Mathematical Methods for Automation Engineering M
– *Random Variables* –

Andrea Mentrelli

Department of Mathematics &
Alma Mater Research Center on Applied Mathematics, AM^2
University of Bologna

andrea.mentrelli@unibo.it

A.Y. 2024/25

Random Variables

Frequently, when an experiment is performed, we are interested mainly in some **function of the outcome** as opposed to the outcome itself

- in tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die
- in flipping a coin, one might be interested in the total number of heads that occur and not care about the heads/tails sequence

These quantities of interest or, more formally, these **real-valued functions** defined on the sample space Ω, are known as **random variables**

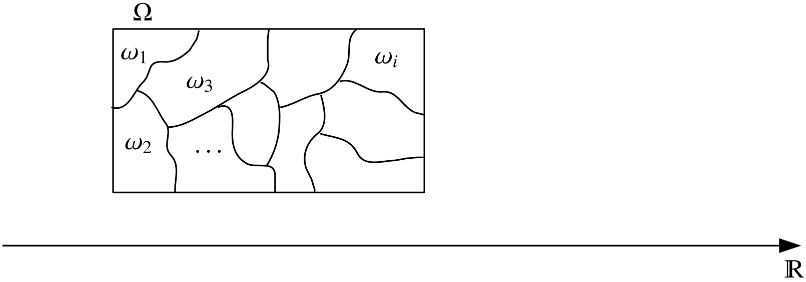
Because the value of a random variable is associated to an outcome of a random experiment (*elementary event*), which in turn can be associated to a probability value, **we may associate probabilities to the possible values of the random variable**

Definition (Random variable)

A *random variable* X is a measurable real-valued function defined on the sample space Ω

$$X : \Omega \to \mathbb{R}$$

In order to ensure that X is measurable, we assume that the set $E = \{\omega \in \Omega : X(\omega) \leqslant x_0\}$ is an event for any choice of x_0

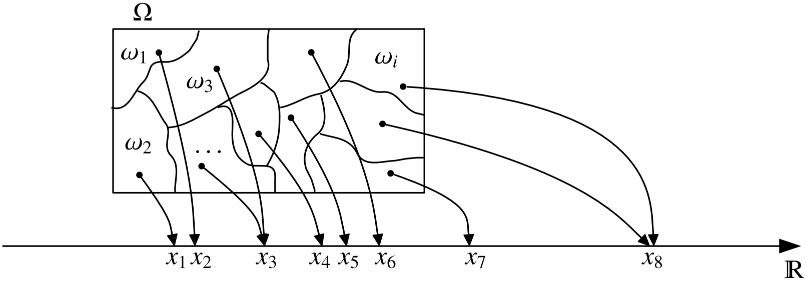


Definition (Random variable)

A *random variable* X is a measurable real-valued function defined on the sample space Ω

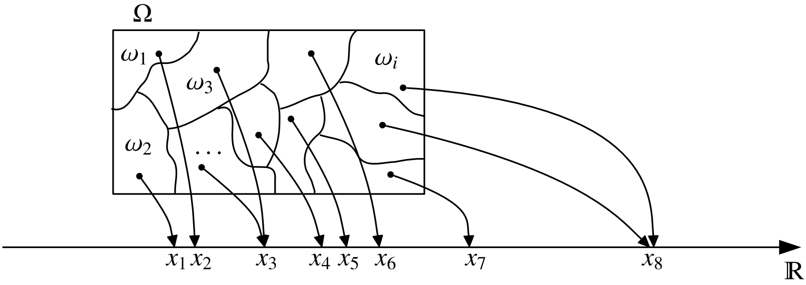
$$X : \Omega \to \mathbb{R}$$

In order to ensure that X is measurable, we assume that the set $E = \{\omega \in \Omega : X(\omega) \leqslant x_0\}$ is an event for any choice of x_0



Random Variables

- We are dealing with **one-dimensional** random variables
- A *random variable* is in fact. . . a *function*!
- We indicate the **image** (or, **range**) of X with the symbol Ω_X:
  $X : \Omega \to \Omega_X \subseteq \mathbb{R}$
- When Ω_X is countable, X is a **discrete random variable**; otherwise X is a **continuous random variable**



Example

In the experiment "tossing a die" the sample space is

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$$

It seems natural to introduce the **random variable** X defined as the function that for any possible outcome of the experiment, i.e. to each side of the die, returns the points of that side

$$X(\omega_i) = i, \qquad i = 1, 2, \ldots 6 \qquad \Rightarrow \qquad \begin{cases} X(\omega_1) = 1 \\ X(\omega_2) = 2 \\ \quad \ldots \\ X(\omega_6) = 6 \end{cases}$$

In this case, X is a bijective function

Example

In the experiment "tossing a die" the sample space is

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$$

In addition to the random variable X previously defined, it is possible to define also a **random variable** Y that for any elementary event of Ω returns, for instance, "double the points shown on the side minus 2"

$$Y(\omega_i) = 2i - 2, \qquad i = 1, 2, \ldots 6 \qquad \Rightarrow \qquad \begin{cases} Y(\omega_1) = 0 \\ Y(\omega_2) = 2 \\ \ldots \\ Y(\omega_6) = 10 \end{cases}$$

The random variable Y is bijective, too (but this property is not required)

