260

Espérance, Variance, moments d'une variable aléatoire

R. Dubourg
R. Vannier

Dans cette leçon, on considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et des variables aléatoires qui peuvent être discrètes ou à densité, à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^d$.

I Moments d'une variable aléatoire

1. Espérance

Définition 1 : X est une v.a réelle. Alors X est dite intégrable si le réel $\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P}$ est défini. Ce réel est appelé espérance de X.

Rq : si $\mathbb{E}(X) = 0$, X est dite centrée. $X - \mathbb{E}(X)$ est centrée.

Propriété 2 : (linéarité). $\forall a \in \mathbb{R}$, X, Y v.a.

$$\mathbb{E}(aX + Y) = a\,\mathbb{E}(X) + \mathbb{E}(Y)$$

Exemples : $X \sim$ Binomiale $(n, p) \Rightarrow \mathbb{E}(X) = np$

$X \sim$ Géométrique $(p) \Rightarrow \mathbb{E}(X) = 1/p$

$X \sim$ Poisson $(\lambda) \Rightarrow \mathbb{E}(X) = \lambda$

Contre-exemple : la loi de Cauchy, de densité $f_X(t) = \frac{1}{\pi(1+t^2)}$ n'a pas d'espérance.

Théorème 3 (transfert). Si g est une fonction mesurable réelle, alors $\int_\Omega (g \circ X) \, d\mathbb{P} = \int_{\mathbb{R}} g \, d\mathbb{P}_X$ (X à valeurs réelles)

2. Moments d'ordre k.

Définition 4 (moment d'ordre k). $M_X(k) = \mathbb{E}(X^k)$

Rq : le moment d'ordre 1 est l'espérance

le moment d'ordre 2 de la v.a centrée $X - \mathbb{E}(X)$ est sa variance.

Déf 5 (variance)

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \qquad \mathrm{Var}(X) > 0.$$

Propriété 6 : Egalité de Koenig

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

Remarques : . l'écart-type $\sigma = \sqrt{\mathrm{Var}(X)}$ permet d'avoir un ordre de grandeur des variations d'une v.a.

. Une v.a est dite réduite si $\mathrm{Var}(X) = 1$.

Propriétés 7 : . $\forall a \in \mathbb{R}$ $\mathrm{Var}(X + a) = \mathrm{Var}(X)$

. $\forall \alpha \in \mathbb{R}$ $\mathrm{Var}(\alpha X) = \alpha^2 \mathrm{Var}(X)$

. $\mathrm{Var}(X) = 0 \Leftrightarrow X$ est constante p.s.

Définition 8 (covariance).

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big) = \mathbb{E}(XY - \mathbb{E}(X) \cdot \mathbb{E}(Y))$$

Propriétés 9 . $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

. $\mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \mathrm{Cov}(X_i, Y_j)$ (bilinéarité)

$\hookrightarrow \mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)$

Définition 10 : matrice de covariance

$X = (X_1, X_2 \dots X_d)$ un vecteur aléatoire. Sa matrice de covariance est $\mathrm{Cov}(X) = \left(\mathrm{Cov}(X_i, X_j)\right)_{1 \leq i,j \leq d}$

Cette matrice est symétrique, et définie positive si aucun des X_i n'est constante p.s.

Propriété 11 (distinction entre dépendance et corrélation)

Si X et Y sont indépendantes, alors elles sont non-corrélées.

$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$ et $\mathrm{Cov}(X, Y) = 0$. La réciproque est fausse

Contre-exemple 11

	X	Y
ω_1	-1	0
ω_2	0	1
ω_3	1	0

avec $\mathbb{P}(\omega_i) = \frac{1}{3}$ $\forall i$

$\mathrm{Cov}(X, Y) = 0$ alors que X et Y ne sont pas indépendantes.

