Chapitre 1. Random Variables

Prof. REMITA Mohamed Riad

National School of Artificial Intelligence.

2024-2025

Definitions and notations

We have a probabilized space $(\Omega, \mathcal{A}, \mathbb{P})$ and a probabilizable space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Definition

We call a real random variable, noted r.r.v. any application X of $(\Omega, \mathcal{A})$ in $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that:
$\forall B \in \mathcal{B}_{\mathbb{R}}, X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$.

B can be presented in several forms

- If $B =]a, b] : X^{-1}(B) = \{a < X \leq b\}$
- If $B = \{a\} : X^{-1}(B) = \{X = a\}$
- If $B = [a, +\infty[: X^{-1}(B) = \{X \geq a\}$
- If $B =]-\infty, a] : X^{-1}(B) = \{X \leq a\}$

Definitions and notations

Since the Borel $\sigma-$algebra is generated by all the intervals $]-\infty, x]$, then a random variable can be defined by the following definition:

Definition

The application X of $(\Omega, \mathcal{A})$ in $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is a real random variable if for all $x \in \mathbb{R}$ and $\forall B \in \mathcal{B}_{\mathbb{R}}$ the subset $A_x = X^{-1}(]-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$.

Example

We throw two coins and let X be the number of tails obtained. We know that $\Omega = \{(F, F); (F, P); (P, F); (P, P)\}$ and the values of X are $\{0, 1, 2\}$.

1. Show that X is a random variable on Ω endowed with the algebra $\mathcal{P}(\Omega)$.
2. Show that X is not a random variable on Ω endowed with the algebra $\mathcal{A}_1 = \{\Omega, \emptyset, \{(F, F)\}, \{(F, F); (F, P); (P, F)\}\}$.

Definitions and notations

Let Ω be the space of trials associated to a Bernoulli random experiment and let $I_A(\cdot)$ be the function from Ω to $\{0,1\}$ defined by

$$I_A(\cdot) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases}.$$

$I_A(\cdot)$ is called indicator function (or Dirac measure) of the event A. We can show easily that $I_A(\cdot)$ is a random variable for the algebra $\mathcal{A}_{I_A(\cdot)} = \{\Omega, \emptyset, A, \overline{A}\}$.

Properties. The indicator function $I_A(\cdot)$ satisfies the following properties:

1. $I_A(\omega) = 1 - I_{\overline{A}}(\omega), \forall A \in \mathcal{A}$,
2. $I_{\cap A_i}(\omega) = \prod_i I_{A_i}(\omega), \forall A_i \in \mathcal{A}$,
3. $\mathbb{P}(I_A(\omega) = 1) = \mathbb{P}(A), \forall A \in \mathcal{A}$,
4. $\mathbb{P}(I_A(\omega) = 0) = 1 - \mathbb{P}(A) = \mathbb{P}(\overline{A}), \forall A \in \mathcal{A}$.

Induced probability

Theorem

Let X be a r.r.v. defined on probabilized space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a probabilizable space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The application $\mathbb{P}_X$ of $\mathcal{B}_{\mathbb{R}}$ in $\mathbb{R}$ defined by $\mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(B)\right)$, is a probability on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Remark

The definition is due to the existence of $\mathbb{P}$ on $(\Omega, \mathcal{A})$, hence the notion of induced probability.

Cumulative distribution function of a random variable

Definition

The cumulative ditribution function of a r.r.v. is the function F or F_X defined by:

$$F\left(x\right) = F_X\left(x\right) = \mathbb{P}\left(X \leq x\right).$$

Properties of a cumulative distribution function

Definition

A sequence of events $\left(A_n\right)_{n\geq 1}$ is increasing (resp. decreasing) if $A_n \subset A_{n+1}$ (resp. $A_{n+1} \subset A_n$) for all $n \geq 1$.
$\left(A_n\right)_{n\geq 1}$ is said to be monotonic if it is increasing or decreasing. In this case we put $\lim\limits_{n\to\infty} A_n = \bigcup\limits_{n\geq 1} A_n$ if it is increasing (resp. $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{n\geq 1} A_n$ if it is decreasing).

