02 - Random variables - 01

Bayesian Statistics

Spring 2022-2023

02 - Random variables - 01

The random variable (r.v.) concept

Discrete r.v. - pmf and cdf

Continuous r.v. - pdf and cdf

Expectation, variance, higher moments

Asymptotics: LLN and CLT

Bivariate r.v. - Bayes' rule for r.v.

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.

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)
```

Example

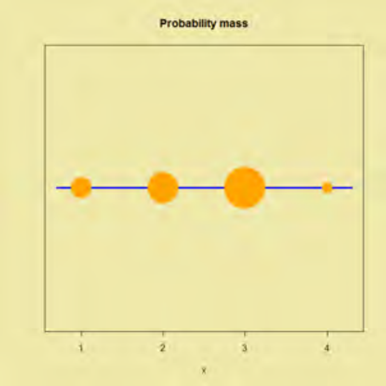
X is a r.v. taking the values:

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

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)

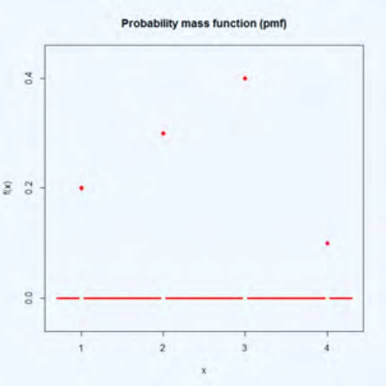
The *probability mass function (pmf)* of a r.v. X is:

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

defined by:

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

Example



Cumulative distribution function (cdf)

cdf of a r.v. X, F or F_X, maps each $x \in \mathbb{R}$

to the sum of probabilities of X values

smaller than or equal to x.

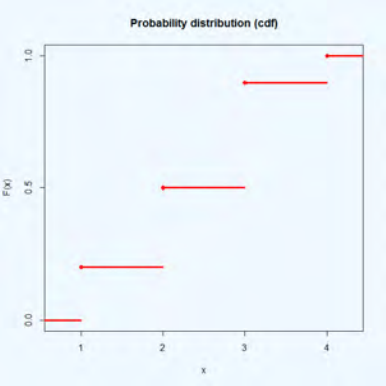
$$F(x) = \mathrm{P}\{X \leq x\}.$$

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



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:

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

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

F values are the cumulative sums of f 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 has all the information about X.

Quantile function - Pseudoinverse of the cdf

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

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

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 $X \sim \mathrm{Ber}(p)$

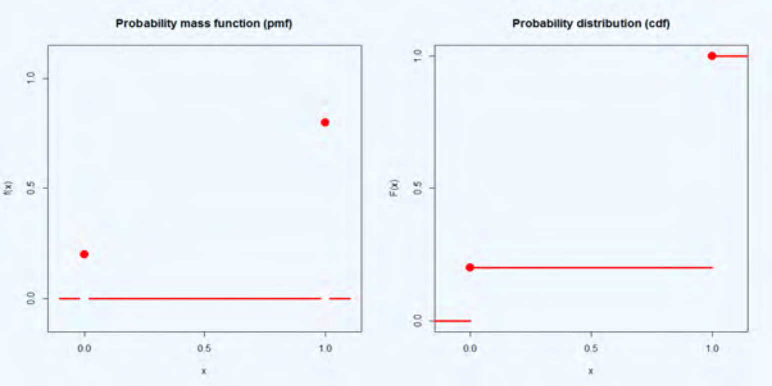
The pmf is:

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

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.

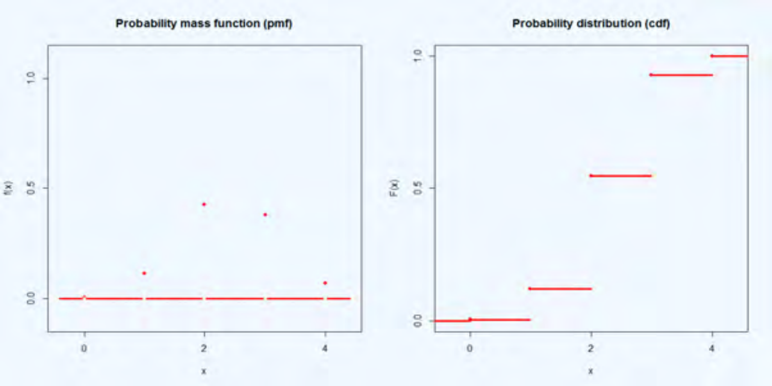
Hypergeometric pmf

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

$N = N_1 + N_2$.

