Data Science Fundamentals

Part I: Probability theory

ISEP 2nd year
2023-2024

Based on the course given by Nathalie Colin & Jean-Claude Guillerot

Probability theory

- **3rd session (October 13th 2023):**

Chapter 4 : TYPICAL VALUES OF A REAL-VALUED RANDOM VARIABLE.

Chapter 5: CHARACTERISTIC FUNCTION

Probability theory

Chapter 4 : TYPICAL VALUES OF A REAL RANDOM VARIABLE.

- 4.1 Expected value (mean)
- 4.2 The Median (the middle value)
- 4.3 The mode (the most frequent value)
- 4.4 Percentiles
- 4.5 Moments of a random variables

TYPICAL VALUES OF A REAL RANDOM VARIABLE

Introduction :

A random variable is entirely defined by either :

- Its cumulative distribution function (CDF), $F_X(x)$ OR
- Its probability density function, $f_X(x)$

In practice, these functions depend on some parameters which are usually unknown. However, we can and need to describe a random variable by certain parameters which are easy to measure:

- Expected value (also known as mean or average)
- Median (middle value)
- Mode: most probable value
- Percentiles
- Variance

Expected value

Definition (for a continuous random variable):

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) \cdot dx$$

$E(X)$: expected value of the random variable X

$f_X(x)$: probability density of the random variable X

Definition (for a discrete random variable) :

$$E(X) = \sum_i P_i \cdot x_i$$

P_i : probability of obtaining value x_i

Remarks:

- If the probability density function is symmetric around a point, $x = a$ with $a \in \mathbb{R}$, the expected value is equal to this value:

$$E(X) = a \text{ if } f_X(x) = f_X(2a - x) \; \forall \, x \in \mathbb{R}$$

Example: the normal distribution N(m,σ^2) is symmetric around the point $x = m$

- For a continuous random variable, the expected value is undefined if the integral is not finite.

  Example: The Cauchy distribution's expected value is undefined

- The expected value and the random variable have the same unit of measurement/dimension.

  Example: if X represents time in seconds, E(X) is time in [s] *as well.*

- If the random variable is discrete X, the expected value can be equal to a value that does not belong to the domain of X.

  Example: X:number of tails obtained after tossing a coin, E(X)=0.5

Interpretation of the expected value based on the relative frequency

Consider an experiment where there are N possible outcomes $\{\omega_i\}_{i=1,\dots,N}$ to which N numerical values correspond $i = 1, \dots, N$ by the application $X(\omega_i)$

Outcome	**frequency**	**As a result…**
ω_1	k_1 times	k_1 times, $x_1=\varphi(\omega_1)$
ω_2	k_2 times	k_2 times, $x_2=\varphi(\omega_2)$
…	…	…
ω_N	k_N times	k_N times, $x_N=\varphi(\omega_N)$
possible outcomes	trials	

By taking the weighted average:

$$\bar{x} = \frac{k_1 x_1 + \cdots + k_N x_N}{n} = \sum_i^N \frac{k_i}{n} x_i$$

$\frac{k_i}{n}$ relative frequency of the value x_i, this can be assimilated to the probability of obtaining $X = x_i$ as n becomes huge and approaches infinity.

$$P_i = P(X = x_i) \approx \frac{k_i}{n} \qquad \text{therefore} \qquad \bar{x} \approx \sum_i^N P_i x_i = E(X)$$

Expected value of a deterministic function of a random variable

Let the random variable Y be a transformation of the random variable X given by $Y=g(X)$.

We want to calculate $\mathrm{E}(Y)$

There are two possibilities:

1. First calculate $f_Y(y)$ and deduce E(Y) (next session)

2. Simpler, directly from g(x), for instance, for a continuous random variable:

$$\boxed{E(Y) = E\big(g(x)\big) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx}$$

The median x_e ("the middle" value) (1/2)

The median x_e is the value that separates the higher half from the lower half of a probability distribution. Formally, the median x_e is any value that satisfies:

$$P(X \leq x_e) = P(X > x_e) = 0.5 = F_X(x_e)$$

Warning : Do not confuse with the expected value!