Cumulative distribution function of a random variable

Remark

$\lim_{n\to\infty} A_n$ exists if and only if the sequence $(A_n)_{n\geq 1}$ is monotonic.

Lemma

(Property of the continuity of $\mathbb{P}$)
If $(A_n)_{n\geq 1}$ is a monotonic sequence of events, then we have:

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right).$$

Cumulative distribution function of a random variable

Theorem

If F is the cumulative distibution function of X then

1. $\forall x \in \mathbb{R}\ 0 \leq F(x) \leq 1$;
2. F *is an increasing function;*
3. F *is right continuous;*
4. $\lim_{x \to \infty} F(x) = 1$ *and* $\lim_{x \to -\infty} F(x) = 0$.

Support of a real random variable

We call the support of an r.v. X the set $X(\Omega)$. This support comes in several forms:

- If $X(\Omega)$ is finite or infinite countable X is said to be a discrete (or discontinuous) random variable, denoted d.r.v.
- If $X(\Omega)$ is infinite uncountable X is said to be a continuous random variable, denoted c.r.v.
  Moreover a c.r.v. is said to be absolutely continuous if it admits a continuous and derivable distribution function (except possibly at some points).

Discrete random variables

Definition

The random variable X is said to be discrete if it takes a finite or infinite countable number of values.

Notation: When the r.v. X takes the value x we write $\{X = x\}$ to describe the event $\{\omega \in \Omega, X(\omega) = x\}$.

Probability distribution of a discrete random variable

Definition

Let X be a d.r.v. one calls probability distribution or mass function of the r.v. X the application

$$\begin{aligned} p \quad : \quad & \mathbb{R} \longrightarrow [0,1] \\ & x \longmapsto p(x) = \mathbb{P}(X = x). \end{aligned}$$

Properties:

1.$\forall x \in \mathbb{R}, p(x) \geq 0;$

2. $\sum_{x \in \mathbb{R}} p(x) = 1.$

Discrete random variables

Cumulative distribution function

1. If X is a discrete r.v. then

$$F_X(x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i) = \sum_{x_i \leq x} p(x_i).$$

2. The cumulative distribution function allows to determine the probability law of the r.v. X.

Indeed, $\forall x_j \in X(\Omega)$

$$\mathbb{P}(X = x_j) = \sum_{i=1}^{j} \mathbb{P}(X = x_i) - \sum_{i=1}^{j-1} \mathbb{P}(X = x_i) = F_X(x_j) - F_X(x_{j-1}).$$

Continuous random variables

Definition

A real random variable X is said to be absolutely continuous if its cumulative distribution function $F_X(\cdot)$ satisfy the two following conditions:

1. F_X is continuous on $\mathbb{R}$;
2. F_X is derivable in every point $x \in \mathbb{R}$ except perhaps on a finite set D.

Continuous random variables

Theorem

Let X be an absolutely continuous random variable, with cumulative distribution function F_X, then for any pair $(a, b) \in \mathbb{R}^2$ such that $a < b$, we have

1. $\mathbb{P}(X = a) = 0.$
2. $\mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in]a, b[) = \mathbb{P}(X \in [a, b[) = \mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a).$
3. $\mathbb{P}(X \in]a, \infty[) = \mathbb{P}(X \in [a, \infty[) = 1 - F_X(a).$
4. $\mathbb{P}(X \in]-\infty, b]) = \mathbb{P}(X \in]-\infty, b[) = F_X(b).$

Continuous random variables

Definition

A real random variable X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with cumulative distribution function F_X is said to be absolutely continuous random variable, if there exists a real function f_X satisfying the following conditions:

1. $f_X(x) \geq 0; \forall x \in \mathbb{R}$;
2. f_X is continuous on $\mathbb{R}$, except perhaps on a finite number of points where it has a finite left limit and finite right limit.
3. The integral $\int_{-\infty}^{+\infty} f_X(x)\, dx$ exists and is equal to 1.
4. The cumulative distribution function F_X can be written, for all $x \in \mathbb{R}$ in the form

$$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds.$$

Continuous random variables

Definition

A function f that satisfies the four previous conditions is called a probability density function or distribution function of an absolutely continuous random variable X.