Notation: 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 a binary experiment.

In each of them we register an event A of probability p.

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

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

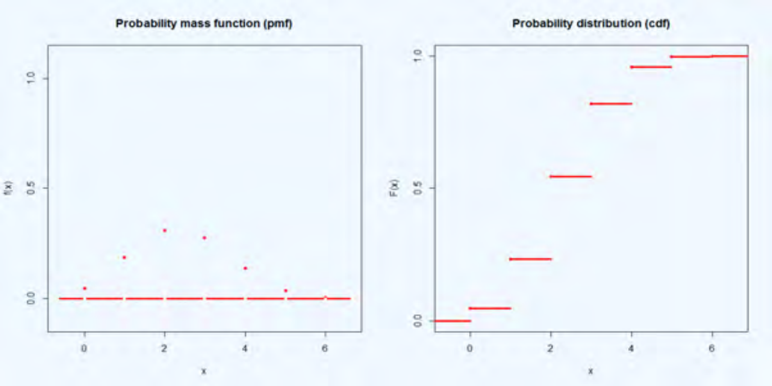
Notation: $X \sim \mathrm{Binom}(n, p)$.

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

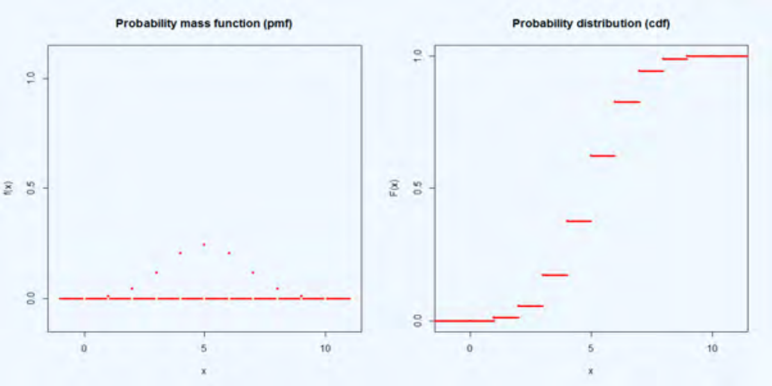
For $0 \leq k \leq n$,

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

Pmf and cdf of a Binom(6, 0.4)



Pmf and cdf of a Binom(10, 0.5)



Infinite discrete variables

Countably infinite set of values $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{N}}$.

Everything is *"almost"* like 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 a binary experiment.

Stop on first occurrence of A, $p = \mathrm{P}(A)$.

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

Pmf is:

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

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

Geometric r.v. (alternative notation)

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

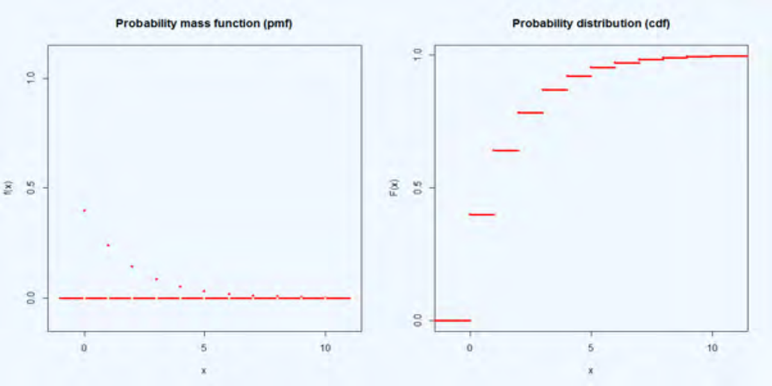
Posible values are $0, 1, 2, \ldots$

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

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. , BN(r, p) or NB(r, p)

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

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

When $r \in \mathbb{N}$, extension of 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. , BN(r, p) or 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 values $y = 0, 1, 2, \ldots$, or

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

For $r \in \mathbb{N}$: $f_Y(y) = \binom{r+y-1}{r-1} p^r \, (1-p)^y, \quad y = 0, 1, \ldots$

Definition of a Poisson(λ) r.v.

