

Variables aléatoires et espérance

Cours de Probabilités — Licence 3

Plan du chapitre

- Variables aléatoires

 - Cadre général

 - Variable aléatoire réelle

- Espérance

 - Définition

 - Rappels de théorie de la mesure

 - Théorème de transfert

 - Moments

 - Espérance et queue de distribution

- Synthèse

Variables aléatoires

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(E, \mathcal{E})$ un espace mesurable. Une **variable aléatoire (v.a.)** à valeurs dans E est une application mesurable

$$X: (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E}), \quad \text{c'est-à-dire} \quad \forall B \in \mathcal{E}, X^{-1}(B) \in \mathcal{A}.$$

- Ω décrit le hasard ; X transporte ce hasard dans E ;
- la mesurabilité garantit que l'on peut parler de $\mathbb{P}(X \in B)$ pour $B \in \mathcal{E}$.

Exemple : somme de deux dés

Exemple

On lance deux dés équilibrés : $\Omega = \llbracket 1, 6 \rrbracket^2$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ uniforme, donc $\mathbb{P}(A) = |A|/36$.

On définit

$$X(i, j) = i + j, \quad (i, j) \in \Omega,$$

une v.a. à valeurs dans $\llbracket 1, 12 \rrbracket$ (ou dans $\mathbb{R}$) représentant la somme des deux dés.

Notation

Si $f: E \rightarrow F$ est mesurable, on note $f(X) := f \circ X$: c'est encore une v.a., comme composée de fonctions mesurables.

Définition

La **loi de** X , notée P_X , est la mesure image de $\mathbb{P}$ par X :

$$\forall B \in \mathcal{E}, \quad P_X(B) = \mathbb{P}(X^{-1}(B)).$$

Plus généralement, une **loi** sur $(E, \mathcal{E})$ désigne une mesure de probabilité sur $(E, \mathcal{E})$.

- Connaître P_X suffit à calculer la probabilité de tout événement dépendant de X ;
- à chaque $\omega \in \Omega$ correspond une réalisation $X(\omega)$, et $P_X(B)$ donne la probabilité que cette réalisation tombe dans B .

Notation

On note $\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$, et l'on omet les accolades dans la probabilité :

$$P_X(B) = \mathbb{P}(X \in B).$$

De même : $\{X = x\}$, ou pour X, Y réelles, $\{X < Y\}$, avec $\mathbb{P}(X = x)$, $\mathbb{P}(X < Y)$.

Cette convention d'écriture sera utilisée dans tout le cours.

Exemple : loi de la somme de deux dés

Exemple

Dans l'exemple précédent, la loi de X est

$$P_X = \sum_{k=2}^{12} \mathbb{P}(X = k) \delta_k, \quad \text{avec } \mathbb{P}(X = 5) = \frac{|\{(1, 4), (2, 3), (3, 2), (4, 1)\}|}{36} = \frac{4}{36}.$$

Connaître $P_X(\{k\})$ pour tout $k \in \llbracket 1, 12 \rrbracket$ caractérise P_X (car E est fini, ou dénombrable).

Deux façons de calculer $\mathbb{P}(X \leq 4)$:

Sur Ω , avec $\mathbb{P}$:

$$\mathbb{P}(X \leq 4) = \frac{6}{36} = \frac{1}{6}.$$

Sur E , avec P_X :

$$P_X(\{2, 3, 4\}) = \frac{1 + 2 + 3}{36} = \frac{1}{6}.$$

La loi ne caractérise pas la v.a.

Exemple

Avec X comme ci-dessus, posons $Y = 14 - X$. Alors Y a la même loi que X , mais $Y \neq X$ (on a $X = Y$ ssi la somme des dés vaut 7).

À retenir

De nombreuses v.a. différentes peuvent partager la même loi : la loi ne détermine pas la v.a. elle-même, seulement son comportement probabiliste.