Mathematical expectation and variance

Definition

Let X be a d.r.v. with possible values $x_1, x_2, \cdots$ and mass function $p(x)$. The mathematical expectation of X is

$$\mathbb{E}[X] = \sum_{i \geq 1} x_i p(x_i) = \sum_{i \geq 1} x_i \mathbb{P}(X = x_i)$$

provided that the above serie is absolutely convergent, otherwise we will say that X does not have a mathematical expectation.

Remark

If X has a finite number of values then $\mathbb{E}[X]$ exists.

Mathematical expectation and variance

Definition

Let X be a c.r.v. with distribution function f, the mathematical expectation of X is

$$\mathbb{E}\left[X\right] = \int_{-\infty}^{+\infty} xf\left(x\right)dx$$

provided that the above integral is absolutely convergent, otherwise we will say that X does not have a mathematical expectation.

Definition

Let G be a function of a random variable X, the expectation of $G\left(X\right)$ is given by

$$\mathbb{E}\left[G\left(X\right)\right] = \begin{cases} \sum_{x\in\mathbb{R}} G\left(x\right)p\left(x\right) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{+\infty} G\left(x\right)f\left(x\right)dx & \text{if } X \text{ is continuous} \end{cases}$$

provided that the above serie and integral are absolutely convergent

Mathematical expectation and variance

Theorem

Let X be a random variable, then

1. $\mathbb{E}\left[c\right] = c$ *where c is a constant,*
2. $\mathbb{E}\left[\alpha H\left(X\right) + \beta G\left(X\right)\right] = \alpha\mathbb{E}\left[H\left(X\right)\right] + \beta\mathbb{E}\left[G\left(X\right)\right]$ *where H and G are functions of X and α, β are reals. Provided that the different expectations exist.*

Definition

Let X be a random variable, we call moment of order k $\left(k \in \mathbb{N}\right)$ the following value

$$\mathbb{E}\left[X^k\right] = \begin{cases} \sum_{x\in\mathbb{R}} x^k p\left(x\right) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{+\infty} x^k f\left(x\right) dx & \text{if } X \text{ is continuous} \end{cases}$$

provided that the above serie and integral are absolutely convergent.

Mathematical expectation and variance

Definition

Let X be a random variable, the variance of X, noted σ_X^2 or $Var\left(X\right)$ is

$$\sigma_X^2 = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 .$$

We call standard deviation of X the number

$$\sigma_X = \sqrt{Var\left(X\right)}.$$

If $\mathbb{E}\left[X\right] = 0$ we say that the random variable is centrend.
If $Var\left(X\right) = 1$ we say that the random variable is reduced.

Mathematical expectation and variance

Theorem

Let X be a random variable with expectation $\mathbb{E}[X]$ and variance σ_X^2. If $Y = aX + b$ where a and b are real constants, then

$$\mathbb{E}[Y] = a\mathbb{E}[X] + b \text{ and } \sigma_Y^2 = a^2\sigma_X^2.$$

Definition

The r.v. X has the discrete uniform distribution on the set of real numers $\{x_1, \cdots, x_n\}$ if $\mathbb{P}_X$ is the equiprobability on this set i.e.: $X \in X(\Omega) = \{x_1, \cdots, x_n\}$ and $\forall k \in \{1, \cdots, n\}, \mathbb{P}(X = x_k) = \frac{1}{n}$

We note $X \rightsquigarrow \mathcal{U}(\{x_1, \cdots, x_n\})$.

