Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Définition et distribution jointe

Le plus souvent, on s'intéresse à plusieurs v.a.

Dans ce cas, les outils du chapitre 2 apportent de l'information sur chaque v.a, mais ils ne permettent pas de capturer le lien entre ces diverses v.a.

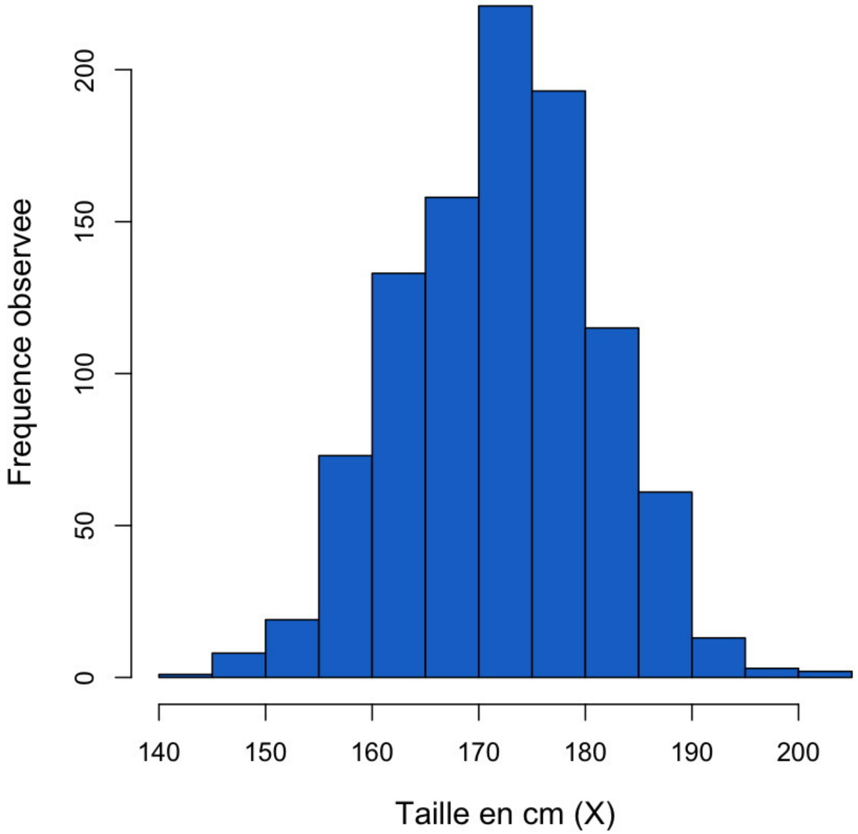
Or ce lien est souvent d'un grand intérêt, que ce soit

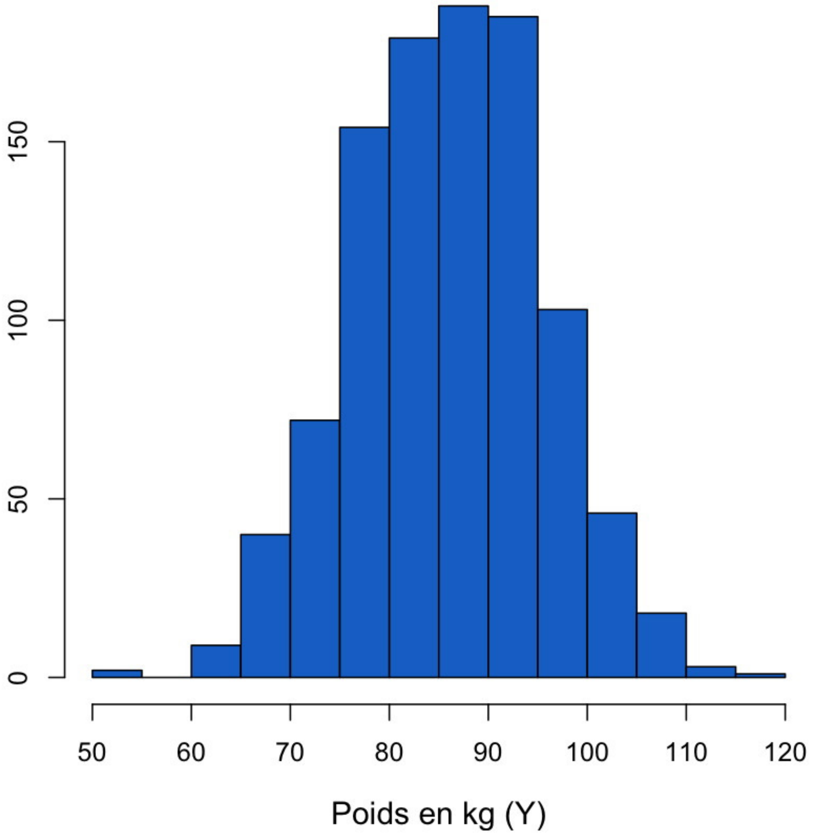
- pour faire de la prévision d'une variable sur la base des autres x
- pour détecter d'éventuelles observations particulières
- ...

