Chapter 2. Expectation and limit theorems

Prof. REMITA Mohamed Riad

National School of Artificial Intelligence.

2024-2025

Properties of the expectation

Definition

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probabilised space and $A \in \mathcal{A}$. We say that A is true **almost surely (as),** if for $\mathbb{P}(A) = 1$.

Theorem

Let X and Y be two random variables defined on the probabilised space $(\Omega, \mathcal{A}, \mathbb{P})$. We have the following properties

- $\mathbb{E}[X]$ *is finite if and only if* $\mathbb{E}[|X|]$ *is finite;*
- $|X| \leq Y$ *and* $\mathbb{E}[Y]$ *finite leads to* $\mathbb{E}[X]$ *finite;*
- $-\infty < a \leq X \leq b < \infty \Longrightarrow a \leq \mathbb{E}[X] \leq b;$
- $X = a$ *a.s.* $\Longrightarrow \mathbb{E}[X] = a.$
- $\mathbb{E}[X]$ *finite* $\Longrightarrow |\mathbb{E}[X]| \leq \mathbb{E}[|X|]$. *(due to Jensen inequality: for a convex function h we have $h(\mathbb{E}[X]) \leq \mathbb{E}[h(X)]$).*

Properties of the expectation

Theorem

Let X and Y be two r.v. defined on the probabilised space $(\Omega, \mathcal{A}, \mathbb{P})$. If $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$, we have the following properties:

A. *Linearity*

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
- $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X], (\lambda \in \mathbb{R})$

B. *Monotony*

- $X \geq 0 \Longrightarrow \mathbb{E}[X] \geq 0$;
- $X \geq Y \Longrightarrow \mathbb{E}[X] \geq \mathbb{E}[Y]$;
- $X = Y$ *a.s.* $\Longrightarrow \mathbb{E}[X] = \mathbb{E}[Y]$;

C. *Independence. If X and Y are independent, then $\mathbb{E}[XY]$ is finite and we have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.*

Properties of the expectation

Definitions

For all $r > 0$, we define, if there exists, the moment of order r by

$$m_r = \mathbb{E}(X^r)$$

and the centred moment of order r by

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

Property 1. For all r.v. X, the r.v. $Y = \frac{X - \mathbb{E}(X)}{\sigma_X}$ is centred $(\mathbb{E}[X] = 0)$ and of de variance unity (reduced r.v.) $(Var(X) = 1)$.
2. For all $A \in \mathcal{A}$ we have $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}_A]$.

Properties of the expectation

Proposition (Markov inequality)
Let X be a r.v. For all $a > 0$, we have

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

and more generally, for all $a > 0$ and $r > 0$

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^r)}{a^r}.$$

Proposition (Bienaymé-Tchebychev inequality)
Let X be a r.v. For all $a > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{Var\,(X)}{a^2}.$$

Convergence

Let $(X_n)_{n\geq 1}$ be a sequence of r.v. on the same probabilized space

1. The sequence $(X_n)_{n\geq 1}$ converges in probability to the variable X, noted $X_n \xrightarrow{\mathbb{P}} X$, if

$$\forall \varepsilon > 0, \lim_{n\to +\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1.$$

2. Let F_n the cumulative distribution functions of the sequence $(X_n)_{n\geq 1}$. Suppose that there exists a r.v. X having the cumulative distribution function F. We say that (X_n) converges distribution to X, noted $X_n \xrightarrow{\mathcal{L}} X$, if

$$\lim_{n\to +\infty} |F_n(x) - F(x)| = 0.$$

3. We say that the sequence $(X_n)_{n\geq 1}$ converges almost surely to the r.v. X, noted $X_n \xrightarrow{p.s} X$, if

$$\mathbb{P}(\left\{\omega \in \Omega, \lim_{n\to +\infty} X_n(\omega) = X(\omega)\right\}) = 1.$$

Convergences

4. We say that the sequence $(X_n)_{n\geq 1}$ converges in mean of order r $(r \geq 1)$ to the r.v. X, noted $X_n \xrightarrow{r} X$, if $\mathbb{E}\left[|X_n|^r\right] < \infty$ for all n and
$$\lim_{n\to+\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

Remarques. 1. If $X_n \xrightarrow{1} X$, we say that X_n converges in mean.
2. If $X_n \xrightarrow{2} X$, we say that X_n converges in quadratic mean.

Theorem

The following implications are verified

$$\begin{array}{ccccc} X_n \xrightarrow{p.s} X & \Longrightarrow & X_n \xrightarrow{\mathbb{P}} X & \Longleftarrow & X_n \xrightarrow{s} X \overset{1\leq s<r}{\Longleftarrow} X_n \xrightarrow{r} X \\ & & \Downarrow & & \\ & & X_n \xrightarrow{\mathcal{L}} X & & \end{array}$$

Independent random variables

The r.v. $X_1, X_2, \cdots, X_n$ are independent if for all $B_1, B_2, \cdots, B_n \subset \mathbb{R}$,

$$\mathbb{P}\left(X_1 \in B_1 \cap \cdots \cap X_n \in B_n\right) = \mathbb{P}\left(X_1 \in B_1\right) \cdot \cdots \cdot \mathbb{P}\left(X_n \in B_n\right).$$

