02 - Random variables - 01

Master in Foundations of Data Science

Bayesian Statistics and Probabilistic Programming

Fall 2018-2019

Josep Fortiana

Wednesday, September 26, 2018

Universitat de Barcelona

02 - Random variables - 01

The random variable (r.v.) concept

Discrete r.v. - pmf and cdf

Computing with discrete r.v.

Continuous r.v. - pdf - Computations

02 - Random variables - 01

Purpose of the (r.v.) concept

A r.v. is a mathematical object we use to model (numerical or more general) quantities whose value depends on the outcome of a random experiment.

Example 1

We toss a coin.

The indicator of “coin falls heads”.

If the coin falls heads, value is 1;

if it falls tails, value is 0.

Example 2

We toss a coin 10 times.

Number of heads.

It takes values in: $\{0, 1, 2, 3, \ldots, 10\}$.

Example 3

A die is thrown repeatedly until a 6 is obtained. Then the experiment is stopped.

Number of throws needed.

It takes values in the set of positive integers: $1, 2, 3, \ldots$.

Discrete r.v.

Examples 1, 2, and 3 are *discrete variables*, taking values in a discrete set.

Discrete set means it consists of “separated points”.

A discrete set can be finite (examples 1 and 2) or countably infinite (example 3).

Example 4

Time since the last maintenance/repair to the first malfunctioning of a conditioned air equipment.

Example 5

Height (or weight, or any numerical biometrical measurement) of an individual from a given population.

Continuous r.v.

Examples 4 and 5 are *continuous variables*, taking values in an interval of real numbers.

Example 3 is a discrete r.v. with an infinite set of values.

Historical remark

Technically speaking, a r.v. is a function:

$$\Omega \to \{\text{a set of numbers or more general objects}\}.$$

The *variable* name is just an atavism.

Often r.v. are written with capital letters: X, Y, ..., their values with the same letters, lowercased: x, y, ...

Usual notations

Given a r.v. $X : \Omega \to \mathbb{R}$, and real numbers $a, b \in \mathbb{R}$:

$$\{X = a\}, \qquad \{X \leq a\}, \qquad \{a < X \leq b\},$$

represent subsets of Ω.

Namely,

$$\{X = a\} \stackrel{\text{def}}{=\!=} X^{-1}(a) = \{\omega \in \Omega : X(\omega) = a\}.$$

02 - Random variables - 01

The random variable (r.v.) concept

Discrete r.v. - pmf and cdf

Computing with discrete r.v.

Continuous r.v. - pdf - Computations

Description of a discrete r.v. X

Determined by two vectors of equal length $m \leq \infty$:

- Values: $\boldsymbol{x} = (x_j)$,
- Probabilities: $\boldsymbol{d} = (d_j)$,

Where $d_j = \mathrm{P}\{X = x_j\}$ and the d_j add up to 1.

Example

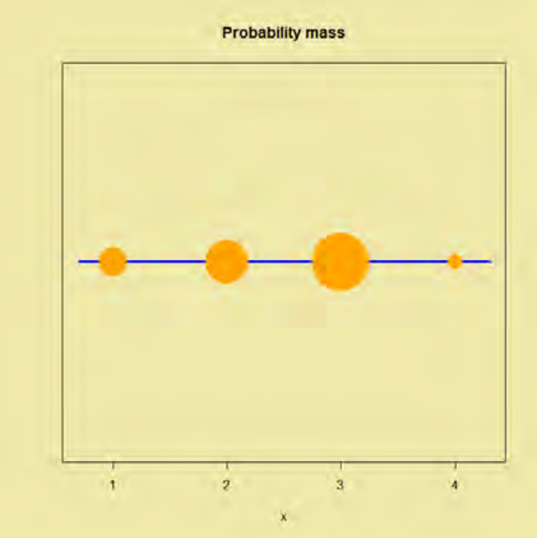
X is a r.v. taking the values:

$$\boldsymbol{x} = (1, \quad 2, \quad 3, \quad 4),$$

with probabilities:

$$\boldsymbol{d} = (0.2, \; 0.3, \; 0.4, \; 0.1).$$

Example



Probability mass function (pmf)

The *probability mass function (pmf)* of a r.v. X maps each value x_j of X to its probability:

$$d_j = \mathrm{P}\{X = x_j\},$$

and the remaining real numbers to 0.

Probability mass function (pmf)

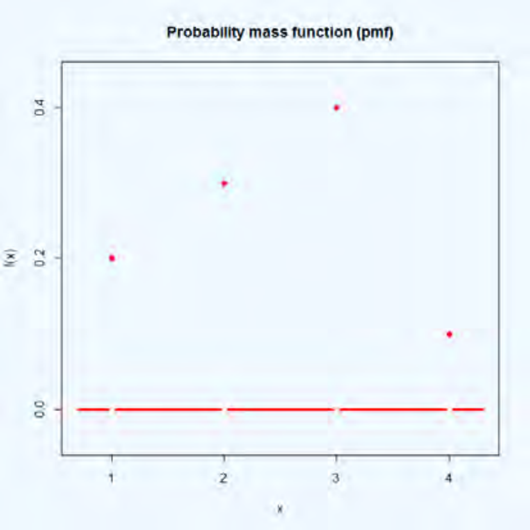
The *probability mass function (pmf)* of a r.v. X is:

$$f : \mathbb{R} \longrightarrow [0, 1],$$

defined by:

$$f(x) = \begin{cases} d_j, & \text{if} \quad x = x_j, \quad 1 \leq j \leq m, \\ 0, & \text{if} \quad x \notin \{x_1, \dots, x_m\}. \end{cases}$$

Example



(Cumulative) Probability distribution function (cdf)

The *cumulative probability distribution function* of a r.v. X, F, maps each real number x to the sum of probabilities of the (values of X) x_j smaller than or equal to x.

