Chapitre 3. Vecteurs aléatoires

Prof. REMITA Mohamed Riad

National School of Artificial Intelligence.

2023-2024

Introduction

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $V = (X_1, \cdots, X_n)$. Une application de Ω dans $\mathbb{R}^n$, qui à tout $\omega \in \Omega$ fait correspondre une suite $V(\omega) = (X_1(\omega), \cdots, X_n(\omega))$. On dit que V est un vecteur aléatoire, si pour tout $i = 1, \cdots, n$, l'application X_i est une variable aléatoire.

Definition

V est un vecteur aléatoire ou une variable aléatoire à n dimension si et seulement si $X^{-1}(x) \in \mathcal{F}$ ou $x = (x_1, \cdots, x_n)$ i.e. si $\mathbb{P}(V(\omega) \leq x)$ existe.

$\{V(\omega) \leq x\} = \{X_1(\omega) \leq x_1, \cdots, X_n(\omega) \leq x_n\} \Longleftrightarrow \{\omega \in V^{-1}(x)\}$

$V^{-1}(x)$ étant l'image inverse du pavé de $\mathbb{R}^n$: $(]-\infty, x_1], \cdots,]-\infty, x_n])$.

Fonction de répartition

Le v.a. V est caractérisé par sa fonction de répartition F définie par $F(x) = \mathbb{P}(V(\omega) \leq x)$.

Cas discret

La loi du vecteur $V = (X_1, \cdots, X_n)$ est définie par

$$\mathbb{P}_X(k_1, \cdots, k_n) = \mathbb{P}(X_1 = k_1, \cdots, X_n = k_n),$$

et pour tout $A_1 \in \mathcal{F}_1, \cdots, A_n \in \mathcal{F}_n$

$$\mathbb{P}(X_1 \in A_1, \cdots, X_n \in A_n) = \sum_{k_1 \in A_1, \cdots, k_n \in A_n} \mathbb{P}_X(k_1, \cdots, k_n).$$

La fonction de répartition du vecteur aléatoire V est définie par

$$F_V(x_1, \cdots, x_n) = \sum_{k_1 \leq x_1} \cdots \sum_{k_n \leq x_n} \mathbb{P}_X(k_1, \cdots, k_n).$$

Fonction de répartition

Cas absolument continu

F_X est définie par

$$F_V(x_1, \cdots, x_n) = \mathbb{P}(X_1 \leq x_1, \cdots, X_n \leq x_n).$$

D'ou la densité du vecteur aléatoire V est définie par

$$f_V(x_1, \cdots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_V(x_1, \cdots, x_n).$$

Alors, pour tout $A_1 \in \mathcal{F}_1, \cdots, A_n \in \mathcal{F}_n$

$$\mathbb{P}(X_1 \in A_1, \cdots, X_n \in A_n) = \int_{A_1} \cdots \int_{A_n} f_V(x_1, \cdots, x_n)\, dx_1 \cdots dx_n.$$

Definition

On dit que les variables aléatoires $X_1, \cdots, X_n$ sont mutuellement indépendantes si

$$\mathbb{P}(X_1 \in A_1, \cdots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

Lois conjointes. Lois marginales - Cas discret

Nous considérons à partir de maintenant le cas $n = 2$, **c'est-à-dire on s'intéressera au couple aléatoire** (X, Y).

Soit p_{ij} la probabilité liée au point (x_i, y_j) i.e. $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$ et $\sum_i \sum_j p_{ij} = 1$.
p_{ij} est appelée loi conjointe de X et Y.

La loi de X et la loi de Y sont appelées lois marginales. On a
$\mathbb{P}(X = x_i) = \sum_j \mathbb{P}(X = x_i, Y = y_j) = \sum_j p_{ij} = p_{i.}$ et $\mathbb{P}(Y = y_i) = p_{.j}$.
On a $\sum_i p_{i.} = \sum_j p_{.j} = 1$.

Lois conjointes. Lois marginales - Cas discret

Fonction de répartition

Soit le couple $V = (X, Y)$. On a

$$F_V(v) = F_V(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}.$$

Alos la fonction de répartition de la marginale X est

$$F_X(x) = \mathbb{P}(X \leq x) = F_V(x, +\infty).$$

De même que la fonction de répartition de la marginale Y est

$$F_Y(y) = \mathbb{P}(Y \leq y) = F_V(+\infty, y).$$

Lois conjointes. Lois marginales - Cas discret

Loi conditionnelle

Soit l événement $\{X = x_i\}$ tel que $\mathbb{P}(X = x_i) \neq 0$ i.e. $p_{i\cdot} \neq 0$. On appelle variable aléatoire conditionnelle Y liée par $X = x$, notée $Y|_{X=x_i}$ la variable aléatoire discrète dont les probabilités p_j^i sont données par

$$p_j^i = \mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(X = x_i)} = \frac{p_{ij}}{p_{i\cdot}},$$

de même

$$p_i^j = \frac{p_{ij}}{p_{\cdot j}}.$$

Remarque

On peut alors déterminer la fonction de répartition, l'espérance et la variance de $Y|_{X=x_i}$ *et* $X|_{Y=y_j}$.