Proposition If $X_1, X_2, \cdots, X_n$ are independent r.v., then

1. The r.v. $f_1\left(X_1\right), f_2\left(X_2\right), \cdots, f_n\left(X_n\right)$ are independent for all functions $f_1, f_2, \cdots, f_n$.
2. If the expectations are well defined, then we have
$$\mathbb{E}(X_1 X_2 \cdot \cdot X_n) = \mathbb{E}(X_1)\mathbb{E}(X_2) \cdot \cdot \mathbb{E}(X_n).$$
3. If the variances are well defined, then we have $Cov\left(X_i, X_j\right) = \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] = 0, \forall i \neq j$, thus
$$Var\left(X_1 + X_2 + \cdots + X_n\right) = Var\left(X_1\right) + Var\left(X_2\right) + \cdots + Var\left(X_n\right).$$

The law of large numbers

The weak law of large numbers

Theorem

Let $(X_n)_{n\geq 1}$ be an independent, identically ditributed (i.i.d.) sequence having an expectation m and a variance σ^2. We define the r.v. $\overline{X}_n$ called empirical mean, by $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$, we have

$$\forall \varepsilon > 0, \mathbb{P}(\left|\overline{X}_n - m\right| < \varepsilon) \underset{n\to+\infty}{\longrightarrow} 1.$$

The law of large numbers

The strong law of large numbers

Theorem

Let $(X_n)_{n\geq 1}$ be an independent, identically ditributed (i.i.d.) sequence having an expectation m and a variance σ^2. We have the strong law of large numbers when the empirical mean $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$ converges almost surely to m.

You have to show that $\mathbb{P}(\{\omega \in \Omega, \lim_{n\to+\infty} \overline{X}_n(\omega) = m\}) = 1$.

The central limit theorem

Theorem

Let $(X_n)_{n\geq 1}$ *be an independent, identically-distributed (i.i.d.) a.v. sequence having a second moment. Let*

$$\begin{aligned} \mu &= \mathbb{E}\left[X_1\right], \sigma^2 = Var\left(X_1\right), S_n = \sum_{i=1}^{n} X_i, \\ et\ Y_n &= \frac{S_n - n\mu}{\sigma\sqrt{n}}. \end{aligned}$$

Then, the sequence $(Y_n)_{n\geq 1}$ *converges in law to a normal r.v.* $\mathcal{N}(0,1)$.

Generating Function

Definition

Let X be an integer r.v.. The generating function of the r.v. X with distribution $p_k = \mathbb{P}\left(X = k\right)$, denoted G_X, is the function

$$G_X\left(t\right) = \mathbb{E}\left[t^X\right] = \sum_{k=0}^{\infty} t^k p_k, \forall t \in \left[-1, 1\right].$$

Generating Function

Propriétés.

1. G_X characterize the distribution of X

$$\forall k \in N; \mathbb{P}\left(X = k\right) = \frac{1}{k!} G_X^{(k)}\left(0\right),$$

where $G_X^{(k)}\left(0\right)$ is the derivative of order k at the point $t = 0$.

2. If X and Y have the same generating function, then they have the same distribution.

3. If for $m \geq 1$, the moments of order m of X exist $\left(\mathbb{E}\left[|X|^m\right] < \infty\right)$; we have

$$G_X^{(m)}\left(1\right) = \mathbb{E}\left[X\left(X-1\right)\cdots\left(X-m+1\right)\right].$$

In particular we have

$$\mathbb{E}\left[X\right] = G_X'\left(1\right); Var\left(X\right) = G_X''\left(1\right) + G_X'\left(1\right) - \left(G_X'\left(1\right)\right)^2.$$

Moments generating function (Laplace transform)

Definition

We call moments generating function of the r.v. X, the function M_X defined by

$$M_X\left(t\right) = \mathbb{E}\left[e^{tX}\right].$$

Moments generating function (Laplace transform)

Propriétés.

1. If X is bounded, then M_X is defined and continuous on $\mathbb{R}$.
2. If X have positive values, then M_X is continuous and bounded on $]-\infty, 0]$. In this case, we often change the variable $s = -t$ to obtain the Laplace transform of X,

$$\mathcal{L}_X(s) = M_X(-s) = \mathbb{E}\left[e^{-sX}\right].$$

The function $\mathcal{L}_X$ is continuous and bounded on $[0, \infty[$.

Remark. All the moments of order n can be calculated using the derivatives of this function at point $t = 0$.

Characteristic function

Definition

The characteristic function of the r.v. X is the function φ_X defined by

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right].$$

$\varphi_X(t) = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]$.

Remark. If X is continuous with density f_X then φ_X is the Fourier transform of f_X.

1. If $\varphi_X = \varphi_Y$ then X and Y follow the same distribution.
2. $\varphi_X(0) = 1$ and $\forall t, |\varphi_X(t)| \leq 1$.
3. $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$
   φ_X is real if and only if X is symetric.

Characteristic function

4. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at), \forall (a,b) \in \mathbb{R}^2$.
5. If $\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt < \infty$, then the distribution of X admits the continuous bounded density f_X, given by the Fourier inversion formula
$$f_X(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{itx}\varphi_X(t)\, dt.$$
6. If X admits moments of order k $\left(\mathbb{E}\left[X^k\right] < \infty\right)$, then φ_X is continuously derivable and for $m = 1, \cdots, k$ we have
$$\varphi_X^{(m)}(t) = i^m \mathbb{E}\left[X^m e^{itX}\right] \text{ and } \varphi_X^{(m)}(0) = i^m \mathbb{E}\left[X^m\right].$$
In particular
$$\mathbb{E}[X] = -i\varphi_X'(0) \text{ and } \mathbb{E}\left[X^2\right] = -\varphi_X''(0).$$