Exemple :

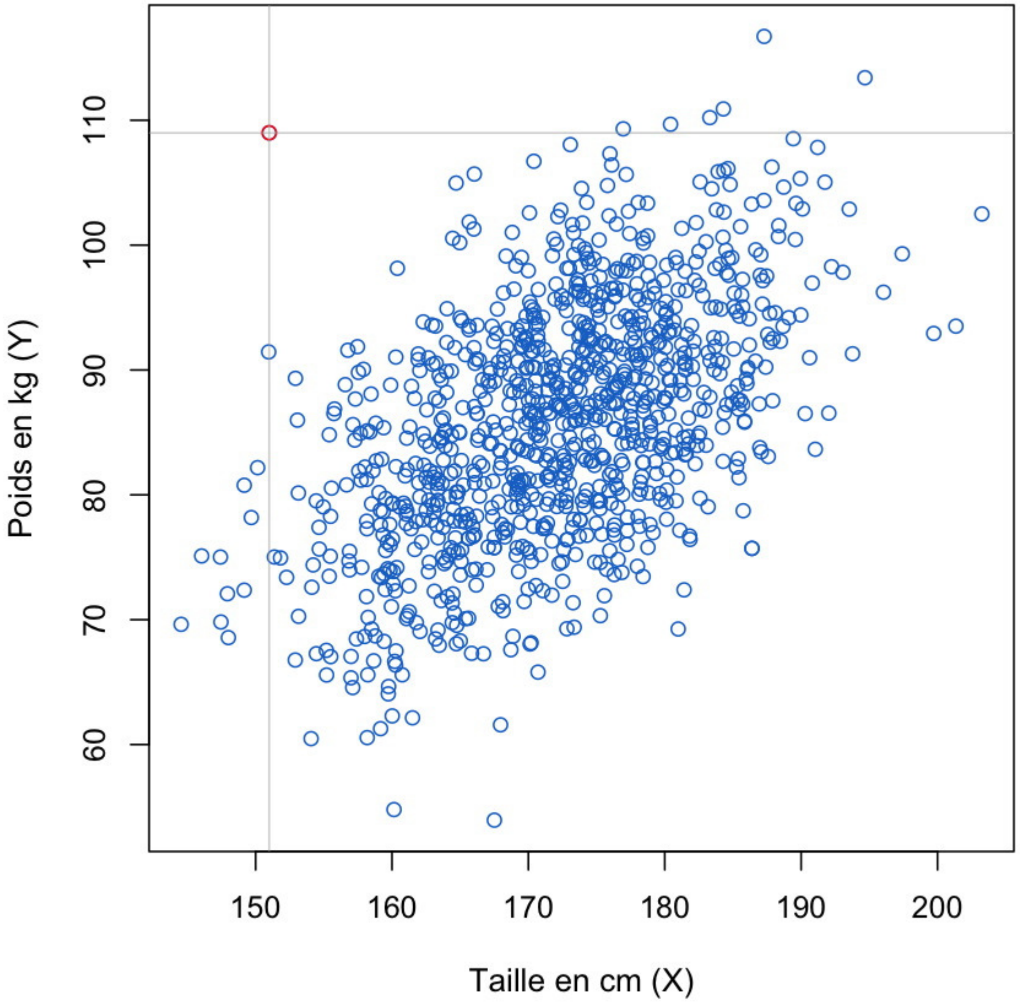
$X =$ la taille d'un être humain (en cm)
$Y =$ le poids de la même personne (en kg)
On mesure X et Y sur 1000 personnes...