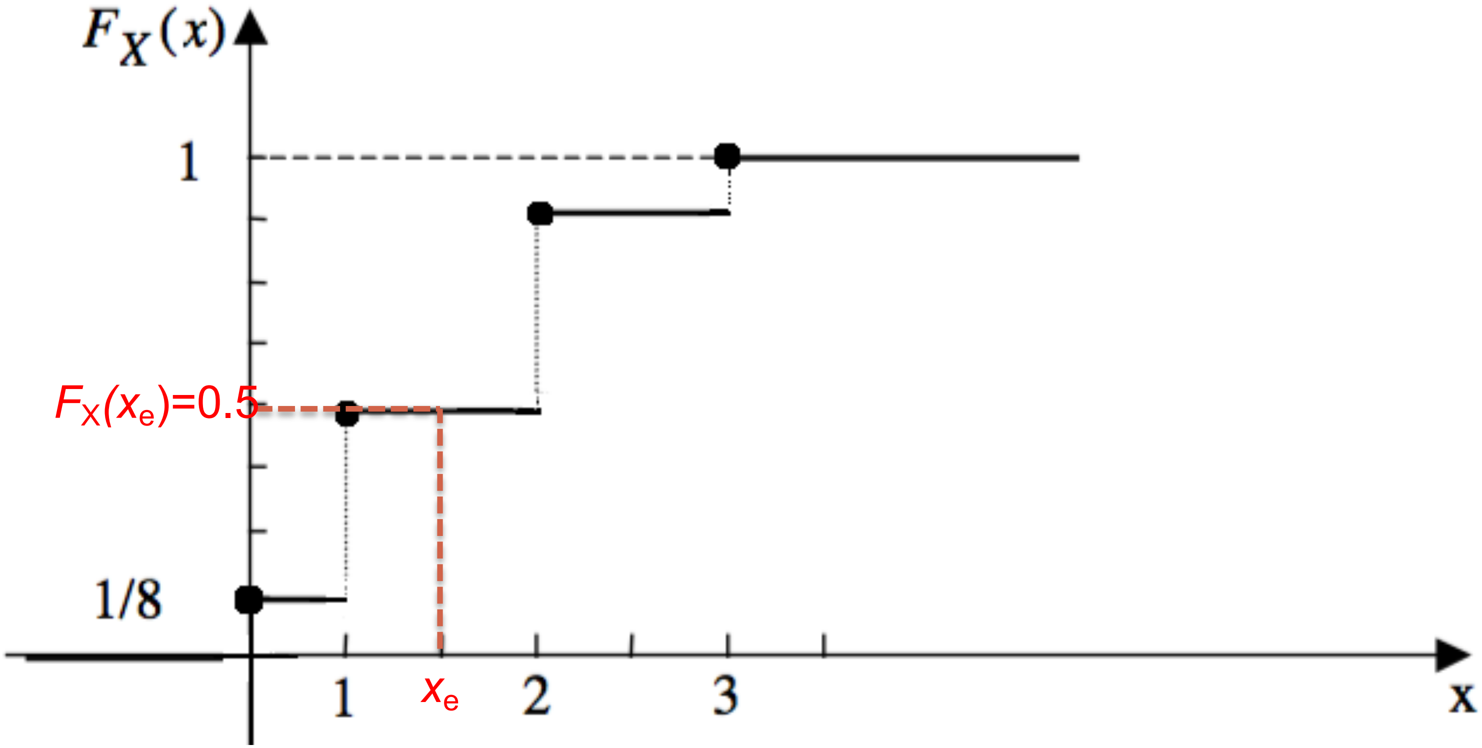
Remark: *if the probability density function is symmetric* $E(X) = x_e$

An interesting property of the median compared to the expected value (mean) is that it is not *skewed* by a small proportion of extremely large or small values, and therefore provides a better representation of a "typical" value to characterize the distribution.

The median x_e ("the middle" value) (2/2)

Example of the median x_e for a discrete random variable. x_e can be retrieved using the CDF diagram:

X: number of *tails* after tossing a coin 3 times



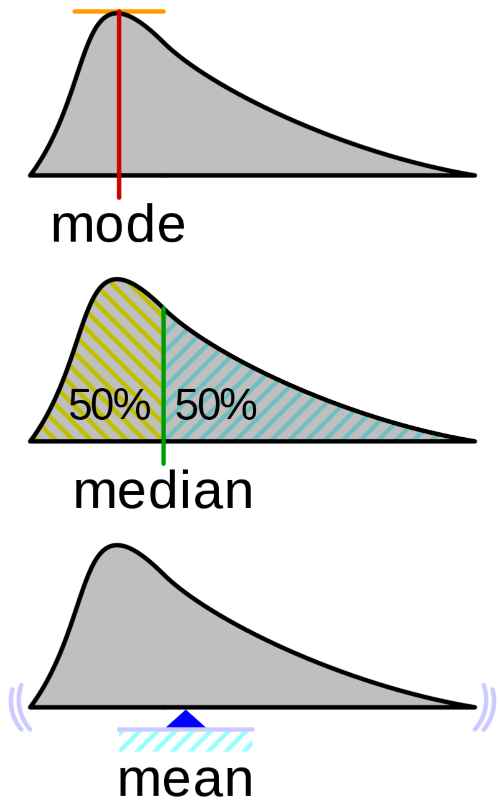
For this example, x_e can be any value in the interval $1 \leq x_e < 2$