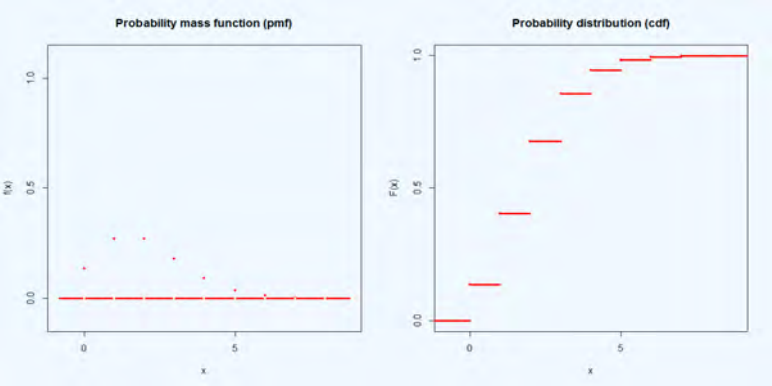
Poisson distribution with parameter $\lambda \in \mathbb{R}_+$

Values $x = \{0, 1, 2, \dots\}$. Pmf:

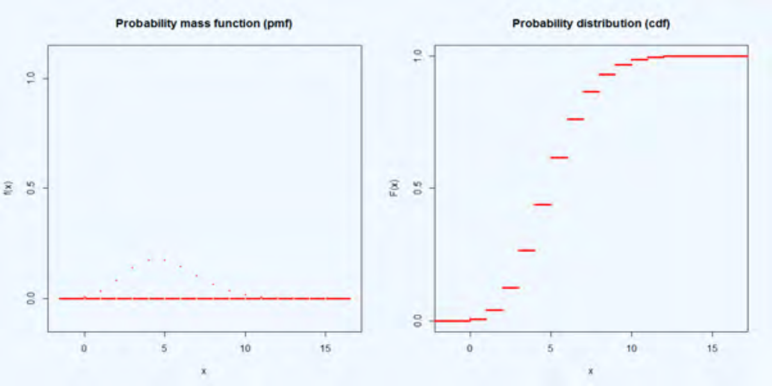
$$f(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Notation: Poisson(λ).

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



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



Discrete uniform distribution – Generalized die

Discrete uniform r.v. Values: $\boldsymbol{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$.

Probabilities:

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

When $\boldsymbol{x} = (1, 2, 3, \dots, m)$,

we have a *generalized die*, with m faces.

Pmf of a discrete uniform r.v.

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

defined by:

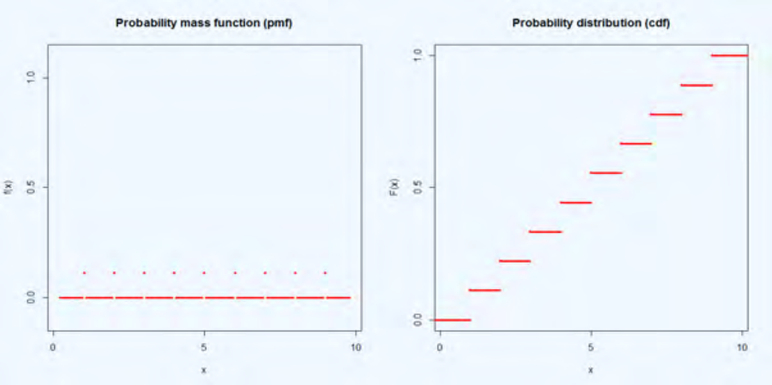
$$f(x) = \begin{cases} \dfrac{1}{m}, & \text{if} \quad x = x_j, \quad 1 \leq j \leq m, \\ 0, & \text{if} \quad x \notin \{x_1, \ldots, x_m\}. \end{cases}$$

Cdf of a discrete uniform r.v.

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

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

pmf and cdf of a discrete uniform r.v.



Hypergeometric pmf

n extractions without replacement from an urn with
$N = N_1(\text{white}) + N_2(\text{black})$ balls.

$X = \sharp\ (\text{white balls})$.

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

$N = N_1 + N_2$.

02 - Random variables - 01

General r.v.

Every r.v. has a cdf:

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

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

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

Characterizing discrete r.v. by their cdf

Discrete r.v. are those with F a step function.

Discontinuities of F are (finite) jumps, on a finite or countable set of points.

The r.v. has a non-null probability on jump points:

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

Absolutely continuous r.v.

When a cdf F, is the integral of another function f,

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

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

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

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. .

We use the same symbol f for both.