Définition et distribution jointe



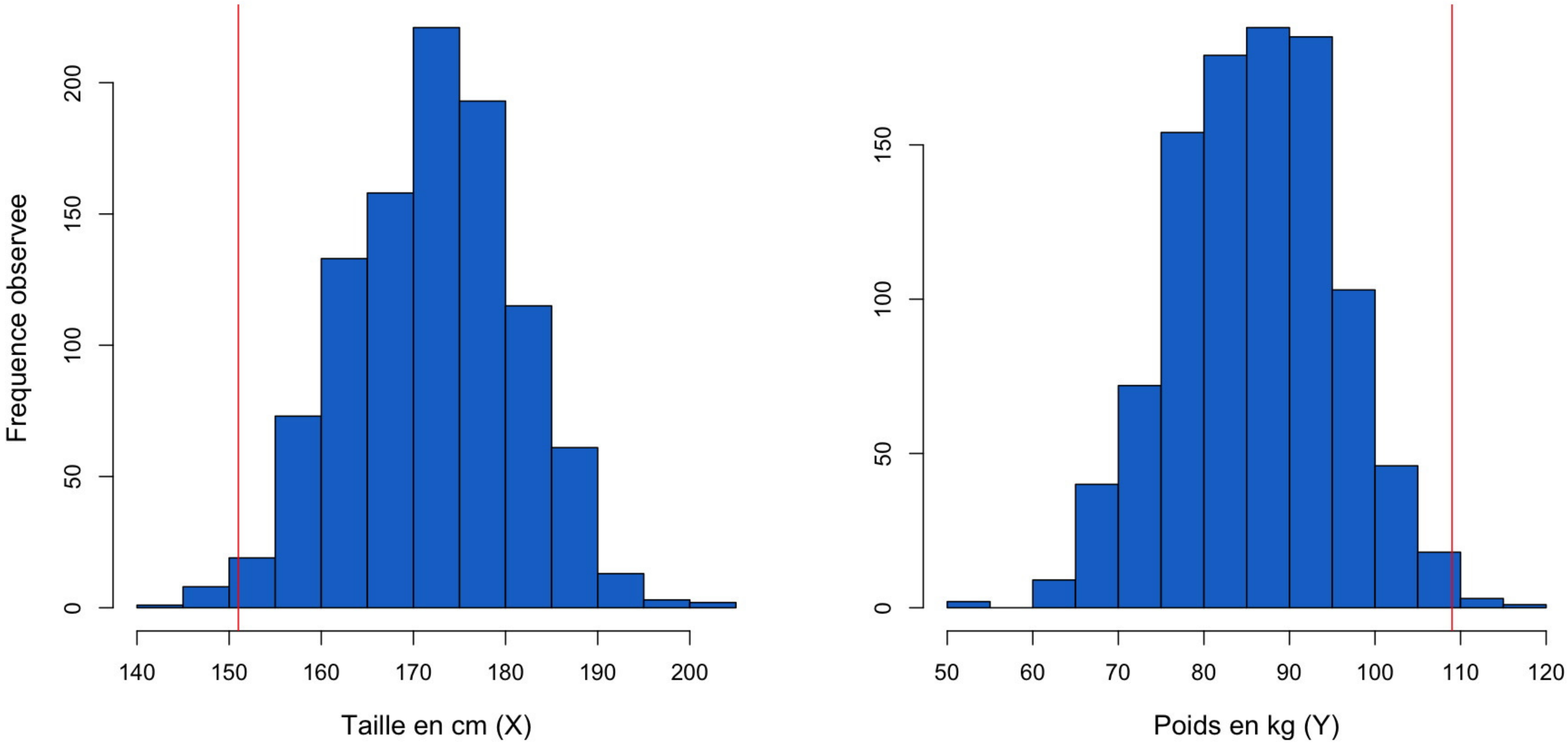


Définition et distribution jointe



Un individu "à risque"...

Définition et distribution jointe



... Masqué dans les distributions individuelles de X et de Y

Définition et distribution jointe

Un autre exemple :

"En bourse, il faut diversifier pour diminuer le risque".

Soient $A_1, \ldots, A_k$ des actifs financiers.
Soient $X_1, \ldots, X_k$ les valeurs (aléatoires !) de ces actifs (en €).
Un portefeuille est une quantité (aléatoire) agrégée, du type

$$Z = c_1 X_1 + c_2 X_2 + \ldots + c_k X_k,$$