Random Variables

Example

In the experiment "flipping 3 coins" the sample space is

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

We can introduce the following **random variable** X

$$X = \text{"number of heads (H) of the three flipped coins"}$$

The image of the random variable X is the following

$$\Omega_X = \{0, 1, 2, 3\}$$

ω_i	$X(\omega_i)$	ω_i	$X(\omega_i)$	ω_i	$X(\omega_i)$	ω_i	$X(\omega_i)$
TTT	0	THT	1	HHT	2	THH	2
HTT	1	TTH	1	HTH	2	HHH	3

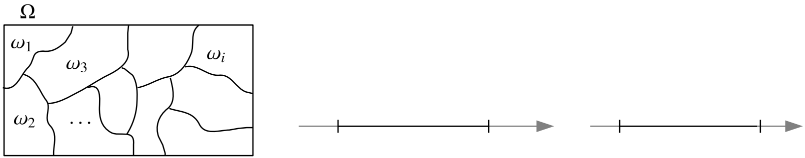
Definition (Probability mass function)

A *probability mass function* $p_X(x)$ is a function that takes a value returned by a **discrete** random variable X and returns a probability

$$p_X : \Omega_X \to [0,1] \qquad (\Omega_X \subseteq \mathbb{R})$$

Therefore, $p_X(x)$ represents the **probability** of the event E sent into x by the function (random variable) X, namely: $E = X^{-1}(x)$

$$p_X(x) \equiv \mathcal{P}(E) \qquad \text{where} \qquad X(E) = x$$



Random Variables

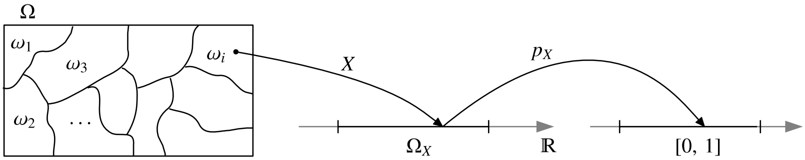
Definition (Probability mass function)

A *probability mass function* $p_X(x)$ is a function that takes a value returned by a **discrete** random variable X and returns a probability

$$p_X : \Omega_X \to [0, 1] \qquad (\Omega_X \subseteq \mathbb{R})$$

Therefore, $p_X(x)$ represents the **probability** of the event E sent into x by the function (random variable) X, namely: $E = X^{-1}(x)$

$$p_X(x) \equiv \mathcal{P}(E) \qquad \text{where} \qquad X(E) = x$$



Random Variables

Definition (Probability mass function)

A *probability mass function* $p_X(x)$ is a function that takes a value returned by a **discrete** random variable X and returns a probability

$$p_X : \Omega_X \to [0,1] \qquad (\Omega_X \subseteq \mathbb{R})$$

Therefore, $p_X(x)$ represents the **probability** of the event E sent into x by the function (random variable) X, namely: $E = X^{-1}(x)$

$$p_X(x) \equiv \mathcal{P}(E) \qquad \text{where} \quad X(E) = x$$

In the case of a **discrete** sample space Ω, over which a **discrete** random variable X is defined, the following properties must hold

- $p_X(x_i) \geqslant 0 \quad i = 1, 2, \ldots$
- $\sum_i p_X(x_i) = 1$

Random Variables

$X : \Omega \rightarrow \mathbb{R}$

Example

$\Omega_X \{0, 1, 2, 3\}$

In the experiment "flipping three coins", the random variable X defined as "number of resulting heads" is defined as follows

ω_i	$X(\omega_i)$	ω_i	$X(\omega_i)$	ω_i	$X(\omega_i)$	ω_i	$X(\omega_i)$
TTT	0	THT	1	HHT	2	THH	2
HTT	1	TTH	1	HTH	2	HHH	3

The associated probability mass function can be defined (assuming that the coins are fair)

$X : \Omega \rightarrow \Omega_X$

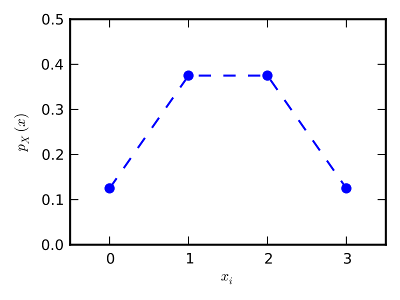
$P_X : \Omega_X = [0, 1]$

$$
\begin{aligned}
p_X (X = 0) &\equiv \mathcal{P}(\{TTT\}) = 1/8 \\
p_X (X = 1) &\equiv \mathcal{P}(\{HTT,\ THT,\ TTH\}) = 3/8 \\
p_X (X = 2) &\equiv \mathcal{P}(\{HHT,\ HTH,\ THH\}) = 3/8 \\
p_X (X = 3) &\equiv \mathcal{P}(\{HHH\}) = 1/8
\end{aligned}
$$

Random Variables

Example

ω_i	X	$p_X(X)$
TTT	0	1/8
HTT, THT, TTH	1	3/8
HHT, HTH, THH	2	3/8
HHH	3	1/8



$$p_X(X=0) \equiv \mathcal{P}(\{TTT\}) = 1/8$$
$$p_X(X=1) \equiv \mathcal{P}(\{HTT,\ THT,\ TTH\}) = 3/8$$
$$p_X(X=2) \equiv \mathcal{P}(\{HHT,\ HTH,\ THH\}) = 3/8$$
$$p_X(X=3) \equiv \mathcal{P}(\{HHH\}) = 1/8$$

