the probability of observing proportions of white balls with a distance lower or equal to 0.1 from the expected one ($p = 0.5$) is 0.656

LLN in practice

If we perform 10 draws, the probability of observing proportions of white balls with a distance lower or equal to 0.1 from the expected one ($p = 0.5$) is 0.656

What if we perform 20 draws?

The proportion of white balls S_2 falling in the same ± 0.1 interval from the expected proportion corresponds to a number of white balls between 8 and 12

The probability is then equal to

$$P(0.4 \leq S_2 \leq 0.6) = 0.737$$

LLN in practice

If we perform 10 draws, the probability of observing proportions of white balls with a distance lower or equal to 0.1 from the expected one ($p = 0.5$) is 0.656

What if we perform 20 draws?

The proportion of white balls S_2 falling in the same ± 0.1 interval from the expected proportion corresponds to a number of white balls between 8 and 12

The probability is then equal to

$$P(0.4 \leq S_2 \leq 0.6) = 0.737$$

What if we perform 50 draws?

The proportion of white balls S_3 falling in the same ± 0.1 interval from the expected proportion corresponds to a number of white balls between 20 and 30

$$P(0.4 \leq S_3 \leq 0.6) = 0.881$$

LLN in practice

If we perform 10 draws, the probability of observing proportions of white balls with a distance lower or equal to 0.1 from the expected one ($p = 0.5$) is 0.656

What if we perform 50 draws?

The proportion of white balls S_3 falling in the same ± 0.1 interval from the expected proportion corresponds to a number of white balls between 20 and 30

$$P(0.4 \leq S_3 \leq 0.6) = 0.881$$

From the LLN we note that as n increases, this probability increases

More in detail, the probability that the observed proportion falls in a fixed width interval centered on the expected proportion increases

Convergence in distribution

A sequence of random variables X_n converge in distribution to $X \sim F$, with F cumulative distribution, if and only if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for every $x \in \mathbb{R}$ and we write

$$X_n \underset{d}{\to} X$$

Knowing that a sequence converges in distribution allows us to use the limit distribution as an approximation for large n

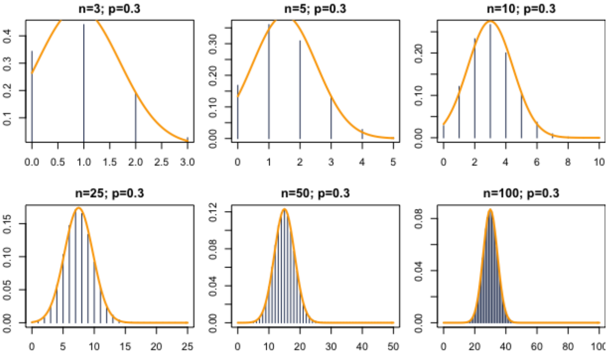
Note that the Normal distribution is frequently used also because it is a valid approximation in several situations

Normal approximation to the Binomial

Consider a random variable S following a Binomial distribution with dimension n and probability p

$$P(S = s) = \binom{n}{s} p^s (1-p)^{n-s}$$

It converges in distribution to a Normal with mean np and variance $np(1-p)$



Normal approximation to the Binomial

What are the practical implications?

If $S \sim \text{Binom}(n, p)$, with n large enough, then approximately

$$P(S \leq s) \approx \Phi\left(\frac{s - np}{\sqrt{np(1-p)}}\right)$$

that is, we can use the Normal distribution to compute probabilities related to a Binomial

We can extend the same result to the proportion $X = S/n$

$$P(X \leq x) = P(S \leq nx) \approx \Phi\left(\frac{x - p}{\sqrt{p(1-p)/n}}\right)$$

Example: *overbooking*

An airline accepts reservations for a plane with 200 seats, knowing that 10% of the people who booked the flight do not show up at the airport,

1. What is the probability that the plane will travel fully if there have been 200 reservations?

The company accepts more than 200 reservations, hoping that no more than 200 people show up.

2. If 220 reservations have been accepted, what is the probability that some passengers who have booked will stay ashore?
3. If 210 reservations have been accepted, what is the probability that some passengers who have booked will stay ashore?

The number of passengers that show up at the airport X is

$$X \sim \text{Binom}(m, p) \underset{d}{\rightarrow} N\left(mp, mp(1-p)\right)$$

where m is the number of reservations and p is the probability to show up

1. With 200 bookings, the plane will travel full with probability $0.9^{200} = 7.0550791 \times 10^{-10}$ (without approximation)

The number of passengers that show up at the airport X is

$$X \sim \text{Binom}(m, p) \underset{d}{\rightarrow} N\left(mp, mp(1-p)\right)$$

where m is the number of reservations and p is the probability to show up

1. With 200 bookings, the plane will travel full with probability $0.9^{200} = 7.0550791 \times 10^{-10}$ (without approximation)
2. If the company accepts $m = 220$ reservations, the probability that 200 (or less) passengers will show up is

$$\Phi\left(\frac{200.5 - mp}{\sqrt{mp(1-p)}}\right) = \Phi\left(\frac{200.5 - 220 \cdot 0.9}{\sqrt{220 \cdot 0.9 \cdot 0.1}}\right) = 0.71289$$