Intuitively, we are "replacing sums with integrals".

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

If F is a step function cdf (of a discrete r.v. X), its derivative is 0 except on X values, where F is discontinuous.

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

For $A \subset \mathbb{R}$
P(A) is the sum of P$\{X = x_i\}$, for $x_i \in A$.

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

Values of a pdf f are not probabilities, but $f \geq 0$.

Values of f can be arbitrarily large, on a sufficiently small interval, provided that $\displaystyle\int_{\mathbb{R}} f = F(+\infty) = 1$.

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

Computing probabilities with continuous r.v.

For X continuous, with pdf f and cdf F, and $a, b \in \bar{\mathbb{R}}$,

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

$-\infty \leq a \leq b \leq +\infty$.

In particular,

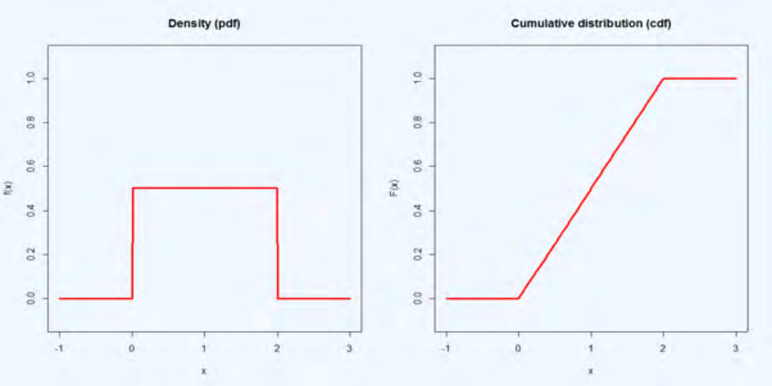
$$P(X = a) = 0, \quad \text{for } a \in \mathbb{R}.$$

Uniform (rectangular) distribution

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

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

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



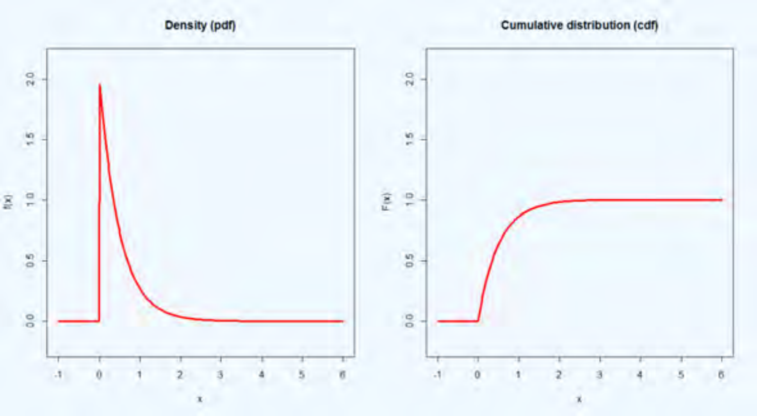
Exponential distribution

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

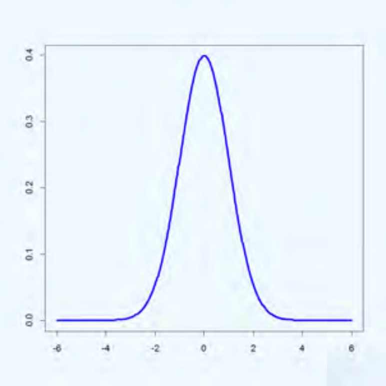
$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ \lambda \exp(-\lambda x), & \text{if } 0 \leq x. \end{cases}$$

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

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



Normal pdf *(Gaussian bell-shaped curve)*



Definition

A r.v. X is *normal or Gaussian*, 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: $X \sim \text{Normal}(\mu, \sigma^2)$.

Meaning of parameters

μ is the *mean (expectation)* of X and the symmetry axis.

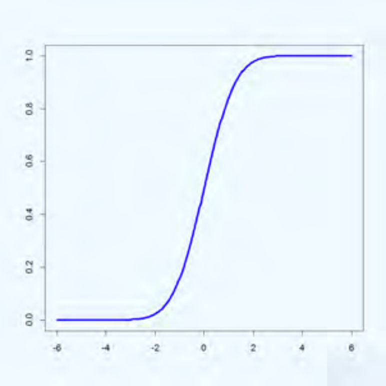
σ^2 is the *variance*, a measure of dispersion of X.

