Les variables aléatoires multivariées

Introduction

Souvent, nous nous intéressons à des statistiques avec plus d'une Variable.

Définition

Une variable aléatoire multivariée est un vecteur de fonctions $Z(\cdot) : S \rightarrowtail \mathbb{R}^k$, $k > 1$.

Nous commencerons avec $k = 2$ ou $Z = (X, Y)$; on peut généraliser pour $k > 2$ Points d'intérêts:

- La distribution jointe (X, Y);
- La distribution marginale de X et Y ;
- La distribution conditionnelle de $X|Y$ et $Y|X$;
- Leurs moments et transformations.

Examples

- Distribution jointes de l'âge et des années d'études, ou des salaires et de l'expérience
- Distribution conditionnelle des salaires par âge.
- Distribution marginale des salaires.

La Distribution Jointe

Définition

La distribution jointe (X, Y) notée $P_{X,Y}$ est :

$$P_{X,Y}(X, Y \in A),$$

Avec $A \subseteq \mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. la fonction

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

Représente la PDF de (X, Y).

En prenant $A = (-\infty, x] \times (-\infty, y]$ on obtient la CDF jointe de (X, Y)

$$F_{X,Y}(x, y) = P_{X,Y}(X \leq x, Y \leq y),$$

La Distribution Jointe

De plus, en fonction de X et Y, l'a probabilité d'un évènement, A
Dans $\mathbb{R}^2$ est :

$$P_{X,Y}((X,Y) \in \mathcal{A}) = \sum_{(x,y)\in\mathcal{A}} f_{X,Y}(x,y) = \sum_{(x,y)\in\mathcal{A}} P_{X,Y}(X = x, Y = y),$$

$$\begin{aligned} P_{X,Y}((X,Y) \in \mathcal{A}) &= \int_{\mathcal{A}} f_{X,Y}(x,y)\, \mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} 1_{\{(x,y)\in\mathcal{A}\}}\, f_{X,Y}(x,y)\, \mathrm{d}x\mathrm{d}y, \end{aligned}$$

avec $f_{X,Y}(x,y) = \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$ Sous l'hypothèse de continuité.

Exemple

Soit les probabilités jointes :

		Y			
		0	1	2	
	0	.20	.05	.10	.35
X	1	.10	.10	.05	.25
	2	.05	.20	.15	.40
		.35	.35	.30	1

On a , $P_{X,Y}(X = 0, Y = 0) = .20$. on remarque que

$$F_{X,Y}(2,2) = \sum_{i=0,1,2} \sum_{j=0,1,2} P_{X,Y}(X = i, Y = j) = 1,$$

Propriétés

Toute fonction $f(x, y)$ de telle sorte que $f(x, y) \geq 0$ pour tout $(x, y) \in \mathbb{R}^2$ et

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1$$

Est la PDF jointe d'un vecteur bivarié

La distribution marginale

La distribution marginale (maginal pdf) de (X,Y) en X=x est :

$$f_X(x) = \sum_y f_{X,Y}(x,y), \qquad f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}y$$

La CDF marginale est :

$$F_X(a) = P_X(X \leq a) = \sum_{x \leq a} f_X(x), \qquad F_X(a) = \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x$$

Exemple

Les mêmes probabilités que plus haut

		Y 0	1	2	
	0	.20	.05	.10	.35
X	1	.10	.10	.05	.25
	2	.05	.20	.15	.40
		.35	.35	.30	1

Ici,

$$P_X(X=0) = \sum_{j=0,1,2} P_{X,Y}(X=0, Y=j) = .20 + .05 + .10 = .35$$

Et $P_X(X=1) = .25$, $P_X(X=2) = .40$.

donc,

$$F_X(x) = \begin{cases} 0 & \text{if} \quad -\infty < x < 0 \\ .35 & \text{if} \quad 0 \leq x < 1 \\ .60 & \text{if} \quad 1 \leq x < 2 \\ 1 & \text{if} \quad 2 \leq x < \infty \end{cases}$$

Les probabilité conditionnelles

Souvent, les valeurs des variables aléatoires sont corrélées : connaitre la valeur que prend Y peut nous donner des informations sur la valeur que prend X, sauf si X et Y sont indépendants.