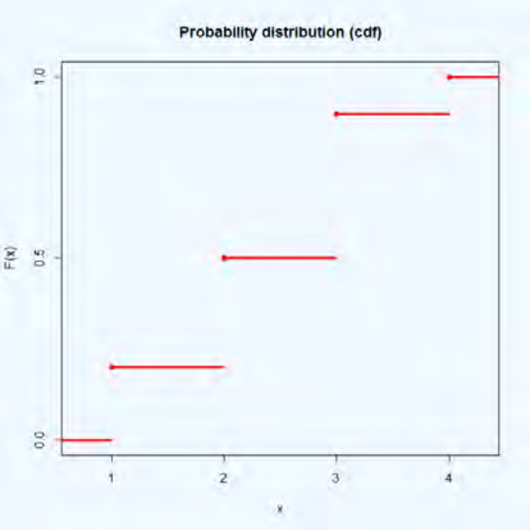
$$F(x) = \mathrm{P}\{X \leq x\}.$$

(Cumulative) Probability distribution function (cdf)

For a r.v. X, with values $x_1 < \cdots < x_m$, and probabilities $(d_1, \ldots, d_m)$. The cdf is $F : \mathbb{R} \to [0, 1]$, defined by:

$$F(x) = \begin{cases} 0, & \text{if} \quad x < x_1, \\ d_1, & \text{if} \quad x_1 \leq x < x_2, \\ d_1 + d_2, & \text{if} \quad x_2 \leq x < x_3, \\ \vdots & \vdots \quad \vdots \\ 1, & \text{if} \quad x_m \leq x. \end{cases}$$

Example



Description

The cdf of a discrete r.v. X is an increasing, right continuous, step function F, with jumps of height d_j at those points x_j which are values of X.

The vector of values of F is $\boldsymbol{p} = (0, p_1, \ldots, p_m)$, where:

$$p_j = \sum_{i=1}^{j} d_i, \quad 1 \leq i \leq m. \quad \text{In particular, } p_m = 1.$$

For the above example

X is a r.v. taking the values:

$$\boldsymbol{x} = (1, \quad 2, \quad 3, \quad 4),$$

with probabilities:

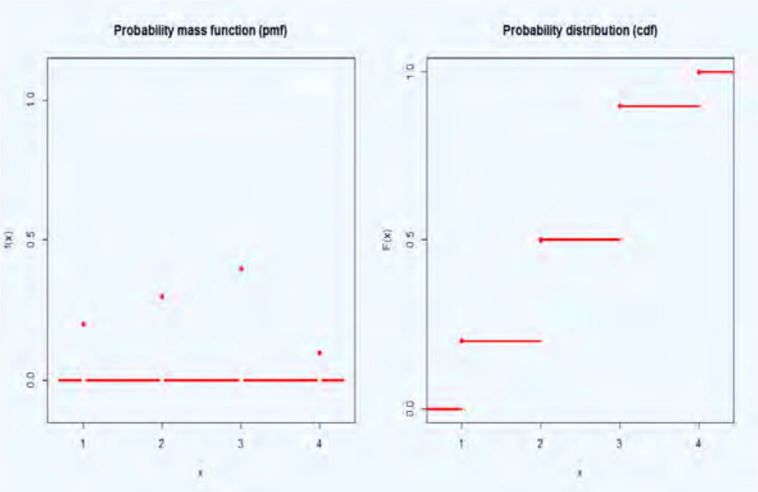
$$\boldsymbol{d} = (0.2, 0.3, 0.4, 0.1).$$

For the above example

Cdf:

$$F(x) = \begin{cases} 0, & \text{if} \quad x < 1, \\ 0.2, & \text{if} \quad 1 \leq x < 2, \\ 0.5, & \text{if} \quad 2 \leq x < 3, \\ 0.9, & \text{if} \quad 3 \leq x < 4, \\ 1, & \text{if} \quad 4 \leq x. \end{cases}$$

pmf and cdf for the example r.v.



From pmf (f) to cdf (F) and back

F's values are the cumulative sums of f's values:

$$F(x) = \sum_{t \leq x} f(t) = \mathrm{P}\{X \leq x\}, \quad x \in \mathbb{R}.$$

Given F, we recover f as its jumps function.

Each of both f and F provides all the information about the r.v.

Quantile function - Pseudoinverse of the cdf

For a r.v. with values $x_1 < \cdots < x_m$, and cumulative probabilities $\boldsymbol{p} = (0, p_1, \ldots, p_m)$, the *quantile function*, $Q : (0, 1] \to \mathbb{R}$ is defined as:

$$Q(t) = \begin{cases} x_1, & \text{if} \quad 0 < t \leq p_1, \\ x_2, & \text{if} \quad p_1 < t \leq p_2, \\ \cdot & \qquad \cdots \\ x_j, & \text{if} \quad p_{j-1} < t \leq p_j, \qquad 1 \leq j \leq m \\ \cdot & \qquad \cdots \\ x_m, & \text{if} \quad p_{m-1} < t \leq p_m = 1. \end{cases}$$

Degenerate distribution (constant r.v.)

The constant function $C : \Omega \to \mathbb{R}$, with value $c \in \mathbb{R}$ for all $\omega \in \Omega$, is a r.v.

$$\{C = x\} = \begin{cases} \emptyset, & \text{if } x \neq c, \\ \Omega, & \text{if } x = c, \end{cases}$$

for $x \in \mathbb{R}$.

pmf and cdf of a constant r.v.

The pmf is:

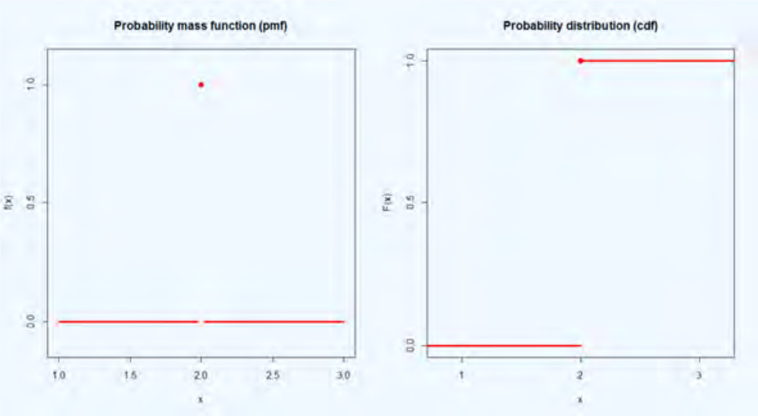
$$f(x) = \mathrm{P}\{C = x\} = \begin{cases} 0, & \text{if } x \neq c, \\ 1, & \text{if } x = c, \end{cases}$$

for $x \in \mathbb{R}$. The cdf is:

$$F(x) = \mathrm{P}\{C \leq x\} = \begin{cases} 0, & \text{if } x < c, \\ 1, & \text{if } c \leq x. \end{cases}$$

for $x \in \mathbb{R}$.

pmf and cdf of a constant r.v.