Remarque : le rôle secondaire de Ω

Remarque importante

Dans ce cours, $(\Omega, \mathcal{A}, \mathbb{P})$ est surtout un *prétexte* pour définir des v.a. On s'intéresse en réalité à $(E, \mathcal{E})$ et à P_X ; le choix de Ω est souvent laissé implicite.

Exemple type : « Soit X une v.a. de Poisson de paramètre 3 » suffit pour calculer

$$\mathbb{P}(X \in 2\mathbb{N}) = \sum_{n=0}^{\infty} \frac{e^{-3} 3^{2n}}{(2n)!} = e^{-3} \cosh(3) = \frac{1 + e^{-6}}{2}.$$

- Réciproquement, pour toute loi μ sur $(E, \mathcal{E})$, il existe $(\Omega, \mathcal{A}, \mathbb{P}, X)$ de loi μ : prendre $\Omega = E$, $\mathbb{P} = \mu$, $X = \text{id}$.
- Un espace $(\Omega, \mathcal{A}, \mathbb{P})$ convenable existe donc toujours, mais le préciser n'aide généralement pas.

Définition

Une **variable aléatoire réelle** est une v.a. à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition

Une loi P sur $\mathbb{R}$ est **discrète** si $P = \sum_{i \in \mathcal{I}} p_i \delta_{x_i}$, avec $\mathcal{I}$ fini ou dénombrable non vide, $p_i > 0$. Les x_i sont les **atomes** de P . Une v.a. réelle X est **discrète** si sa loi l'est.

On a alors $p_i = \mathbb{P}(X = x_i)$ et $\sum_i p_i = 1$, et pour toute $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable :

$$\int_{\mathbb{R}} f \, dP = \sum_{i \in \mathcal{I}} p_i f(x_i).$$

Vocabulaire

- X a un **atome** en x si $\mathbb{P}(X = x) > 0$;
- une loi **discrète** est caractérisée par ses atomes ;
- une loi est **diffuse** si elle n'a aucun atome.

Définition

Une loi sur $\mathbb{R}$ est **continue** si elle a une densité p par rapport à Lebesgue :

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad P(B) = \int_B p(x) dx, \quad \text{où } \int_{\mathbb{R}} p(x) dx = 1.$$

On note alors $dP(x) = p(x) dx$. Une v.a. réelle continue admet une densité p_X .

Les lois continues sont un cas particulier de lois diffuses (voir remarque plus loin).

Exemple

Soit X de loi $\frac{x}{2}1_{[0,2]}(x) dx$. Alors

$$\mathbb{P}(X \leq 1)$$

Exemple

Soit X de loi $\frac{x}{2}1_{[0,2]}(x) dx$. Alors

$$\mathbb{P}(X \leq 1) = \int_{-\infty}^1 \frac{x}{2} 1_{[0,2]}(x) dx = \int_0^1 \frac{x}{2} dx = \left[\frac{x^2}{4} \right]_0^1 = \frac{1}{4}.$$

Comme la densité s'annule hors de $[0, 2]$, on a $X \in [0, 2]$ p.s.

Remarque

Il existe des v.a. réelles ni discrètes ni continues, par exemple de loi

$$\frac{1}{2}\delta_0 + \frac{1}{2}\lambda \quad \text{sur } [0, 1].$$

Interprétation : avec probabilité $1/2$ on tire 0, avec probabilité $1/2$ un nombre uniforme dans $[0, 1]$.

- Il existe aussi des lois **diffuses mais non continues** : la mesure uniforme sur l'ensemble de Cantor n'a aucun atome mais n'a pas de densité (le Cantor est de mesure de Lebesgue nulle).

Espérance

Définition

Soit X une v.a. réelle. L'**espérance** de X est

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega),$$

dès que cette intégrale a un sens :

- si $X \geq 0$ (alors $\mathbb{E}[X]$ peut être infinie) ;
- si X est intégrable, i.e. $\mathbb{E}[|X|] < \infty$.

Exemple

Sur $\Omega = \llbracket 1, n \rrbracket$ avec $\mathbb{P}$ uniforme, pour toute v.a. réelle X :

$$\mathbb{E}[X] = \sum_{i=1}^n X(i) \mathbb{P}(\{i\}) = \frac{X(1) + \cdots + X(n)}{n}.$$

L'espérance retrouve ici la **moyenne classique**.