où c_i est le nombre de titres A_i possédés.

Le risque $\mathrm{Var}[Z]$ du portefeuille dépend

- des risques individuels $\mathrm{Var}[X_i]$, mais aussi
- de la dépendance entre les X_i

(c'est une très mauvaise idée de ne posséder que des actions du secteur bancaire !)

Définition et distribution jointe

Soit une expérience aléatoire E. Soit un espace probabilisé $(\Omega, \mathcal{A}, P)$ associé.

Définition

Un vecteur aléatoire (v.a.!) (bivarié) est une fonction

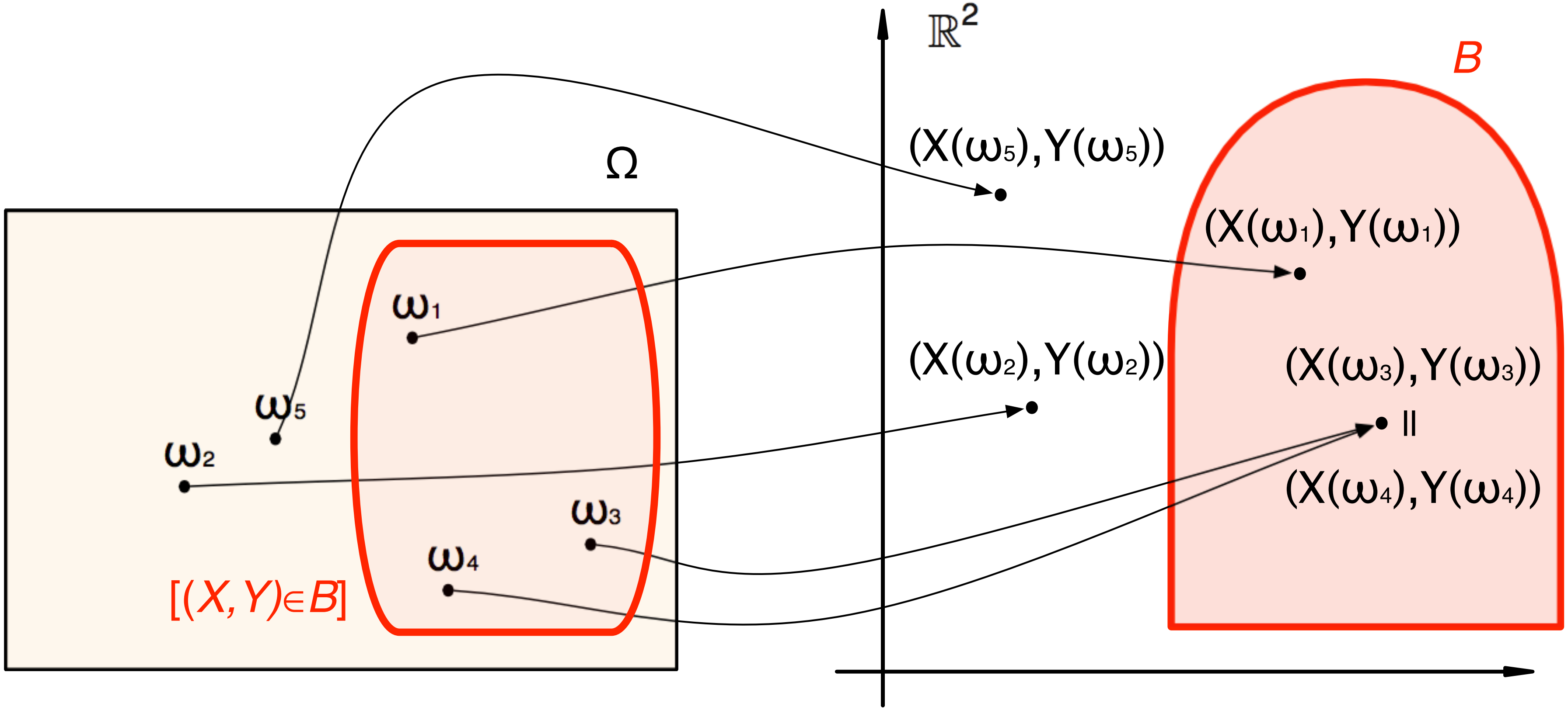
$$\begin{aligned}(X, Y) : \Omega &\rightarrow \mathbb{R}^2 \\ \omega &\mapsto (X(\omega), Y(\omega))\end{aligned}$$