Example

In the experiment "tossing two dice", let us define

$$U : \Omega \rightarrow \Omega_U \qquad U\left(\omega_i^1, \omega_j^2\right) = \underbrace{\max(i,j)}_{\text{random variable = max pairs}} \qquad i, j = 1, 2, \ldots 6$$

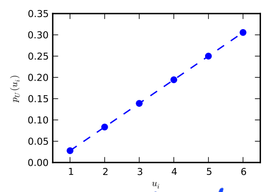
U → random variable

$P_U \geq 0$, $P_U \leq 1$, $\sum_{i=1} P_i = 1$

We have $\Omega_U = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$

$\left(\omega_i^1, \omega_j^2\right)$	u	$p_U\left(u_i\right)$
(1,1)	1	$1/36 \simeq 0.028$
(1,2), (2,1), (2,2)	2	$3/36 \simeq 0.083$
(1,3), (3,1), (2,3), (3,2), (3,3)	3	$5/36 \simeq 0.139$
(1,4), (4,1), (2,4), (4,4) . . .	4	$7/36 \simeq 0.194$
(1,5), (5,1), (2,5), (5,2) . . .	5	$9/36 = 0.25$
(1,6), (6,1), (2,6), (6,2) . . .	6	$11/36 \simeq 0.306$

Ω_U $\qquad$ + $\qquad$ 1



not continuous!

$P_i \equiv P_S(s_i)$

Example

In the experiment "tossing two dice", let us define

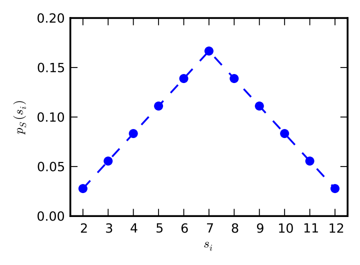
$$S : \Omega \to \Omega_S \qquad S\left(\omega_i^1, \omega_j^2\right) = i + j \qquad i, j = 1, 2, \dots 6$$

sum of two dice

$$\sum_{i=1}^{n} P_i = 1$$

We have $\Omega_S = \{2,\ 3, \dots 12\}$ from 2 to 12

s	$p_S(s)$	s	$p_S(s)$
2	$1/36 \simeq 0.028$	8	$5/36 \simeq 0.139$
3	$2/36 \simeq 0.056$	9	$4/36 \simeq 0.111$
4	$3/36 \simeq 0.083$	10	$3/36 \simeq 0.083$
5	$4/36 \simeq 0.111$	11	$2/36 \simeq 0.056$
6	$5/36 \simeq 0.139$	12	$1/36 \simeq 0.028$
7	$6/36 \simeq 0.167$		



Random Variables

Definition (Cumulative distribution function (CDF))

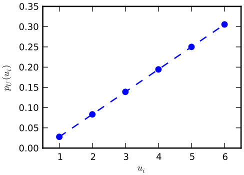
Given a random variable X, the *cumulative distribution function*, $\mathcal{F}_X$, is defined as follows

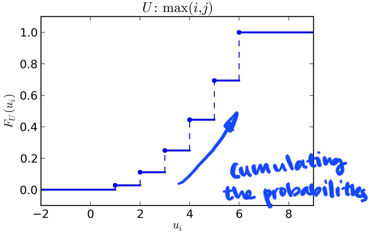
$$\mathcal{F}_X(x) = \mathcal{P}(X \leqslant x) \qquad \underbrace{(-\infty < x < +\infty)}_{\text{real values}}$$

$F(x): \mathbb{R} \to [0,1]$

$$U\left(\omega_i^1, \omega_j^2\right) = \max(i,j) \qquad i,j = 1,2,\dots 6$$

u	$p_U(u)$
1	$1/36 \simeq 0.028$
2	$3/36 \simeq 0.083$
3	$5/36 \simeq 0.139$
4	$7/36 \simeq 0.194$
5	$9/36 = 0.25$
6	$11/36 \simeq 0.306$





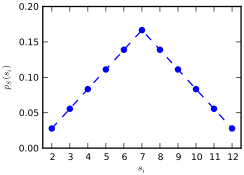
Definition (Cumulative distribution function)

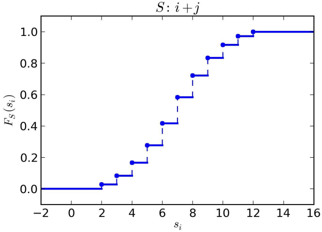
Given a random variable X, the *cumulative distribution function*, $\mathcal{F}_X$, is defined as follows

$$\mathcal{F}_X(x) = \mathcal{P}(X \leqslant x) \qquad x \in \mathbb{R}$$

$$S\left(\omega_i^1, \omega_j^2\right) = i + j \qquad \forall i, j = 1, 2, \dots 6$$

s	$p_S(s)$
2	$1/36 \simeq 0.028$
3	$2/36 \simeq 0.056$
4	$3/36 \simeq 0.083$
5	$4/36 \simeq 0.111$
6	$5/36 \simeq 0.139$
7	$6/36 \simeq 0.167$
	......