3 - Inégalités

Inégalité 12 (Markov). Si X est une v.a intégrable positive. $\forall t>0$

$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$ Généralisation $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(|X|^p)}{t^p}$ si $X \in \mathcal{L}^p$

Inégalité 13 (Tchebychev) si $X \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. $\forall t>0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\operatorname{Var} X}{t^2}$$

Application 14 (loi normale) si $X \sim \mathcal{N}(\mu, \sigma)$

$\mathbb{P}(|X-\mu| \leq 2\sigma) \leq \frac{1}{4}$ | Valeur exacte $\simeq 0{,}05$

$\mathbb{P}(|X-\mu| \leq 3\sigma) \leq \frac{1}{9}$ | $\simeq 0{,}003$

int de confiance ⚠ allentri

↳ Ces inégalités, bien que générales sont donc grossières.

Inégalité 15 (Jensen) : si f convexe sur $\mathbb{R}$ et X une v.a. réelle telle que X et $f(X)$ sont intégrables $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$

Inégalité 16 (Hölder) : si $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$, p,q réels conjugués

$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{1/p}\, \mathbb{E}(|Y|^q)^{1/q}$$

Conséquences :
- $p=2$: inégalité de Cauchy-Schwarz.
- $p \longmapsto \mathbb{E}(|X|^p)^{1/p}$ est croissante
- $\mathbb{E}(|\cdot|^p)^{1/p}$ est une norme sur $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ si $p \geq 1$
- $\mathcal{L}^p$ et $\mathcal{L}^q$ sont duaux.
- $\mathcal{L}^2$ est un espace de Hilbert pour le produit scalaire :

$$\langle f, g \rangle = \int_\Omega fg \, d\mathbb{P}$$

II Fonctions caractéristiques et génératrices des moments.

1 - Fonction caractéristique

Définition 17 : X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. La fonction caractéristique de X, ou transformée de Fourier de X est : $\varphi_X : t \in \mathbb{R}^d \longmapsto \mathbb{E}(e^{i\langle t, X\rangle})$

Théorème 18 : si X et Y sont deux variables aléatoires de lois $\mathbb{P}_X$ et $\mathbb{P}_Y$, $\varphi_X = \varphi_Y \iff X$ et Y ont même loi. $\mathbb{P}_X = \mathbb{P}_Y$.

Théorème 19 : lien entre moments et fonction caractéristique

si $\exists n \in \mathbb{N}^*$ tel que $\mathbb{E}|X|^n < \infty$, alors φ_X est n fois dérivable et $\varphi_X^{(k)}(t) = i^k\, \mathbb{E}[X^k e^{itX}]$ $\forall k \leq n$.

De plus, $\varphi_X^{(k)}(0) = i^k\, \mathbb{E}[X^k]$

Réciproquement, si φ_X est k fois dérivable en 0, $k \geq 2$, X admet des moments jusqu'à l'ordre $2\lfloor \frac{k}{2} \rfloor$

DEV 1

Théorème [20] **des moments** : si X et Y sont des v.a à valeurs dans un intervalle borné $[a,b]$

$$\forall k \in \mathbb{N} \quad \mathbb{E}[X^k] = \mathbb{E}[Y^k] \iff X \sim Y.$$

2 - Fonction génératrice des moments.

Définition 21 (transformée de Laplace).

Si X v.a à valeurs dans $\mathbb{R}^d$, la transformée de Laplace est $L_X(t) = \mathbb{E}(e^{\langle t, X\rangle})$, définie pour les valeurs de $t \in \mathbb{R}$ telles que $e^{\langle t, X\rangle}$ est intégrable.

Proposition 22 : si e^{tX} est intégrable dans un intervalle ouvert contenant 0, alors L_X est analytique dans un voisinage de 0 et

$$L_X(t) = \sum_{n \in \mathbb{N}} \frac{t^n}{n!}\, \mathbb{E}[X^n]$$

Remarque : les moments sont les dérivées successives de $L_X(t)$ en 1.

Proposition 23 : la transformée de Laplace, si elle est définie au voisinage de 0, caractérise la loi.

III Théorèmes de convergence.