Remarque

Pour $A \in \mathcal{A}$, la fonction indicatrice 1_A est une v.a. et

$$\mathbb{E}[1_A] = \mathbb{P}(A).$$

Proposition (Linéarité)

Soit X, Y des v.a. réelles et $a, b \in \mathbb{R}$.

- Si X, Y intégrables, alors $aX + bY$ est intégrable ;
- si $X, Y, a, b \geq 0$, alors $aX + bY \geq 0$.

Dans les deux cas : $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

Théorèmes de convergence

Rappel : limsup / liminf

Pour $(a_n) \in \overline{\mathbb{R}}^{\mathbb{N}}$,

$$\limsup_n a_n = \lim_n \downarrow \sup_{k \geq n} a_k, \quad \liminf_n a_n = \lim_n \uparrow \inf_{k \geq n} a_k.$$

Théorème

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles.

- **Convergence monotone.** $X_n \geq 0$ et $X_n \uparrow X \Rightarrow \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.
- **Lemme de Fatou.** $X_n \geq 0 \Rightarrow \mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.
- **Convergence dominée.** $X_n \rightarrow X$ p.s., et $\exists Z$ avec $\mathbb{E}[|Z|] < \infty$, $|X_n| \leq Z$ p.s. $\Rightarrow \mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$.

Théorème (Continuité)

Soit X à valeurs dans $(E, \mathcal{E})$, $I \subset \mathbb{R}$ un intervalle et $f: I \times E \rightarrow \mathbb{R}$. On suppose :

- pour tout $t \in I$, $x \mapsto f(t, x)$ est mesurable ;
- p.s., $t \mapsto f(t, X)$ est continue sur I ;
- il existe une v.a. réelle Z avec $\mathbb{E}[|Z|] < \infty$ et $|f(t, X)| \leq Z$ p.s. pour tout $t \in I$.

Alors $t \mapsto \mathbb{E}[f(t, X)]$ est continue sur I .

Théorème (Dérivabilité)

Sous les mêmes hypothèses, en remplaçant la continuité par :

- p.s., $t \mapsto f(t, X)$ est dérivable sur I ;
- il existe une v.a. réelle Z avec $\mathbb{E}[|Z|] < \infty$ et $\left| \frac{\partial f}{\partial t}(t, X) \right| \leq Z$ p.s. pour tout $t \in I$;

alors $t \mapsto \mathbb{E}[f(t, X)]$ est dérivable sur I , de dérivée

$$t \longmapsto \mathbb{E} \left[\frac{\partial f}{\partial t}(t, X) \right].$$

Pour $p \in [1, \infty]$, on pose $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ (et $\|X\|_\infty = \inf\{C > 0 : |X| \leq C \text{ p.s.}\}$).

$$L^p(\Omega, \mathcal{A}, \mathbb{P}) = \{X \text{ v.a. réelle} : \|X\|_p < \infty\} / (\text{égalité p.s.})$$

muni de $\|\cdot\|_p$, est un **espace de Banach**.

Proposition (Hölder)

Soit $p, q \in [1, \infty]$ avec $\frac{1}{p} + \frac{1}{q} = 1$. Alors $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.

Cas $p = q = 2$ (Cauchy–Schwarz) :

$$\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{1/2} \mathbb{E}[Y^2]^{1/2}.$$

Théorème de transfert

Théorème (Transfert)

Soit X à valeurs dans $(E, \mathcal{E})$ et $f: E \rightarrow \overline{\mathbb{R}}$ mesurable. Alors

$$f \in L^1(E, \mathcal{E}, P_X) \iff f(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P}),$$

et si $f \in L^1(E, \mathcal{E}, P_X)$ ou $f \geq 0$:

$$\mathbb{E}[f(X)] = \int_E f(x) dP_X(x).$$

Pourquoi c'est important

C'est l'**outil principal** pour calculer une espérance : il ramène un calcul sur Ω (souvent inconnu) à un calcul sur E , à l'aide de la seule loi P_X .

