

Intitulé de la matière : Probabilités 1

Objectifs de l'enseignement

Introduire les notions de base sur les théorèmes limites

Connaissances préalables recommandées

Notions de probabilité, variables aléatoires, lois discrètes et lois continues.

Contenu de la matière

- Vecteurs aléatoires gaussiens
- Loi des grands nombres
- Convergence en loi
- Théorème central limite
- Fonctions caractéristiques
- Espérance conditionnelle
- Espaces filtrés discrets
- Temps d'arrêt discrets
- Martingales discrètes

Cours: Probabilités 1

Masters: Proba/Stat et Actuariat

I. Vecteurs aléatoires gaussiens

Soit (Ω, \mathcal{F}, P) un espace probabilisé.

I.1 Généralités

Définitions:

-On appelle vecteur aléatoire de dimension n toute variable aléatoire $X = (X_1, X_2, \dots, X_n)^T$ défini sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^n munit de sa tribu borélienne $\mathcal{B}_{\mathbb{R}^n}$.

-Les variables aléatoires (*v.a.r.* en abrégé) X_1, X_2, \dots, X_n s'appellent les variables aléatoires marginales de X .

-L'application $P_X : \mathcal{B}_{\mathbb{R}^n} \rightarrow [0, 1]$ définie par:

$$P_X(A) = P(X^{-1}(A))$$

la loi s'appelle de probabilité de X .

Définition:

Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)^T$ de dimension n est dit discret si l'ensemble des valeurs possibles D_X de X est un ensemble fini ou dénombrable. L'ensemble D_X s'appelle le support de X et la fonction (qu'on notera encore P_X) $P_X : \mathbb{R}^n \rightarrow \mathbb{R}$, définie par

$$P_X(x) = P(X = x)$$

s'appelle la fonction de masse (ou loi) de X .

Remarque:

1. $P_X(x) = 0$ pour tout $x \notin D_X$.

2. $\{X = x_1\} \cap \{X = x_2\} = \emptyset$ pour tout $x_1 \neq x_2$.

Propriétés de P_X :

1. $P_X(x) \geq 0$.

2. $\sum_{x \in \mathbb{R}^n} P_X(x) = \sum_{x \in D_X} P_X(x) = 1$.

En effet

$$\sum_{x \in D_X} P_X(x) = \sum_{x \in D_X} P(X = x) = P\left(\bigcup_{x \in D_X} \{X = x\}\right) = P(\Omega) = 1.$$

Notons que toute fonction vérifiant les propriétés 1. et 2. peut être considérée comme une fonction de masse d'un certain vecteur aléatoire discret X .

Remarques:

1. $D_{X_i} = \pi_i(D_X)$, où $\pi_i(x_1, x_2, \dots, x_n) = x_i$ est $i^{\text{ième}}$ projection sur \mathbb{R}^n .

2. $P_{X_i}(x_i) = \sum_{(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) : (x_1, x_2, \dots, x_n) \in D_X} P_X(x_1, x_2, \dots, x_n)$, pour

tout $x_i \in D_{X_i}$.

Exemple:

Soit $X = (X_1, X_2)^T$ un couple aléatoire défini par le tableau suivant:

$x_2 \rightarrow$ $x_1 \downarrow$	-1	0	2	5	$P_{X_1}(x_1)$
1	0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
3	0	$\frac{1}{4}$	0	$\frac{1}{16}$	$\frac{5}{16}$
4	$\frac{3}{16}$	0	$\frac{1}{4}$	0	$\frac{7}{16}$
$P_{X_2}(x_2)$	$\frac{3}{16}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{16}$	/

Cela signifie que $D_X = \{(1, 0), (1, 2), (3, 0), (3, 5), (4, -1), (4, 2)\}$,

$$P_X(x_1, x_2) = \begin{cases} \frac{1}{8} & \text{si } (x_1, x_2) \in \{(1, 0), (1, 2)\} \\ \frac{1}{4} & \text{si } (x_1, x_2) \in \{(3, 0), (4, 2)\} \\ \frac{1}{16} & \text{si } (x_1, x_2) \in \{(3, 5), (4, -1)\} \\ 0 & \text{ailleurs} \end{cases},$$