Définition 24 (convergence en loi)
Soit X une v.a et $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. Alors
$X_n \xrightarrow{\text{loi}} X$ si $\forall \phi : \mathbb{R}\to\mathbb{R}$ continue bornée,
$\mathbb{E}(\phi(X_n)) \longrightarrow \mathbb{E}(\phi(X))$

Théorème 25 (Paul Lévy). si $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X, alors la suite des fonctions caractéristiques $(\varphi_{X_n})_{n\in\mathbb{N}}$ converge uniformément vers φ_X sur tout intervalle fini.

Théorème limite central (26)
$(X_i)_{1\le i\le n}$ une famille de v.a i.i.d de loi $X\in\mathcal{L}^2(\Omega,\mathcal{A},\mathbb{P})$
alors $\dfrac{S_n - n\,\mathbb{E}(X)}{\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, \mathrm{Var}(X))$

Application : (approximation de loi binomiale)
$(X_i)_{1\le i\le n} \sim \frac{1}{n}\mathcal{B}(1,p)$, $S_n = \sum X_i$

$$\lim_{n\to+\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \int_a^b \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$$

Théorème 27 (Bernstein) si $f : [0,1] \longrightarrow \mathbb{R}$ continue.
$$B_n(f,x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}$$
Alors $B_n \longrightarrow f$ uniformément sur $[0,1]$
Conséquence : théorème d'approximation de Weierstrass.

DEV 2

Définition 28 : (convergence $\mathcal{L}^p$)
$(X_n)_{n\in\mathbb{N}}$ une suite de v.a de $\mathcal{L}^p(\Omega,\mathcal{A},\mathbb{P})$ et $X\in\mathcal{L}^p(\Omega,\mathcal{A},\mathbb{P})$
$X_n \xrightarrow{\mathcal{L}^p} X$ si $\mathbb{E}(|X_n - X|^p) \xrightarrow[n\to+\infty]{} 0$

Définition 29 (convergence en probabilité)
$X_n \xrightarrow{\text{proba}} X$ si $\forall \varepsilon > 0$ $\mathbb{P}(|X_n - X| \ge \varepsilon) \xrightarrow[n\to+\infty]{} 0$

Proposition 30
cvg $\mathcal{L}^p$ $\Rightarrow$ cvg en proba $\Rightarrow$ cvg en loi.

Théorème 31 (loi faible des grands nombres) $(X_i)_{i\in\mathbb{N}}$ suite de v.a de même loi que X i.i.d
si $\mathbb{E}[|X|] < +\infty$ alors $\dfrac{S_n}{n} \xrightarrow{\text{proba}} \mathbb{E}[X]$

Définition 32 (convergence presque sûre)
$X_n \xrightarrow{\text{p.s}} X$ $\mathbb{P}\{\omega\in\Omega : \lim_{n\to+\infty} X_n(\omega) = X(\omega)\} = 1$

Proposition 33 : cvg p.s $\Rightarrow$ cvg en proba

Théorème 34 : (loi forte des grands nombres)
$\mathbb{E}(X) < \infty \iff \dfrac{S_n}{n} \xrightarrow{\text{p.s}} \mathbb{E}(X)$ p.s

biblio
Barbe
Ouvrard. Proba 2.
Cottrel. Exo de proba
Ross - Initiation aux proba → mauvais sauf ex.
Ledoux Proba.

Théorème de Weierstrass

démontré à l'aide des polynômes de Bernstein et d'un peu de probas

Romain Dubourg et Rémi Vannier

Théorème 1. *Théorème de Bernstein*

Soit f une fonction réelle continue sur l'intervalle fermé $I = [0,1]$. Pour $n \in \mathbb{N}$, on note B_n le polynôme de Bernstein de degré n associé à f.

$$B_n(x) = \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}x^k(1-x)^k$$

Pour tout $x \in]0,1[$, on se donne une suite de variables aléatoires i.i.d $(X_k) \sim \mathcal{B}(x)$ et on note $S_n = \sum_{k=1}^{n} X_k$. Alors,

$$||B_n - f|| \to 0$$

Lemme 1. *Théorème de transfert discret*

Soit X une v.a discrète, D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Soit g une fonction quelconque de $\overline{D}$ dans $\mathbb{R}$. Alors la v.a $Y = g(X)$ est intégrable ssi

$$\sum_{i \in D} |g(i)| \; \mathbb{P}(X = i) \; < +\infty$$

De plus, si cette somme est finie, on a

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_{i \in D} g(i) \; \mathbb{P}(X = i)$$

