Mathematics for Data Science

Lecture 1: Random Variables

Mohamad GHASSANY

EFREI PARIS

Mohamad GHASSANY

- Associate Professor at EFREI Paris, head of Data & Artificial Intelligence Master program.
- Phd in Computer Science Université Paris 13.
- Master 2 in Applied Mathematics & Statistics from Université Grenoble Alpes.
- Personal Website: mghassany.com

Introduction to probability theory

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Events: In this random experiment, one can be interested in more complex events than just a simple result of the experiment.
- The The Power set Ω, called $\mathcal{P}(\Omega)$, is the set of all subsets of Ω.
- A family of subsets $\mathcal{A}$ of Ω. These subsets are called events. We say that the event A has occured if and only if the result ω of Ω that has occurred belongs to A.
- σ-**Algebra**: We call σ-Algebra any family $\mathcal{A}$ of subsets of Ω satisfying:
  1. $\Omega \in \mathcal{A}$.
  2. if $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$.
  3. if $(A_n)_{n\in\mathbb{N}}$ is a sequence of elements in $\mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.
- $(\Omega, \mathcal{A})$ is a measurable space (or a Borel space).

- Let $(\Omega, \mathcal{A})$ be a measurable space:
  - The set $\mathcal{A}$ is called σ-Algebra of events. The elements of $\mathcal{A}$ are called the events.
  - The event Ω is called certain event. The event $\emptyset$ is called impossible event.
- Operations on events. Let A and B be two events:
  - $\bar{A}$ is the complement event of A (we also note A^c). $barA=Om \setminus A$.
    $barA$ occurs if and only if A does not occur.
  - $A \cap B$ is the event «A and B».
    $A \cap B$ occurs when both events occur.
  - $A \cup B$ is the event «A or B».
    $A \cup B$ occurs when at least one of the two events occurs.
- Mutually Exclusive Events: A and B are mutually exclusive if their simultaneous realization is impossible: $A \cap B = \emptyset$.
- Implication: A implies B means that if A occurs, then B also occurs: $A \subset B$.

- Let $(\Omega, \mathcal{A})$ a measurable space. A probability function on $(\Omega, \mathcal{A})$, is any map

$$P : \mathcal{A} \to \mathbb{R}$$

such that:

1. $\forall A \in \mathcal{A}, P(A) \geqslant 0$.
2. $P(\Omega) = 1$.
3. $\forall (A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$, a familty of pairwise disjoint (mutually exclusive) events, we have:

$$P(\bigcup_{n \in \mathbb{N}^*} A_n) = \sum_{n=1}^{+\infty} P(A_n)$$

- The triplet $(\Omega, \mathcal{A}, P)$ is called a **probability space**.

1. $P(\emptyset) = 0$.
2. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
3. If A_1 and A_2 are mutually exclusive, $A_1 \cap A_2 = \emptyset$, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.
4. $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$.
5. $P(\bar{A}) = 1 - P(A)$.
6. $P(B \setminus A) = P(B) - P(B \cap A)$.
7. $A \subset B \Rightarrow P(A) \leqslant P(B)$.

Uniform probability on finite Ω

- Let Ω be a finite sample space. We say that P is the **uniform probability** on the measurable space $(\Omega, P(\Omega))$ if:

$$\forall \omega, \omega' \in \Omega, \qquad P(\omega'\}) = P(\omega'\})$$

  One also says that there is **equiprobability** of elementary events.

- Let $(\Omega, \mathcal{P}(\Omega), P)$ be a finite probability space. If P is the uniform probability, then

$$\forall A \in \mathcal{A}, \qquad P(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$$

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.

- Remark: $(A|B)$ is not an event! We use the notation $P(A|B)$ for simplicity, but the notation $P_B(A)$ is the correct one.
- Chain rule:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

- Law of total probability:
  - $\forall A \in \mathcal{A}, \quad P(A) = P(A \cap B) + P(A \cap \bar{B})$
  - We call partition of Ω, a set of events that are pairwise disjoint and whose union is the sample space Ω. The partition is said to be "countable" if its cardinality is at most equal to that of $\mathbb{N}$.
  - Let $(B_n)_{n \geqslant 0}$ a partition of Ω. We have:

$$\forall A \in \mathcal{A}, \qquad P(A) = \sum_{n \geqslant 0} P(A \cap B_n)$$

- Independence: Events A and B are independent iff $P(A \cap B) = P(A)P(B)$.

First Bayes' theorem

Let $(\Omega, \mathcal{A}, P)$ a probability space. For all events A and B such that $P(A) \neq 0$ and $P(B) \neq 0$, we have:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Second Bayes' theorem

Let $(\Omega, \mathcal{A}, P)$ a probability space and $(B_n)_{n \geqslant 0}$ a partition of Ω s.t. for all $n \geqslant 0$ $P(B_n) \neq 0$. We have for all $A \in \mathcal{A}$ s.t. $P(A) \neq 0$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{n \geqslant 0} P(A|B_n)P(B_n)} \qquad \forall i \geqslant 0$$

Real Random Variable

Concept of Real Random Variable

Definition

Let ε an experiment and $(\Omega, \mathcal{A}, \mathrm{P})$ the associated probability space. . In many situations, one associates to each result $\omega \in \Omega$ a real number denoted $\mathrm{X}(\omega)$; In this way, one builds a map $\mathrm{X} : \Omega \to \mathbb{R}$. Historically, ε was a game and X représented the earning of a player.

Example: a die throw

A player throws a fair six faces dice and we observe the obtained number:

- If the result is 1,3 or 5, the player earns 1 euro.
- If the result is 2 or 4, the player earns 5 euros.
- If the result is 6, the player loses 10 euros.

Analysis

- ε: "throw a fair die".
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{A} = \mathcal{P}(\Omega)$.
- P is the equiprobability on $(\Omega, \mathcal{A})$.

Let X the map function from Ω to $\mathbb{R}$ that associates to each result the corresponding earning.

So we have

- $X(1) = X(3) = X(5) = 1$
- $X(2) = X(4) = 5$
- $X(6) = -10$

We say that X is a **random variable** on Ω.

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

Similarly, we have:

- $P(X = 5) = 1/3$.
- $P(X = -10) = 1/6$.

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

This is tantamount of considering a new sample space:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

equipped with the probability P_X defined in the table above. This new probability is called the **probability distribution** of X.

Notice that

$$P\Big(\bigcup_{x_i \in \Omega_X} \{X = x_i\}\Big) = \sum_{x_i \in \Omega_X} P(X = x_i) = 1$$

In this chapter:

- We treat the case where $X(\Omega)$ is countable.
- The random variable in this case is ***discrete***.
- We define its probability law by its probability distribution.
- We will define the two main numerical characteristics of a discrete random variable:
  - Expected value: characteristic of centrality (the *mean*).
  - Variance: characteristic of dispersion.
- We will also define the couples of random variables.

Discrete Random Variables

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \dots, x_i, \dots\}$ with $x_1 \leqslant x_2 \leqslant \dots \leqslant x_i \leqslant \dots$.*
- *The probability distribution defined on $X(\Omega)$ by*

$$p_i = p(x_i) = P(X = x_i) \quad \forall\, i = 1, 2, \dots$$

Remarks:

- $B \subset \mathbb{R}$, $P(X \in B) = \sum_{i/x_i \in B} p(x_i)$.
- $P(a < X \leqslant b) = \sum_{i/a < x_i \leqslant b} p(x_i)$.
- $p(x_i) \geqslant 0$ and $\sum_{i=1}^{\infty} p(x_i) = 1$.
- If the number of possible values of X is small enough, the probability distribution of X is often presented as a table.

Cumulative distribution function

Definition

Given a discrete random variable X, we call cumulative distribution function of X (or simply distribution function), denoted F_X, the function defined by: for any real a,

$$F(a) = P(X \leqslant a) = \sum_{i/x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.

Properties

1. It is a staircase function.
2. $F(a) \leqslant 1$ since it is a probability.
3. $F(a)$ is continuous from the right.
4. $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$

The distribution function characterizes the distribution of X. In other words, if X and Y are two random variables, we have $F_X = F_Y$ if and only if their probability distributions are the same.

Distribution function and probabilities over X

All the computations of probabilities about X can be carried out using the distribution function. For example,

$$P(a < X \leqslant b) = F(b) - F(a) \qquad \text{pour tout } a < b$$

This is easier to understand if one writes the event $\{X \leqslant b\}$ as a union of two incompatible events $\{X \leqslant a\}$ and $\{a < X \leqslant b\}$, Let

$$\{X \leqslant b\} = \{X \leqslant a\} \cup \{a < X \leqslant b\}$$

In this way,

$$P(X \leqslant b) = P(X \leqslant a) + P(a < X \leqslant b)$$

which proves the equality above.