On en déduit que $D_{X_1} = \pi_1(D_X) = \{1, 3, 4\}$, $D_{X_2} = \pi_2(D_X) = \{-1, 0, 2, 5\}$

$$P_{X_1}(x_1) = \begin{cases} \frac{1}{4} & \text{si } x_1 = 1 \\ \frac{5}{16} & \text{si } x_1 = 3 \\ \frac{7}{16} & \text{si } x_1 = 4 \\ 0 & \text{ailleurs} \end{cases} \quad \text{et} \quad P_{X_2}(x_2) = \begin{cases} \frac{3}{16} & \text{si } x_2 = -1 \\ \frac{3}{8} & \text{si } x_2 = 0 \\ \frac{3}{8} & \text{si } x_2 = 2 \\ \frac{1}{16} & \text{si } x_2 = 5 \\ 0 & \text{ailleurs} \end{cases}$$

Définitions:

-Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)^T$ de dimension n est dit absolument continu s'il existe une fonction positive et intégrable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que

$$P_X(A) = \int_A f(x) dx \text{ pour tout } A \in \mathcal{B}_{\mathbb{R}^n}$$

-L'ensemble $D_X := \{x \in \mathbb{R}^n : f(x) \neq 0\}$ est appelé support de X et la fonction f est appelée densité de X .

Propriétés de f :

1. $f(x) \geq 0$
2. $\int_{\mathbb{R}^n} f(x) dx = \int_{D_X} f(x) dx = 1$.

On notera que toute fonction vérifiant les deux propriétés peut être considérée comme une densité d'un certain vecteur aléatoire absolument continu X .

I.2 Matrice des covariances

On munit \mathbb{R}^n du produit scalaire usuel

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \text{ pour tous } x = (x_1, x_2, \dots, x_n)^T, y = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$$

et de la norme correspondante

$$\|x\| = \langle x, x \rangle^{\frac{1}{2}}.$$

Définition:

-Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)^T$ de dimension n est dit intégrable si la *v.a.* réelle $\|X\|$ est intégrable.

- X est dit à carré intégrable si la *v.a.* réelle $\|X\|^2$ est intégrable.

On notera par $\mathcal{L}_{\mathbb{R}^n}^1(\Omega, \mathcal{F}, P)$ (resp. $\mathcal{L}_{\mathbb{R}^n}^2(\Omega, \mathcal{F}, P)$) l'espace vectoriel des vecteurs aléatoires intégrables (resp. à carré intégrable) de dimension n .

Remarques:

1. $X \in \mathcal{L}_{\mathbb{R}^n}^1(\Omega, \mathcal{F}, P) \Rightarrow X_i \in \mathcal{L}_{\mathbb{R}}^1(\Omega, \mathcal{F}, P)$ pour tout $i = 1, 2, \dots, n$. La réciproque est fautive en général.

En effet, si $X \in \mathcal{L}_{\mathbb{R}^n}^1(\Omega, \mathcal{F}, P)$, alors $\mathbb{E}(|X_i|) \leq \mathbb{E}(\|X\|) < \infty$, d'où $X_i \in \mathcal{L}_{\mathbb{R}}^1(\Omega, \mathcal{F}, P)$.

2. $X \in \mathcal{L}_{\mathbb{R}^n}^2(\Omega, \mathcal{F}, P) \iff X_i \in \mathcal{L}_{\mathbb{R}}^2(\Omega, \mathcal{F}, P)$ pour tout $i = 1, 2, \dots, n$.

En effet, l'inégalité $|X_i|^2 \leq \|X\|^2$ entraîne que $X \in \mathcal{L}_{\mathbb{R}^n}^2(\Omega, \mathcal{F}, P) \Rightarrow X_i \in \mathcal{L}_{\mathbb{R}}^2(\Omega, \mathcal{F}, P)$. L'implication dans l'autre sens provient du fait que si $X_i \in \mathcal{L}_{\mathbb{R}}^2(\Omega, \mathcal{F}, P)$ alors $X_i^2 \in \mathcal{L}_{\mathbb{R}}^1(\Omega, \mathcal{F}, P)$, qui est une espace vectoriel, d'où $\mathbb{E}(\|X\|^2) = \sum_{i=1}^n \mathbb{E}(|X_i|^2)$.