Lois conjointes. Lois marginales - Cas discret

Loi d'une somme de variables aléatoires

La probabilté $\mathbb{P}(Z = k)$ de la somme $Z = X + Y$ de deux variables aléatoires X et Y est la somme des probabilités $\mathbb{P}(X = x_i, Y = y_j)$ étendue à tous les couples (x_i, y_j) liés par la relation $k = x_i + y_j$.

$$\mathbb{P}(Z = k) = \sum_{i+j=k} \mathbb{P}(X = x_i, Y = y_j).$$

Si les variables aléatoires X et Y sont indépendantes, on a

$$\mathbb{P}(Z = k) = \sum_{i+j=k} \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j).$$

Lois conjointes. Lois marginales - Cas discret

Somme de deux variables aléatoires binomiales.

Example

Soient X_1 et X_2 deux variables aléatoires indépendantes telles que $X_1 \rightsquigarrow \mathcal{B}(n_1, p)$ et $X_2 \rightsquigarrow \mathcal{B}(n_2, p)$.
On a $\mathbb{P}(X_1 + X_2 = k) = \sum_A \mathbb{P}(X_1 = i, X_2 = j)$ où
$A = \{(i, j) : 0 \leq i \leq n_1, 0 \leq j \leq n_2; i + j = k\}$.

$$
\begin{aligned}
\mathbb{P}(X_1 + X_2 = k) &= \sum_A \mathbb{P}(X_1 = x_1)\, \mathbb{P}(X_2 = x_2) \\
&= \sum_A C_{n_1}^i p^i (1-p)^{n_1 - i} C_{n_2}^j p^j (1-p)^{n_2 - j} \\
&= \sum_A C_{n_1}^i C_{n_2}^j p^{i+j} (1-p)^{n_1 + n_2 - i - j},
\end{aligned}
$$

on a $\sum_A C_{n_1}^i C_{n_2}^j = \sum_{k=0}^{n_1+n_2} C_{n_1}^i C_{n_2}^j = C_{n_1+n_2}^k$,
d'où $\mathbb{P}(X_1 + X_2 = k) = C_{n_1+n_2}^k p^k (1-p)^{n_1+n_2-k}$, alors
$X_1 + X_2 \rightsquigarrow \mathcal{B}(n_1 + n_2, p)$.

Lois conjointes. Lois marginales - Cas absolument continu

Loi du couple (loi conjointe)

Une variable aléatoire $Z = (X, Y)$ est dite absolument continue s'il existe une application $f(x, y)$ appelée densité de probabilité du couple aléatoire (X, Y) vérifiant

a. $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$

b. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dxdy = 1$

La fonction de répartition du couple (X, Y) est définie par

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dudv$$

et l'on a

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Les fonctions

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(u, v)\, dudv$$

et

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{+\infty} f(u, v)\, dudv$$

sont appelées fonctions de répartitions marginales des variables aléatoires X et Y respectivement.

Les fonctions

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, v)\, dv \text{ et } f_Y(y) = \int_{-\infty}^{+\infty} f(u, y)\, du$$

sont les densités marginales de X et Y respectivement.

La densité conditionnelle de X sachant $\{Y = y\}$ est définie par

$$f_{X|Y}(x|\, y) = \frac{f(x, y)}{f_Y(y)} \text{ si } f_Y(y) \neq 0.$$

La densité conditionnelle de Y sachant $\{X = x\}$ est définie par

$$f_{Y|X}(y|\, x) = \frac{f(x, y)}{f_X(x)} \text{ si } f_X(x) \neq 0.$$

Les variables aléatoires X et Y sont indépendantes si

$$\forall x \in \mathbb{R}, \forall y \in \mathbb{R}, \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\,\mathbb{P}(Y \leq y)$$

ou bien

$$F(x, y) = F_X(x)\,F_Y(y),$$

ou encore dans le cas continu, en dérivant deux fois par rapport à x et à y

$$f(x, y) = f_X(x)\,f_Y(y).$$

Dans ce cas

$$f_{X|Y}(x|\,y) = f_X(x) \text{ et } f_{Y|X}(y|\,x) = f_Y(y).$$

Lois conjointes. Lois marginales - Cas absolument continu

Indépendance de deux variables aléatoires

Example

Un couple de variable aléatoire $Z = (X, Y)$ a pour densité

$$f(x, y) = \begin{cases} kxye^{-\left(x^2+y^2\right)} & \text{si } x > 0 \text{ et } y > 0 \\ 0 & \text{sinon} \end{cases}.$$

1. Trouver le coefficient k.
2. Trouver les lois marginales de X et Y.
3. Calculer les densités conditionnelles de X sachant $\{Y = y\}$ et de Y sachant $\{X = x\}$.