Remark

One can compute the individual probabilities by:

$$p_i = P(X = x_i) = F(x_i) - F(x_{i-1}) \qquad \text{pour } 1 \leqslant i \leqslant n$$

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $\mathrm{card}(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow$ $X(\Omega) = \{0, 1, 2, 3\}$.

- Let's calculate $P(X = 1)$.
- $X^{-1}(1) = \{(P, F, F), (F, P, F), (F, F, P)\}$.

$\Rightarrow$ $P(X = 1) = \frac{3}{8}$

Using the same method we obtain the probability distribution of X:

k	0	1	2	3
$P(X = k)$	$1/8$	$3/8$	$3/8$	$1/8$

The distribution function X is therefore given by:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/8 & \text{si } 0 \leqslant x < 1 \\ 1/2 & \text{si } 1 \leqslant x < 2 \\ 7/8 & \text{si } 2 \leqslant x < 3 \\ 1 & \text{si } x \geqslant 3 \end{cases}$$

One can represent both the probability distribution and the distribution function of X in the same table:

k	0	1	2	3
$P(X = k)$	$1/8$	$3/8$	$3/8$	$1/8$
$F_X(x)$	$1/8$	$1/2$	$7/8$	1

The graph of the distribution function is represented below:

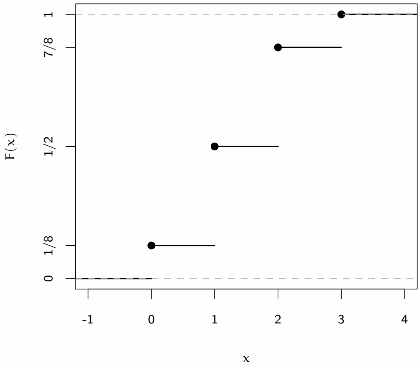


Figure 1: Distribution function

Here is another slightly different representation of the distribution function:

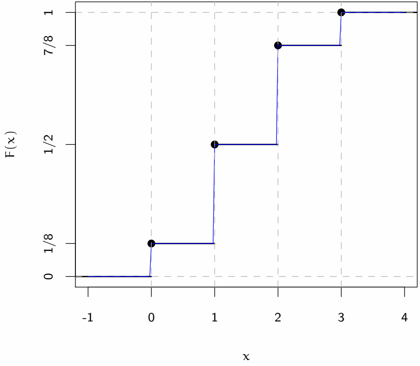


Figure 2: Distribution function

Definition

Let A an event. We call indicator random variable of the event A, the random variable $X = \mathbb{1}_A$ defined by:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in \bar{A} \end{cases}$$

Therefore:

- $P(X = 1) = P(A) = p$
- $P(X = 0) = P(\bar{A}) = 1 - p$

The distribution function of the indicator random variable is therefore:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - p & \text{si } 0 \leqslant x < 1 \\ 1 & \text{si } x \geqslant 1 \end{cases}$$

Example

- Let $\mathcal{U}$ an urn containing two white ball and three red balls.
- We randomly take one ball out of the box.
- Let A : "take one white ball out".
- Let X be the indicator random variable of A.

Find the probability distribution and the distribution function of X.

The probability distribution of X is

k	0	1
$P(X=k)$	$\frac{3}{5}$	$\frac{2}{5}$

and its distribution function is:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 3/5 & \text{si } 0 \leqslant x < 1 \\ 1 & \text{si } x \geqslant 1 \end{cases}$$

Moments of a discrete random variable

Definition

For a discrete random variable X with probability distribution $p(.)$, we define the expected value of X, called $E(X)$, by

$$E(X) = \sum_{i \in \mathbb{N}} x_i p(x_i)$$

In concrete terms, the expected value of X is the weighted mean of the values of X, the weights being the probabilities associated to the values of X.

Examples

1. In the previous example where we play three times to "heads or tails", the expected value of X is:

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

2. For the indicator random variable of A:

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = P(A) = p$$

   which means that the expected value of the indicator of an event A corresponds to the probability that A occurs.

Theorem

Let X be a discrete random variable whose possible values are x_i, $i \geqslant 1$, and denote by $p(x_i)$ the probability that $X = x_i$ occurs. Then, for any real function g, we have

$$E(g(X)) = \sum_i g(x_i)p(x_i)$$

Example

Let X be a random variable that can take three values $\{-1, 0, 1\}$ with the following probabilities:

$$P(X = -1) = 0.2 \qquad P(X = 0) = 0.5 \qquad P(X = 1) = 0.3$$

Calculate $E(X^2)$.

Solution

First method: Let $Y = X^2$. The probability distribution of Y is given by

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.5$$
$$P(Y = 0) = P(X = 0) = 0.5$$

So

$$E(X^2) = E(Y) = 1(0.5) + 0(0.5) = 0.5$$

Second method: Using the theorem

$$\begin{aligned} E(X^2) &= (-1)^2(0.2) + 0^2(0.5) + 1^2(0.3) \\ &= 1(0.2 + 0.3) + 0(0.5) = 0.5 \end{aligned}$$

Remark

$$0.5 = E(X^2) \neq (E(X))^2 = 0.01$$

Properties

1. $E(X+a) = E(X) + a, \quad a \in \mathbb{R}$

   results which follows from:

$$\sum_i p_i(x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

2. $E(aX) = aE(X), \quad a \in \mathbb{R}$

   to prove it, just write:

$$\sum_i p_i a x_i = a \sum_i p_i x_i$$

3. $E(X+Y) = E(X) + E(Y)$, X and Y being two random variables.

All these three properties are summarised in the claim that the expected value is linear:

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y), \quad \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}.$$

Definition

Let X be a discrete random variable. We call variance of X, denoted $V(X)$, the quantity defined by, when it exists,

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

Thus, the variance is the expected value of the square of the centered random variable $X - E(X)$. The variance can be interpreted as a measure of the dispersion of the possible values of X around its expected value.

Remark

Equivalently, the variance might be defined by the following formula:

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

Indeed:

$$\begin{aligned} V(X) &= E\left[X^2 - 2XE(X) + E^2(X)\right] \\ &= E(X^2) - E[2XE(X)] + E[E^2(X)] \\ &= E(X^2) - 2E^2(X) + E^2(X) \end{aligned}$$

Example

OLet us compute $V(X)$ in the case where X is the number obtained when throwing a fair die.

Previously, we saw that $E(X) = \frac{7}{2}$. Moreover,

$$
\begin{aligned}
E(X^2) &= \sum_i x_i^2 p(x_i) \\
&= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\
&= \left(\frac{1}{6}\right)(91) = \frac{91}{6}
\end{aligned}
$$

And therefore

$$
\begin{aligned}
V(X) &= E(X^2) - E^2(X) \\
&= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}
\end{aligned}
$$

Properties

1. $V(X) \geqslant 0$

2. $\forall\, a \in \mathbb{R}, \quad V(X+a) = V(X)$

   en effet:

$$\begin{aligned} V(X+a) &= E\left[\,[X+a-E(X+a)]^2\,\right] \\ &= E\left[\,[X+a-E(X)-a]^2\,\right] \\ &= E\left[\,[X-E(X)]^2\,\right] = V(X) \end{aligned}$$

3. $\forall\, a \in \mathbb{R}, \quad V(aX) = a^2 V(X)$

   en effet:

$$\begin{aligned} V(aX) &= E\left[\,[aX-E(aX)]^2\,\right] \\ &= E\left[\,[aX-aE(X)]^2\,\right] \\ &= E\left[a^2\,[X-E(X)]^2\right] \\ &= a^2\left[E\,[X-E(X)]^2\right] = a^2 V(X) \end{aligned}$$

Standard deviation

Definition

*Let X be a discrete random variable. The square root of the variance is called the **standard deviation** of X and is denoted*

$$\sigma_X = \sqrt{V(X)}$$

σ_X has the same physical units as the random variable X.

- The standard deviation allows to measure the dispersion of a set of data.
- The smaller sigma is, the closer to each other the values of the data are.
- Example: the dispersion of the grades in an exam. The smaller sigma is, the more homogeneous the class is.
- - Expected value and standard deviation are linked through *Bienaymé-Tchebychev inequality*.

Theorem

Let X a random variable of expected value μ and variance σ^2. For all $\varepsilon > 0$, We have:

$$P(|X - E(X)| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

Remark

This inequality can be written in a slightly different fashion. Let $k = \varepsilon/\sigma$.

$$P(|X - E(X)| \geqslant k\sigma) \leqslant \frac{1}{k^2}$$

Importance

This inequality relates the probability for X to deviate from its expected value $E(X)$ to its variance, which is precisely an indicator of the dispersion around the expected value. The inequality makes quantitatively precise the statement "the smaller the variance is, the less likely it is to find values far away from the expected value".