In the particular case where the discrete uniform distribution is defined on the set $\{1, 2, \cdots, n\}$ we have :

$\mathbb{E}(X) = \frac{n+1}{2}$; $Var(X) = \frac{n^2-1}{12}$.

Definition

The r.v. X follows the Bernoulli distribution of parameter p, $(p \in [0,1])$ if it takes only two values 0 and 1 with $\mathbb{P}(X=1)=p$ and $\mathbb{P}(X=0)=1-p=q$ (with: $p+q=1$).

We note $X \rightsquigarrow \mathcal{B}(p)$.
$\mathbb{E}(X)=p$; $Var(X)=p(1-p)=pq$.

Let be an urn containing:

- white balls W in proportion p;
- red balls R in proportion $q = 1 - p$.

One carries out n successive draws of a ball with delivery. We define the r.v. X as the number of white balls obtained during the n draws (It can take the values: $0, 1, \cdots, n$).

Remark

The r.v. X can be defined as a sum of n independent Bernoulli r.v. $X_1, X_2, \cdots, X_n$ $(X = X_1 + X_2 + \cdots + X_n)$. Such that $\mathbb{P}(X_i = 1) = p$.

Definition

A r.v. X follows a binomial distribution of parameters (n, p) where $n \geq 0$ and $(p \in [0, 1])$ if $X(\Omega) = \{0, 1, \cdots, n\}$ and $\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \forall k = 0, 1, \cdots, n$ (with: $p + q = 1$).

We note $X \rightsquigarrow \mathcal{B}(n, p)$.

$\mathbb{E}(X) = np$; $Var(X) = np(1-p)$.

Discrete probability distribution (finite case)

Hypergeometric distribution

One carries out n successive drawings of a ball, without handing-over, which is the same as when one takes a sample of n balls in only one blow, in an urn containing N balls of two categories:

- N_p white balls W in proportion p;
- N_q red balls R in proportion $q = 1 - p$.

Let be the r.v. X, representing the number of balls W obtained.

Remark

The possible values of X *are* $\max(0, n - N_q) \leq k \leq \min(n, N_p)$

Discrete probability distribution (finite case)

Hypergeometric distribution

Definition

The r.v. X follows the hypergeometric distribution of parameters N, n, p, where $n \leq N$, if $X(\Omega) = \{0, 1, \cdots, n\}$ we have

$$\forall k \in X(\Omega), \mathbb{P}(X = k) = \frac{C_{N_p}^{k} C_{N_q}^{n-k}}{C_N^n}$$

We note $X \rightsquigarrow \mathcal{H}(N, n, p)$, with $p = \frac{N_p}{N}, p + q = 1$.

$\mathbb{E}(X) = np$; $Var(X) = npq\frac{N-n}{N-1}$.

Discrete probability distribution (finite case)

Geometric distribution

The geometric distribution is the law of expectation of the first success of a sequence of independent trials each of which has a probability p of success, i.e. $\mathbb{P}(X = k)$ is the probability that the k^{th} trial is the first success.

Definition

A r.v. X follows a geometric distribution of parameter p, where $0 \leq p \leq 1$ if

- $X(\Omega) = \mathbb{N}^*$;
- $\mathbb{P}(X = k) = pq^{k-1}$ with $p + q = 1$.

We note $X \rightsquigarrow \mathcal{G}(p)$. $\mathbb{E}(X) = \frac{1}{p}$; $Var(X) = \frac{q}{p^2}$.

Discrete probability distribution (finite case)

Negative Binomial (Pascal) distribution

If a r.v. represents the number of fails before the r^{th} success of a sequence of independent Bernoulli trials each of which has a probability p of success.

Definition

A r.v. X follows a Negative Binomial distribution of parameters r and p, where $0 \leq p \leq 1$ if

- $X(\Omega) = \mathbb{N}$;
- $\mathbb{P}(X = k) = C_{k+r-1}^{k} p^r \left(1-p\right)^k$ with $p + q = 1$.

