CHAPITRE 31

Variable

Hugo SALOU MP2I

Dernière mise à jour le 15 juin 2022

TABLE DES MATIÈRES

I	Définitions	2
II	Exemples de lois	4
III	Couples de variables aléatoires	6
IV	Familles de variables aléatoires	8
V	Espérance d'une variable aléatoire	10
VI	Variance	12
VII	Covariance (HORS-PROGRAMME)	15
VIII	Loi des grands nombres	18

Première partie

Définitions

Définition: Une variable aléatoire est une application sur un espace probabilisé (Ω, P) dans E où E est un ensemble quelconque :

$$X : \Omega \longrightarrow E.$$

Si $E \subset \mathbb{R}$, on dit que X est réelle.

Si E est un espace vectoriel, on dit que X est un vecteur aléatoire.

Si $E \subset \mathcal{M}_{n,p}(\mathbb{K})$, on dit que X est une matrice aléatoire.

Dans la suite du chapitre, (Ω, P) est un espace probabilisé.

Proposition: Soit $X : \Omega \rightarrow E$ une variable aléatoire avec $E = X(\Omega)$. L'application

$$\begin{aligned} P_X : \mathcal{P}(E) &\longrightarrow [0, 1] \\ A &\longmapsto P(X^{-1}(A)) \end{aligned}$$

est une probabilité sur E . ■

Définition: Soit $X : \Omega \rightarrow E$ une variable aléatoire. P_X est la loi de X .

REMARQUE (Notations):

Soit $X : \Omega \rightarrow E$ une variable aléatoire.

- Soit $A \in \mathcal{P}(E)$. $X^{-1}(A)$ est noté $(X \in A)$. D'où $P_X(A) = P(X \in A)$.
- Soit $x \in E$. $X^{-1}(\{x\})$ est noté $(X = x)$; $P_X(\{x\}) = P(X = x)$.
- Si $E \subset \mathbb{R}$, $X^{-1}(]-\infty, x])$ est noté $(X \leq x)$ et donc $P(X^{-1}(]-\infty, x])) = P(X \leq x)$.
De même avec les autres inégalités strictes et larges.

REMARQUE (Notation):

On note $X \sim Y$ si X et Y suivent la même loi, i.e. $P_X = P_Y$.

Deuxième partie

Exemples de lois

Définition: Soit $X : \Omega \rightarrow E$ une variable aléatoire. On dit que X suit une loi uniforme si P_X est l'équiprobabilité sur $X(\Omega)$. On note alors $X \sim \mathcal{U}(X(\Omega))$.

Définition: Soit $X : \Omega \rightarrow \{0, 1\}$. On note $p = P(X = 1) \in [0, 1]$. On dit que X suit la loi de Bernoulli de paramètre p . On note alors $X \sim \mathcal{B}(p)$.

Définition: Soit $X : \Omega \rightarrow \llbracket 0, n \rrbracket$. On dit que X suit la loi binomiale de paramètres n et p si

$$\forall k \in \llbracket 0, n \rrbracket, P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On note alors $X \sim \mathcal{B}(n, p)$.

Théorème: On considère une expérience aléatoire qui n'a que deux issues possibles : succès ou échec. On répète n fois à l'identique cette expérience, et on note X le nombre de succès. Alors $X \sim \mathcal{B}(n, p)$ où P est la probabilité de succès. $\square$

Troisième partie

Couples de variables aléatoires

Définition: Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ deux variables aléatoires.

La loi du couple (X, Y) est P_Z où

$$\begin{aligned} Z : \Omega &\longrightarrow E \times F \\ \omega &\longmapsto (X(\omega), Y(\omega)) \end{aligned}$$

i.e. c'est la donnée de

$$\forall x \in X(\Omega), \forall y \in Y(\Omega), P(X = x, Y = y) = P((X = x) \cap (Y = y)).$$

Définition: Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$. Les lois marginales du couple (X, Y) sont P_X et P_Y .

Proposition: Avec les notations précédentes,

$$\begin{aligned} \forall x \in X(\Omega), P(X = x) &= \sum_{y \in Y(\Omega)} P(X = x, Y = y); \\ \forall y \in Y(\Omega), P(Y = y) &= \sum_{x \in X(\Omega)} P(X = x, Y = y). \end{aligned}$$

■

REMARQUE:
Si

$$\forall y \in Y(\Omega), P(Y = y) \neq 0$$

alors

$$\forall x \in X(\Omega), P(X = x) = \sum_{y \in Y(\Omega)} P_{(Y=y)}(X = x) P(Y = y).$$

Définition: Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ deux variables aléatoires. On dit que X et Y sont indépendantes si

$$\forall A \subset X(\Omega), \forall B \subset Y(\Omega), P(X \in A, Y \in B) = P(X \in A) P(Y \in B).$$

On note alors $X \perp\!\!\!\perp Y$.