3. If $m = 210$ reservations are accepted,

$$\Phi\left(\frac{200.5 - (210 \cdot 0.9)}{\sqrt{210 \cdot 0.9 \cdot 0.1}}\right) = 0.99592$$

U.S. presidential elections I

In Prob3 slide 10/27 we computed the probability that a voter, selected at random, supported the Democratic party is equal to 0.513

Assuming that we select 1000 subjects at random among the voters, the number of voters supporting the Democratic party follows a Binomial distribution with mean 513 and variance 249.831

(this approximation assumes an extraction with replacement, but the effect is negligible)

Among the 1000 subjects, what is the probability that more than 500 supported the Democratic party?

Using the Normal approximation, we have

$$1 - \Phi\left(\frac{500.5 - 513}{\sqrt{249.831}}\right) = 1 - \Phi\left(-0.7908368\right) = 0.7854804$$

U.S. presidential elections II

What is the probability that the proportion of votes for the Democratic party is between 50 % and 52.6 %?

We observe that if the number of voters is

$$V \dot{\sim} N(np, np(1-p))$$

then the sampling percentage is

$$V/n \dot{\sim} N(p, p(1-p)/n)$$

the requested probability is then

$$\Phi\left(\frac{0.526-0.513}{\sqrt{0.00025}}\right) - = \Phi\left(\frac{0.5-0.513}{\sqrt{0.00025}}\right) =$$
$$= \Phi(0.8222) - \Phi(-0.8222) =$$
$$= 0.7945162 - 0.2054838 = 0.589$$

Central Limit Theorem

If $X_1, \ldots, X_n, \ldots$ are independent and identically distributed with same mean $E(X_i) = \mu$ and (finite) variance $V(X_i) = \sigma^2$, then the sample mean

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \underset{d}{\to} N(0,1)$$

In practice, this means that

$$\bar{X}_n \overset{\bullet}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

where $\overset{\bullet}{\sim}$ stands for "is approximately distributed as"

Note that we are not making assumptions on the distribution of X_i

Example

Suppose that the weight (in kg) of individuals of a population has mean $\mu = 70$ and variance $\sigma^2 = 50$

What is the probability that the average weight for 25 subjects is between 68 and 72 kg?

From the CLT the average weight of the 25 subjects $\bar{X}_{25}$ is approximately distributed as a

$$\bar{X}_{25} \overset{\bullet}{\sim} N\left(70, \frac{50}{25}\right)$$

then the requested probability is ...

Example: sheeps

Imagine you have a huge flock of sheep and you want to estimate its value, thus you need to know the sheep's average weight (let μ be the flock's average weight)

Since you have many sheep, you decide to weigh some of them to get an estimate of the weight

For the CLT, the average weight of n sheep is approximately distributed as

$$\bar{X}_n \overset{\bullet}{\sim} N(\mu, \sigma^2/n)$$

where σ^2 is the variance of the flock, assumed equal to 100

Example: sheeps

Using the notions about the Normal distribution, we are sure at 95% that the average weight of the n sheep is in the interval

$$\left[\mu - 2\frac{\sigma}{\sqrt{n}}, \mu + 2\frac{\sigma}{\sqrt{n}}\right]$$

If we weigh 100 sheeps, then the average weight will fall at 95% in the interval

$$[\mu - 2, \mu + 2]$$

that is, $\pm 2kg$ around the average weight of the flock

Joint continuous random variables

The joint distribution of continuous random variables is of great interest

Joint continuous distribution function

Given two r.v. X and Y the joint continuous distribution function $f(x, y)$ is a non-negative valued function such that

$$P((X, Y) \in A) = \iint_A f(x, y) dxdy$$

Conditional and marginal distributions

From the joint density function, we can find the marginal densities of each r.v.

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy, \qquad f_Y(y) = \int_{\mathbb{R}} f(x,y)dx$$

and the conditional density functions

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \qquad f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

The two variables X and Y are independent if and only if

$$f(x,y) = f_X(x)f_Y(y)$$

Expected value

The expected value of a function $h(X, Y)$ with (X, Y) with joint density function $f(x, y)$, is the quantity

$$E(h(X, Y)) = \iint_{\mathbb{R}^2} h(x, y) f(x, y) dxdy$$

Exercise

(Agresti, Finlay ex. 4.33)
The Psychomotor Development Index scores for measuring childhood development are approximately Normal with a mean of 100 and a standard deviation of 15.

- For a child selected at random, find the probability that the index is less than 90.
- One study refers to a random sample of 25 children. What is the distribution of the sample mean? What is the probability that the sample mean is less than 90?
- Would you be surprised to find an individual with a score of 90? Would you be surprised to find a sample average of 90?