Definition

We call non centered moment of order $r \in \mathbb{N}^*$ of X the quantity, when it exists:

$$m_r(X) = \sum_{i \in \mathbb{N}} x_i^r p(x_i) = E(X^r).$$

Definition

The centered moment of order $r \in \mathbb{N}^*$ the quantity, when it exists:

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i \left[x_i - E(X)\right]^r = E\left[X - E(X)\right]^r.$$

Remark

The first moments are:

- $m_1(X) = E(X), \quad \mu_1(X) = 0.$
- $\mu_2(X) = V(X) = m_2(X) - m_1^2(X).$

Two Random Variables

Two Discrete Random Variables

So far, we have dealt with one random variable. However, it is often necessary to consider events related to two variables simultaneously, or even to more than two variables.

Definition

Let X *and* Y *two discrete random variables, defined on probability space* $(\Omega, \mathcal{A}, P)$ *and that* $X(\Omega) = \{x_1, x_2, \ldots, x_l\}$ *and* $Y(\Omega) = \{y_1, y_2, \ldots, y_k\}$, l *and* $k \in \mathbb{N}$.

The ***probability law of*** (X, Y) *is defined by joint probabilities:*

$$p_{ij} = P(X = x_i; Y = y_j) = P(\{X = x_i\} \cap \{Y = y_j\})$$

We have

$$p_{ij} \geqslant 0 \quad \textit{et} \quad \sum_{i=1}^{l} \sum_{j=1}^{k} p_{ij} = 1$$

The pair (X, Y) *is called two dimensional random vector and can have* $l \times k$ *valeurs.*

The probabilities p_{ij} can be presented in a two dimensional table than we call joint probability distribution table:

Table 1: Joint probability distribution table

$X\backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_k
x_1	p_{11}	p_{12}		p_{1j}		p_{1k}
x_2	p_{21}	p_{22}		p_{2j}		p_{2k}
$\vdots$						
x_i	p_{i1}	p_{i2}		p_{ij}		p_{ik}
$\vdots$						
x_l	p_{l1}	p_{l2}		p_{lj}		p_{lk}

In the header we have the possible values of Y and in the first column the possible values of X. The probability $p_{ij} = P(X = x_i; Y = y_j)$ is at the intersection of i^{th} line and j^{th} column.

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".
- $|\Omega| = C_{12}^3 = 220$.
- $X(\Omega) = \{0, 1, 2, 3\}$ and $Y(\Omega) = \{0, 1, 2, 3\}$.
- $p(X = 0, Y = 0) = p(0, 0) = C_5^3 / C_{12}^3 = \frac{10}{220}$.
- $p(0, 1) = C_4^1 C_5^2 / C_{12}^3 = \frac{40}{220}$.
- $p(1, 0) = C_3^1 C_5^2 / C_{12}^3 = \frac{30}{220}$.

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

Table 2: Joint probability distribution table

$X \backslash Y$	0	1	2	3
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0
3	$\frac{1}{220}$	0	0	0

When we know the joint distribution of the random variables X and Y, we can also look at the probability distribution of X alone and Y alone. These are the marginal probability distributions.

- Marginal distribution of X:

$$p_{i.} = P(X = x_i) = P[\{X = x_i\} \cap \Omega] = \sum_{j=1}^{k} p_{ij} \qquad \forall\, i = 1, 2, \ldots, l$$

- Marginal distribution of Y:

$$p_{.j} = P(Y = y_j) = P[\Omega \cap \{Y = y_j\}] = \sum_{i=1}^{l} p_{ij} \qquad \forall\, j = 1, 2, \ldots, k$$

We can calculate the marginal distributions directly from the table of the joint distribution.

Marginal Probability Distributions

Table 3: Joint distribution table with marginal distributions

$X\backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_k	Marginal of X
x_1	p_{11}	p_{12}		p_{1j}		p_{1k}	$p_{1.}$
x_2	p_{21}	p_{22}		p_{2j}		p_{2k}	$p_{2.}$
$\vdots$							
x_i	p_{i1}	p_{i2}		p_{ij}		p_{ik}	$p_{i.}$
$\vdots$							
x_l	p_{l1}	p_{l2}		p_{lj}		p_{lk}	$p_{l.}$
Marginal of Y	$p_{.1}$	$p_{.2}$		$p_{.l}$		$p_{.k}$	1

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

Table 4: Joint distribution table

$X \backslash Y$	0	1	2	3	$p_{i.} = P(X = x_i)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$p_{.j} = P(Y = y_j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

Definition

For each value y_j *of* Y *such that* $p_{.j} = P(Y = y_j) \neq 0$ *we can define the conditional distribution of* X *given* $Y = y_j$ *by*

$$p_{i/j} = P(X = x_i / Y = y_j) = \frac{P(X = x_i ; Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}} \qquad \forall i = 1, 2, \ldots, l$$

Same for Y *given* $X = x_i$*:*

$$p_{j/i} = P(Y = y_j / X = x_i) = \frac{P(X = x_i ; Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}} \qquad \forall j = 1, 2, \ldots, k$$

Definition

We say that two random variables are independent iff

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \qquad \forall i = 1, 2, \ldots, l \textit{ and } j = 1, 2, \ldots, k$$

One demonstrates that

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) \qquad \forall\ A \textit{ and } B \in \mathcal{A}$$

Properties

Let two random variables X and Y,

1. $E(X + Y) = E(X) + E(Y)$
2. If X and Y are independent so $E(XY) = E(X)E(Y)$. But the reciprocal is not always true.

Definition

*Let two random variables X and Y. The **covariance** of X and Y, when it exists, is*

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

that we can calculate using the formula

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Properties

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX_1 + bX_2, Y) = aCov(X_1, Y) + bCov(X_2, Y)$
- $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$
- If X and Y are independant so
  - $Cov(X,Y) = 0$ (the reciprocal is not always true)
  - $V(X+Y) = V(X) + V(Y)$ (the reciprocal is not always true)

Linear Correlation (Correlation of Spearman)

Definition

The correlation between X and Y is defined by

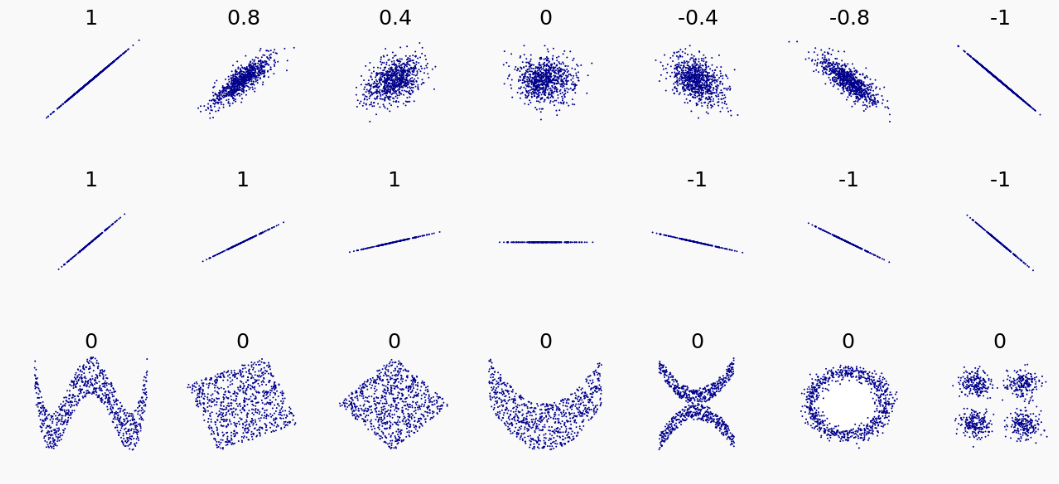
$$\rho = \rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

We can demonstrate that

$$-1 \leqslant \rho(X, Y) \leqslant 1$$

Interpretation of ρ

- The correlation coefficient is a measure of the degree of linearity between X and Y.
- Values of rho close to 1 or -1 indicate an almost rigorous linearity between X and Y.
- Values of rho close to 0 indicate the absence of any linear relationship.
- When $\rho(X, Y)$ is positive, Y tends to increase if X does the same.
- When $\rho(X, Y) < 0$, Y tends to decrease if X increases.


1
0.8
0.4
0
-0.4
-0.8
-1
1
1
1
-1
-1
-1
0
0
0
0
0
0
0

Loi Uniforme Discrète $\mathcal{U}(n)$

Definition

Une distribution de probabilité suit une loi uniforme lorsque toutes les valeurs prises par la variable aléatoire sont équiprobables. Si n est le nombre de valeurs différentes prises par la variable aléatoire alors on a:

$$P(X = x_i) = \frac{1}{n} \qquad \forall i \in \{1, \ldots, n\}$$

On dit $X \sim \mathcal{U}(n)$.