Bernoulli distribution

Distribution of $X = \mathbb{1}_A : \Omega \to \mathbb{R}$, indicator of $A \subset \Omega$.

$$X(\omega) = \begin{cases} 1, & \text{if} \quad \omega \in A, \\ 0, & \text{if} \quad \omega \notin A. \end{cases}$$

Values: $(0, 1)$. Probabilities: $(1 - p, p)$, $p = \mathrm{P}(A)$.

Notation: $X \sim \mathrm{Ber}(p)$.

pmf and cdf of an $X \sim \mathrm{Ber}(p)$

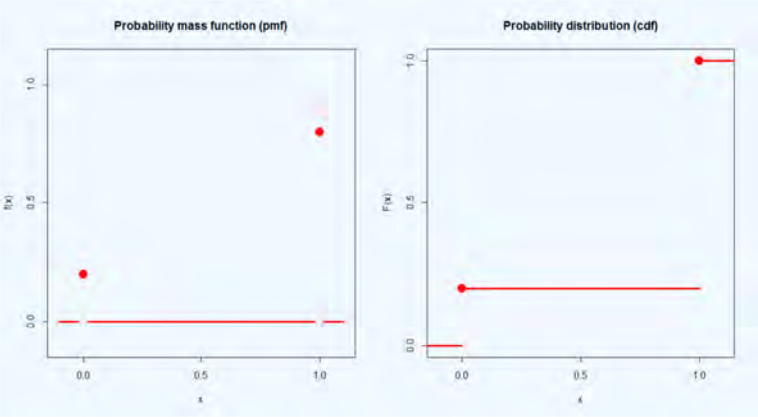
The pmf is:

$$f(x) = \begin{cases} 0, & \text{if } x \notin \{0, 1\}, \\ 1 - p, & \text{if } x = 0, \\ p, & \text{if } x = 1, \end{cases} \qquad \text{for } x \in \mathbb{R}.$$

The cdf is:

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - p, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } 1 \leq x, \end{cases} \qquad \text{for } x \in \mathbb{R}.$$

pmf and cdf of a Bernoulli r.v. with $p = 0.8$



Hypergeometric distribution

Defined as the distribution of the number X of white balls drawn when extracting without replacement n balls from an urn containing $N = N_1(\text{white}) + N_2(\text{black})$ balls.

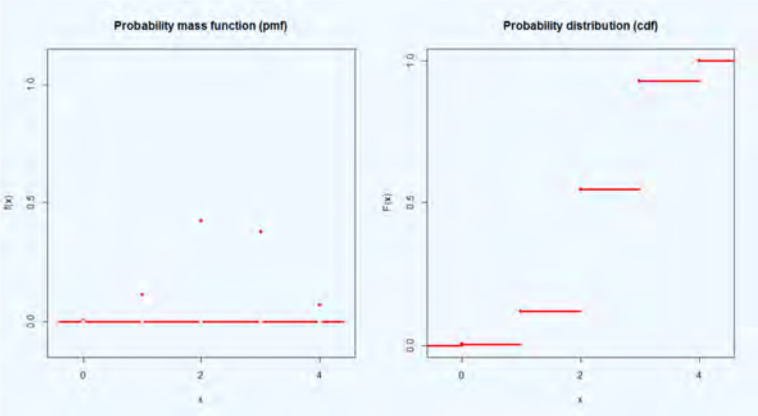
Hypergeometric pmf

$$f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots, n.$$

$N = N_1 + N_2$.

Notation: $\mathrm{Hyper}(N_1, N_2, n)$.

pmf and cdf Hyper$(N_1 = 6, N_2 = 4, n = 4)$



Binomial distribution

$n \geq 1$ independent repetitions of an experiment.

In each of them we record occurrence of an event A of probability p.

$$\text{The r.v.} \quad X = \text{Number of occurrences of } A,$$
$$\textit{(absolute frequency of } A\textit{)},$$

has a *binomial distribution* with parameters n, p.

Notation: $X \sim \mathrm{B}(n, p)$.

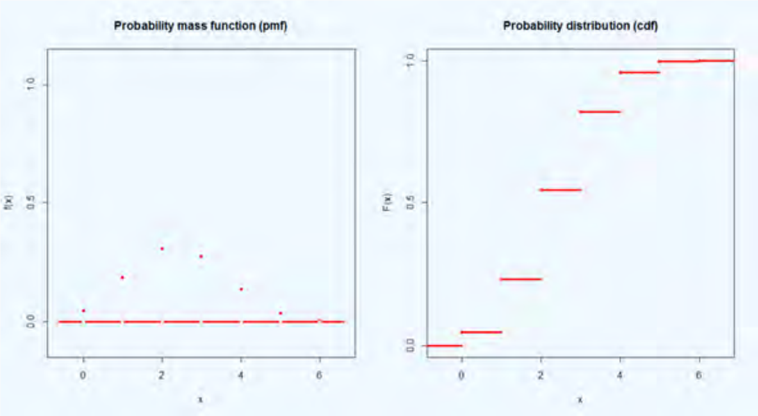
Pmf of $X \sim \mathrm{B}(n, p)$

For $0 \leq k \leq n$,

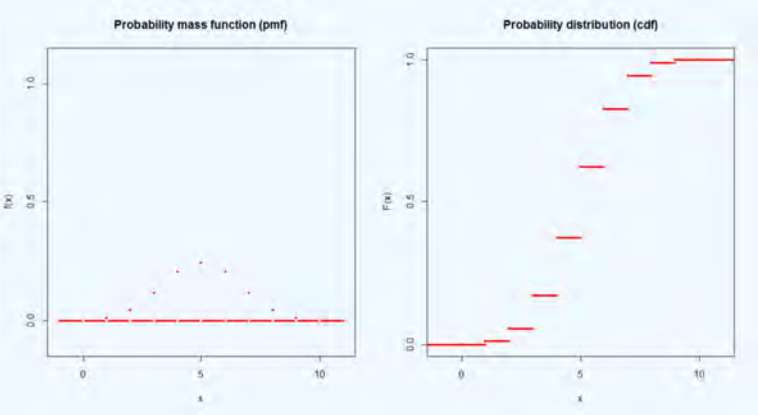
$$f(k) = \mathrm{P}(X = k)$$

$$= \binom{n}{k} p^k \,(1 - p)^{(n-k)}.$$

Pmf and cdf of a $B(6, 0.4)$



Pmf and cdf of a $B(10, 0.5)$



Sum of binomial probabilities

From Newton's binomial theorem it follows that the sum of all probabilities for $X \sim \mathrm{B}(n, p)$ is 1.

$$\begin{aligned}\sum_{k=0}^{n} f(k) = \sum_{k=0}^{n} \mathrm{P}(X = k) \\ = \sum_{k=0}^{n} \binom{n}{k} p^k \, (1-p)^{(n-k)} \\ = (p + (1-p))^n = 1.\end{aligned}$$

Infinite discrete variables

The set $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{N}}$ of values is countably infinite, e.g. $\boldsymbol{x} = \mathbb{Z}_+$.