We note $X \rightsquigarrow \mathcal{BN}(r, p)$.
$\mathbb{E}(X) = \frac{rq}{p}$; $Var(X) = \frac{rq}{p^2}$.

Discrete probability distribution (infinite case)

Poisson Distribution

We observe the realization of random events in time and space obeying the following conditions:

- The probability of realization in a small period Δt is proportional to Δt.
- It is independent of what has happened previously.

Definition

X follows a Poisson Distribution of parameter $\lambda (\lambda > 0)$, noted $\mathcal{P}(\lambda)$ if its values are in $\mathbb{N}$ and if:

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

We note $X \rightsquigarrow \mathcal{P}(\lambda)$.
$\mathbb{E}(X) = \lambda$; $Var(X) = \lambda$.

Definition

X follows a continuous uniform distribution on $[a, b]$ if it has the following density

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{elswhere} \end{cases}.$$

We note $X \rightsquigarrow \mathcal{U}([a, b])$.
The cumulative distribution function:

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

$\mathbb{E}(X) = \frac{a+b}{2}$; $Var(X) = \frac{(b-a)^2}{12}$.

Definition

X follows an exponential distribution of parameter $\lambda > 0$ if it has the following density

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{elswhere} \end{cases}.$$

We set $X \rightsquigarrow \mathcal{E}(\lambda)$.
The cumulative distribution function:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{elswhere} \end{cases}.$$

$\mathbb{E}(X) = \frac{1}{\lambda};\; Var(X) = \frac{1}{\lambda^2}$.

Continuous probability distribution

Gamma distribution

Definition

X follows a Gamma distribution of parameters $\alpha > 0$ and $\beta > 0$ if it has the following density

$$f(x) = \begin{cases} \frac{\beta^{\alpha} e^{-\beta x} x^{\alpha - 1}}{\Gamma(\alpha)} & \text{if } x \geq 0 \\ 0 & \text{elswhere} \end{cases}.$$

Where $\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha - 1} dt$.

We set $X \rightsquigarrow \mathcal{G}amma(\alpha, \beta)$.

The cumulative distribution function:

$$F(x) = \frac{\beta^{\alpha} \int_0^{x} t^{\alpha - 1} e^{-\beta t} dt}{\Gamma(\alpha)}.$$

$\mathbb{E}(X) = \frac{\alpha}{\beta}; \qquad Var(X) = \frac{\alpha}{\beta^2}.$

Remark: If $\alpha = 1$ we find the exponential distribution with parameter $\lambda = 1$.

Properties of the function $\Gamma(\alpha)$

a. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
b. $\Gamma(1) = 1$.
c. $\Gamma(n) = (n - 1)!$.

Definition

X follows the Gaussian (Normal) distribution of parameters μ and σ if it has the following density

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

We note $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$.
$\mathbb{E}(X) = \mu;\ Var(X) = \sigma^2$.

The standard normal distribution

The random variable $U = \frac{X-\mu}{\sigma}$ follows the normal distributuion $\mathcal{N}(0, 1)$.

Any problem concerning X is reduced to U and we have several tables concerning the standard normal distribution.

Approximations

Approximation of the hypergeometric distribution by a binomial distribution

The hypergeometric distribution can be approximated by the binomial distribution as soon as the size N of the population is large compared with the size n of the sample.

Approximation of the binomial distribution by a Poisson distribution

If n is large and p small enough (in practice if $n \geq 30$ and $p \leq 0,1$ with $np \leq 10$) we can replace the binomial distribution $\mathcal{B}(n, p)$ with the Poisson distribution $\mathcal{P}(np)$, $(\lambda = np)$.

Approximation of the binomial distribution by a normal distribution

If n is large and $\not p$ not too close to 0 and 1 (in practice if $n \geq 30$, $np \geq 5$ and $nq \geq 5$) we can replace the binomial distribution $\mathcal{B}(n, p)$ with the normal distribution $\mathcal{N}(np, \sqrt{npq})$.