Définition

Pour y tel que $f_Y(y) > 0$, la probabilité conditionnelle pmf/pdf de X sachant $Y = y$ à x est x notée $f_{X|Y}(x|y)$ et vaut :

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

La CDF associée est :

$$F_{X|Y}(a|y) = \sum_{x \leq a} f_{X|Y}(x|y) = P(X \leq a | Y = y)$$

$$F_{X|Y}(a|y) = \int_{-\infty}^{a} f_{X|Y}(x|xy)\, dx$$

Exemple

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Toujours les mêmes probabilités :

		Y 0	1	2	
	0	.20	.05	.10	.35
X	1	.10	.10	.05	.25
	2	.05	.20	.15	.40
		.35	.35	.30	1

ici,

$$f_{X|Y}(0|0) = \frac{.20}{.35}, \qquad f_{X|Y}(1|0) = \frac{.10}{.35}, \qquad f_{X|Y}(2|0) = \frac{.05}{.35},$$

La somme fait toujours 1.

Moments

Les distributions jointes, conditionnelles et marginales ont les même moments que les distributions univariées.

<u>Exemples:</u> soit Z =

$$Z = \binom{X}{Y}$$

$$\mu_Z = \mathrm{E}[Z] = E\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = \begin{pmatrix} \mathrm{E}[X] \\ \mathrm{E}[Y] \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$$

$$\mathrm{V}[Z] = \mathrm{E}[(Z - \mu_Z)(Z - \mu_Z)'] = \mathrm{E}[ZZ'] - \mu_Z \mu_Z'$$

C'est-à-dire ,

$$\begin{pmatrix} \mathrm{E}[XX] & \mathrm{E}[XY] \\ \mathrm{E}[YX] & \mathrm{E}[YY] \end{pmatrix} - \begin{pmatrix} \mu_X \mu_X & \mu_X \mu_Y \\ \mu_Y \mu_X & \mu_Y \mu_Y \end{pmatrix},$$

On appelle ce terme la matrice des covariances :
La diagonales représentes les variances ; les valeurs non diagonales les covariances;

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] = \mathrm{E}[XY] - \mu_X \mu_Y$$

La loi des espérances itérées

La moyenne conditionnelle de X sachant $Y = y$ est

$$\mu_{X/Y=y} = \mathrm{E}[X/Y = y] = \int_{-\infty}^{+\infty} x \, f_{X/Y}(x/y)\,\mathrm{d}x$$

La Loi des espérances itérées nous donne :

$$\begin{aligned}
\mathrm{E}_Y\left[\mathrm{E}[X|Y]\right] &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x \, f_{X|Y}(x|y)\,\mathrm{d}x\right] f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} x \left[\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}y\right] dx \\
&= \int_{-\infty}^{+\infty} x\,[f_X(x)]\,\mathrm{d}x = \mathrm{E}[X]
\end{aligned}$$

C'est un résultat fondamental. Details

L'espérance de l'espérance conditionnelle de X sachant Y est l'espérance de X.

Pour toute fonction g:

$$\mathrm{E}[g(X)] = \mathrm{E}_Y\left[\mathrm{E}[g(x)/Y]\right]$$

Loi de la variance totale

La loi de la variance totale nous donne la décomposition suivante :

$$\mathrm{V}[X] = \mathrm{E}[\mathrm{V}[X/Y]] + \mathrm{V}[\mathrm{E}[X/Y]].$$

Le premier terme représente la variance de X *sachant* Y ; le second terme la variance de X dans chaque groupe.