Example

La distribution des chiffres obtenus au lancer de dé (si ce dernier est non pipé) suit une loi uniforme dont la loi de probabilité est la suivante :

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Moments de loi uniforme discrète

Cas particulier

Dans le cas particulier d'une loi uniforme discrète où chaque valeur de la variable aléatoire X correspond à son rang, i.e. $x_i = i \ \forall i \in \{1, \ldots, n\}$, on a:

$$E(X) = \frac{n+1}{2} \quad \text{et} \quad V(X) = \frac{n^2-1}{12}$$

Démonstration

La démonstration de ces résultats est établie en utilisant les égalités:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \quad \text{et} \quad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Vous avez la démonstration de ces égalités dans l'Annexe.

Example

L'exemple du lancer du dé: on peut calculer directement les moments de X:

$$E(X) = \frac{6+1}{2} = 3.5 \quad \text{et} \quad V(X) = \frac{6^2-1}{12} = \frac{35}{12} \simeq 2.92.$$

Loi de Bernoulli $\mathcal{B}(p)$

Loi de Bernoulli $\mathcal{B}(p)$

Variable Indicatrice

Soit A un événement quelconque; on appelle v.a. indicatrice de l'événement A, la v.a. définie par $X = \mathbb{1}_A$, c'est à dire:

$$X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 0 & \text{si } \omega \in \bar{A} \\ 1 & \text{si } \omega \in A \end{cases}$$

Ainsi $X(\Omega) = \{0, 1\}$ avec:

$$\begin{aligned} & P(X=1) = P\{\omega \in \Omega / X(\omega) = 1\} = P(A) = p \\ & P(X=0) = P\{\omega \in \Omega / X(\omega) = 0\} = P(\bar{A}) = 1 - P(A) = q \\ & \text{avec } p + q = 1 \end{aligned}$$

Definition

On dit que X suit une loi de Bernoulli de paramètre $p = P(A)$, ce qu'on écrit symboliquement $X \sim \mathcal{B}(p)$. Une distribution de Bernoulli est associée à la notion "épreuve de Bernoulli", qui est une épreuve aléatoire à deux issues: succès $(X = 1)$ et échec $(X = 0)$.

Loi de Bernoulli $\mathcal{B}(p)$

Fonction de répartition de loi de Bernoulli

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \leqslant x < 1 \\ 1 & \text{si } x \geqslant 1. \end{cases}$$

Espérance de loi de Bernoulli

$$E(X) = 1 \times P(A) + 0 \times P(\bar{A}) = P(A) = p$$

Variance de loi de Bernoulli

$$V(X) = E(X^2) - E^2(X) = p - p^2 = p(1-p) = pq$$

car

$$E(X^2) = 1^2 \times P(A) + 0^2 \times P(\bar{A}) = P(A) = p$$

Loi Binomiale $\mathcal{B}(n, p)$

Loi Binomiale $\mathcal{B}(n, p)$

- Décrite pour la première fois par *Isaac Newton* en 1676 et démontrée pour la première fois par le mathématicien suisse *Jacob Bernoulli* en 1713.
- La loi binomiale est l'une des distributions de probabilité les plus fréquemment rencontrées en statistique appliquée.
- On exécute n épreuves **indépendantes** de Bernoulli.
- Chaque épreuve a p pour probabilité de succès et $1 - p$ pour probabilité d'échec.

$$A \quad A \quad \bar{A} \quad A \quad \bar{A} \quad \dots \quad \bar{A} \quad A \quad A$$

$$S \quad S \quad E \quad S \quad E \quad \dots \quad E \quad S \quad S$$

- $X =$ **le nombre de succès** sur l'ensemble des n épreuves.
- X dépend de deux paramètres n et p.

$$S \quad S \quad E \quad S \quad E \quad \dots \quad E \quad S \quad S$$

- $X =$ **le nombre de succès** sur l'ensemble des n épreuves.
- $X(\Omega) = \{0, 1, \dots, n\}$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad 0 \leqslant k \leqslant n$$

- $\binom{n}{k}$ est le nombre d'échantillons de taille n comportant exactement k succès, de probabilité p^k, indépendamment de l'ordre, et donc $n - k$ échecs, de probabilité $(1 - p)^{n-k}$.
- On écrit $X \sim \mathcal{B}(n, p)$.

Remark

Une variable de Bernoulli n'est donc qu'une variable binomiale de paramètres $(1, p)$.

$$X \sim \mathcal{B}(p) \iff X \sim \mathcal{B}(1, p)$$

Triangle de Pascal

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \quad \forall n \geqslant 1 \text{ et } 1 \leqslant k \leqslant n-1$$

$n = 0$:					1				
$n = 1$:				1		1			
$n = 2$:			1		2		1		
$n = 3$:		1		3		3		1	
$n = 4$:	1		4		6		4		1

Binôme de Newton

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

Cette formule permet de vérifier que la loi Binomiale est une loi de probabilité:

$$\sum^{n} P(X=k) = \sum^{n} \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

Example

On jette cinq pièces équilibrées. Les résultats sont supposés indépendants. Donner la loi de probabilité de la variable X qui compte le nombre de piles obtenus.

Solution

- X = nombre de piles (*succès*).
- $n = 5$.
- $p = 1/2$.
- $X \sim \mathcal{B}(5, \frac{1}{2})$.
- $X(\Omega) = \{0, 1, \ldots, 5\}$
- $P(X = 0) = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(1 - \frac{1}{2}\right)^{5-0} = \frac{1}{32}$
- $P(X = 1) = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(1 - \frac{1}{2}\right)^4 = \frac{5}{32}$
- $P(X = 2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(1 - \frac{1}{2}\right)^3 = \frac{10}{32}$
- $P(X = 3) = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(1 - \frac{1}{2}\right)^2 = \frac{10}{32}$
- $P(X = 4) = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(1 - \frac{1}{2}\right)^1 = \frac{5}{32}$
- $P(X = 5) = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(1 - \frac{1}{2}\right)^0 = \frac{1}{32}$

Moments de la loi Binomiale

Si $X \sim \mathcal{B}(n,p)$ alors $E(X) = np$ et $V(X) = np(1-p)$

Démonstration

Première approche: On associe à chaque épreuve i, $1 \leqslant i \leqslant n$, une v.a. de Bernoulli.

$$\mathbb{1}_A = X_i = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } \bar{A} \text{ est réalisé} \end{cases}$$

On peut écrire alors: $X = \sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$

Donc

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

et

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

car les v.a. X_i sont indépendantes.

Deuxième approche: Calcul direct.

- $E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \ldots = np$
- $V(X) = E(X^2) - E^2(X)$
- Pour obtenir $E(X^2)$ par un procédé de calcul identique, on passe par l'intermédiaire du moment factoriel $E[X(X-1)]$.
- $V(X) = E(X^2) - E^2(X) = E[X(X-1)] + E(X) - E(X^2)$
- $E[X(X-1)] = \sum_{k=0}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \ldots = n(n-1)p^2$
- $V(X) = n(n-1)p^2 + np - (np)^2 = np(1-p)$

Example

Le nombre de résultats pile apparus au cours de n jets d'une pièce de monnaie suit une loi binomiale $\mathcal{B}(n, 1/2)$:

$$P(X=k) = \binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k} = \frac{\binom{n}{k}}{2^n}, \quad 0 \leqslant k \leqslant n$$

avec $E(X) = n/2$ et $V(X) = n/4$.

Example

Le nombre N de boules rouges apparues au cours de n tirages avec remise dans une urne contenant deux rouges, trois vertes et une noire suit une loi binomiale $\mathcal{B}(n, 1/3)$:

$$P(N=k) = \binom{n}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{n-k} = \binom{n}{k}\frac{2^{n-k}}{3^n}, \quad 0 \leqslant k \leqslant n$$

avec $E(X) = n/3$ et $V(X) = 2n/9$.

Remark

Si $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$, les v.a. X_1 et X_2 étant **indépendantes**, alors $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$. Ceci résulte de la définition d'une loi binomiale puisqu'on totalise ici le résultat de $n_1 + n_2$ épreuves indépendantes.

Loi de Poisson $\mathcal{P}(\lambda)$

Definition

Une v.a. X suit une loi de Poisson de paramètre $\lambda > 0$ si c'est une variable à valeurs entières, $X(\Omega) = \mathbb{N}$, donc avec une infinité de valeurs possibles, de probabilité:

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}$$

Cette loi ne dépend qu'un seul paramètre réel positif λ, avec l'écriture symbolique $X \sim \mathcal{P}(\lambda)$.

Remark

$$e^x = \sum_{i=0}^{+\infty} \frac{x^i}{i!}$$

Donc

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

Si $X \sim \mathcal{P}(\lambda)$ alors $E(X) = \lambda$ et $V(X) = \lambda$

Espérance de loi de Poisson

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kP(X = k) \\ &= \dots \\ &= \lambda. \end{aligned}$$

Variance de loi de Poisson

- On calcule d'abord $E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \dots = \lambda(\lambda + 1)$.