Everything is *"almost"* like in the finite case.

The infinite sequence $\boldsymbol{d} = \{d_n\}_{n\in\mathbb{N}}$ of probabilities must be summable, with sum equal to 1.

Geometric r.v.

Independent repetitions of an experiment. In each of them we record occurrence of event A with probability p. Stopped on the first occurrence. Define:

$$X = \text{“Number of repetitions needed until } A \text{ occurs”}.$$

The pmf is:

$$f(x) = d_x = \mathrm{P}\{X = x\} = (1-p)^{x-1}\, p, \qquad x \in \mathbb{N}.$$

Notation: $X \sim \mathsf{Geom}(p)$.

Geometric r.v. (alternative notation)

Number $Y = X - 1$ of A^c results obtained before A.

Posible values are $0, 1, 2, \ldots$

In terms of Y, the pmf is:

$$f_Y(y) = \mathrm{P}\{Y = y\} = (1-p)^y\, p, \qquad y = 0, 1, \ldots$$

In `R (stats)`, `dgeom` & related functions use this convention

Sum of geometric probabilities

$\{d_k : k \in \mathbb{N}\}$ is a geometric progression with ratio $r = 1 - p$ and first term $d_1 = p$.

The sum of n terms is:

$$s_n = \frac{d_n\, r - d_1}{r - 1} = 1 - (1 - p)^n.$$

The $\{d_k\}$ sequence is sumable, with sum:

$$\lim_{n \to \infty} s_n = 1.$$

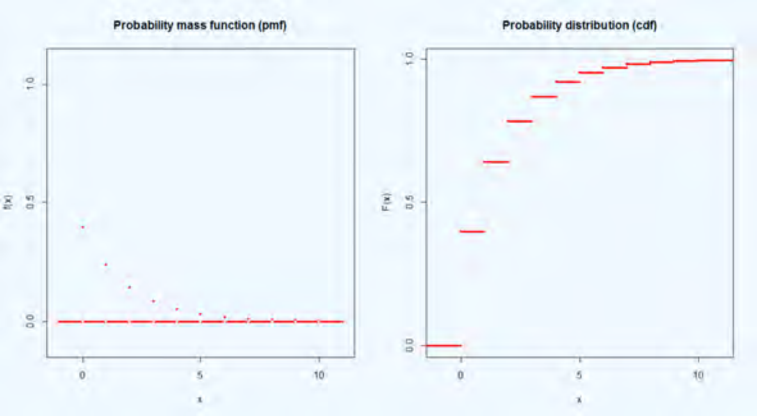
Cdf of a geometric r.v.

The cdf is:

$$F(x) = \begin{cases} 0, & \text{if} \quad x < 0, \\ 1 - (1-p)^k, \ \text{on } k = [x], & \text{if} \quad 0 \leq x, \end{cases}$$

for $x \in \mathbb{R}$, where $[\,\cdot\,]$ is the floor (integer part) function.

pmf and cdf of a Geom(0.4) r.v.



[1] In this plot x takes values $0, 1, \ldots$, using R convention.

Negative binomial r.v. , $\mathrm{NegBin}(r,p)$ or $\mathrm{NB}(r,p)$

For $r \in \mathbb{R}_+$, defined by its pmf:

$$f_Y(y) = \frac{\Gamma(r+y)}{\Gamma(r)\, y!}\, p^r\, (1-p)^y, \qquad y = 0, 1, 2, \ldots$$

If $r \in \mathbb{N}$ it generalizes the geometric distribution: Number of independent repetitions of a binary experiment with outcomes $\{A, A^c\}$, $p = \mathrm{P}(A)$, needed to obtain $r \in \mathbb{N}$ times A, then stop the experiment.

Negative binomial r.v. , $\mathsf{NegBin}(r, p)$ **or** $\mathsf{NB}(r, p)$

As in the Geom case the variable is either:

$Y =$ number of A^c outcomes needed to obtain r A's, with possible values $y = 0, 1, 2, \ldots$, or

$X = Y + r$, total number of repetitions, with possible values $x = r, r + 1, 2, \ldots$.

If $r \in \mathbb{N}$ the pmf is:

$$f_Y(y) = \binom{r+y-1}{r-1} p^r \, (1-p)^y, \qquad y = 0, 1, 2, \ldots$$

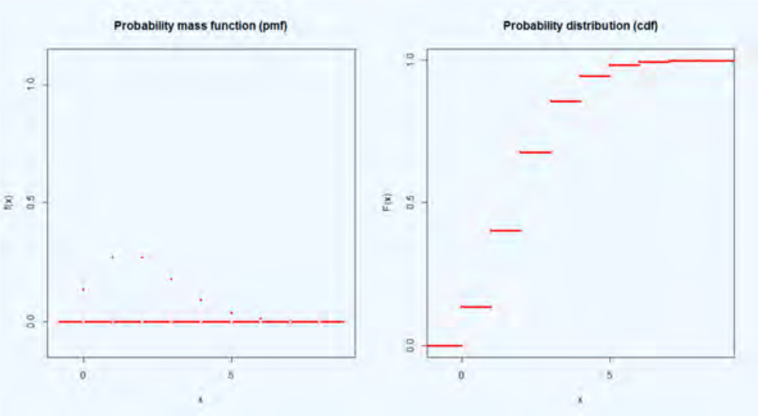
Definition of a Poisson(λ) r.v.