Random Variables

From discrete to continuous random variables

- **Discrete random variables** have a **discontinuous** (step) CDF
- **Continuous random variables** are those with a **continuous** CDF

Remarkable properties of CDFs

In general, for both discrete/continuous random variables

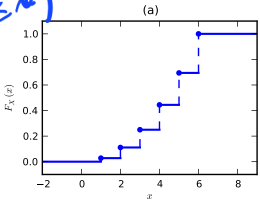
CDF

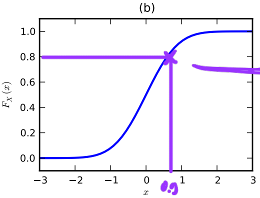
- $0 \leqslant \mathcal{F}_X(x) \leqslant 1 \quad \forall x \in \mathbb{R} \quad (\mathcal{F}_X(x)$ is a probability!)
- $\mathcal{F}_X(x) = \mathcal{P}(X \leqslant x) \quad \Rightarrow \quad \mathcal{P}(X \geq x) = 1 - \mathcal{F}_X(x)$
- $\boxed{\mathcal{P}(a < X \leqslant b) = \mathcal{F}_X(b) - \mathcal{F}_X(a)}$ probability $\quad \forall a \leqslant b$

Valid for discrete and cont.

(From the III axiom: $A =]-\infty, a]$, $B =]a, b]$, $A \cup B =]-\infty, b]$)

$F_X(x) = P(X \leq x)$





$P(X \leq 0.9) \simeq 0.8$

Random Variables

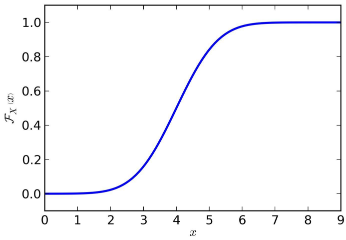
Definition (Probability density function)

Given a continuous random variable X and its cumulative distribution function $\mathcal{F}_X(x)$, we call *probability density function* the function defined as follows

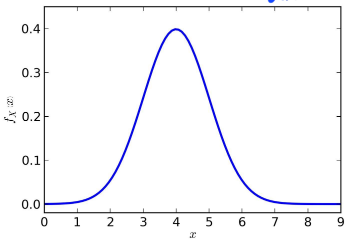
$$f_X(x) = \frac{d}{dx}\mathcal{F}_X(x)$$

derivative

$\mathcal{F}_X(x)$



$f_X(x) = \frac{d}{dx}\mathcal{F}_X(x)$



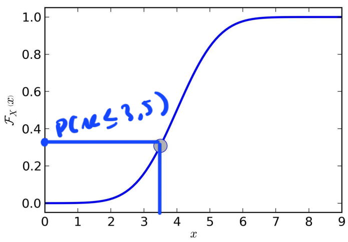
Random Variables

Some properties of probability density functions $f_X(x)$

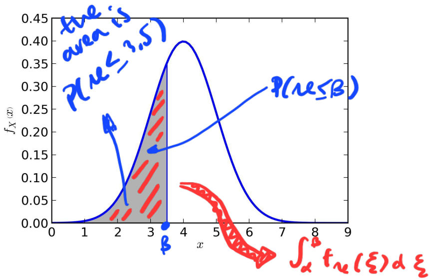
- The probability that a random variable X takes on values $X \leqslant \beta$ corresponds to the area between the x-axis and the curve $f_X(x)$ over the interval $[-\infty, \beta]$

$$\mathcal{P}(X \leqslant \beta) = \int_{-\infty}^{\beta} f_X(\xi)\, d\xi = \mathcal{F}_X(\beta) - \mathcal{F}_x(-\infty)$$

$\mathcal{F}_X(x)$



$f_X(x)$

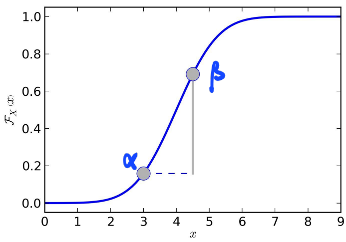


Some properties of probability density functions $f_X(x)$

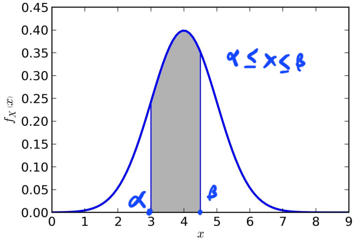
- The probability that a random variable X takes on values $\alpha \leqslant X \leqslant \beta$ corresponds to the area between the x-axis and the curve $f_X(x)$ over the interval $[\alpha, \beta]$

$$\mathcal{P}(\alpha \leqslant X \leqslant \beta) = \int_{\alpha}^{\beta} \underbrace{f_X(\xi)\, d\xi}_{\text{accumulation function}} = \mathcal{F}_X(\beta) - \mathcal{F}_X(\alpha)$$

$\mathcal{F}_X(x)$



$f_X(x)$



Random Variables

For a **continuous** random variable X

0 doesn't indicate impossibility

$$\boxed{\mathcal{P}(X = a)} = \int_a^a f_X(\xi)\, d\xi = 0 \qquad \forall a \in \mathbb{R}$$

However, that does not mean that the event $X = a$ cannot occur. Rather, its probability is *infinitesimal* ($\rightarrow$ mass of a material point)

in discrete case, maybe we could say $P(x) = 0$ is impossible. But in cont. case it does not!