Ensuite

$$V(X) = \lambda(\lambda + 1) - \lambda^2 = \lambda$$

Loi de Poisson $\mathcal{P}(\lambda)$

Example

- $X =$ nombre de micro-ordinateurs vendus chaque jour dans un magasin.
- On suppose $X \sim \mathcal{P}(5)$.
- La probabilité associée à la vente de 5 micro-ordinateurs est

$$P(X=5) = e^{-5}\frac{5^5}{5!} = e^{-5} \simeq 0.1755$$

- La probabilité de vendre au moins 2 micro-ordinateurs est

$$P(X \geqslant 2) = 1 - \left(e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!}\right) \simeq 0.9596$$

- Le nombre moyen de micro-ordinateurs vendus chaque jour dans le magasin est égal à 5 puisque $E(X) = \lambda = 5$.

Properties

Si X et Y sont deux variables **indépendantes** suivant des lois de Poisson, $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$, alors leur somme suit aussi une loi de Poisson: $X + Y \sim \mathcal{P}(\lambda + \mu)$.

Approximation d'une loi binomiale

$$\text{Si } n \to \infty \text{ et } p \to 0 \text{ alors } X : \mathcal{B}(n,p) \sim \mathcal{P}(\lambda).$$

Remark

Une bonne approximation est obtenue si $n \geqslant 50$ et $np \leqslant 5$.

Dans ce contexte, la loi de Poisson est souvent utilisée pour modéliser le nombre de succès lorsqu'on répète un très grand nombre de fois une expérience ayant une chance très faible de réussir par une loi de Poisson.

Applications de la loi de Poisson

- Le nombre d'individus dépassant l'âge de 100 ans dans une communauté.
- Le nombre de faux numéros téléphoniques composés en un jour.
- Le nombre de clients pénétrant dans un bureau de poste donné en l'espace d'un jour.
- Le nombre de particules α émises par un matériau radioactif pendant un certain laps de temps.

La v.a. dans ces exemples est répartie de manière approximativement poissonienne car: on approxime par là une variable binomiale.

Loi Géométrique ou de Pascal $\mathcal{G}(p)$

Loi Géométrique $\mathcal{G}(p)$

- ε : "On répéte l'épreuve de Bernoulli jusqu'à avoir le premier succès".
- Exemple:

$$\begin{array}{ccccccccc} \bar{A} & \bar{A} & \bar{A} & \bar{A} & \bar{A} & \dots & \bar{A} & \bar{A} & A \\ E & E & E & E & E & \dots & E & E & S \end{array}$$

- Chaque épreuve a p pour probabilité de succès et $1-p$ pour probabilité d'échec.
- $X =$"le nombre d'épreuves effectuées".

$$\underbrace{E \quad E \quad E \quad E \quad E \quad \dots \quad E \quad E}_{k-1} \quad S$$

- $X(\Omega) = \mathbb{N}^* = \{1, 2, 3, \dots\}$. On dit $X \sim \mathcal{G}(p)$.
- $\forall\, k \in \mathbb{N}^* \quad P(X = k) = (1-p)^{k-1} p$
- Attention: Parfois $X =$ "nombre d'épreuves effectuées avant obtenir le premier succès". Dans ce cas $X(\Omega) = \mathbb{N}$. On dit $X \sim \mathcal{G}(p)$ sur $\mathbb{N}$.
- Cette loi peut servir à modéliser des temps de vie, ou des temps d'attente, lorsque le temps est mesuré de manière discrète (nombre de jours par exemple).
- Série entière : $\sum_{k=0}^{\infty} x^k = 1/(1-x)$ pour $|x| < 1$
- $\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{j=0}^{\infty} (1-p)^j \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{j=0}^{\infty} (1-p)^j = p \frac{1}{1-(1-p)} = 1$

Espérance de loi Géométrique

- $E(X) = \sum_{k=1}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{\infty} k(1-p)^{k-1}$
- Série entière: $\sum_{k=0}^{\infty} x^k = 1/(1-x) \quad \text{pour} \quad |x| < 1$
- Dérivée première de la série entière: $\sum_{k=1}^{\infty} kx^{k-1} = 1/(1-x)^2$
- Donc $E(X) = \frac{p}{[1-(1-p)]^2} = \frac{1}{p}$

En d'autres termes, si des épreuves indépendantes ayant une probabilité p d'obtenir un succès sont réalisés jusqu'à ce que le premier succès se produise, le nombre espéré d'essais nécessaires est égal à $1/p$. Par exemple, le nombre espéré de jets d'un dé équilibré qu'il faut pour obtenir la valeur 1 est 6.

Moments de loi Géométrique

Variance de loi Géométrique

- $V(X) = E(X^2) - E^2(X) = E[X(X-1)] + E(X) - E^2(X)$. Or,

$$E[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)p(1-p)^{k-1}$$
$$= p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2}$$

- Dérivée première de la série entière: $\sum_{k=1}^{\infty} kx^{k-1} = 1/(1-x)^2$
- Dérivée seconde de la série entière: $\sum_{k=2}^{\infty} k(k-1)x^{k-2} = 2/(1-x)^3$
- Donc $E[X(X-1)] = \frac{2p(1-p)}{[1-(1-p)]^3} = \frac{2(1-p)}{p^2}$
- Et alors $V(X) = E[X(X-1)] + E(X) - E^2(X) = \frac{1-p}{p^2}$.

Loi Binomiale Négative $\mathcal{BN}(r, p)$

- ε : "On répéte l'épreuve de Bernoulli jusqu'à obtenir un total de r succès".
- Exemple avec $r = 3$:

$$\begin{matrix} \bar{A} & A & \bar{A} & \bar{A} & \bar{A} & A & \bar{A} & \bar{A} & A \\ E & S & E & E & E & S & E & E & S \end{matrix}$$

- Mais on peut obtenir r succès d'autres façons:

$$S \quad E \quad E \quad E \quad E \quad E \quad S \quad E \quad S$$

$$E \quad E \quad E \quad E \quad S \quad E \quad S \quad E \quad S$$

- Chaque épreuve a p pour probabilité de succès et $1 - p$ pour probabilité d'échec.
- Désignons $X =$"le nombre d'épreuves nécessaires pour attendre ce résultat".

$$\underbrace{\overbrace{E \quad S \quad E \quad E \quad E \quad S \quad E \quad E}^{r-1 \text{ succès et } k-r \text{ échecs}} \quad S}_{X=k}$$

- $X(\Omega) = \{r, r+1, r+2, \ldots\}$. On dit $X \sim \mathcal{BN}(r, p)$.
- $\forall k \in X(\Omega)$,

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

$\mathcal{G}(p) = \mathcal{BN}(1, p)$

- ε : "On répéte l'épreuve de Bernoulli jusqu'à obtenir un total de r succès".
- Soit,

$$E \quad \dots \quad E \quad S \quad E \quad \dots \quad E \quad S \dots \quad E \dots \quad E \quad S$$

- Soit, Y_1 le nombre d'épreuves nécessaires jusqu'au premier succès, Y_2 le nombre d'épreuves supplémentaires nécessaires pour obtenir un deuxième succès, Y_3 celui menant au 3ème et ainsi de suite.
- Càd,

$$\underbrace{E \quad \dots \quad E \quad S}_{Y_1} \; \underbrace{E \quad \dots \quad E \quad S}_{Y_2} \; \underbrace{\dots}_{\dots} \; \underbrace{E \quad \dots \quad E \quad S}_{Y_r}$$

- Les tirages étants indépendantes et ayant toujours la même probabilité de succès, chacune des variables $Y_1, Y_2, \dots, Y_r$ est géométrique $\mathcal{G}(p)$.
- $X =$"le nombre d'épreuves nécessaires à l'obtention de r succès"$= Y_1 + Y_2 + \dots + Y_r$.
- Donc,

$$E(X) = E(Y_1) + E(Y_2) + \dots + E(Y_r) = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

et

$$V(X) = \sum_{i=1}^{r} V(Y_i) = \frac{r(1-p)}{p^2}$$

car les Y_i sont indépendantes.

Variables Aléatoires Continues

- Précédemment nous avons traité des variables aléatoires discrètes, c'est-à-dire de variables dont l'univers est fini ou infini dénombrable.
- Il existe cependant des variables dont l'univers est infini non dénombrable.
- Exemples:
  - L'heure d'arrivée d'un train à une gare donnée.
  - La durée de vie d'un transistor.

Definition

X est une variable aléatoire continue s'il existe une fonction f non négative définie pour tout $x \in \mathbb{R}$ et vérifiant pour tout ensemble B de nombres réels la propriété

$$P(X \in B) = \int_B f(x)dx$$

La fonction f est appelée densité de probabilité de la variable aléatoire X.