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

σ is a measure of the bell width.

Also, the measurement unit or scale of its x-axis.

cdf of a normal r.v.



The standard normal distribution

Is the Normal$(0, 1)$ distribution, with $\mu = 0$ and $\sigma^2 = 1$.

Its pdf is:

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

Standardizing a normal r.v.

f, the pdf of a Normal(μ, σ^2) r.v. , can be obtained by performing a translation and a scale change on ϕ, the pdf of a Normal$(0, 1)$ r.v.

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

Standardizing a normal r.v.

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

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

   is the *standardized* X.

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

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

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

02 - Random variables - 01

Expectation of a discrete r.v.

For a discrete r.v. X, with values $\boldsymbol{x} = (x_1, \ldots, x_m)$ and probabilities $\boldsymbol{d} = (d_1, \ldots, d_m)$,

The *expectation* or *mean* of X is the average of the x values, weighted by their respective probabilities:

$$\mathrm{E}(X) = \sum_{j=1}^{m} x_j \, d_j.$$

Properties of expectation

$\mathrm{E}(\,\cdot\,)$ is a linear operator:

$$\boxed{\begin{aligned} \mathrm{E}(X+Y) &= \mathrm{E}(X)+\mathrm{E}(Y), \\ \mathrm{E}(cX) &= c\,\mathrm{E}(X), \end{aligned}}$$

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

Expectation of a product of r.v.

If X and Y are independent r.v., then:

$$\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Non-independent r.v.'s do not satisfy this equality.

Example: expectation of a constant r.v.

If c is a constant r.v. , with value $c \in \mathbb{R}$, then

$$E(c) = c.$$

Expectation of an infinite discrete r.v.

When the set x is not finite, we substitute the sum of a series for the finite sum.

When the series $\sum x_n\, d_n$ is not (abs.) summable the r.v. has no expectation.

Expectation of a function of a discrete r.v.

X discrete r.v. , with values $\{x_j\}$ and probabilities $\{d_j\}$, g a function, *the expectation of* $g(X)$ is:

$$\mathrm{E}(g(X)) = \sum_j g(x_j)\, d_j,$$

if this sum exists, (i.e. is finite).

Variance of a discrete r.v.

X discrete r.v. , with values $\{x_j\}$, probabilities $\{d_j\}$, $\mu_X = \mathrm{E}(X)$, *the variance of X* is defined by:

$$\mathrm{var}(X) = \mathrm{E}((X - E(X))^2) = \sum_j (x_j - \mu_X)^2 \, d_j,$$

if this sum exists (i.e. is finite).

Expectation of X^2 and squared $\mathrm{E}(X)$

$\mathrm{E}(X^2)$ and $(\mathrm{E}(X))^2$ do not coincide. Always:

$$\mathrm{E}(X^2) \geq (\mathrm{E}(X))^2.$$

Equality if, and only if, X is constant. When both exist:

$$\boxed{\operatorname{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.}$$

Properties of the variance

$\mathrm{var}(\,\cdot\,)$ is a quadratic operator:

$$\mathrm{var}(c\,X) = c^2\ \mathrm{var}(X), \qquad \text{for } c \in \mathbb{R}.$$

When X and Y are <u>independent r.v.</u>, then

$$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

Expectation and variance of a Bernoulli r.v.

If $X \sim \mathrm{Ber}(p)$,

$$\mathrm{E}(X) = p,$$

$$\mathrm{var}(X) = p\,(1-p).$$

Expectation and variance of a binomial r.v.

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

$$\text{E}(X) = n\,p,$$
$$\text{var}(X) = n\,p\,(1 - p).$$

$X =$ number of A in n independent repetitions of a binary experiment, i.e., sum of n independent $\text{Ber}(p)$.

$E(X)$ and $\text{var}(X)$ are n times those of a Bernoulli r.v.

Expectation of a continuous r.v.

If X is a continuous r.v. with pdf f, we define:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx,$$

if this integral is abs. convergent.

There are continuous r.v. with no expectation.

Expectation of a function of a continuous r.v.

If X is a continuous r.v. with pdf f, and g is a function,

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx,$$

if this integral is abs. convergent.

Variance of a continuous r.v.

If X is a continuous r.v. with pdf f,

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx,$$

where $\mu = \text{E}(X)$.