Dans le cas où $n = 1$, la dispersion de X autour de la moyenne $\mathbb{E}(X)$ est mesurée par sa variance $\sigma^2(X)$. Dans le cas général, la connaissance de $\sigma^2(X_i)$ est insuffisante car elle ne concerne que les dispersions suivant les axes de coordonnées. Nous aurons une bien meilleure évolution de la dispersion de X en considérant la dispersion de X sur une droite quelconque.

On suppose que le vecteur aléatoire X est à carré intégrable. Soit $u = (u_1, u_2, \dots, u_n)^T \in \mathbb{R}^n$. $\langle u, X \rangle = \sum_{i=1}^n u_i X_i$ est une *v.a.r.* à carré intégrable de

variance

$$\begin{aligned}
\sigma^2(\langle u, X \rangle) &= \mathbb{E}(\langle u, X \rangle^2) - (\mathbb{E}(\langle u, X \rangle))^2 \\
&= \mathbb{E}\left(\sum_{i=1}^n u_i X_i \sum_{j=1}^n u_j X_j\right) - \left(\sum_{i=1}^n u_i \mathbb{E}(X_i)\right) \left(\sum_{j=1}^n u_j \mathbb{E}(X_j)\right) \\
&= \sum_{i,j=1}^n u_i u_j \mathbb{E}(X_i X_j) - \sum_{i,j=1}^n u_i u_j \mathbb{E}(X_i) \mathbb{E}(X_j) \\
&= \sum_{i,j=1}^n u_i u_j \text{cov}(X_i, X_j)
\end{aligned}$$

L'application $u \rightarrow \sigma^2(\langle u, X \rangle)$ est une forme quadratique positive sur \mathbb{R}^n . Sa matrice relativement à base canonique de \mathbb{R}^n est notée $C(X) = C_{ij}$, où $C_{ij} := \text{cov}(X_i, X_j)$.

Définition:

La matrice $C(X)$ est appelée la matrice des covariances de X .

Remarque:

On notera la relation matricielle $\sigma^2(\langle u, X \rangle) = u^T C(X) u$

Proposition:

On suppose que le vecteur aléatoire X est à carré intégrable. et que A est une matrice carré d'ordre n . Alors $Y := AX$ est un vecteur aléatoire à carré intégrable, admettant pour matrice des covariance, la matrice

$$C(Y) = AC(X)A^T.$$

Démonstration:

Soit $A = (a_{ij})$. Pour tout $i = 1, 2, \dots, n$, $Y_i = \sum_{j=1}^n a_{ij} X_j$ est une v.a.r. à carré intégrable. Il s'en suit que Y est un vecteur aléatoire de dimension n à carré intégrable. La relation $\langle u, Av \rangle = \langle A^T u, v \rangle$ nous permet d'écrire:

$$\sigma^2(\langle u, Y \rangle) = \sigma^2(\langle A^T u, X \rangle) = (A^T u)^T C(X) A^T u = u^T (AC(X)A^T) u,$$

d'où

$$C(Y) = AC(X)A^T.$$

■

I.3 Vecteurs aléatoires gaussiens

Définition:

Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)^T$ de dimension n est dit gaussien (ou normal) si toute combinaison linéaire de ses composantes $X_j, j = 1, 2, \dots, n$ est une v.a.r. gaussienne, c'est à dire que pour tout $u \in \mathbb{R}^n$ la v.a. $\langle u, X \rangle$ est gaussienne (En particulier, toutes les marginales X_j sont gaussiennes).

Remarque:

Si $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une application linéaire et X est un vecteur gaussien de dimension n alors la composée $Y := g(X)$ est un vecteur gaussien de dimension m .