Proposition: Avec les notations précédentes,

$$X \perp\!\!\!\perp Y \iff \forall x \in X(\Omega), \forall y \in Y(\Omega), P(X = x, Y = y) = P(X = x)P(Y = y).$$

■

Quatrième partie

Familles de variables aléatoires

Définition: Soit $(X_i)_{i \in I} \in \prod_{i \in I} E_i^\Omega$ une famille de variables aléatoires. On dit que ces variables sont indépendantes si, pour toute partie finie J de I , et pour tout $(A_j)_{j \in J} \in \prod_{j \in J} \mathcal{P}(E_j)$,

$$P\left(\bigcap_{j \in J} (X_j \in A_j)\right) = \prod_{j \in J} P(X_j \in A_j).$$

Proposition: Avec les notations précédentes, $(X_i)_{i \in I}$ est une famille de variables aléatoires indépendantes si et seulement si

$$\forall J \subset I, J \text{ finie}, \forall (x_j)_{j \in J} \in \prod_{j \in J} X_j(\Omega), P\left(\bigcap_{j \in J} (X_j = x_j)\right) = \prod_{j \in J} P(X_j = x_j).$$

□

Proposition: Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires suivant la même loi de Bernoulli $\mathcal{B}(p)$. Alors,

$$\forall n \in \mathbb{N}^*, \sum_{i=1}^n X_i \sim \mathcal{B}(n, p).$$

Proposition: Soient $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(m, p)$.

Si $X \perp Y$, alors $X + Y \sim \mathcal{B}(n + m, p)$.

□

Cinquième partie

Espérance d'une variable aléatoire

Dans ce paragraphe, toutes les variables aléatoires sont à valeurs réelles.

Définition: Soit $X : \Omega \rightarrow \mathbb{R}$. L'espérance de X est

$$E(X) = \sum_{x \in X(\Omega)} x P(X = x).$$

Lemme: Soit $X : \Omega \rightarrow \mathbb{R}$.

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

Théorème (linéarité de l'espérance): Soient X et Y deux variables aléatoires réelles, et α, β deux réels. Alors,

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Corollaire: Soit X une variable aléatoire suivant la loi binomiale $\mathcal{B}(n, p)$. On a $E(X) = np$.

Définition: Soient $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire et $f : \mathbb{R} \rightarrow \mathbb{R}$.

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathbb{R} \\ & \searrow & \downarrow f \\ & & \mathbb{R} \end{array}$$

La composée $f \circ X$ est notée $f(X)$.

Théorème (formule de transfert): Avec les notations précédentes,

$$E(f(X)) = \sum_{x \in X(\Omega)} f(x) P(X = x).$$

Sixième partie

Variance

Dans ce paragraphe, toutes les variables aléatoires sont à valeurs réelles.

Définition: Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. La variance de X est

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

Proposition: Avec les notations précédentes,

$$V(X) \geq 0.$$

■

Définition: L'écart-type de X est

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

Définition: On dit que X vérifie presque sûrement une propriété si la probabilité que X vérifie cette propriété vaut 1.

Proposition:

$$V(X) = 0 \iff X \text{ est presque sûrement constante.}$$

□

Proposition (Kœnig-Huygens):

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

■

Théorème: Soit $X, Y : \Omega \rightarrow \mathbb{R}$.

$$V(X + Y) = V(X) + V(Y) + 2 \operatorname{Cov}(X, Y)$$

où $\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y)$ est la covariance de X et Y .