Lois conjointes. Lois marginales - Cas absolument continu

Loi d'une somme de deux variables aléatoires

Soit $Z = X + Y$, on peut déterminer la fonction de répartition de la variable aléatoire Z.On a

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}(X + Y \leq z) \\ &= \int\int_A f(x, y)\, dxdy, \end{aligned}$$

ou $A = \left\{(x, y) \in \mathbb{R}^2 : x + y \leq z\right\}$
Si X et Y sont indépendantes, on a

$$\begin{aligned} F_Z(z) &= \int\int_A f(x, y)\, dxdy = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-x} f(x, y)\, dy\right] dx \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-x} f_Y(y)\, dy\right] f_X(x)\, dx \\ &= \int_{-\infty}^{+\infty} F_Y(z - x)\, f_X(x)\, dx. \end{aligned}$$

On peut trouver de la même façon

$$F_Z(z) = \int_{-\infty}^{+\infty} F_X(z-y)\, f_Y(y)\, dy.$$

En dérivant $F_Z(z)$ par rapport à z on trouve

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)\, f_Y(y)\, dy \text{ et } f_Z(z) = \int_{-\infty}^{+\infty} f_Y(z-x)\, f_X(x)\, dx.$$

f_Z est appelée produit de convoution de f_X et f_Y noté $f_Z = f_X * f_Y$.

Lois conjointes. Lois marginales - Cas absolument continu

Loi d'une somme de deux variables aléatoires

Example

Soient n variables aléatoires $X_i \rightsquigarrow \mathcal{N}(0,1)$ $(i = 1, \cdots, n)$. Déterminer la loi de la variable aléatoire $Y = X_1^2 + \cdots + X_n^2$.

1. Soit $X_1 \rightsquigarrow \mathcal{N}(0,1)$. Déterminons la loi de $Z = X_1^2$. On a

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}\left(X_1^2 \leq z\right) = \mathbb{P}\left(-\sqrt{z} \leq X_1 \leq \sqrt{z}\right) \\ &= F_X\left(\sqrt{z}\right) - F_X\left(-\sqrt{z}\right). \end{aligned}$$

   La densité de probabilité de Z est alors

$$\begin{aligned} f_1(z) &= \frac{1}{2\sqrt{z}}\left(f_X\left(\sqrt{z}\right) + f_X\left(\sqrt{z}\right)\right) = \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}} \\ &= \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} z^{\frac{1}{2}-1} e^{-\frac{z}{2}}. \end{aligned}$$

Example

2. Soit $X_2 \rightsquigarrow \mathcal{N}(0,1)$ et déterminons la loi de $Z = X_1^2 + X_2^2$.
   Soit f_2 la densité de Z. On a

$$
\begin{aligned}
f_2(z) &= \int_0^z f_1(y) f_2(z-y)\, dy \\
&= \int_0^z \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \frac{1}{\sqrt{2\pi}} (z-y)^{-\frac{1}{2}} e^{-\frac{(z-y)}{2}} dy \\
&= \frac{e^{-\frac{z}{2}}}{2\pi} \int_0^z y^{-\frac{1}{2}} (z-y)^{-\frac{1}{2}} dy = \frac{1}{2\Gamma\left(\frac{2}{2}\right)} z^{\frac{2}{2}-1} e^{-\frac{z}{2}}.
\end{aligned}
$$

En généralisant à $Z = X_1^2 + \cdots + X_n^2$ par l'intermédiaire d'un raisonnement par récurence, on aura $f_Z(z) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}$,alors Z suit la loi du Khi deux à n degrés de libertés notée $Z \rightsquigarrow \chi_n^2$.

Example

Soient deux variables aléatoires indépendantes X et Y telles que $X \rightsquigarrow \mathcal{N}(0,1)$ et $Y \rightsquigarrow \chi^2_n$.On peut montrer que la variable aléatoire $T_n = \frac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi de Student à n degrés de libertés notée $T_n \rightsquigarrow \mathcal{T}_n$ de densité

$$f(x) = \frac{1}{\sqrt{n}B\left(\frac{1}{2}, \frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \forall x \in \mathbb{R}.$$

Pour $n = 1$ on a $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ qui est la densité de la loi de Cauchy.

Example

Soient deux variables aléatoires indépendantes X et Y telles que $X \rightsquigarrow \chi^2_p$ et $Y \rightsquigarrow \chi^2_q$.

Alors, la variable aléatoire $F(x,y) = \frac{\frac{X}{p}}{\frac{Y}{q}} \rightsquigarrow \mathcal{F}(p,q)$ loi de Fisher à p et q degrés de libertés de densité

$$f(x) = \frac{1}{B\left(\frac{p}{2}, \frac{q}{2}\right)} \frac{\left(\frac{p}{q}\right)^{\frac{p}{2}} x^{\frac{p}{2}-1}}{\left(1 + \frac{p}{q}x\right)^{\frac{p+q}{2}}}, x \geq 0,$$

avec la fonction béta $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \alpha > 0, \beta > 0.$