Exemple

Soit X uniforme sur $[0, \pi]$, densité $p_X(x) = \frac{1}{\pi} 1_{[0, \pi]}(x)$. On veut $\mathbb{E}[\sin(X)]$.

Méthode : déterminer une loi

Méthode

Pour trouver P_X , on exprime $\mathbb{E}[f(X)]$, pour $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable quelconque, sous la forme $\int_{\mathbb{R}} f(x) d\mu(x)$: on conclut alors $P_X = \mu$.

Cas utile : $X = \psi(Y)$ avec Y continue de densité p_Y . Alors

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(\psi(y)) p_Y(y) dy,$$

et un **changement de variable** $x = \psi(y)$ ramène à $\int_{\mathbb{R}} f(x) p(x) dx$: X a densité p .

Si ψ ne prend qu'un nombre fini de valeurs, X est discrète : μ contient alors des masses de Dirac.

Exemple : loi de $X = Y^2$

Exemple

Soit Y de loi $e^{-y}1_{[0,\infty[}(y) dy$. On cherche la loi de $X = Y^2$.

Définition

Le **moment d'ordre** n de X est $\mathbb{E}[X^n]$, dès qu'il est bien défini ($\mathbb{E}[|X|^n] < \infty$ ou $X \geq 0$).

Proposition

Soit $1 \leq p < q \leq \infty$. Alors $L^q(\Omega, \mathcal{A}, \mathbb{P}) \subset L^p(\Omega, \mathcal{A}, \mathbb{P})$.

Conséquence

Si X a un moment d'ordre n fini, alors tous ses moments d'ordre $\leq n$ sont finis.

Attention : propriété spécifique aux mesures *finies*.

Démonstration de l'inclusion $L^q \subset L^p$

Soit $p' = q/p > 1$ et q' tel que $\frac{1}{p'} + \frac{1}{q'} = 1$. Par Hölder appliqué à $|X|^p$ et 1 :

$$\mathbb{E}[|X|^p] = \mathbb{E}[|X|^p \cdot 1] \leq \mathbb{E}\left[(|X|^p)^{p'} \right]^{1/p'} \mathbb{E}[1]^{1/q'} = \mathbb{E}[|X|^q]^{p/q}.$$

Ainsi $X \in L^q \Rightarrow \mathbb{E}[|X|^q] < \infty \Rightarrow \mathbb{E}[|X|^p] < \infty \Rightarrow X \in L^p$.

□

Définition

Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La **variance** de X est

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2],$$

et l'**écart type** de X est $\sigma_X = \sqrt{\text{Var}(X)}$.

Proposition (formule de König–Huygens)

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Preuve. Avec $m = \mathbb{E}[X]$:

$$\text{Var}(X) = \mathbb{E}[(X - m)^2] = \mathbb{E}[X^2] - 2m\mathbb{E}[X] + m^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

□

Proposition

Soit X une v.a. positive. Alors

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > x) \, dx = \int_0^{\infty} \mathbb{P}(X \geq x) \, dx.$$

Corollaire

Pour $X \geq 0$ et $p > 0$: $\mathbb{E}[X^p] = \int_0^{\infty} \mathbb{P}(X > x) \, p x^{p-1} \, dx.$

Synthèse

Ce qu'il faut retenir

- Une v.a. est une application mesurable ; sa loi P_X suffit presque toujours à travailler avec elle.
- Discret, continu, ou ni l'un ni l'autre : la théorie de la mesure unifie ces cas.
- L'espérance est une intégrale : tous les théorèmes de convergence (monotone, Fatou, dominée) s'appliquent.
- Le théorème de transfert est l'outil central pour calculer $\mathbb{E}[f(X)]$ et pour déterminer une loi.
- Variance, moments, inégalité de Hölder : des outils à maîtriser pour la suite du cours.

Prochain chapitre : fonction caractérisant la loi