vérifiant la condition technique que, pour tout borélien $B \in \mathcal{B}^2$,

$$[(X, Y) \in B] \in \mathcal{A} \qquad (*)$$

où $[(X, Y) \in B] := \{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}$.

Plus tard, on considérera aussi des v.a. k-variés $(X_1, X_2, \ldots, X_k)$.



$\mathbb{R}^2$
B
Ω
$(X(\omega_5),Y(\omega_5))$
$(X(\omega_1),Y(\omega_1))$
$(X(\omega_2),Y(\omega_2))$
$(X(\omega_3),Y(\omega_3))$
$\parallel$
$(X(\omega_4),Y(\omega_4))$
ω_1
ω_5
ω_2
ω_3
ω_4
$[(X,Y)\in B]$

Distribution jointe, fonction de répartition

Comme dans le cas des variables aléatoires, l'ensemble des valeurs possibles $\{(X(\omega), Y(\omega)) : \omega \in \Omega\}$ ne décrit le v.a. (X, Y) que très partiellement.

Ce qui décrit complètement (X, Y) est

- sa distribution jointe, c'est-à-dire la mesure de probabilité

$$\begin{array}{rccl} P^{(X,Y)} : & \mathcal{B}^2 & \to & [0,1] \\ & B & \mapsto & P^{(X,Y)}[B] = P[(X,Y) \in B], \end{array}$$