Expectation and variance of a Unif(0, 1) r.v.

Pdf of $X \sim \text{Unif}(0, 1)$:

$$f(x) = \mathbb{1}_{(0,1)}(x), \qquad x \in \mathbb{R}$$

For $k \neq -1$,

$$\mathrm{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)\, dx = \int_0^1 x^k\, dx = \frac{1}{k+1}.$$

In particular, $\mathrm{E}(X) = \dfrac{1}{2}$, $\mathrm{var}(X) = \dfrac{1}{12}$.

Expectation and variance of a normal r.v.

If $X \sim \text{Normal}(\mu, \sigma^2)$,

- μ is the *expectation* of X, $\mathrm{E}(X) = \mu$.
- σ^2 is the *variance* of X, $\mathrm{var}(X) = \sigma^2$.

02 - Random variables - 01

Law of Large Numbers (LLN) for frequencies

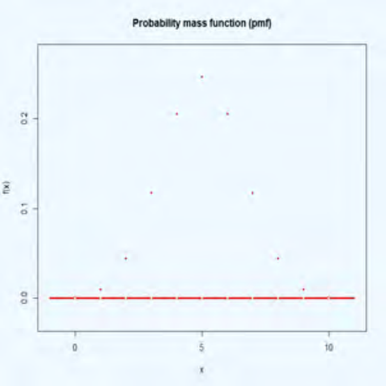
Dividing $X_n \sim \text{Binom}(n, p)$ by n, we obtain the *relative frequency* of an event in n independent repetitions:

$$f_n = \frac{X_n}{n}.$$

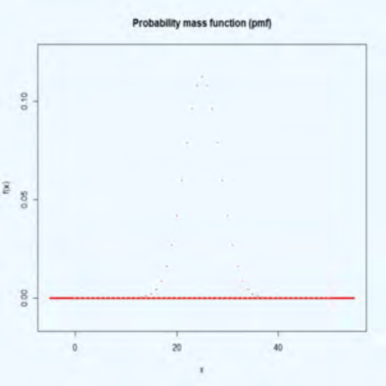
Its values are: $\{k/n : 0 \leq k \leq n\}$.

Values closer to p have higher probability.

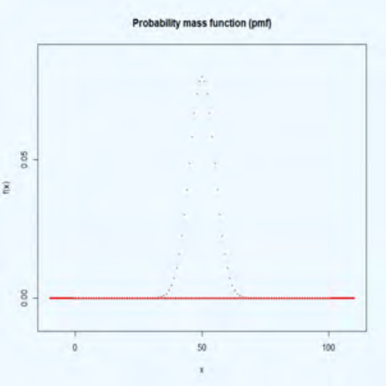
Pmf of f_n for $p = 0.5$ and $n = 10$



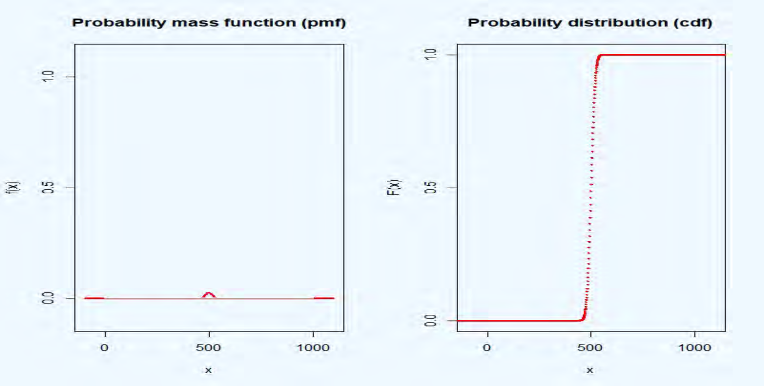
Pmf of f_n for $p = 0.5$ and $n = 50$



Pmf of f_n for $p = 0.5$ and $n = 100$



pmf and cdf of f_n for $p = 0.5$ and $n = 1000$



LLN for frequencies

When $n \to \infty$ values with a significant probability are few, around $p \cdot n$, and its proportion is close to p.

E.g., when $p = 0.5$ and $n = 1000$ only 81 values of f_n have $\mathrm{P} > 0.001$.

Sequence converges to a constant r.v. with value p.