The Poisson distribution with parameter $\lambda \in \mathbb{R}_+$ has values $\boldsymbol{x} = \{0, 1, 2, \ldots\}$ amd pmf:

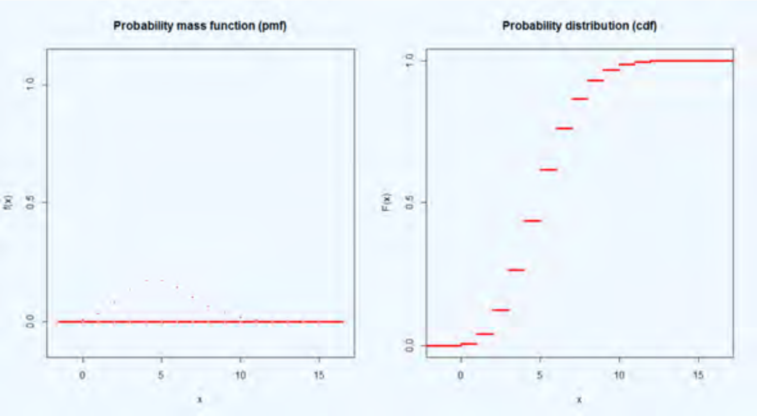
$$f(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Notation: Poisson(λ).

pmf and cdf of a Poisson(2) r.v.



pmf and cdf of a Poisson(5) r.v.



Binomial with a large n and a very small p

The Poisson pmf can be derived as a limit of a binomial pmf, letting:

$$n \to \infty, \qquad p \to 0, \qquad n\,p \to \lambda.$$

Setting $p = \lambda/n$ in the binomial pmf:

$$\frac{n!}{k!\,(n-k)!}\,\left(\frac{\lambda}{n}\right)^{k}\,\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Binomial with a large n and a very small p

$$= \frac{\lambda^k}{k!} \cdot A_n \cdot B_n \cdot C_n, \qquad \text{where:}$$

$$A_n = \frac{n\,(n-1)\,\cdots(n-k+1)}{n^k} \to 1,$$

$$B_n = \left(1 - \frac{\lambda}{n}\right)^n \quad \to e^{-\lambda},$$

$$C_n = \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1.$$

Example: number of accidents

This limiting process is natural in accident statistics.

As a first approximation, accident counting is a binomial experiment (independent repetitions of a random experiment).

The probability p that a given individual suffers an accident is very small, while the number n of individuals is large.

Siméon-Denis Poisson (1781 – 1840)

Discoverer of his namesake distribution, from the real forensic problem of evaluating the probability of a correct outcome from a trial where a jury decides the verdict.

Recherches sur la probabilité des jugements (1837).

Ladislaus Bortkiewicz (1868 – 1931)

Rediscovered Poisson's law, in his book *Das Gesetz der kleinen Zahlen (The law of small numbers)* (1898).

Including the classic analysis of data on deaths of soldiers in the Prussian army caused by horsekicks during the 1875-1894 period.

02 - Random variables - 01

Description of a discrete r.v.

Vector of values:

$$\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m,$$

which we assume ordered, $x_1 < \cdots < x_m$,

Vector of probabilities:

$$\boldsymbol{d} = (d_1, \ldots, d_m), \quad d_j \in (0, 1), \quad \sum_{j=1}^{m} d_j = 1.$$

R syntax

The `cumsum` and `diff` functions.

Given `d`:

```
p<-c(0,cumsum(d))
```

Given `p` (including the initial 0):

```
d<-diff(p)
```

Probability of an event

For a subset $A \subset \{x_1, \ldots, x_m\}$,

$$\mathrm{P}(A) = \sum_{j : x_j \in A} d_j,$$

is the sum of probabilities of the elements in A.

Probability of an interval

With the cdf function F:

Since, by definition, $F(a) = \mathrm{P}\{X \leq a\}$, $a \in \mathbb{R}$.

Aplying properties of a probability:

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

$$\mathrm{P}\{a < X \leq b\} = F(b) - F(a),$$

$$\mathrm{P}\{a \leq X \leq b\} = F(b) - F(a) + \mathrm{P}\{X = a\},$$

The discrete uniform distribution – Generalized die

Given $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$, the *discrete uniform r.v. with values* $\boldsymbol{x}$ has probabilities:

$$\boldsymbol{d} = (d_1, \ldots, d_m), \quad d_j = \frac{1}{m}, \quad 1 \leq j \leq m.$$

In particular, when $\boldsymbol{x} = (1, 2, 3, \ldots, m)$, we have a *generalized die,* with m faces.

Pmf of a discrete uniform r.v.

$$f:\mathbb{R}\longrightarrow[0,1],$$

defined by:

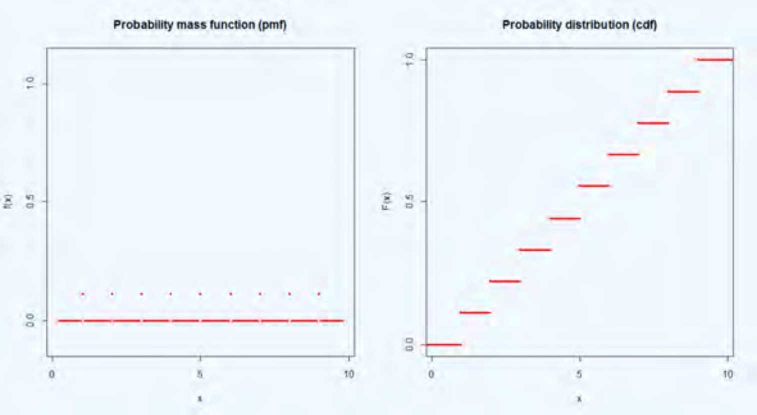
$$f(x)=\begin{cases}\dfrac{1}{m}, & \text{if} \quad x=x_j, \quad 1\le j\le m,\\[2ex] 0, & \text{if} \quad x\notin\{x_1,\dots,x_m\}.\end{cases}$$

Cdf of a discrete uniform r.v.

Assuming $x_1 < \cdots < x_m$, the cdf is:

$$F(x) = \begin{cases} 0, & \text{if} \quad x < x_1, \\ \vdots & \vdots \quad\; \vdots \\ \dfrac{i}{m}, & \text{if} \quad x_i \leq x < x_{i+1}, \quad 1 \leq i \leq m-1, \\ \vdots & \vdots \quad\; \vdots \\ 1, & \text{if} \quad x_m \leq x. \end{cases}$$

pmf and cdf of a discrete uniform r.v.



Binomial computations with `R`

Pmf:

`dbinom(x,size=n,prob=p)`

Cdf:

`pbinom(x,size=n,prob=p)` .

Quantile function (the pseudoinverse of the cdf)

`qbinom(t,size=n,prob=p)`.

Example

50 tosses of a perfect coin.

Probability of obtaining a number of heads comprised between 23 and 27, including both borders:

`pbinom(27,size=50,prob=0.5)`
`-pbinom(22,size=50,prob=0.5)=0.520`.