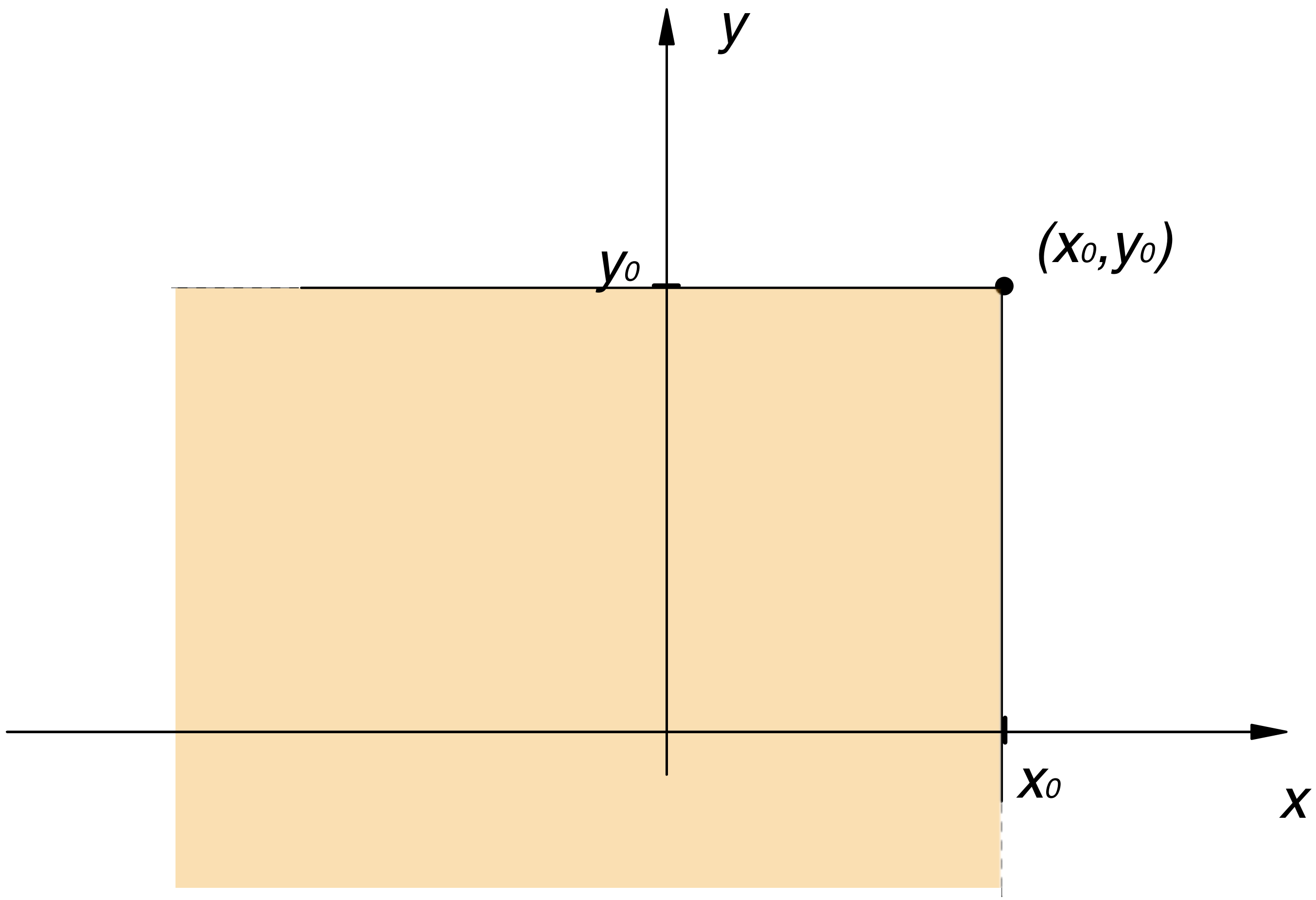
- ou, de manière équivalente, sa fonction de répartition.

Définition

La fonction de répartition de (X, Y) est la fonction

$$\begin{array}{rccl} F : & \mathbb{R}^2 & \to & [0,1] \\ & (x,y) & \mapsto & P[[X \leq x] \cap [Y \leq y]] \stackrel{\text{not}}{=} P[X \leq x, Y \leq y]. \end{array}$$

Distribution jointe, fonction de répartition



$$F(x_0, y_0) = P[X \leq x_0, Y \leq y_0]$$

Distribution jointe, fonction de répartition

Comme annoncé, $P^{(X,Y)}$ et F fournissent la même information :