$$\{f_n\} \xrightarrow{n \to \infty} p.$$

LLN for r.v. with a finite expectation

$\{X_n\}_{n\in\mathbb{N}}$ sequence of r.v. i.i.d.$\sim X$

Assume $\mu \equiv \mathrm{E}(X) < +\infty$.

Then the sequence $\{\bar{X}_n\}_{n\in\mathbb{N}}$ of arithmetic means, $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$, is convergent to the constant μ.

$$\{\ \bar{X}_n\ \}_{n\in\mathbb{N}} \xrightarrow[n\to\infty]{q.s.} \mu .$$

CLT context: Sum of independent normal r.v.

Normal r.v. are closed under *"Sum of independent r.v."*.

If $X_1 \sim \text{Normal}(\mu_1, \sigma_1^2)$, $X_2 \sim \text{Normal}(\mu_2, \sigma_2^2)$ are independent:

$$X_1 + X_2 \sim \text{Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

The sum of n r.v., i.i.d.$\sim$ $\text{Normal}(\mu, \sigma^2)$

is a normal r.v., with mean $n\,\mu$ and variance $n\,\sigma^2$.

Central Limit Theorem (CLT)

What happens when non-normal r.v. are added?

The sum of n independent r.v. is approximately normal, assuming some regularity conditions.

The approximation improves as $n \to \infty$.

In particular, if $X_n \sim \text{Binom}(n, p)$,

$$X_n \approx \text{Normal}(np, np(1-p)),$$

the larger is n the better is the approximation.

02 - Random variables - 01

Joint pmf for two discrete variables

The joint pmf of a pair of discrete r.v. (X, Y) is:

$$h(x, y) \equiv \mathrm{P}\left(\{X = x, Y = y\}\right), \quad (x, y) \in \mathbb{R}^2.$$

	$X = 1$	$X = 2$	$X = 3$	
$Y = 0$	0.90	0.03	0.02	0.95
$Y = 1$	0.01	0.02	0.02	0.05
	0.91	0.05	0.04	1

As a matrix, each (i, j)-th entry is $\mathrm{P}(X = j, Y = i)$.

Marginal univariate pmf's

f, *marginal pmf* of X, is obtained by adding the columns, giving:

$$\mathrm{P}(X = 1) = 0.91, \quad \mathrm{P}(X = 2) = 0.05, \quad \mathrm{P}(X = 3) = 0.04.$$

g, *marginal pmf* of Y, is obtained by adding the rows, giving:

$$\mathrm{P}(Y = 0) = 0.95, \qquad \mathrm{P}(Y = 1) = 0.05.$$

Independent random variables

X, Y are (stochastically) independent if, and only if,

$$h(x, y) = h_0(x, y) \equiv f(x) \cdot g(y), \quad (x, y) \in \mathbb{R}^2.$$

Joint pmf for two independent r.v.

h_0, the *independence joint pmf* with same marginals as h:

	$X = 1$	$X = 2$	$X = 3$	
$Y = 0$	0.8645	0.0475	0.038	0.95
$Y = 1$	0.0455	0.0025	0.002	0.05
	0.91	0.05	0.04	1

Conditional table, given Y

The *table of conditional probabilities (to Y)* is obtained dividing each row in h its total (marginal probability of that Y value).

	$X = 1$	$X = 2$	$X = 3$	
$Y = 0$	0.9474	0.0316	0.0211	1
$Y = 1$	0.2000	0.4000	0.4000	1

Now the (i, j)-th entry is $\mathrm{P}(X = j \mid Y = i)$.

This is the *the matrix of row profiles.*

Conditional table, given X

The *table of conditional probabilities (to X)*, *matrix of column profiles*, is obtained dividing each column by its total (marginal probability of that value of X).

	$X = 1$	$X = 2$	$X = 3$
$Y = 0$	0.9890	0.6000	0.5000
$Y = 1$	0.0110	0.4000	0.5000
	1	1	1

Now the (i, j)-th entry is $\mathrm{P}(Y = i \mid X = j)$.

Covariance of two discrete r.v. (X, Y)

Given h, joint pmf, we get $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{var}(X)$, $\mathrm{var}(Y)$, from the marginal pmf's f and g.

The *covariance* of (X, Y) is:

$$\mathrm{cov}(X, Y) \overset{\text{def}}{=\!=} \sum_x \sum_y h(x, y)\,(x - \mathrm{E}(X))\,(y - \mathrm{E}(Y)),$$

where x and y take all values of X and Y, respectively.

