Probabilité et statistiques

L2 prépa

William Hodonou

Novembre 2020

Outline

Introduction

Nous n'avons traité jusqu'ici que de distributions de variables isolées. Or, il est souvent nécessaire de considérer des événements relatifs à deux variables simultanément, ou même à plus de deux variables.

Example:

Un site Web mobile est accessible à partir d'un smart phone; X est la force du signal, en nombre de barres, et Y est le temps de réponse, à la seconde près.

On considère un arrêt de bus. On va définir 2 v.a. :
$X =$ « heure d'arrivée du bus »
$Y =$ « heure d'arrivée de la personne »
Pour estimer la probabilité que la personne rate ou prenne le bus, on va travailler sur le couple (X, Y)

Loi de probabilité jointe (cas discret)

Dans le cas où X et Y sont deux variables discrètes, il est commode de définir la fonction p suivante, dite loi de probabilité simultanée ou jointe de X et Y:

$$P(x, y) = P(X = x, Y = y); \text{(la virgule équivaut à « et »).}$$

$$0 \leq P(X = x, Y = y) \leq 1$$
$$\sum_x \sum_y P(X = x, Y = y) = 1$$

Loi de probabilité jointe (cas discret)

La loi de probabilité marginale de X s'en déduit ainsi:

Since X is a random variable, it also has its own probability distribution, ignoring the value of Y , called its marginal probability distribution.

$$P_X(x) = P(X = x) = \sum_y P(x, y)$$

De façon similaire

$$P_Y(y) = P(Y = y) = \sum_x P(x, y)$$

NB :

Connaissance de $P(X = x, Y = y) => P(X = x)$ et $P(Y = y)$ LA RECIPROQUE EST FAUSSE

Loi de probabilité jointe (cas discret)

X / Y	y_1	y_2	...	y_m	Somme
x_1	$P(X = x_1, Y = y_1)$	$P(X = x_1, Y = y_2)$	...	$P(X = x_1, Y = y_m)$	$P(X = x_1)$
x_2					$P(X = x_2)$
...					
x_n	$P(X = x_n, Y = y_1)$			$P(X = x_n, Y = y_n)$	$P(X = x_n)$
Somme	$P(Y = y_1)$	$P(Y = y_2)$	...	$P(Y = y_m)$	1 C.N.

Exercice :

Un site Web mobile est accessible à partir d'un smartphone; X est la force du signal, en nombre de barres, et Y est le temps de réponse, à la seconde près.

Y/X	1	2	3
4+	0.15	0.10	0.05
3	0.02	0.10	0.05
2	0.02	0.03	0.20
1	0.01	0.02	0.25

Exercice :

Y/X	1	2	3	Marginale
4+	0.15	0.10	0.05	0.30
3	0.02	0.10	0.05	0.17
2	0.02	0.03	0.20	0.25
1	0.01	0.02	0.25	0.28
Marginale	0.20	0.25	0.55	1.00

Cas des v.a continues

On parle de densité de probabilité jointe.

soit X et Y 2 v.a.c, (X,Y) prennent des valeurs dans un domaine D du plan. La densité de probabilité jointe du couple (X, Y) est la fonction f(x, y) telle que :

$$P(x \leq X \leq x + dx, y \leq y + dy) = f(x, y)dxdy$$

NB : Soit A un domain inclus dans D

$$p((x, y) \in A) = \iint_{(x,y)\in A} f(x, y).dxdy$$

Propiété de f(x,y):

$$f(x, y) \geq 0$$

$$\int \int_D f(x, y).dxdy = 1$$

Densités de probabilités marginales de X et Y :

Enfin, si X et Y sont des variables aléatoires conjointement continues, alors elles sont individuellement continues, également. On obtient leurs densités marginales ainsi:

$$f_X(x) = \int_{y_1}^{y_2} f(x, y).dy$$

$$f_Y(y) = \int_{x_1}^{x_2} f(x, y).dx$$

NB :

Connaissance de$f(x, y) => f_X(x)$ et $f_Y(y)$ LA RECIPROQUE EST FAUSSE

Exemple

Supposons que $K(x^2 + y^2)$ est une fonction de densité pour la distribution conjointe de la variable aléatoire continue X et Y définie sur le carré unitaire délimité par les points (0,0) (1,0) (1,1) (0,1) trouver K , trouvez $P(X + Y \geq 1)$. Trouver les distributions marginales de X et Y.

Fonction de répartition conjointe

On définit pour traiter de tels problèmes une fonction F de répartition simultanée, ou conjointe, pour toute paire de variables aléatoires X et Y:

$$F(x, y) = P(X \leq x, Y \leq y) = P[(X \leq x) \cap (Y \leq y)]$$

Dans le cas des v.a continues:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dtds$$

Fonction de répartition conjointe

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

cas discret

$$F(x,y) = \sum_{s=-\infty}^{x} \sum_{t=-\infty}^{y} f(s,t)$$

V.a. indépendantes

Les v.a. X et Y sont indépendantes si :

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

ou

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

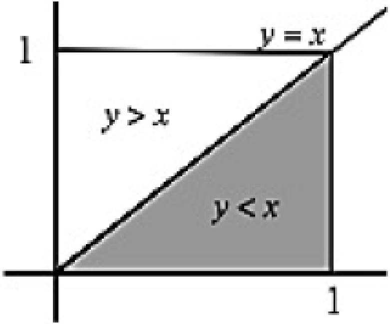
Exemple

Supposons que X et Y sont des variables aléatoires continues indépendantes avec les fonctions de densité suivantes$f_X(x) = 1$ si $0 < x < 1$ et $f_Y(y) = 2y$ si $0 < y < 1$. Trouver $P(Y < X)$

V.a. indépendantes

solution

$P[Y < X] = \int_0^1 \int_0^x 2y\,dy\,dx = \frac{1}{3}.$



Espérance de h(X, Y)

Dans le cas des v.a. doubles :

Si X et Y discrètes : $E[h(X,Y)] = \sum_x \sum_y h(x,y) P(X = x, Y = y)$

Si X et Y continues : $E[h(X,Y)] = \iint_D h(x,y) f(x,y).dxdy$

Variance de h(X, Y)

$$Var[h(X,Y)] = E[(h(X,Y) - E[h(X,Y)])^2] = E[h^2(X,Y)] - (E[h(X,Y)])^2$$

Covariance

Soit X et Y deux v.a. alors la covariance de X et Y est :

$$\boxed{cov(X,Y) = E(XY) - E(X)E(Y)}$$

Propriété

Si X et Y indépendantes, alors $cov(X,Y) = 0$

Application :

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abcov(X,Y)$$

Coefficient de corrélation

Définition

Soit X et Y deux v.a. Le coefficient de corrélation est défini par :

$$\boxed{\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}}$$

On a $-1 \leq \rho \leq 1$