Appliquons le théorème de transfert à $g : k \mapsto f\left(\frac{k}{n}\right)$, $X = S_n$, $Y = f\left(\frac{S_n}{n}\right)$, $D = [\![1,n]\!]$

$$\mathbb{E}\left(f\left(\frac{S_n}{n}\right)\right) = \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^k = B_n(x)$$

Lemme 2. *Soit, pour tout $\epsilon > 0$, le réel $\delta(\epsilon)$ défini par :*

$$\delta(\epsilon) = \sup\left\{|f(x) - f(y)|, (x,y) \in [0,1]^2, |x-y| \leq \epsilon\right\}$$

Alors

$$\lim_{\epsilon \to 0} \delta(\epsilon) = 0$$

Démonstration. f est continue sur un compact donc uniformément continue

Démonstration. **Théorème de Bernstein**

On va montrer que :

$$\sup_{x \in [0,1]} |B_n(x) - f(x)| \leq \delta(\epsilon) + \frac{||f||_\infty}{2n\epsilon^2}$$

$$\forall x, B_n(x) - f(x) = \int_\Omega \left[f\left(\frac{S_n}{n}\right) - f(x) \right] d\mathbb{P}$$

En séparant l'intégrale, on obtient :

$$\int_{|\frac{S_n}{n}-x|\leq\epsilon} \left[f\left(\frac{S_n}{n}\right) - f(x) \right] d\mathbb{P} + \int_{|\frac{S_n}{n}-x|>\epsilon} \left[f\left(\frac{S_n}{n}\right) - f(x) \right] d\mathbb{P}$$

Le membre de gauche est majoré par :

$$\delta(\epsilon)\, \mathbb{P}\left(\left| \frac{S_n}{n} - x \right| \leq \epsilon \right) \leq \delta(\epsilon)$$

Le membre de droite est majoré grossièrement par :

$$2\, \mathbb{P}\left(\left| \frac{S_n}{n} - x \right| > \epsilon \right) ||f||_\infty$$

Par l'inégalité de Tchebychev,

$$\mathbb{P}\left(\left| \frac{S_n}{n} - x \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \mathrm{Var}\left(\frac{S_n}{n}\right)$$

Or, S_n suit une loi binomiale, donc $\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{nx(1-x)}{n^2} = \frac{x(1-x)}{n} \leq \frac{1}{4n}$.

D'où :

$$\mathbb{P}\left(\left| \frac{S_n}{n} - x \right| > \epsilon \right) \leq \frac{1}{4n\epsilon^2}$$

$$\forall \epsilon > 0\ \forall n \geq 1, ||B_n - f|| \leq \delta(\epsilon) + \frac{||f||_\infty}{2n\epsilon^2}$$

$$\forall \epsilon > 0\ \limsup_{n\to+\infty} ||B_n - f||_\infty \leq \delta(\epsilon)$$

Grâce au Lemme 2, on a

$$\limsup_{n\to+\infty} ||B_n - f||_\infty = 0$$

La suite des polynômes B_n converge donc uniformément vers f. □

260 (I)

Référence : Probabilités 2, Jean-Yves Ouvrard

Fonction caractéristique et moments :

Lemme : (Lemme de Fatou)
Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n\in\mathbb{N}}$, des éléments de $\mathcal{M}^+$.
On a l'inégalité dans $\overline{\mathbb{R}}_+$:

$$\int_\Omega \liminf_n f_n \, d\mu \leqslant \liminf_n \int_\Omega f_n \, d\mu$$

Théorème : (Lien entre fonction caractéristique et moments)
Soient X une variable aléatoire réelle et φ_X sa fonction caractéristique.
Alors :