Covariance of two discrete r.v. (X, Y)

Alternatively, compute the expectation of the product:

$$\mathrm{E}(X \cdot Y) \overset{\text{def}}{=\!=} \sum_x \sum_y h(x, y)\, x\, y,$$

and then use the equality:

$$\operatorname{cov}(X, Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Covariance of two discrete r.v. (X, Y)

For the example:

$$E(X) = 1.13, \qquad E(Y) = 0.05,$$

$$\text{var}(X) = 0.1931, \qquad \text{var}(Y) = 0.0475,$$

and

$$\text{cov}(X, Y) = 0.0535.$$

Correlation coefficient of two discrete r.v. (X, Y)

$$\operatorname{cor}(X, Y) \stackrel{\text{def}}{=} \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X) \cdot \operatorname{var}(Y)}}.$$

Normalization of $\operatorname{cov}(X, Y)$, $\boxed{-1 \leq \operatorname{cor}(X, Y) \leq 1}$.

$\operatorname{cor}(X, Y)$ measures linear dependence between X, Y.

For the example, $\operatorname{cor}(X, Y) = 0.5586$.

Bayes formula for two discrete variables

Entries in the conditional table given X, result from Bayes' formula,

$$\mathrm{P}(Y = i \mid X = j) = \frac{\mathrm{P}(X = j \mid Y = i)\,\mathrm{P}(Y = i)}{\mathrm{P}(X = j)},$$

operating on the conditional table given Y and both marginals in h.

Denominator:

$$\mathrm{P}(X = j) = \sum_i \mathrm{P}(X = j \mid Y = i) \cdot \mathrm{P}(Y = i) = \sum_i \mathrm{P}(X = j,\ Y = i).$$

Interpreting Bayes formula

Each j-th column in the conditional table, given X,

is the *final*, *posterior* pmf, the transform of

the marginal *initial, a priori* pmf of Y,

after merging it with the observed evidence $\{X = j\}$.

Bivariate continuous r.v. - Joint pdf

A pair (X, Y) of r.v. with an absolutely continuous joint probability distribution has a joint pdf:

$$h(x, y), \quad \text{for } (x, y) \in \mathbb{R}^2.$$

Probability of a region $A \subset \mathbb{R}^2$,

$$\mathrm{P}(A) = \iint_A h(x, y)\, dx\, dy.$$

Marginal pdf's. Independent continuous r.v.

Marginal pdf of X:

$$f(x) = \int_{\mathbb{R}} h(x,y)\,dy, \quad x \in \mathbb{R},$$

marginal pdf of Y:

$$g(y) = \int_{\mathbb{R}} h(x,y)\,dx, \quad y \in \mathbb{R}.$$

(X, Y) are *independent:* $\Leftrightarrow$ $h(x,y) = f(x) \cdot g(y)$.

Covariance of two continuous r.v. (X, Y)

$$\operatorname{cov}(X, Y) = \iint_{\mathbb{R}^2} (x - \mathrm{E}(X)) \cdot (y - \mathrm{E}(Y)) \cdot h(x, y)\, dx\, dy.$$

As in the discrete case,

$$\mathrm{E}(X \cdot Y) = \iint_{\mathbb{R}^2} x \cdot y \cdot h(x, y)\, dx\, dy,$$

and

$$\operatorname{cov}(X, Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Conditional probability density function

The pdf of X, conditional to $\{Y = y\}$ is:

$$f(x \mid y) = \frac{h(x,y)}{g(y)}, \quad x \in \mathbb{R}.$$

The pdf of Y, conditional to $\{X = x\}$ is:

$$g(y \mid x) = \frac{h(x,y)}{f(x)}, \quad y \in \mathbb{R}.$$

Bayes' formula for pdf's

Combining both expressions we obtain:

$$
\begin{aligned}
g(y \mid x) &= \frac{h(x,y)}{f(x)} = \frac{f(x \mid y) \cdot g(y)}{f(x)} \\
&= \frac{f(x \mid y) \cdot g(y)}{\int_{\mathbb{R}} f(x \mid y) \cdot g(y)\, dy}
\end{aligned}
$$

Interpretation of Bayes' formula

The *final,* or *a posteriori* pdf $g(y \mid x)$

is the transform of

the *initial,* or *a priori* pdf $g(y)$,

as a result of merging it

with the observed evidence $\{X = x\}$.