By convention x_e=1.5

The mode: (the Most frequent value)

The mode is the value x_{mode} at which the probability mass function takes its maximum value.

- For the continuous cas, it is the abscissa of the maximum value in the density curve.
- For the discrete case, it is the value x_i which has the greatest probability.



Comparison of the mode, median and mean of an arbitrary probability density function.

Source: wikipedia

P^{th} percentiles or P% percentiles

The percentile P%, also called centile, is the value x_p for which the probability of not being exceeded is P/100 :

$$F(x_p) = \int_{-\infty}^{x_p} f_X(x) \cdot dx = \frac{P}{100}$$

Quartiles : percentiles 25 % , 50% and 75 %

Moments of a random variable : Definition

- Moments : generalization of the expected value

The moment of order *n* of a random variable is the expected value of the function $g(X) = X^n$ and denoted m_n

For a continuous random variable

$$\boxed{m_n = \int_{-\infty}^{+\infty} x^n \cdot f_X(x) \cdot dx}$$

For a discrete random variable

$$\boxed{m_n = \sum_i x_i^n . P(X = x_i)}$$

In particular : $m_0 = 1$ and $m_1 = E(X)$

Second central moment also called « Variance »

Second order moment: m_2

$E(X^2)$ (important in physics)

$$m_2 = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx$$ Warning : m_2 can not be negative

Variance: second central moment Var(X) or σ^2_X

It allows to measure how spread out from the expected value are the values taken by the random variable. It is a measure of dispersion.

$$Var(X) = \sigma_X^2 = E\,(X - m_1)^2$$

For a continuous random variable $$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - m_1)^2 \cdot f_X(x) \cdot dx$$

For a discrete random variable $$\sigma_X^2 = \sum_i (x_i - m_1)^2 \cdot P_i$$

Important properties

Expected value:

- The expected value of a constant *b* is equal to this constant E(b) = b
- *Linearity* : the expected value is a linear operator. For any two random variables X and Y, and for any two real numbers a and *b:*

$$E\,(aX + bY) = aE(X) + b\,E(Y)$$

- *Product* : in general, E(XY) ≠ E(X) E(Y). However, if X and Y are independant the equality is true.

Variance:

$$\sigma_X^2 = Var(X) = E(X^2) - E^2(X) = m_2 - m_1^2$$

If *c* is a constant:

$$Var(cX) = c^2 Var(X)$$ and $$Var(X \pm c) = Var(X)$$

Standard deviation

$$\sigma_X = \sqrt{E(X^2) - E^2(X)} > 0$$

Exercise

Given the random variable X with expected value m_1 and standard deviation σ_X Calculate the expected value and the variance of Y.

$$Y = \frac{X - m_1}{\sigma_X}$$

Bienaymé–Tchebychev's theorem (1/2) :

(also called inequality)

Let X be a random variable with finite expected value m and finite non-zero variance σ^2. Then for any real number $\varepsilon > 0$,

$$\boxed{P\left\{\left|X - m_1\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{\varepsilon^2} \qquad \varepsilon > \sigma \geq 0}$$

Interpretation : The smaller is the variance of a random variable, the smaller is the probability that it deviates from its mean by more than ε units.

- This theorem highlights the role of the expected value and of the variance in the description of the random variable.

Bienaymé–Tchebychev's theorem (2/2) :

Another version :

$$\varepsilon = n\sigma \quad \Rightarrow \quad P\left(\left\{\left|X - m_1\right| \geq n\sigma\right\}\right) \leq \frac{1}{n^2} \qquad (n > 1)$$

Interpretation: This theorem guarantees no more than $1/n^2$ of the distribution's values can be n or more *standard deviations away from the mean (or equivalently, over* $1 - 1/n^2$ *of the distribution's values are less* than n standard deviations away from the mean).