A somewhat *more intuitive interpretation*

$$\mathcal{P}\left(a - \frac{\delta x}{2} \leqslant X \leqslant a + \frac{\delta x}{2}\right) = \int_{a-\frac{\delta x}{2}}^{a+\frac{\delta x}{2}} f_X(x)\, dx \approx f_X(a)\, \delta x$$

when δx is *small* and when $f_X(x)$ is continuous at $x = a$, the probability that X will be contained in an interval of length δx around the point a is approximately $f_X(a)\, \delta x$

As a consequence, for continuous random variables:

the maxima and minima do not matter

$$\boxed{\mathcal{P}(a < X \leqslant b) = \mathcal{P}(a \leqslant X \leqslant b) = \mathcal{P}(a \leqslant X < b) = \mathcal{P}(a < X < b)}$$

Example

Given the function $\mathcal{F}_X(x)$ defined in the following, verify that it is a cumulative distribution function and determine the corresponding probability density function $f_X(x)$

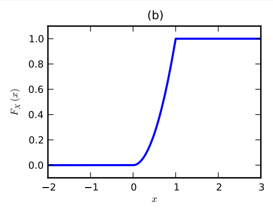
$$\mathcal{F}_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leqslant x < 1 \\ 1 & x \geqslant 1 \end{cases} \qquad f_X(x) = \begin{cases} 0 & x < 0 \\ 2x & 0 \leqslant x < 1 \\ 0 & x \geqslant 1 \end{cases}$$

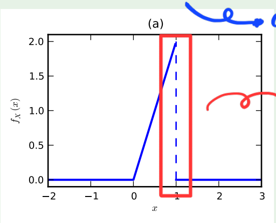
$\frac{d\mathcal{F}_x(x)}{dx} = f_x(x)$

constant, which is 0

constant = 0

visually, it's a CDF because it's increasing from left to right





discontinued

CDF does not require continuity

$\mathcal{F}_X(x)$ is right-continuous and it is non-decreasing. Moreover, $\mathcal{F}_X(-\infty) = 0$ and $\mathcal{F}_X(\infty) = 1$, therefore it is a legitimate CDF

Central tendency and dispersion parameters

A random variable X is completely characterized by either

- its cumulative distribution function, $\mathcal{F}_X(x)$
- its probability mass function (discrete case) or density function (continuous case)

It is useful to introduce scalar quantities that describe the behavior/main properties of random variables

- central tendency parameters – **expected value (mean value)**, **mode**, **median**, quartiles, deciles, percentiles
- dispersion parameters – initial moments, central moments, **variance**, **standard deviation**

Definition (Expected value, Mean value, Expectation)

Given a random variable X, we call *expected value* (or *mean value*, or *expectation*), denoted by $E(X)$ or μ_X (or simply μ, when it is clear from the context the reference to X), the quantity

- for a discrete random variable X

weighted sum / average

$$\boxed{\mu_X \equiv E(X) = \sum_{i=1}^{n} x_i \, p_X(x_i)}$$

- for a continuous random variable X $\int_{\mathbb{R}} \underbrace{x f_X(x) dx}_{p_X(x_i)}$

$$\boxed{\mu_X \equiv E(X) = \int_{-\infty}^{+\infty} x \, f_X(x) \, dx}$$

Say: $f_X(x) = 0, \ x < 0$

Example

Given the experiment "tossing a die" and the random variable X defined as "points shown by the upside face of the die", determine the expected value of the random variable X

mutually equal /same

$$X(\omega_i) = i \qquad p_X(x_i) = 1/6 \qquad i = 1, 2, \ldots, 6$$

therefore

the expected value is not a possible outcome of the exp.

$$E(X) = \sum_{i=1}^{6} x_i \, p_X(x_i)$$
$$= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

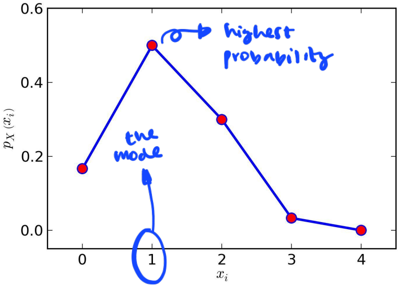
Random Variables

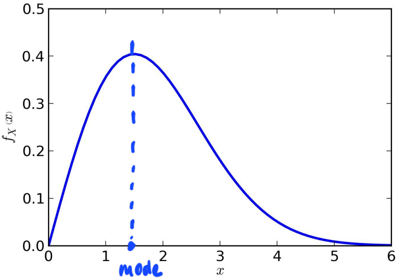
Definition (Mode)

Given a random variable X, we call *mode*, denoted by the symbol $\mathcal{M}_X$, the value(s) for which the probability mass/density function has a maximum (i.e. the points of maximum)

- if $\boxed{X \in C^2(\Omega_X): \; f'(\mathcal{M}_X) = 0, \quad f''(\mathcal{M}_X) < 0}$
- If there is **only one** maximum: **unimodal** distribution

→ one mode



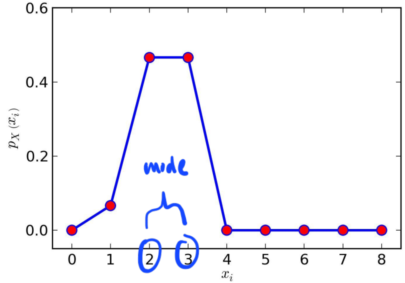


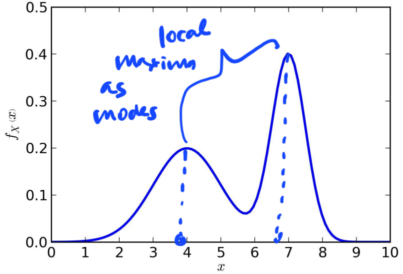
Definition (Mode)