■

Proposition: Si $X \perp\!\!\!\perp Y$, alors $\operatorname{Cov}(X, Y) = 0$. D'où

$$V(X + Y) = V(X) + V(Y).$$

■

Proposition: Soit $(X_i)_{1 \leq i \leq n}$ des variables aléatoires définies sur le même espace probabilisé (Ω, P) . Alors

$$\begin{aligned} V\left(\sum_{i=1}^n X_i\right) &= \sum_{1 \leq i, j \leq n} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \end{aligned}$$

□

Corollaire: Avec les notations précédentes, si les X_i sont deux à deux non corrélés (i.e. $\text{Cov}(X_i, X_j) = 0$ pour $i \neq j$), alors

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i).$$

□

Corollaire: Avec les notations précédentes, si les X_i sont indépendantes deux à deux, alors

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i).$$

□

Corollaire: Soit $X \sim \mathcal{B}(n, p)$. Alors,

$$V(X) = np(1-p).$$

■

Proposition (transfert): Soient $X_1, \dots, X_n$ des variables aléatoires à valeurs réelles toutes définies sur un même espace probabilisé (Ω, P) et $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$E(f(X_1, \dots, X_n)) = \sum_{x_1 \in X_1(\Omega)} \cdots \sum_{x_n \in X_n(\Omega)} f(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n).$$

■

Septième partie

Covariance (HORS-PROGRAMME)

On se place dans une optique de Big Data. On dispose d'un tableau à N lignes : chaque ligne correspond à une observation et chaque colonne à une "mesure" (ou caractéristique).

Ces caractéristiques peuvent être corrélées plus ou moins fortement et contenir plus ou moins d'information.

Plus la variance est grande, plus il y a d'information.

Soient X et Y deux colonnes. D'après l'inégalité de Cauchy-Schwarz :

$$|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y \quad \text{donc} \quad -1 \leq \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \leq 1.$$

Si $Y = \alpha X + \beta$:

$$\left| \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \right| = 1.$$

L'objectif est de modifier les colonnes de façon à extraire le plus d'informations possible sur le moins de colonnes possibles.

La matrice de covariance est

$$A = (\text{Cov}(X_i, X_j))_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}}.$$

On aimerait que la matrice A soit diagonale avec de grands coefficients diagonaux.

L'année prochaine, nous verrons le théorème suivant :

Théorème (théorème spectral): Toute matrice symétrique réelle est diagonalisable dans une base orthonormée de vecteurs propres.

Il existe donc des variables aléatoires $Y_1, \dots, Y_k$ combinaisons linéaires des $X_1, \dots, X_k$ telles que

$$(\text{Cov}(Y_i, Y_j)) = \begin{pmatrix} \lambda_1 & & & & \\ & \dots & & & \\ & & (0) & & \\ & & & \dots & \\ & (0) & & & \lambda_k \end{pmatrix}.$$

De plus, le maximum de $V(\alpha_1 X_1 + \dots + \alpha_k X_k)$ avec la condition que $\alpha_1^2 + \dots + \alpha_k^2 = 1$ est $V(Y_1) = \lambda_1$.

Définition (Multiplicateurs de Lagrange): Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ où U est un ouvert de $\mathbb{R}^n$. On cherche $\max_{(x_i)_{i \in [1, n]} \in \mathbb{R}^n} f(x_1, \dots, x_n)$ avec la contrainte $g(x_1, \dots, x_n)$.

On pose

$$F : \mathbb{R}^{n+1} \rightarrow \mathbb{R} \\ (x_1, \dots, x_n, \lambda) \mapsto f(x_1, \dots, x_n) + \lambda g(x_1, \dots, x_n).$$

On cherche les points critiques de F :

$$\begin{aligned} \nabla F(x_1, \dots, x_n, \lambda) = 0 &\iff \begin{cases} \forall i, \frac{\partial F}{\partial x_i}(x_1, \dots, x_n, \lambda) = 0 \\ \frac{\partial F}{\partial \lambda}(x_1, \dots, x_n, \lambda) = 0 \end{cases} \\ &\iff \begin{cases} \forall i, \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) + \lambda \frac{\partial g}{\partial x_i}(x_1, \dots, x_n) = 0 \\ g(x_1, \dots, x_n) = 0 \end{cases} \end{aligned}$$

Si $(x_1, \dots, x_n, \lambda)$ est le maximum de F , alors

$$\forall y_1, \dots, y_n, \mu \in \mathbb{R}^{n+1} \quad F(x_1, \dots, x_n, \lambda) \geq F(y_1, \dots, y_n, \mu)$$

$$f(x_1, \dots, x_n) + \underbrace{\lambda g(x_1, \dots, x_n)}_{=0} \geq f(y_1, \dots, y_n)$$

Huitième partie

Loi des grands nombres

Lemme (inégalité de Марков ^{a)}: Soit $X : \Omega \rightarrow \mathbb{R}^+$ une variable aléatoire positive.