Example: If n=10, for any random variable X, the probability that it deviates from its mean by more than 10σ is less than 1%.

Generalization :

Non-central moments : $$m_n = \int_{-\infty}^{+\infty} x^n \cdot f_X(x) \cdot dx$$

Central moments : $$\mu_n = E\left((X - m_1)^n\right) = \int_{-\infty}^{+\infty} (x - m_1)^n \cdot f_X(x) \cdot dx$$

Some properties between moments of any order :

(proof by using linearity of the operator E(.))

$$C_n^r = \frac{n!}{r! \cdot (n-r)!}$$

$$\mu_n = \sum_{r=0}^{n} C_n^r \cdot (-1)^r \cdot m_1^r \cdot m_{n-r}$$

$$m_n = \sum_{r=0}^{n} C_n^r \cdot m_1^r \cdot \mu_{n-r}$$

Probability theory

Chapitre 5: CHARACTERISTIC FUNCTION OF A REAL RANDOM VARIABLE

- Definition
- Related theorems

Definition of characteristic function

The characteristic function of the real random variable X is the complex-valued function of t defined by:

$$\boxed{\begin{cases} R & \rightarrow & C \\ t & \mapsto & \varphi_X(t) = E\{e^{jtX}\} \end{cases}}$$

where $t \in \mathbb{R}$ is a deterministic parameter and $j^2 = -1$ is the imaginary unit.

For a continuous random variable X with density $f_x(x)$:

$$\boxed{\varphi_X(t) = \int_{-\infty}^{+\infty} e^{jtx} f_X(x) dx}$$

For a discrete random variable X :

$$\boxed{\begin{gathered}\varphi_X(t) = \sum_i P_i \cdot e^{j \cdot t \cdot x_i} \\ \text{with } P_i = P(X = x_i)\end{gathered}}$$

Remarks:

- the characteristic function is the Fourier transform of the probability density function.
- the characteristic function completely defines the probability distribution.

Example 1 : Let X be a random variable X with Bernoulli distribution of parameter *p.*

$$X = \begin{cases} 1 & \text{with} \quad P\{X = 1\} = p \\ 0 & \text{with} \quad P\{X = 0\} = q \end{cases} \quad \text{et} \quad p + q = 1$$

$$\varphi_X(t) = \sum_{i=0}^{1} P_i e^{jtx_i} = qe^{jt0} + pe^{jt1}$$

$$\varphi_X(t) = pe^{jt} + q$$

Example 2: Let X be a exponential random variable:

$$f_X(x) = \alpha e^{-\alpha x}, \; \alpha > 0, \; x \geq 0$$

$$\varphi_X(t) = \int_0^{+\infty} \alpha \, e^{-\alpha x} e^{jtx} dx$$

$$\varphi_X(t) = \varphi_{\exp}(t) = \frac{\alpha}{\alpha - jt}$$

Moment-generating function :

If the random variable *X* has moments up to *n*-th order, then the characteristic function φ_X is *n* times continuously differentiable on the entire real line and we can use it to find moments

$$\boxed{\left.\frac{d^n \varphi_X(t)}{dt^n}\right|_{t=0} = j^n m_n}$$

Example : Calculation of E(X) and Var (X)

Binomial distribution : $P\{X = k\} = C_n^k p^k q^{n-k} \quad 0 \le X \le n$

with $p + q = 1$

The characteristic function is : $\varphi_X(t) = \left(pe^{jt} + q\right)^n$

The 1st and 2nd derivative are:

$$\varphi_X'(t) = n\left(pe^{jt} + q\right)^{n-1} .jpe^{jt}$$

$$\varphi_X''(t) = jnp\left[(n-1)\left(pe^{jt} + q\right)^{n-2} .jpe^{2jt} + \left(pe^{jt} + q\right)^{n-1} je^{jt}\right]$$

At point t = 0 $\boxed{\varphi_X'(0) = jnp = jm_1 \Rightarrow m_1 = np}$

$$\varphi_X''(0) = j^2\left[n^2p^2 + npq\right] = j^2 m_2 \qquad m_2 = n^2p^2 + npq$$

$$\boxed{Var\{X\} = \sigma_X^2 = m_2 - m_1^2 = npq}$$

Theorem: Inversion formulae:

Fourier inversion theorem

If characteristic function φ_X is integrable, then F_X is absolutely continuous, and therefore X has a probability density function.

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi_X(t) e^{-jxt} dt$$

There is a one-to-one correspondence between cumulative distribution functions and characteristic functions.