Details

Propriété des espérances conditionelles

Theoreme

Pour chaque fonction $r(\cdot)$ sur $S \rightarrow \mathrm{R}$, une propriété fondamentale est

$$\mathrm{E}[g(x)\mathrm{E}[Y/X]] = \mathrm{E}[g(x)Y]$$

Cela implique

$$\mathrm{E}[g(X)Y/X] = g(X)\mathrm{E}[Y/X]$$

et

$$\mathrm{E}[Y+Z/X] = \mathrm{E}[Y/X] + \mathrm{E}[Z/X]$$

$$\mathrm{E}[cY/X] = c\mathrm{E}[Y/X]$$

Exemple

$$f_{X,Y}(x,y) = \begin{cases} 6(x-y) & 0 \le y \le x \le 1 \\ 0 & \text{elsewhere} \end{cases} \quad \text{and} \quad \begin{array}{l} f_X(x) = 3x^2 \\ f_Y(y) = 3(y-1)^2 \end{array}$$

Il faut remarquer que

$$\mathrm{E}[Y|X=x] = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x)dy = \int_0^x y \left[\frac{2(x-y)}{x^2}\right] \mathrm{d}y = \frac{x}{3}.$$

En appliquant la LEI :

$$\begin{aligned} \mathrm{E}[Y] = E\left[\mathrm{E}[Y|X]\right] &= \int_{-\infty}^{+\infty} \mathrm{E}[Y|x] f_X(x)dx \\ &= \int_0^1 \mathrm{E}[Y|x][3x^2]\ \mathrm{d}x = \int_0^1 \frac{x}{3}[3x^2]\ \mathrm{d}x = \left[\frac{1}{4}x^4\right]_0^1 = \frac{1}{4} \end{aligned}$$

Si l'on calcule la moyenne depuis la pdf marginale de Y :

$$\mathrm{E}[Y] = \int_{-\infty}^{+\infty} y f_Y(y)dy = \int_0^1 y\ [3(y-1)^2]\ \mathrm{d}y = \left[\frac{3}{4}y^4 - 2y^3 + \frac{3}{2}y^2\right]_0^1 = \frac{1}{4},$$

Propriétés

Si X et Y deux r.v. alors :

$$\mathrm{V}[X+Y] = \mathrm{V}[X] + \mathrm{V}[Y] + 2\mathrm{Cov}(X, Y)$$
$$\mathrm{V}[X-Y] = \mathrm{V}[X] + \mathrm{V}[Y] - 2\mathrm{Cov}(X, Y)$$

Pour une matrice A,

$$\mathrm{V}[AX] = A\,\mathrm{V}[X]\,A^j$$

Pour des constantes, $a,\ b,\ c,\ d$,

$$\mathrm{Cov}(a+bX,\ c+dY) = bd\ \mathrm{Cov}(X, Y)$$

À noter : $\mathrm{Cov}(X, X) = \mathrm{V}[X]$.

Indépendance

Définition

X et Y sont indépendants si

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

Cela implique $f_{X|Y}(x|y) = f_X(x)$.

EN pratique , X et Y sont indépendants si connaitre la distribution de l'un ne nous donne aucune information sur la distribution de l'autre .

<u>Observation:</u> (sans démonstration)

Si X et Y sont indépendants, $g_1(X)$ et $g_2(Y)$ aussi.
Si la PDF jointe de X et Y est une fonction de X et Y seulement, X et Y sont indépendants.
Si X et Y sont indépendants, alors $E[XY] = E[X]E[Y]$

L'indépendance de la moyenne

Définition

L'indépendance de la moyenne :

$$E[X/Y = y] = E[X] \text{ pout tout } y$$

C'est une hypothèse plus faible mais qui implique par LIE :

$$E[g(Y)X] = E\left[g(Y)E[X|Y]\right] = E[g(Y)]\, E[X]$$

En particulier : soit $U = X - E[X/Y]$ la déviation de X de sa moyenne conditionnelle, on a :

$$E[U/Y] = 0, \qquad E[U] = 0, \qquad E[Y\,U] = 0,$$

Et donc $Cov(Y, U) = 0$.

C'est un concept plus faible que l'indépendance, l'indépendance l'implique mais pas la réciproque. De même ce n'est pas parce que X est m-indépendant par rapport à Y que Y l'est par rapport à X.

Indépendance et Covariance

L'indépendance implique zéro covariance, la réciproque est fausse.
L'indépendance de la moyenne implique zéro covariance, la réciproque est fausse.

Si $\mathrm{Cov}(X, Y) = 0$, alors

$$\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$$

Par définition de la covariance.