Probability of obtaining 45 or more heads:

`1-pbinom(44,size=50,prob=0.5)=2.10e-09`.

Hypergeometric pmf

$X =$ number of white balls obtained from n extractions without replacement from an urn containing $N = N_1(\text{white}) + N_2(\text{black})$ balls.

$$f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots, n.$$

$N = N_1 + N_2.$

Hypergeometric computations with R

Pmf: `dhyper(x, m, n, k)`.

- `m` $\equiv N_1$, number of white balls in the urn.
- `n` $\equiv N_2$, number of black balls in the urn.
- `k` $\equiv n$, number of extractions.

Cdf: `phyper(x, m, n, k)`.

Quantile function: `qhyper(t,size=n,prob=p)`.

Example

An urn contains 50 white balls and 50 black balls.

50 balls are extracted without replacement.

Probability of obtaining a number of white balls comprised between 23 and 27, including both borders:

```
phyper(27,m=50,n=50,k=50)
      -phyper(22,m=50,n=50,k=50)
 =0.68266
```

A remarkable property of the geometric distribution

Assume we model the *waiting time* until the occurence of a certain event A with a geometric r.v. X.

We will see that this distribution is *memoryless:* meaning the probability of waiting for k time units longer is always the same, even when we already have been waiting an arbitrary number n of time units.

Waiting for k time units longer

With a r.v. $X \sim \mathrm{Geom}(p)$,

$$d_n \equiv \mathrm{P}\{X = n\} = q^{n-1}\, p, \qquad q \equiv (1 - p),$$

$$p_n \equiv \mathrm{P}\{X \leq n\} = \sum_{m=1}^{n} d_m = \frac{q^n\, p - p}{q - 1} = 1 - q^n,$$

$n = 1, 2, \ldots$

Waiting for k time units longer

The complementary event:

$$P\{X > n\} = q^n, \quad P\{X > n+k\} = q^{n+k},$$

The conditional probability:

$$\begin{aligned} P\{X > n+k | X > n\} &= \frac{P\left(\{X > n+k\} \cap \{X > n\}\right)}{P\{X > n\}} \\ &= \frac{P\{X > n+k\}}{P\{X > n\}} = q^k. \end{aligned}$$

Interpretation of the result

We have obtained:

$$\mathrm{P}\{X > n + k | X > n\} = \mathrm{P}\{X > k\}$$

That is, the information that we have been waiting for n time units does not affect the probability of waiting k time units more.

02 - Random variables - 01

General r.v.

Every r.v. has a cdf:

$$F : \mathbb{R} \longrightarrow [0, 1],$$

defined as:

$$F(x) = \mathrm{P}\{X \leq x\}, \quad x \in \mathbb{R},$$

F is a non decreasing, right continuous function such that $F(-\infty) = 0$, $F(+\infty) = 1$.

Characterizing discrete r.v.

Discrete r.v. are exactly those whose cdf F is a step function.

Its discontinuities are (finite) jumps, which occur on a finite or countable set of points.

Jump points are those where the r.v. has a non-null probability:

$$\mathrm{Jump}(F, a) = F(a) - \lim_{x \to a-} F(x) = \mathrm{P}\{X = a\}.$$

Absolutely continuous r.v.

When F, the cdf of a r.v. X, is equal to the integral of another function f,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \quad x \in \mathbb{R},$$

then this $f = F'$ is the *probability density function (pdf)* of the *absolutely continuous* r.v. X.

Necessarily $f \geq 0$ and $\int_{-\infty}^{\infty} f = 1$.

Analogies [abs. cont. r.v.] $\leftrightarrow$ [discrete r.v.]

The pdf of a continuous r.v. has properties analogous to those of the pmf of a discrete r.v. .

This is why we use the same symbol f for both.

Intuitively, the analogy corresponds to "replacing sums with integrals".

Differences [abs. cont. r.v.] $\leftrightarrow$ [discrete r.v.]

If F is a step function (hence the r.v. is discrete) its derivative is 0 except for X values, where it is discontinuous.

The values of a pmf are probabilities. In particular they lie between 0 and 1.

The probability of $A \subset \mathbb{R}$ is a sum (of probabilities of $x_i \in A$).

Differences [abs. cont. r.v.] $\leftrightarrow$ [discrete r.v.]

Values of a pdf f are not probabilities.

$f \geq 0$ but its values can be arbitrarily large -on a sufficiently small interval- provided that $\int_{\mathbb{R}} f = F(+\infty) = 1$.

The probability of $A \subset \mathbb{R}$ is the *integral* of the pdf on A.

Computing probabilities with continuous r.v.

X continuous, with pdf f and cdf F:

For $a, b \in \bar{\mathbb{R}}$, $-\infty \leq a \leq b \leq +\infty$,

$$\mathrm{P}(a < X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a).$$

In particular,

$$\mathrm{P}(X = a) = 0, \quad \text{for } a \in \mathbb{R}.$$

Uniform (rectangular) distribution

Given $a, b \in \mathbb{R}$, $a < b$, a r.v. $X \sim \mathrm{Unif}(a, b)$ if it is continuous with pdf:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } x \in (a, b), \\ 0, & \text{if } x \notin (a, b). \end{cases}$$

Cdf of a r.v. $X \sim \mathrm{Unif}(a, b)$

$$F(x) = \begin{cases} 0, & \text{if } x < a, \\ \dfrac{x-a}{b-a}, & \text{if } x \in [a, b), \\ 1, & \text{if } b \leq x. \end{cases}$$

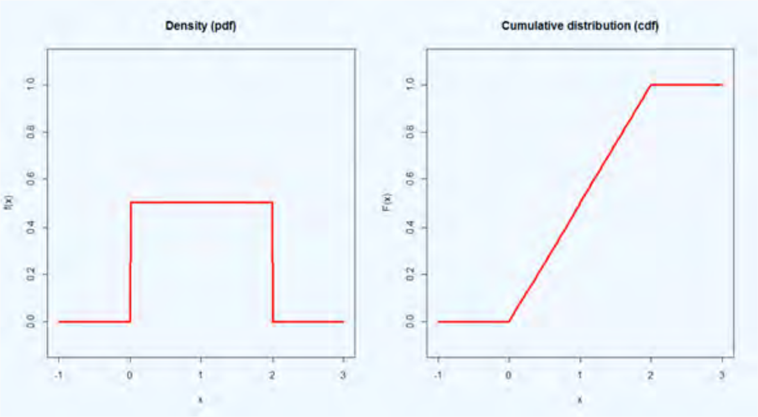
Probability of an interval with $X \sim \mathrm{Unif}(a, b)$

For an interval $(x_1, x_2) \subset (a, b)$,

$$\mathrm{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = \frac{x_2 - x_1}{b - a}.$$

The probability is proportional to the interval length.