En effet, si $g_k(x) = \sum_{j=1}^n g_{kj}x_j$, alors pour tout $u \in \mathbb{R}^m$, la v.a.

$$\langle u, Y \rangle = \sum_{k=1}^m u_k Y_k = \sum_{k=1}^m u_k \sum_{j=1}^n g_{kj} X_j = \sum_{j=1}^n \left(\sum_{k=1}^m u_k g_{kj} \right) X_j$$

est gaussienne comme étant une combinaison linéaire des marginales de X qui est gaussien.

Notation:

Soit X est un vecteur gaussien de dimension n . On note par $\mathbb{E}(X)$ le vecteur $(\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n))$ et on écrit

$$X \rightsquigarrow \mathcal{N}_n(m, C),$$

où $m := \mathbb{E}(X)$ et $C := C(X)$.

Remarques:

1. $X \rightsquigarrow \mathcal{N}_n(m, C)$ et $b \in \mathbb{R}^n$, alors

$$X + b \rightsquigarrow \mathcal{N}_n(m + b, C).$$

En effet

$$\langle u, X + b \rangle = \sum_{i=1}^n u_i (X_i + b_i) = \sum_{i=1}^n u_i X_i + \sum_{i=1}^n u_i b_i,$$

est une v.a.r. gaussienne, d'où $X + b$ est gaussien. De plus $\mathbb{E}(X + b) = \mathbb{E}(X) + b = m + b$. On a aussi

$$\begin{aligned} \text{cov}(X_i + b_i, X_j + b_j) &= \mathbb{E}((X_i + b_i)(X_j + b_j)) - \mathbb{E}(X_i + b_i)\mathbb{E}(X_j + b_j) \\ &= \mathbb{E}(X_i X_j + b_i X_j + b_j X_i + b_i b_j) \\ &\quad - (\mathbb{E}(X_i)\mathbb{E}(X_j) + b_i \mathbb{E}(X_j) + b_j \mathbb{E}(X_i) + b_i b_j) \\ &= \mathbb{E}(X_i X_j) + b_i \mathbb{E}(X_j) + b_j \mathbb{E}(X_i) + b_i b_j \\ &\quad - \mathbb{E}(X_i)\mathbb{E}(X_j) - b_i \mathbb{E}(X_j) - b_j \mathbb{E}(X_i) - b_i b_j \\ &= \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &= \text{cov}(X_i, X_j), \end{aligned}$$

qui signifie que $C(X + b) = C(X) = C$, d'où

$$X + b \rightsquigarrow \mathcal{N}_n(m + b, C).$$

2. Si $X \rightsquigarrow \mathcal{N}_n(m, C)$, alors $X - m \rightsquigarrow \mathcal{N}_n(0, C)$.

3. Si $X \rightsquigarrow \mathcal{N}_n(m, C)$, A est une matrice carré d'ordre n et $b \in \mathbb{R}^n$, alors $AX + b$ est également un vecteur aléatoire gaussien.

Lemme:

Soit \sum une matrice $n \times n$, symétrique et semi-définie positive. Alors, il existe une matrice carré A de dimension $n \times n$ telle que $\sum = A^T A$. On dit que A est une racine carré de \sum . De plus si \sum est inversible, alors il en est de même de A .

Démonstration:

Du fait que \sum est symétrique, on déduit qu'elle est diagonalisable et donc qu'ils existent une matrice orthogonale O (i.e. telle que ${}^T O = O^{-1}$) et une matrice diagonale Λ avec $\Lambda_{ii} = \lambda_i$ la $i^{\text{ème}}$ valeur propre de \sum telles que $\sum = {}^T O \Lambda O$. Puisque \sum est semi-définie positive on a que les valeurs propres λ_i de \sum sont positives pour tout $i = 1, 2, \dots, n$. Soit $\Lambda^{\frac{1}{2}}$ la matrice diagonale telle que $\left(\Lambda^{\frac{1}{2}}\right)_{ii} = \sqrt{\lambda_i}$, donc $\Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} = \Lambda$ et si on pose $A = {}^T O \Lambda^{\frac{1}{2}}$ on aura

$$A {}^T A = {}^T O \Lambda^{\frac{1}{2}} {}^T \left({}^T O \Lambda^{\frac{1}{2}} \right) = {}^T O \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} O = {}^T O \Lambda O = \sum.$$