Exemple

Prenons $Y = X^2$. évidement , Y et X sont dépendants. Cependant, si $\mathrm{E}[X] = 0$ and $\mathrm{E}[X^3] = 0$ (par Example si X suit une loi normale), alors :

$$\mathrm{Cov}(X, X^2) = \mathrm{E}[X^3] - \mathrm{E}[X]\,\mathrm{E}[X^2] = 0.$$

Le signe de la covariance nous indique dans quel sens les variables seront proportionnelles, sa magnitude dans quelle ampleur, mais elle est difficile à interpréter, c'est pourquoi on préfèrera utiliser la corrélation.

La Corrélation

Il s'agit d'une normalisation utile de la covariance

Définition

La corrélation se définit par :

$$\rho_{X,Y} = \mathrm{E}\left[\left(\frac{X-\mathrm{E}[X]}{\sigma_X}\right)\left(\frac{Y-\mathrm{E}[Y]}{\sigma_Y}\right)\right]$$
$$= \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]}{\sigma_X\sigma_Y}.$$

L'avantage est que l'analyse dimensionnelle de la corrélation vaut 1, elle ne dépend pas de l'unité de X et Y et donne une relation linéaire entre les deux

Exemple:
si $Y = a + bX$, then

$$\rho_{X,Y} = \frac{b\sigma_X^2}{|b|\sigma_X\sigma_X} = \begin{cases} -1 & \text{if } b<0 \\ 0 & \text{if } b=0 \\ 1 & \text{if } b>0 \end{cases} = \operatorname{sign} b$$

Cauchy-Schwarz

Théorème

L'inégalité de Cauchy Schwartz énonce :

$$\mathrm{E}[(XY)]^2 \leq \mathrm{E}[X^2]\ \mathrm{E}[Y^2].$$

Tout d'abord :

$$0 \leq \mathrm{E}[(cX+Y)^2] = c^2\mathrm{E}[X^2] + 2c\mathrm{E}[XY] + \mathrm{E}[Y^2].$$

En prenant $c = \mathrm{E}[XY]/\mathrm{E}[X^2]$, on retrouve le

résultat .

CS peut aussi s'écrire : $Cov(X,Y)^2 \leq V(X)V(Y)$

Corrélation

Grace à cette inégalité on sait que la corrélation est bornée :

$$|\rho_{X,Y}| \leq 1$$

En prenant $X - \mu_X$ et $Y - \mu_Y$, on obtient :

$$\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]^2 \leq \mathrm{E}[(X-\mu_X)]^2\mathrm{E}[(Y-\mu_Y)]^2$$

$$\frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]^2}{\mathrm{E}[(X-\mu_X)]^2\mathrm{E}[(Y-\mu_Y)]^2} \leq 1$$

$$\frac{\mathrm{Cov}(X,Y)^2}{\sigma_X\sigma_Y} \leq 1$$

$$\rho^2_{X,Y} \leq 1$$

L'erreur quadratique moyenne [MSE]

On souhaite prédire X, si l'on postule que X vaudra c, le carré de l'erreur associé sera donc $(X - c)^2$.

définition

Soit c une prédiction de X. la MSE pour c sera $E[(X - c)^2]$.

De plus,

$$E[(X - c)^2] = E[(X - \mu_X)^2] + (\mu_X - c)^2,$$

Cela équivaut à la somme de la variance et du Biais au carré.
le biais est un estimateur de la distance entre l'espérance de X et c.

La moyenne minimise la MSE. C'est donc le meilleur predicteur de X selon l'erreur quadratique

Le meilleur prédicteur selon l'erreur absolue résous :

$$\underset{c}{\operatorname{argmin}} E[|X - c|]$$

Il s'agit de la médiane.

Le meilleurs prédicteur

Le meilleurs prédicteur de X sachant $Y = y$ (au sens MSE) est

$$\arg\min_c \mathrm{E}[(X-c)^2 | Y = y] = \mu_{X|Y=y} = \mathrm{E}[X|Y]$$

Par LIE :

$$\mathrm{E}[(X - \mathrm{E}[X/Y])^2] \leq \mathrm{E}[(X - c(Y))^2]$$

Pour toute fonction c $c(\cdot)$.