- Tous les problèmes de probabilité relatifs à X peuvent être traités grâce à f.
- Par exemple pour $B = [a, b]$, on obtient:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

Graphiquement, $P(a \leqslant X \leqslant b)$ est l'aire de la surface entre l'axe de x, la courbe correspondante à $f(x)$ et les droites $x = a$ et $x = b$.

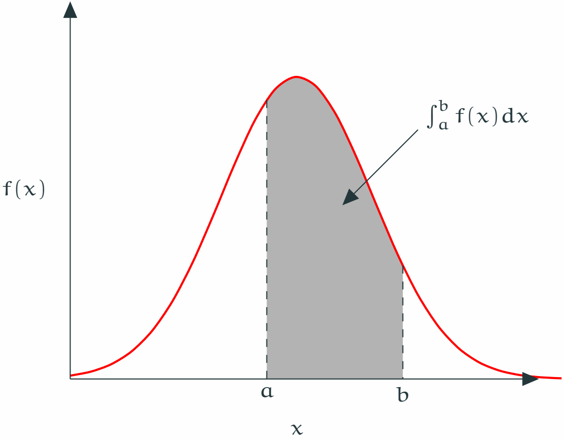


Figure 3: $P(a \leqslant X \leqslant b) =$ *surface grisée*

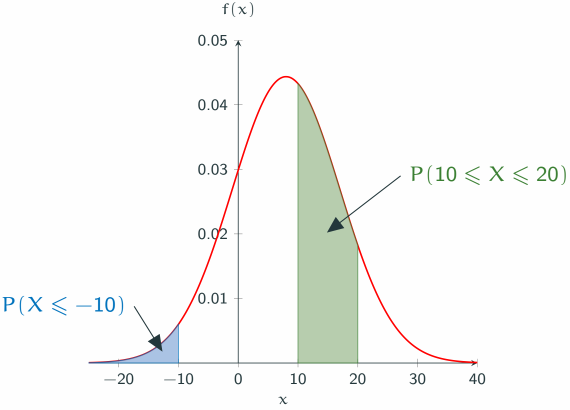
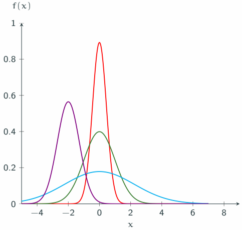
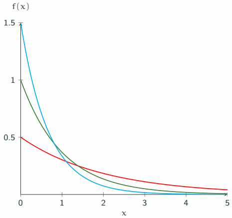
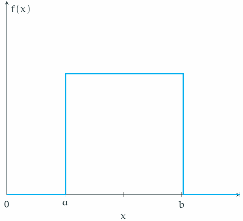


Figure 4: L'aire hachurée correspond à des probabilités. $f(x)$ étant une fonction densité de probabilité.







Properties

Pour toute variable aléatoire continue X de densité f:

- $f(x) \geqslant 0 \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- Comme $P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$, si l'on pose $a = b$ il résulte $P(X = a) = \int_a^a f(x)dx = 0$
- Ceci siginifie que la probabilité qu'une variable aléatoire continue prenne une valeur isolée fixe est toujours nulle.

Exemple

Soit X la variable aléatoire réelle de densité de probabilité

$$f(x) = \begin{cases} kx & \text{si} \quad 0 \leqslant x \leqslant 5 \\ 0 & \text{sinon} \end{cases}$$

1. Calculer k.
2. Calculer: $P(1 \leqslant X \leqslant 3), P(2 \leqslant X \leqslant 4)$ et $P(X < 3)$.

Exemple

Soit X une variable aléatoire réelle continue ayant pour densité de probabilité

$$f(x) = \begin{cases} \frac{1}{6}x + k & \text{si} \quad 0 \leqslant x \leqslant 3 \\ 0 & \text{sinon} \end{cases}$$

1. Calculer k.
2. Calculer $P(1 \leqslant X \leqslant 2)$

Fonction de répartition d'une v.a.c

Definition

Si comme pour les variables aléatoires discrètes, on définit la fonction de répartition de X par:

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F_X(a) = P(X \leqslant a) \end{aligned}$$

alors la relation entre la fonction de répartition F_X et la fonction densité de probabilité $f(x)$ est la suivante:

$$\forall \quad a \in \mathbb{R} \quad F_X(a) = P(X \leqslant a) = \int_{-\infty}^{a} f(x)dx$$

Properties

Pour une variable aléatoire continue X:

- $F'_X(x) = \frac{d}{dx} F_X(x) = f(x)$.
- Pour tous réels $a \leqslant b$,

$$\begin{aligned} P(a < X < b) &= P(a < X \leqslant b) = P(a \leqslant X < b) \\ &= P(a \leqslant X \leqslant b) = F_X(b) - F_X(a) = \int_a^b f(x)dx \end{aligned}$$

La fonction de répartition correspond aux probabilités cumulées associées à la variable aléatoire continue sur l'intervalle d'étude.

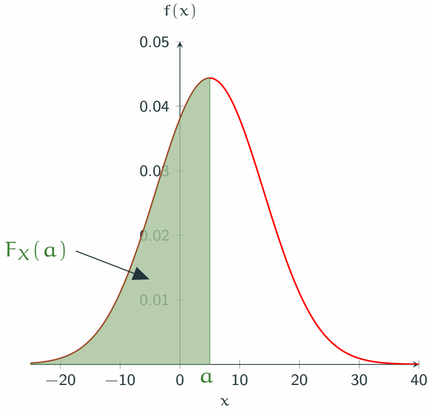


Figure 5: L'aire hachurée en vert sous la courbe de la fonction densité de probabilité correspond à la probabilité $P(X < a) = F_X(a)$ *et vaut* $0,5$ *car ceci correspond exactement à la moitié de l'aire totale sous la courbe.*

Properties

Les propriétés associées à la fonction de répartition sont les suivantes:

1. F_X est continue sur $\mathbb{R}$, dérivable en tout point où f est continue.
2. F_X est croissante sur $\mathbb{R}$.
3. F_X est à valeurs dans $[0, 1]$.
4. $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.

Exemple

Soit X et Y deux variables aléatoires réelles de densités de probabilité

$$f_X(x) = \begin{cases} kx & \text{si} \quad 0 \leqslant x \leqslant 5 \\ 0 & \text{sinon} \end{cases}$$

et

$$f_Y(y) = \begin{cases} \frac{1}{6}y + k & \text{si} \quad 0 \leqslant y \leqslant 3 \\ 0 & \text{sinon} \end{cases}$$

Calculer $F_X(a)$ et $F_Y(a)$ pour tout $a \in \mathbb{R}$.

Fonction d'une variable aléatoire continue

Fonction d'une variable aléatoire continue

- Soit X une variable aléatoire continue de densité f_X et de fonction de répartition F_X.
- Soit h une fonction continue définie sur $X(\Omega)$, alors $Y = h(X)$ est une variable aléatoire.
- Pour déterminer la densité de Y, notée f_Y, on commence par calculer la fonction de répartition de Y, notée F_Y, ensuite nous dérivons pour déterminer f_Y.

Calcul de densités

Soit X une variable aléatoire continue de densité f_X et de fonction de répartition F_X. Calculer la densité des variables aléatoires suivantes:

- $Y = aX + b$
- $Z = X^2$
- $T = e^X$

Example

Soit la v.a.c X ayant la fonction de densité

$$f_X(x) = 2x \times \mathbb{1}_{[0,1]}(x)$$

Déterminer la densité de: $Y = 3X + 1$, $Z = X^2$ et $T = e^X$.

Moments des variables aléatoires continues

Espérance d'une v.a.c

Definition

Si X est une variable aléatoire absolument continue de densité f, on appelle espérance de X, le réel $E(X)$, défini par:

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

si cette intégrale est convergente.

Les propriétés de l'espérance d'une variable aléatoire continue sont les mêmes que pour une variable aléatoire discrète.

Properties

Soit X une variable aléatoire continue,

- $E(aX + b) = aE(X) + b \qquad a \geqslant 0$ et $b \in \mathbb{R}$.
- Si $X \geqslant 0$ alors $E(X) \geqslant 0$.
- Si X et Y sont deux variables aléatoires définies sur un même univers Ω alors

$$E(X + Y) = E(X) + E(Y)$$

Theorem

Si X est une variable aléatoire de densité $f(x)$, alors pour toute fonction réelle g on aura

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\,dx$$

Example

Soit la v.a.c X ayant la fonction de densité

$$f_X(x) = \begin{cases} 2x & \text{si} \quad 0 \leqslant x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Calculer l'espérance des variables aléatoires $Y = 3X + 1$, $Z = X^2$ et $T = e^X$.