Given a random variable X, we call *mode*, denoted by the symbol $\mathcal{M}_X$, the value(s) for which the probability mass/density function has a maximum (i.e. the points of maximum)

- if $X \in C^2(\Omega_X)$: $f'(\mathcal{M}_X) = 0, \quad f''(\mathcal{M}_X) < 0$
- If there is **more than one** maximum: **multimodal** distribution

→ more than one





Definition (Median)

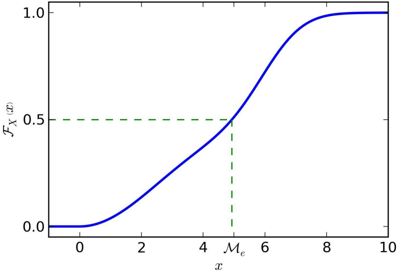
Given a random variable X, we call *median*, denoted by the symbol $\boxed{\mathcal{M}_e}$, the value such that

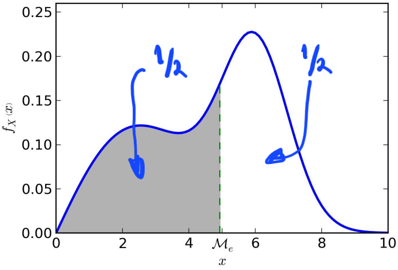
$= 1/2$

$$\boxed{\mathcal{P}\left(X < \mathcal{M}_e\right) = \mathcal{P}\left(X > \mathcal{M}_e\right)}$$

a half of minimum and maximum val

The median is the value $\mathcal{M}_e$ for which it is equally probable for the random variable X to take vales smaller or larger than $\mathcal{M}_e$



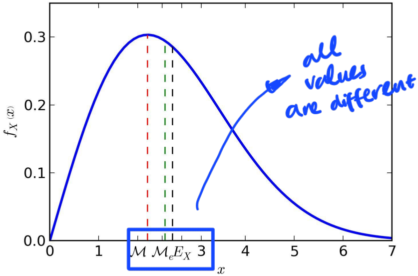


Remark. The **expected value**, the **mode** and the **median** are values **independent** one from the other. Therefore, they are in general three different values.

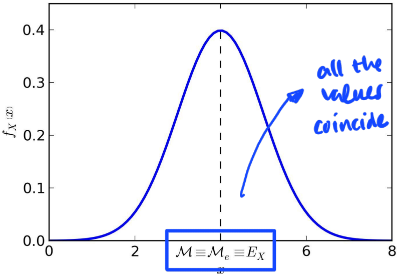
However, there are cases in which they are coincident (for example, for unimodal and symmetric distribution, like the *Gaussian distribution*)

symetric → must be unimodal

Rayleight distribution



Gaussian distribution



Definition (Initial Moments)

Given a discrete/continuous random variable X, we call *initial moment of order* s, α_s, the quantity defined as follows

param, integer

$$\alpha_s(X) = \sum_{i=1}^{n} x_i^s p_X(x_i) \qquad \alpha_s(X) = \int_{-\infty}^{+\infty} x^s f_X(x)\, dx$$

weight / respected prob.

It is easily seen that

$$\boxed{\alpha_s(X) \equiv E(X^s)}$$

it's actually non trivial

and

$$\boxed{\alpha_1(X) = E(X)}$$

x	x^s	P
x_i	x_i^s	$P(x_i)$

i.e. the initial moment of order 1 coincides with the expected value

Definition (Centered random variable)

Given a random variable X with expected value μ_X, the *centered random variable*, X_c, associated to X is the random variable defined as follows

$$\boxed{X_c = X - \mu_X}$$

a new random variable

X	X_c	P
x_i	$x_i - \mu_X$	$P(x_i)$

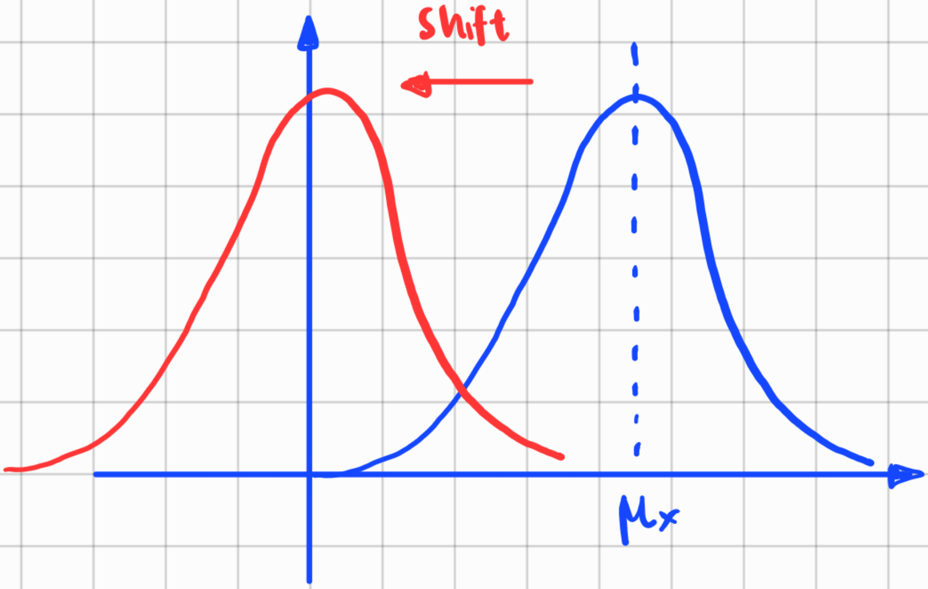
"Centering" a random variable X amounts to translating the origin up to the point coincident with the expected value μ_X