L'espérance conditionnelle est le meilleur prédicteur de X|Y au sens MSE.

Ce théorème sert de fondement à l'analyse par régression. Dans le cadre d'une régression paramétrique, on explique une variable indépendante Y par un vecteur X de variables explicatives (ou indépendantes) selon un modèle de type $Y = g(X;\ \theta_0) + U$, où $g(X;\ \theta)$ est une fonction de x et θ un vecteur de paramètres , U représente le terme d'erreur, que l'on suppose répondre à la condition $\mathrm{E}[U/X] = 0$. Le problème est donc d'estimer le vecteur des paramètres : θ_0.

Le meilleurs prédicteur linéaire

Connaissant des variables aléatoires X , Y , on souhaite prédire Y *depuis* X. Sauf cas particulier, l'espérance conditionnelle peut être une fonction compliquée à définir $\mathrm{E}[Y / X]$ on va donc chercher à l'approximer par une fonction de la forme $Y = \alpha + \beta X$.

Le meilleurs prédicteur linéaire de Y est le prédicteur minimisant la MSE .

Il résout $$\arg\min_{\gamma} \mathrm{E}\left[(Y-(\alpha-\beta X))^2\right]$$

Et vaut :

$$\tilde{\beta} = \frac{\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]}{\mathrm{E}[X^2]-\mathrm{E}[X)^2}$$

$$\tilde{\alpha} = \mu_Y - \tilde{\beta}\mu_X$$

Un estimateur commun du meilleurs prédicteur linéaire est l'OLS

Annexes

Le théorème de la variance totale peut être démontré en utilisant la formule des espérances totales[4]. Tout d'abord

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - [\mathrm{E}[Y]]^2$$

par définition de la variance. On applique ensuite la formule des espérances totales à chaque terme en conditionnant par la variable aléatoire X

$$= \mathrm{E}\left[\mathrm{E}[Y^2 \mid X]\right] - [\mathrm{E}[\mathrm{E}[Y \mid X]]]^2$$

On réécrit alors le moment conditionnel d'ordre 2 de Y en termes de sa variance et de son moment d'ordre 1 :

$$= \mathrm{E}\left[\mathrm{Var}[Y \mid X] + [\mathrm{E}[Y \mid X]]^2\right] - [\mathrm{E}[\mathrm{E}[Y \mid X]]]^2$$

Puisque l'espérance est linéaire, les termes peuvent être regroupés :

$$= \mathrm{E}[\mathrm{Var}[Y \mid X]] + \left(\mathrm{E}[[\mathrm{E}[Y \mid X]]^2] - [\mathrm{E}[\mathrm{E}[Y \mid X]]]^2\right)$$

Finalement, les termes entre parenthèses peuvent être vus comme la variance de l'espérance conditionnelle E[Y|X] :

$$= \mathrm{E}[\mathrm{Var}[Y \mid X]] + \mathrm{Var}[\mathrm{E}[Y \mid X]]$$

Back

$$g(y) = \mathrm{E}[X|Y] = \int_{-\infty}^{+\infty} x \ f_{X|Y}(x|y) \ \mathrm{d}x$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \Rightarrow f_{X|Y}(x|y) f_Y(y) = f_{X,Y}(x,y)$$

$$\begin{aligned}
E_Y\left[\mathrm{E}[X|Y]\right] &= E_Y[g(y)] = \int_{-\infty}^{+\infty} g(y) f_Y(y) dy \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x \ f_{X|Y}(x|y) \ \mathrm{d}x\right] f_Y(y) \ \mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \ f_{X|Y}(x|y) f_Y(y) dx dy \\
&= \int_{-\infty}^{+\infty} x \left[\int_{-\infty}^{+\infty} f_{X,Y}(x,y) \, \mathrm{d}y\right] dx \\
&= \int_{-\infty}^{+\infty} x \left[f_X(x)\right] \mathrm{d}x \\
&= \mathrm{E}[X]
\end{aligned}$$

Back