La variance d'une variable aléatoire $V(X)$ est l'espérance mathématique du carré de l'écart à l'espérance mathématique. C'est un paramètre de dispersion qui correspond au moment centré d'ordre 2 de la variable aléatoire X.

Definition

Si X est une variable aléatoire ayant une espérance $E(X)$, on appelle variance de X le réel

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

Si X est une variable aléatoire continue, on calcule $E(X^2)$ en utilisant le théorème de transfert,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx$$

Example

Calculer la variance de la variable aléatoire X définie dans l'exemple précédent.

Variance et écart-type d'une v.a.c

Properties

Si X est une variable aléatoire admettant une variance alors:

- $V(X) \geqslant 0$, si elle existe.
- $\forall\ a \in \mathbb{R}, V(aX) = a^2 V(X)$
- $\forall\ (a, b) \in \mathbb{R}, V(aX + b) = a^2 V(X)$
- Si X et Y sont deux variables aléatoires indépendantes, $V(X + Y) = V(X) + V(Y)$

Definition

Si X est une variable aléatoire ayant une variance $V(X)$, on appelle écart-type de X, le réel:

$$\sigma_X = \sqrt{V(X)}$$

Lois Usuelles de Variables Aléatoires Continues

Définition

La variable aléatoire X suit une loi uniforme sur le segment $[a, b]$ avec $a < b$ si sa densité de probabilité est donnée par

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad x \in [a, b] \\ 0 & \text{si} \quad x \notin [a, b] \end{cases} = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$$

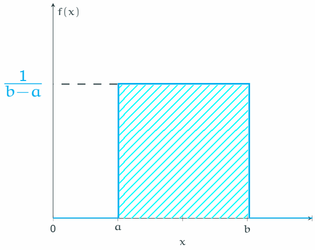


Figure 6: Fonction de densité de $\mathcal{U}([a, b])$

Loi Uniforme $\mathcal{U}(a, b)$

- La *fonction de répartition* associée à la loi uniforme continue est

$$F_X(x) = \begin{cases} 0 & \text{si} \quad x < a \\ \frac{x-a}{b-a} & \text{si} \quad a \leqslant x \leqslant b \\ 1 & \text{si} \quad x > b \end{cases}$$

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

Loi exponentielle $\mathcal{E}(\lambda)$

Définition

On dit qu'une variable aléatoire X est **exponentielle** (ou suit la loi exponentielle) de paramètre λ si sa densité est donnée par

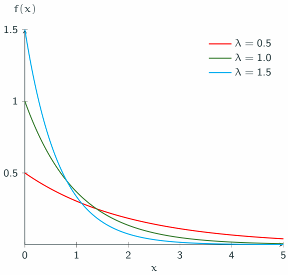
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geqslant 0 \\ 0 & \text{si} \quad x < 0 \end{cases} = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x)$$

On dit $X \sim \mathcal{E}(\lambda)$

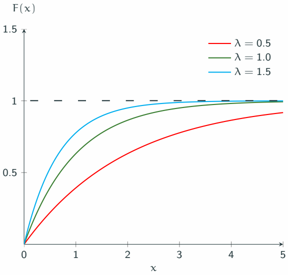
- La fonction de répartition F d'une variable aléatoire exponentielle est donnée par

$$\text{Si } x \geqslant 0 \quad F(x) = P(X \leqslant x) = 1 - e^{-\lambda x}$$

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



(a) Représentation graphique de la densité d'une loi exponentielle



(b) Représentation graphique de la fonction de répartition d'une loi exponentielle

Cas d'utilisations de la loi exponentielle:

- Représenter le temps d'attente avant l'arrivée d'un événement spécifié.
- Modéliser la durée de vie d'un phénomène **sans mémoire**, ou sans vieillissement, ou sans usure.
- On dit qu'une variable aléatoire non négative X est *sans mémoire* lorsque

$$P(X > t + h | X > t) = P(X > h) \qquad \forall \quad t, h \geqslant 0$$

- Par exemple, la durée de vie de la radioactivité ou d'un composant électronique.

Loi Normale ou de Laplace-Gauss

$\mathcal{N}(\mu, \sigma^2)$

Définition

Une variable aléatoire X est dite **normale** avec paramètres μ et σ^2 si la densité de X est donnée par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \forall\, x \in \mathbb{R}$$

Avec $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^+$. On dit que $X \sim \mathcal{N}(\mu, \sigma^2)$.

Moments de la loi Normale $\mathcal{N}(\mu, \sigma^2)$

- $E(X) = \mu$
- $V(X) = \sigma^2$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \forall\, x \in \mathbb{R}$$

- La fonction f est paire autour d'un axe de symétrie $x = \mu$ car $f(x+\mu) = f(\mu - x)$.
- $f'(x) = 0$ pour $x = \mu$, $f'(x) < 0$ pour $x < \mu$ et $f'(x) > 0$ pour $x > \mu$

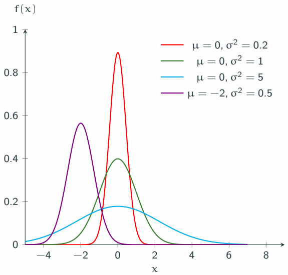


Figure 8: ***Remarque:*** *Le paramètre μ représente l'axe de symétrie et σ le degré d'aplatissement de la courbe de la loi normale dont la forme est celle d'une courbe en cloche.*

Théorème

- $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$
- $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
- X_1 et X_2 sont indépendantes.

Alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Loi Normale centrée réduite $\mathcal{N}(0, 1)$

Loi Normale centrée réduite $\mathcal{N}(0,1)$

Définition

Une variable aléatoire continue X suit une **loi normale centrée réduite** si sa densité de probabilité est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \forall\, x \in \mathbb{R}$$

On dit $X \sim \mathcal{N}(0,1)$.

Moments de la loi Normale $\mathcal{N}(0,1)$

- $E(X) = 0$
- $V(X) = 1$

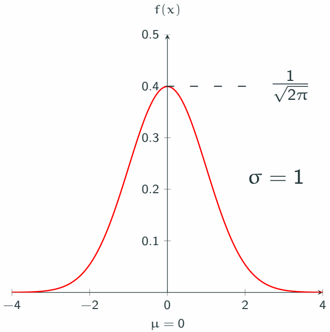


Figure 9: Densité d'une loi normale centrée réduite $\mathcal{N}(0,1)$.

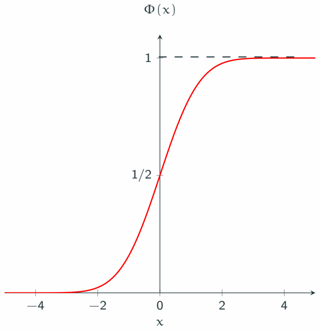


Figure 10: Fonction de répartition de $\mathcal{N}(0,1)$

Relation entre loi normale et loi normale centrée réduite

Relation entre loi normale et loi normale centrée réduite

Théorème

Si X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, alors $Z = \frac{X-\mu}{\sigma}$ est une variable centrée réduite qui suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

La fonction de répartition de la loi normale centrée réduite permet d'obtenir les probabilités associées à toutes variables aléatoires normales $\mathcal{N}(\mu, \sigma^2)$ après transformation en variable centrée réduite.

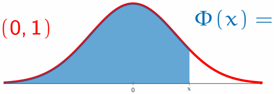
Définition

On appelle fonction Φ, la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$, telle que

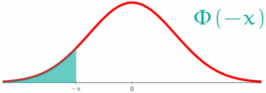
$$\forall\, x \in \mathbb{R} \quad \Phi(x) = P(X \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} f(t)\,dt$$

Fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$

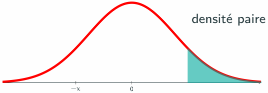
$X \sim \mathcal{N}(0,1)$



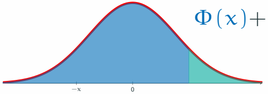
$\Phi(x) = P(X \leqslant x)$



$\Phi(-x) = P(X \leqslant -x)$



densité paire $\Rightarrow$ courbe symétrique % $x = 0$



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

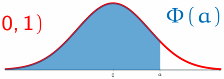
Propriétés de Φ

Les propriétés associées à la fonction de répartition Φ sont:

1. Φ est croissante, continue et dérivable sur $\mathbb{R}$ et vérifie:
   $\lim_{x \to -\infty} \Phi(x) = 0$ et $\lim_{x \to \infty} \Phi(x) = 1$
2. $\forall\, x \in \mathbb{R} \quad \Phi(x) + \Phi(-x) = 1$
3. $\forall\, x \in \mathbb{R} \quad \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

Table de la loi normale centrée réduite $\mathcal{N}(0,1)$

$X \sim \mathcal{N}(0,1)$

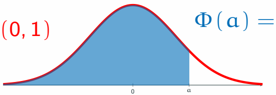


$\Phi(a) = P(X \leqslant a)$

a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Table de la loi normale centrée réduite $\mathcal{N}(0,1)$

$X \sim \mathcal{N}(0,1)$



$\Phi(a) = P(X \leqslant a)$

a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Par exemple, pour $x = 1.23$ (intersection de la ligne 1.2 et de la colonne 0.03), on obtient : $\Phi(1.23) \approx$ 0.8907.