Pdf and cdf of a uniform distribution on $[0, 2]$



Exponential distribution

A r.v. taking values on $(0, \infty)$ is an *exponential with (rate) parameter* $\lambda > 0$ if it is continuous, with pdf:

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ \lambda \exp(-\lambda\, x), & \text{if } 0 \leq x. \end{cases}$$

Notation: $X \sim \mathrm{Exp}(\lambda)$.

Cdf of a r.v. $X \sim \mathrm{Exp}(\lambda)$

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - \exp(-\lambda\, x), & \text{if } 0 \leq x. \end{cases}$$

Probabilities with a r.v. $X \sim \mathrm{Exp}(\lambda)$

Given an interval $(x_1, x_2) \subset \mathbb{R}_+$,

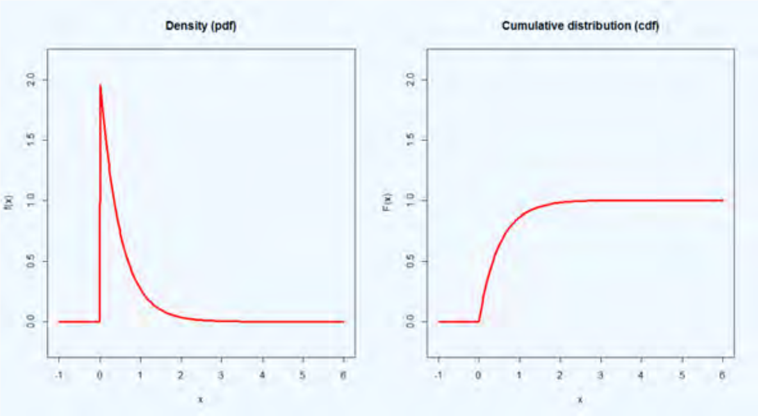
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1)$$

$$= e^{-\lambda\,x_1} - e^{-\lambda\,x_2}.$$

In particular,

$$P(X > x_1) = \int_{x_1}^{\infty} f(x)\,dx = 1 - F(x_1) = e^{-\lambda\,x_1}.$$

Pdf and cdf of an $\mathrm{Exp}(\lambda = 2)$



The memoryless property of the exponential distribution

$X \sim \mathrm{Exp}(\lambda)$.

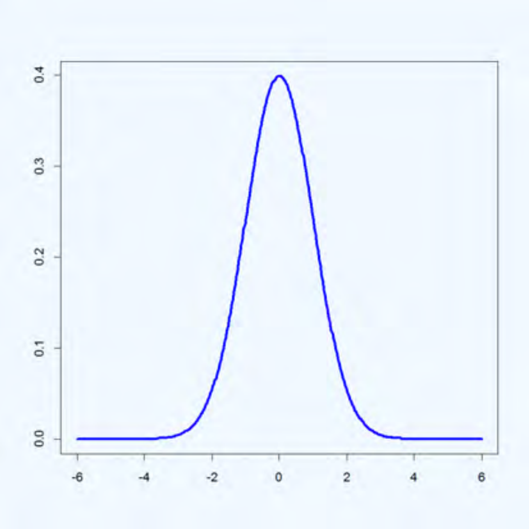
Given $x_0, x_1 > 0$, the following equality is satisfied:

$$\mathrm{P}(X > x_0 + x_1 | X > x_0) = \mathrm{P}(X > x_1).$$

The memoryless property of the exponential distribution

$$
\begin{aligned}
\mathrm{P}(X > x_0 + x_1 | X > x_0)& \\
&= \frac{\mathrm{P}((X > x_0 + x_1) \cap (X > x_0))}{\mathrm{P}(X > x_0)} \\
&= \frac{\mathrm{P}(X > x_0 + x_1)}{\mathrm{P}(X > x_0)} = \frac{e^{-\lambda\,(x_0+x_1)}}{e^{-\lambda\,x_0}} \\
&= e^{-\lambda\,x_1} = \mathrm{P}(X > x_1).
\end{aligned}
$$

Normal pdf (Gaussian bell-shaped curve)



Definition

A r.v. X has a *normal or gaussian distribution*, with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$, if it is continuous with pdf:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

Notation: $X \sim \mathrm{N}(\mu, \sigma^2)$.

Properties of the normal pdf

Clearly $f \geq 0$.

It can be proved that f is integrable and $\int_{-\infty}^{\infty} f = 1$.

$\lim_{x\to-\infty} f(x) = \lim_{x\to+\infty} f(x) = 0$.

f attains a maximum, $1/\sqrt{2\pi\sigma^2}$, for $x = \mu$, it increases for $x < \mu$ and decreases for $x > \mu$. Symmetry axis at $x = \mu$.

Inflection at $\mu \pm \sigma$

Meaning of the parameters

μ is the *mean* of X.

μ is the symmetry axis:

$$f(\mu + a) = f(\mu - a), \quad a \in \mathbb{R}.$$

Meaning of the parameters

σ^2 is the *variance* of X.

$\sigma \equiv \sqrt{\sigma^2}$ is the *standard deviation* of X.

σ is a measure of the relative width of the Gaussian bell-shaped curve (equivalently, the measurement unit or scale of its x-axis).

Computing normal probabilities

If $X \sim \mathrm{N}(\mu, \sigma^2)$, given $a, b \in \mathbb{R}$,

$\mathrm{P}(a < X \leq b)$ is the integral of the pdf on (a, b):

$$\mathrm{P}(a < X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx.$$

The cdf of a normal r.v.

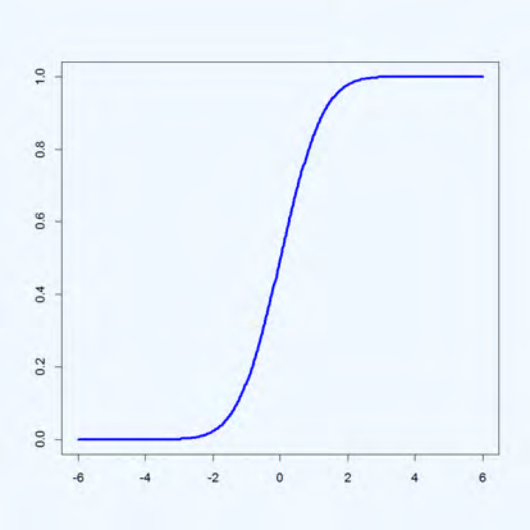
Is given by the indefinite integral of the pdf:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}\right\} dt, \quad x \in \mathbb{R}.$$

F has no expression *in terms of elementary functions.*

It must be evaluated by numerical approximation.