$$\forall t \in \mathbb{R}_+^*, P(X \geq t) \leq \frac{E(X)}{t}$$

ou encore,

$$\forall u \in \mathbb{R}_+^*, P(X \geq u E(X)) \leq \frac{1}{u}.$$

^a. Марков : Markov

■

Lemme (inégalité de Bienaymé-Чебышев ^{a)}: Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle. On pose $\mu = E(X)$ et $\sigma = \sigma_X$. Alors,

$$\forall t \in \mathbb{R}_+^*, P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

ou encore

$$\forall u \in \mathbb{R}_+^*, P(|X - \mu| \geq u\sigma) \leq \frac{1}{u^2}.$$

^a. Чебышев : Tchebychev

■

REMARQUE (Application):

On considère une expérience de Bernoulli de paramètre $p \in]0, 1[$. On répète n fois cette expérience à l'identique et on note Y_n le nombre moyen de succès.

On pose, pour $i \in \llbracket 1, n \rrbracket$,

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ème expérience est un succès} \\ 0 & \text{sinon.} \end{cases}$$

Les (X_i) sont mutuellement indépendants; d'où

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Et donc

$$E(Y_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} np = p.$$

Comme les (X_i) sont deux à deux indépendantes,

$$V(Y_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} npq = \frac{pq}{n} \text{ où } q = 1 - p.$$

D'où, d'après l'inégalité de Bienaymé-Чебышев,

$$\forall \varepsilon > 0, P(|Y_n - p| \geq \varepsilon) \leq \frac{pq}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0.$$

On en déduit que

$$\forall \varepsilon > 0, P(|Y_n - p| < \varepsilon) \geq 1 - \underbrace{\frac{pq}{n\varepsilon^2}}_{\substack{\delta \\ \text{"seuil"}}}.$$

— Par exemple, avec $\varepsilon = 10^{-2}$ et $\delta = 95\%$, on a

$$\begin{aligned} \frac{95}{100} = \delta = 1 - \frac{pq}{n \times 10^{-4}} &\iff \frac{5}{100} = \frac{pq}{n \times 10^{-4}} \\ &\iff 20 = \frac{n \times 10^{-4}}{pq} \\ &\iff n = 20 \underbrace{pq}_{\leq \frac{1}{4}} 10^4 \leq 5 \times 10^4. \end{aligned}$$

En répétant 50 000 fois l'expérience, on est sûr à 95 % que la valeur observée de $Y_{50\,000}$ est dans l'intervalle $[p - 10^{-2}, p + 10^{-2}]$.

— Avec $n = 12$ et $\delta = 95\%$, on a

$$\begin{aligned} \frac{95}{100} = \delta = 1 - \frac{pq}{n\varepsilon^2} &\iff \frac{pq}{n\varepsilon^2} = \frac{5}{100} \\ &\iff \frac{\varepsilon^2 n}{pq} = 20 \\ &\iff \varepsilon^2 = \frac{20}{n} pq \leq \frac{5}{12}. \end{aligned}$$

Donc, $\varepsilon = \sqrt{\frac{5}{12}} \simeq 0,65$.

— Avec $n = 12$ et $\varepsilon = 10^{-2}$,

$$\begin{aligned} P(|Y_{12} - p| < \varepsilon) &\geq 1 - \frac{pq}{12 \cdot 10^{-4}} \\ &\geq 1 - \frac{1}{48 \cdot 10^{-4}} \\ &\geq 1 - \underbrace{\frac{1}{48}}_{< 0} \times 10^4 \end{aligned}$$

Théorème (loi faible des grands nombres — HORS-PROGRAMME): Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes de même loi définies sur le même espace probabilisé (Ω, P) . On note μ leur espérance commune.

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

où $\forall n, \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

■

Proposition: Soient $X_1, \dots, X_n$ des variables aléatoires à valeurs réelles d'un même ensemble probabilisé mutuellement indépendantes. Soient $f: \mathbb{R}^p \rightarrow \mathbb{R}$ et $g: \mathbb{R}^{n-p} \rightarrow \mathbb{R}$. Alors, $f(X_{i_1}, \dots, X_{i_p})$ et $g(X_{j_1}, \dots, X_{j_{n-p}})$ sont indépendantes si

$$\begin{cases} \{i_1, \dots, i_p\} \cap \{j_1, \dots, j_{n-p}\} = \emptyset \\ \forall k \neq \ell, i_k \neq i_\ell \text{ et } j_k \neq j_\ell. \end{cases}$$