Calcul des probabilités d'une loi normale

Exemple 1

Soit X une variable aléatoire normale centrée réduite. Calculer:

1. $P(X > 2)$
2. $P(2 < X < 5)$

Exemple 2

Soit X une variable aléatoire normale de paramètres $\mu = 3$ et $\sigma^2 = 4$. Calculer:

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 4)$

Approximation normale d'une répartition binomiale

Théorème de Moivre Laplace

On suppose que pour tout n, X_n suit la loi binomiale $\mathcal{B}(n,p)$ avec $p \in]0,1[$.

Alors la variable $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ converge en loi vers une loi normale centrée réduite $\mathcal{N}(0,1)$.

- Ce résultat a été progressivement généralisé par Laplace, Gauss et d'autres pour devenir le théorème actuellement connu comme théorème centrale limite qui est un des deux résultats les plus importants de la théorie de probabilités.
- En pratique de très nombreux phénomènes aléatoires suivent approximativement une distribution normale.

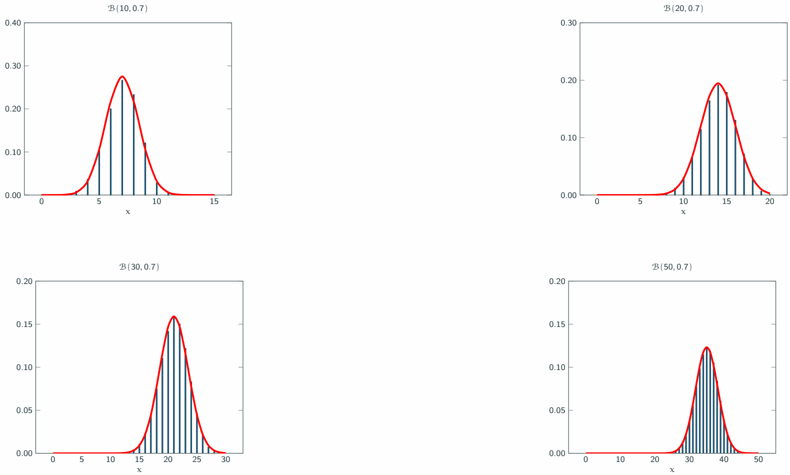


Figure 11: *La loi de probabilité d'une variable aléatoire $\mathcal{B}(n, p)$ devient de plus en plus "normale" à mesure que n augmente.*

Lois déduites de la loi normale

Loi de χ^2 (Khi-deux) de Pearson

Définition

Soit $X_1, X_2, \ldots, X_n$, n variables **normales centrées réduites**, et Y la variable aléatoire définie par

$$Y = X_1^2 + X_2^2 + \ldots + X_i^2 + \ldots + X_n^2 = \sum_{i=1}^{n} X_i^2$$

On dit que Y suit la loi de χ^2 (ou loi de Pearson) à n degrés de liberté, $Y \sim \chi^2(n)$

- La loi de χ^2 trouve de nombreuses applications dans le cadre de la comparaison de proportions, des tests de conformité d'une distribution observée à une distribution théorique et le test d'indépendance de deux caractères qualitatifs. Ce sont les *tests du khi-deux*.
- Remarque: Si $n = 1$, la variable du χ^2 correspond au carré d'une variable normale centrée réduite $\mathcal{N}(0,1)$.

Loi de Student $St(n)$

Définition

Soit U une variable aléatoire suivant une loi **normale centrée réduite** $\mathcal{N}(0,1)$ et V une variable aléatoire suivant une loi de $\chi^2(n)$, U et V étant indépendantes, on dit alors que $T_n = \frac{U}{\sqrt{\frac{V}{n}}}$ suit une **loi de Student** à n degrés de liberté. $T_n \sim St(n)$

- La loi de Student est utilisée lors des tests de comparaison de paramètres comme la moyenne et dans l'estimation de paramètres de la population à partir de données sur un échantillon (Test de Student).

Loi de Fisher-Snedecor $\mathcal{F}(n, m)$

Définition

Soit U et V deux variables aléatoires indépendantes suivant une loi de χ^2 respectivement à n et m degrés de liberté.

On dit que $F = \frac{U/n}{V/m}$ suit une loi de Fisher-Snedecor à (n, m) degrés de liberté. $F \sim \mathcal{F}(n, m)$

- La loi de Fisher-Snedecor est utilisée pour comparer deux variances observées et sert surtout dans les très nombreux tests d'analyse de variance et de covariance.

- A visual introduction to probability and statistics: http://students.brown.edu/seeing-theory/
- Distribution Calculator: https://gallery.shinyapps.io/dist_calc/

Couple de variables aléatoires

Couple de v.a. discrètes

- $X(\Omega) = \{x_1, x_2, \dots, x_l\}$, $Y(\Omega) = \{y_1, y_2, \dots, y_k\}$
- Probabilité conjointe:
  $p_{ij} = P(X = x_i; Y = y_j)$, $\quad p_{ij} \geqslant 0 \quad$ et $\quad \sum_{i=1}^{l} \sum_{j=1}^{k} p_{ij} = 1$
- Loi marginale de X:
  $p_{i.} = P(X = x_i) = \sum_{j=1}^{k} p_{ij} \qquad \forall\, i = 1, 2, \dots, l$
- Loi marginale de Y:
  $p_{.j} = P(Y = y_j) = \sum_{i=1}^{l} p_{ij} \qquad \forall\, j = 1, 2, \dots, k$
- Loi conditionnelle de X sachant $Y = y_j$:
  $p_{i/j} = P(X = x_i / Y = y_j) = \frac{P(X=x_i; Y=y_j)}{P(Y=y_j)} = \frac{p_{ij}}{p_{.j}} \qquad \forall i = 1, 2, \dots, l$
- Indépendance:
  $P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \qquad \forall\, i = 1, 2, \dots, l$ et $j = 1, 2, \dots, k$

Densité conjointe

On dit que (X, Y) est un couple aléatoire continu s'il existe une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ telle que pour tout $D \subseteq \mathbb{R}$ on a

$$P\{(X, Y) \in D\} = \iint_{(x,y)\in D} f(x, y)dxdy$$

- On a la condition de normalité $\iint_{\mathbb{R}^2} f(x, y)dxdy = 1$
- Notons par A et B deux ensembles de nombres réels. En définissant $D = \{(x, y) : x \in A, y \in B\}$, on obtient

$$P(X \in A, Y \in B) = \int_A \int_B f(x, y)dxdy$$

Fonction de répartition

La fonction de répartition du (X, Y) est définie par

$$F(a, b) = P(X \leqslant a, Y \leqslant b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)\,dx\,dy$$

- f est le dérivé de F: $f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$

Densités marginales

- Densité marginale de X:

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

- Densité marginale de Y:

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

Espérance:

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

Exemple: $E(XY) = \iint_{\mathbb{R}^2} xyf(x,y)dxdy$

Rappel: $cov(X,Y) = E(XY) - E(X)E(Y)$.

Indépendance:

Les v.a. X et Y sont indépendantes ssi $\forall (x,y) \in \mathbb{R}^2$ on a

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

Distribution conditionnelle:

Densité conditionnelle de X, sous la condition $Y = y$:

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

Exemple 1

Soit (X, Y) un couple aléatoire continu de densité

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & \text{si} \quad x > 0,\ y > 0 \\ 0 & \text{sinon} \end{cases}$$

Montrer que:

1. f est une densité.
2. $P(X > 1, Y < 1) = e^{-1}(1 - e^{-2})$
3. $P(X < a) = 1 - e^{-a}$
4. $P(X < Y) = 1/3$
5. $E(XY) = 1/2$
6. X et Y sont indépendantes.

Exemple 2

Soit (X, Y) un couple aléatoire continu de densité

$$f(x,y) = \begin{cases} axy^2 & \text{si} \quad 0 \leqslant x \leqslant y \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

1. Trouver la constante a.
2. Trouver les densités marginales de X et Y.
3. X et Y sont elles indépendantes?

Exemple 3

Supposons que X et Y aient pour densité conjointe

$$f(x, y) = \begin{cases} \frac{1}{y} e^{-x/y} e^{-y} & \text{si} \quad x > 0, \ y > 0 \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer la densité conditionnelle de X lorsque $Y = y$.
2. Calculer $P(X > 1 | Y = y)$