The cdf of a normal r.v.



Normal probabilities from the cdf

If $X \sim \mathrm{N}(\mu, \sigma^2)$ and F is its cdf, given $a, b \in \overline{\mathbb{R}}$, $-\infty \leq a < b \leq +\infty$, then:

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

The standard normal distribution

By definition, it is the $\mathrm{N}(0,1)$ distribution, with $\mu = 0$ and $\sigma^2 = 1$.

Its pdf is:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\,x^2\right\}, \quad x \in \mathbb{R}.$$

Relating a normal pdf with the standard normal pdf

f, the pdf of a $\mathrm{N}(\mu, \sigma^2)$ r.v. , can be obtained by performing a translation and a scale change on ϕ, the pdf of a $\mathrm{N}(0, 1)$ r.v.

$$f(x) = \phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}.$$

Relating a normal r.v. with the standard normal r.v.

1. If $X \sim \mathrm{N}(\mu, \sigma^2)$, then the r.v. :

$$Z \equiv \frac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1)$$

   is the *standardized* X.

2. Conversely, if $Z \sim \mathrm{N}(0, 1)$, then the r.v. :

$$X \equiv \mu + \sigma\, Z \sim \mathrm{N}(\mu, \sigma^2).$$

   Every normal r.v. can be obtained in this way.

Computing normal probabilities with `R`

`dnorm(x,mean=m,sd=s)` is the pdf.

`pnorm(x,mean=m,sd=s)` is the cdf.

`qnorm(x,mean=m,sd=s)` is the *quantile function*, the inverse of `pnorm`.

Example 1

Assuming that heights of 800 newborn babies are normally distributed with mean 52 cm and standard deviation 5cm. How many of them can be expected to have a height between 53 cm and 57 cm?

```
p<-pnorm(57,mean=52,sd=5)-pnorm(53,mean=52,sd=5)
0.262085
N<-p*800
209.67
```

Result: aproximately 210 newborn babies.

Example 2

Heights of individuals from two populations A and B are normally distributed with mean 1.70 metres (equal mean for both populations) and standard deviations σ_A and σ_B, respectively, where $\sigma_A < \sigma_B$.

We select a random individual from each population.

Which one has a larger probability of having a height between 1.68 and 1.72 metres?

Example 2

$$X_A \sim \mathrm{N}(1.70, \sigma_A), \quad X_B \sim \mathrm{N}(1.70, \sigma_B).$$

$$p_A = \mathrm{P}(1.68 < X_A < 1.72), \quad p_B = \mathrm{P}(1.68 < X_B < 1.72).$$

$$Z_A = \frac{X_A - 1.70}{\sigma_A} \sim \mathrm{N}(0, 1), \quad Z_B = \frac{X_B - 1.70}{\sigma_B} \sim \mathrm{N}(0, 1)$$

Example 2

$$
\begin{aligned}
p_A &= \mathrm{P}(1.68 < X_A < 1.72) \\
&= \mathrm{P}\left(\frac{1.68 - 1.70}{\sigma_A} < Z_A < \frac{1.72 - 1.70}{\sigma_A}\right) \\
&= \mathrm{P}\left(\frac{-0.02}{\sigma_A} < Z_A < \frac{0.02}{\sigma_A}\right).
\end{aligned}
$$

Similarly: $p_B = \mathrm{P}\left(\dfrac{-0.02}{\sigma_B} < Z_B < \dfrac{0.02}{\sigma_B}\right).$

Example 2

Comparing both intervals:

$$I_A = \left(\frac{-0.02}{\sigma_A},\ \frac{0.02}{\sigma_A} \right),$$

$$I_B = \left(\frac{-0.02}{\sigma_B},\ \frac{0.02}{\sigma_B} \right),$$

the condition $\sigma_A < \sigma_B$ is equivalent to $I_A \supset I_B$.

Since both Z_A i Z_B are $\mathrm{N}(0,1)$ it follows that $p_A \geq p_B$.

Example 3

Assume the total load X of an elevator carrying 4 people is distributed as a $\mathsf{N}(\mu, \sigma^2)$, where $\mu = 270$ kg and $\sigma = 13$ kg.

What is the maximum load the elevator has to accept in order to operate safely 99% of its active time?

Example 3

We want to find c such that:

$$\mathrm{P}(X \leq c) = 0.99,$$

i.e., c such that:

$$F(c) = 0.99,$$

where F is the cdf of X or, equivalently, $c = F^{-1}(0.99)$.

With R: `c<-qnorm(0.99,mean=270,sd=13)`

$$c = 300.24.$$

Example 4

Assume the duration T of a washing machine to the first breakdown, is a normal r.v. with $\mu = 10$ years, with standard deviation $\sigma = 3$ years.

Which duration interval, centered at μ, has 90% probability?

Which warranty period should be given in order to repair for free no more than 1% of machines?

Example 4

We want to compute $a > 0$ such that:

$$P(\mu - a < T \leq \mu + a) = 0.90.$$

The sum of probabilities of both tails, $P(T \leq \mu - a)$ and $P(T > \mu + a)$, is $1 - 0.90 = 0.10$.

Due to the symmetry of the pdf with respect to μ, both probabilities coincide, being equal to 0.05.

Example 4

It follows that:

$$F(\mu + a) = \mathrm{P}(T \leq \mu + a) = 0.95.$$

Hence, $\mu + a = F^{-1}(0.95)$, i.e., $a = F^{-1}(0.95) - \mu$.

With R:

`a<-qnorm(0.95,mean=10,sd=3)-10`,

$a = 4.93$.

Example 4

To answer the second question, we seek b such that:

$$\mathrm{P}(T < b) = F(b) = 0.01.$$

Thus, we must compute $b = F^{-1}(0.01)$. With R:

`b<-qnorm(0.01,mean=10,sd=3)`, giving: $b = 3.02$.

Conclusion: with a 3 years warranty less than 1% machines are expected to breakdown within the guaranteed period.