(i) Si X admet un moment d'ordre $n \in \mathbb{N}^*$, φ_X est de classe C^n et, $\forall k \in [\![1,n]\!]$, on a :

$$\forall t \in \mathbb{R}, \ \varphi_X^{(k)}(t) = i^k \int_\Omega X^k e^{itX} \, d\mathbb{P}.$$

et, en particulier :

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}[X^k]. \quad (*)$$

(ii) Inversement si φ_X est k fois dérivable en 0 $(k \geqslant 2)$, X admet des moments jusqu'à l'ordre $2\left[\frac{k}{2}\right]$; ils sont donnés par la formule $(*)$.

Preuve :

(i) Supposons que X admette un moment d'ordre $n \in \mathbb{N}$. Soit $k \in \mathbb{N}$. On a : $\frac{d^k}{dt^k} e^{itX} = (iX)^k e^{itX}$. Donc on a :

$$\left| \frac{d^k}{dt^k} e^{itX} \right| \leqslant |X|^k$$

De plus, pour tout $t \in \mathbb{R}$, $\omega \longmapsto e^{itX(\omega)}$ est $\mathbb{P}$-intégrable.
On peut donc appliquer n fois le théorème de dérivation d'une intégrale à paramètre, d'où le résultat.

(ii) Supposons que φ_X soit k fois dérivable en 0 ($k \geq 2$). Montrons par récurrence que X admet des moments jusqu'à l'ordre $2\left[\frac{k}{2}\right]$.
Cas $k=2$: φ_X admet un développement limité de Taylor-Young à l'ordre 2, donc $\lim\limits_{t \to 0} \frac{\varphi_X(t)+\varphi_X(-t)-2}{t^2} = \varphi_X''(0)$.
Comme $\varphi_X(t)+\varphi_X(-t) = 2\,\mathrm{Re}(\varphi_X(t)) = 2\,\mathbb{E}[\cos(tX)]$, on a :

$$\lim_{t \to 0} \mathbb{E}\Big[\underbrace{\frac{1-\cos(tX)}{t^2}}_{\geq 0\ \forall t}\Big] = -\frac{1}{2}\varphi_X''(0)$$

Donc si (t_n) est une suite convergeant vers 0, par le lemme de Fatou :

$$\int_\Omega X^2 d\mathbb{P} = \mathbb{E}\left[2 \liminf_n \frac{1-\cos(t_nX)}{t_n^2}\right] \leq 2 \liminf_n \mathbb{E}\left[\frac{1-\cos(t_nX)}{t_n^2}\right] < +\infty$$

Supposons avoir montré l'existence de tous les moments jusqu'à l'ordre $2\left[\frac{k}{2}\right]-2 = 2(n-1)$. On veut montrer l'existence du moment d'ordre $2\left[\frac{k}{2}\right] = 2n$. Par (i) on a :

$$\varphi_X^{(2(n-1))}(t) + \varphi_X^{(2(n-1))}(-t) = (-1)^{n-1}\, 2\,\mathbb{E}\left[X^{2(n-1)}\cos(tX)\right] \text{ et } \varphi_X^{(2(n-1))}(0) = (-1)^{n-1}\,\mathbb{E}\left[X^{2(n-1)}\right]$$

D'ailleurs, $\varphi_X^{(2(n-1))}$ étant par hypothèse deux fois dérivable en 0, $\varphi_X^{(2(n-1))}$ admet un développement limité de Taylor-Young à l'ordre 2, donc :

$$\lim_{t \to 0} \frac{\varphi_X^{(2(n-1))}(t) + \varphi_X^{(2(n-1))}(-t) - 2\varphi_X^{(2(n-1))}(0)}{t^2} = \varphi_X^{(2n)}(0)$$

De ces trois relations on déduit que :

$$\lim_{t \to 0} \mathbb{E}\left[X^{2(n-1)}\,\frac{1-\cos(tX)}{t^2}\right] = \frac{(-1)^n}{2}\varphi_X^{(2n)}(0)$$

On conclut comme précédemment, avec le lemme de Fatou.