Si \sum est inversible, alors $\lambda_i > 0$ pour tout $i = 1, 2, \dots, n$ et donc $\Lambda^{\frac{1}{2}}$ est inversible et $\left(\Lambda^{\frac{1}{2}}\right)^{-1}$ est la matrice diagonale avec sur la diagonale $\frac{1}{\sqrt{\lambda_i}}$. Ainsi A est inversible comme étant le produit de deux matrices inversibles et on a:

$$A^{-1} = \left({}^T O \Lambda^{\frac{1}{2}} \right)^{-1} = \left(\Lambda^{\frac{1}{2}} \right)^{-1} O.$$

Remarque:

Si on considère un vecteur aléatoire $Z = {}^T (Z_1, Z_2, \dots, Z_n)$ de dimension n tel que les *v.a.r.* Z_i sont indépendantes et de loi $\mathcal{N}(0, 1)$ pour tout $i = 1, 2, \dots, n$. Alors le vecteur $Z \rightsquigarrow \mathcal{N}_n(0, Id)$. Soit A une matrice telle que $A {}^T A = \sum$ (une racine carré de \sum donnée par le lemme précédent). Alors le vecteur

$$X := m + AZ \rightsquigarrow \mathcal{N}_n\left(m, \sum\right).$$

D'une famille de *v.a.* gaussiennes indépendantes on peut donc construire n'importe quel vecteur gaussien. Si \sum est inversible, alors il en est de même de A et

$$Z = A^{-1}(X - m).$$

Théorème:

Le vecteur aléatoire $X \rightsquigarrow \mathcal{N}_n(m, C)$ admet une densité si et seulement si C est inversible (i.e. définie positive) et alors

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det C}} \exp\left(-\frac{1}{2} (x - m) C^{-1} (x - m)\right) \quad (*)$$

Démonstration:

On montre seulement que si C est inversible alors X admet une densité donnée en équation (*). On considère le vecteur aléatoire $Z = {}^T (Z_1, Z_2, \dots, Z_n)$ tel que les *v.a.r.* Z_i soient indépendantes et $Z_i \rightsquigarrow \mathcal{N}(0, 1)$ pour tout $i = 1, 2, \dots, n$. La densité de Z est alors donnée par:

$$\begin{aligned} f_Z(z) &= f_{Z_1}(z_1) f_{Z_2}(z_2) \dots f_{Z_n}(z_n) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} (z_1^2 + z_2^2 + \dots + z_n^2)\right) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} z {}^T z\right), \quad z = (z_1, z_2, \dots, z_n) \in \mathbb{R}^n. \end{aligned}$$

Par la remarque précédente, on a que la *v.a.* $X := m + AZ$, où $C = A^T A$, est bien un vecteur gaussien de moyenne m et de matrice des covariances C . Donc la densité de X est donnée par la formule de changement de variables à partir de la densité de Z . Si l'on pose $\Psi(z) = m + Az$ alors $z = \Psi^{-1}(x) = A^{-1}(x - m)$ et on a

$$\begin{aligned} f_X(x) &= f_Z(\Psi^{-1}(x)) J(\Psi^{-1}(x)), \text{ où } J(\Psi^{-1}(x)) \text{ est le jacobien de } \Psi^{-1} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \det(A^{-1}) \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} (\Psi^{-1}(x))^T \Psi^{-1}(x)\right), \end{aligned}$$

or

$$\begin{aligned} {}^T(\Psi^{-1}(x)) \Psi^{-1}(x) &= {}^T(A^{-1}(x - m)) A^{-1}(x - m) = {}^T(x - m)^T (A^{-1}) A^{-1}(x - m) \\ &= {}^T(x - m) ({}^T A)^{-1} A^{-1}(x - m) \\ &= {}^T(x - m) (A^T A)^{-1}(x - m) \\ &= {}^T(x - m) (C)^{-1}(x - m) \end{aligned}$$

Comme $\det C = \det A \det^T A = (\det A)^2$ et $\det A^{-1} = \frac{1}{\det A}$, alors $\det A^{-1} = \frac{1}{\sqrt{\det C}}$, d'où le théorème. ■