Theorem

The expected value of a centered random variable X_c is always null

$$E(X_c) = 0$$

Centering random variable



Random Variables

Definition (Central moment)

Centering: $X_c = X - \mu_X$

Given a discrete/continuous random variable X, we call *central moment* (or *moment with respect to the expected value*) *of order s*, denoted by the symbol μ_s, the quantity defined as follows

central moment

$$\mu_s(X) = \sum_{i=1}^{n} \underline{(x_i - \mu_X)^s}\; p_X(x_i) \qquad \mu_s(X) = \int_{-\infty}^{+\infty} \underline{(x - \mu_X)^s}\, f_X(x)\, dx$$

It is easily seen that the central moment of order s of a random variable X coincides with the mean value of the s^{th} power of the centered random variable $X_c = X - \mu_X$, namely

$$\mu_s(X) \equiv E(X_c^s) = E((X - \mu_X)^s)$$

Among all the possible moments, the most important (meaningful) ones are α_1 (**expected value**) and μ_2 (**variance**)

CENTRAL MOMENT ORDER:

Order 1 $\mu_1(X) = \sum_{i=1}^{n} (x_i - \mu_x)^1 P(x_i)$

$= \sum_{i=1}^{n} \underbrace{x_i - P(x_i)}_{1} - \underbrace{\sum_{i=1}^{n} \mu_x P(x_i)}_{\mu_x \underbrace{\sum_{i=1}^{n} P(x_i)}_{1}} = 0$

always positive

σ^2 variance

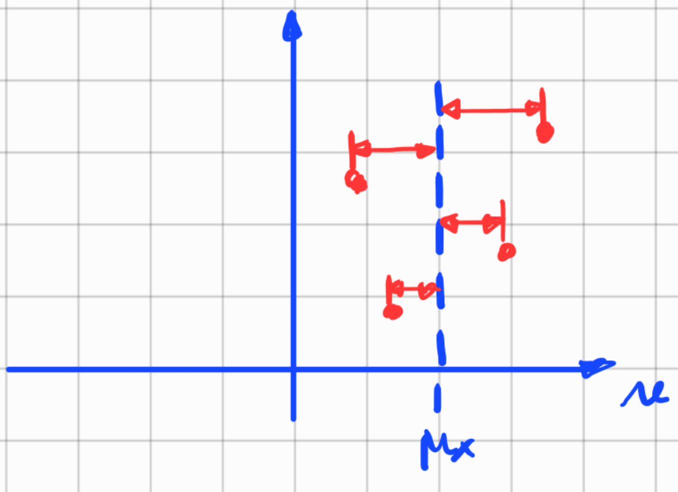
Order 2 $\mu_2(X) = \sum_{i=1}^{n} \underbrace{(x_i - \mu_x)^2}_{\text{weighted sum of distance between rand. value and mean value}} P(x_i) = \sigma^2$

$(x_i - \mu_x)$

if the deviation to the mean val. larger, the variance is higher

$\sigma^2 = \sum (x_i - \mu_x)^2 P(x_i)$

$[x_i] = L$



Random Variables

Definition (Variance & Standard deviation)

Given a random variable X, we call *variance* of X, denoted by the symbol $Var(X)$ or σ_X^2, the central moment of order 2

$$\sigma_X^2 = Var(X) = E\left(X_c^2\right) = \boxed{E\left((X - \mu_X)^2\right)}$$

For discrete/continuous random variables, the variance is

$$\sigma_X^2 = \sum_{i=1}^{n} (x_i - \mu_X)^2 \, p_X(x_i) \qquad \sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \, f_X(x) \, dx$$

The *standard deviation*, σ_X, is defined as the square root of the variance

$$\boxed{\sigma_X = \sqrt{\sigma_X^2} = \sqrt{Var(X)}}$$

The variance of a random variable is a numerical value describing the **spread** of the values taken on by X around the expected value of X. The larger the spread, the larger the variance

Random Variables

Theorem

Given a random variable X, the following property holds equals to mean value rand. v. squared (X^2) minus the squared mean of X

$$\boxed{Var(X) = E\left(X^2\right) - E(X)^2}$$

Proof. Making use of the definitions

$$Var(X) = E\left(X_c^2\right) = E\left(\left(X - \mu_X\right)^2\right) = \sum_{i=1}^{n} \overbrace{\left(x_i - \mu_X\right)^2}^{x_i^2 - 2\mu_x x_1 - \mu_x^2} p_X(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 p_X(x_i) - 2\mu_X \sum_{i=1}^{n} x_i\, p_X(x_i) + \mu_X^2 \sum_{i=1}^{n} p_X(x_i) = E\left(X^2\right) - E(X)^2$$

where it has been taken into account that

$$\sum_{i=1}^{n} x_i\, p_X(x_i) = E(X) \qquad \sum_{i=1}^{n} p_X(x_i) = 1 \qquad \sum_{i=1}^{n} x_i^2 p_X(x_i) = E\left(X^2\right)$$

□

Example

Determine σ_X^2, being X the natural random variable associated to a die roll

The expected value, previously calculate, is $\boxed{E(X) = \frac{7}{2}.}$ Moreover

$$E\left(X^2\right) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) = \frac{91}{6}$$

therefore

calculate this first

$$\sigma_X^2 = E\left(X^2\right) - E(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \simeq 2.917$$

$\sigma_X = \sqrt{\sigma_X^2}$

std. deviation

not a deterministic variable

Definition (Standardized random variable)

Given a non-degenerate random variable X, with expected value μ_X and standard deviation σ_X, we call *standardized random variable*, denoted by the symbol Z, the random variable defined as follows

$\sigma_X > 0$

$\sigma_X^2 = \sum (x_i - \mu_X)^2 P(x_i) > 0$

$$\boxed{Z = \frac{X - \mu_X}{\sigma_X}} = \frac{X_c}{\sigma_X} = a X_c \iff a = \frac{1}{X_c}$$

Remark. A standardized random variable Z is characterized by

- mean $\mu_Z = 0$
- variance $\sigma_Z^2 = \sigma_Z = 1$

stretch/shrink the var until $= 1$

$E(X) = 0 \Rightarrow E(aX) = 0$

$\sum x_i P(x_i) = 0$

$\sum a x_i P(x_i) = a \sum x_i P(x_i) = 0$

$\begin{cases} E(aX) = aE(X) \\ E(X+Y) = E(X) + E(Y) \end{cases}$

Proofing the variance :

$$Var\left(\frac{1}{\sigma_x} x_c\right) = Var\left(\frac{x - \mu_x}{\sigma_x}\right)$$

$$= \sum \left(\frac{x_i - \mu_x}{\sigma_x}\right)^2 P(x_i) = 1 \;//$$

check it by extending the simplification

Random Variables

Definition (Moment coefficient of skewness)

Given a random variable X, we call *moment coefficient of skewness*, or *Pearson's moment coefficient of skewness*, denoted γ_1, the following quantity

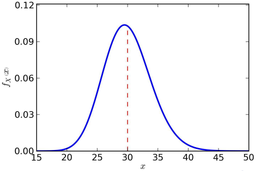
$$\boxed{\gamma_1 \equiv \mathrm{Asym}(X) = \frac{\mu_3}{\sigma_X^3}}$$

third order of central moment

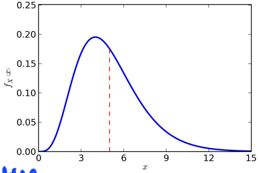
- $\gamma_1 > 0$ ($\gamma_1 < 0$): **positive (negative) skew**: the right (left) tail is longer than the left (right) one

the larger the value, the curve skews to left

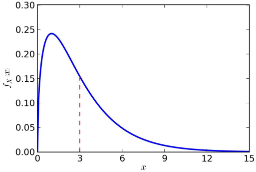
$\gamma_1 = 0.26$



$\gamma_1 = 0.89$



$\gamma_1 = 1.63$



approaching zero, the distribution is more symetric

Theorem (Čebičëv inequality)

Given a $\boxed{\text{random variable } X}$ with $\boxed{\text{expected value } \mu}$ and $\boxed{\text{(finite) variance } \sigma^2}$, the following inequality holds

farther than mean val.

$$\mathcal{P}\left(|X-\mu| \geqslant k\sigma\right) \leqslant \frac{1}{k^2} \qquad \forall k > 0$$

$$\Downarrow$$

or, equivalently,

$$\mathcal{P}\left(|X-\mu| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{\varepsilon^2} \qquad \forall \varepsilon > 0$$

Thus, for any random variable X, the probability of a deviation from the mean value μ of more than k standard deviations σ is not larger than $1/k^2$

Theorem (Čebičëv inequality)

Proof. *Let's consider the case of a discrete random variable X*

$$\sigma^2 = \underbrace{\sum_i (x_i - \mu)^2 \, p_X(x_i)}_{\text{variance formula}} = \sum_{|x_i-\mu|<k\sigma} (x_i - \mu)^2 \, p_X(x_i) + \sum_{|x_i-\mu|\geqslant k\sigma} (x_i - \mu)^2 \, p_X(x_i)$$

Since all the terms in the sum are non-negative

$$\sigma^2 \geqslant \sum_{|x_i-\mu|\geqslant k\sigma} (x_i - \mu)^2 \, p_X(x_i) \geqslant \underbrace{k^2\sigma^2}_{(k\sigma)^2} \underbrace{\left(\sum_{|x_i-\mu|\geqslant k\sigma} p_X(x_i) \right)}_{P(|X-\mu| \geq k\sigma)}$$

but

$$\sum_{|x_i-\mu|\geqslant k\sigma} p_X(x_i) = \mathcal{P}(|X-\mu| \geqslant k\sigma) \quad \Rightarrow \quad \underbrace{\boxed{\sigma^2 \geqslant k^2\sigma^2 \, \mathcal{P}(|X-\mu| \geqslant k\sigma)}}_{\text{Čebičëv inequality}}$$

□