- A partir de $P^{(X,Y)}$, on peut calculer

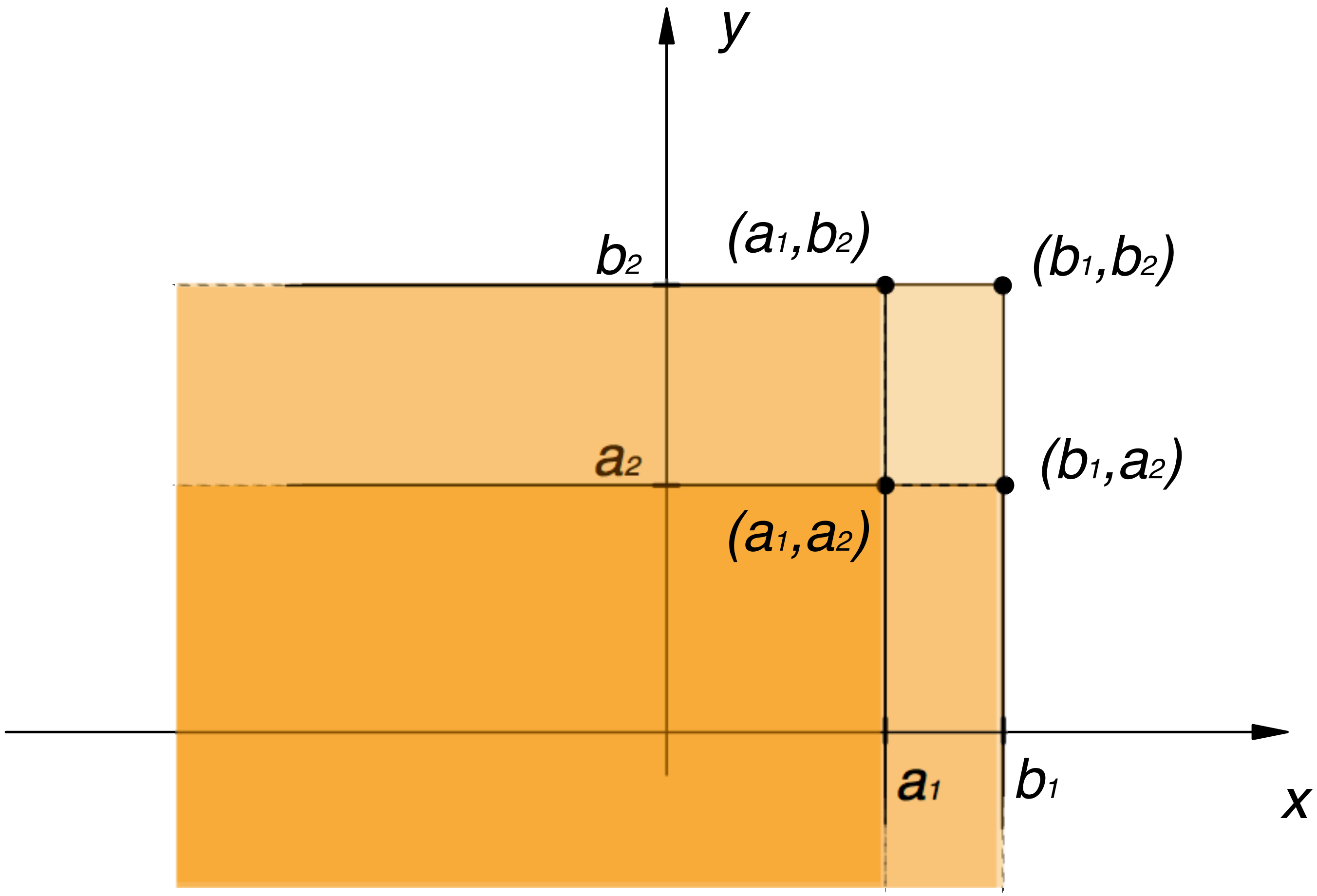
$$F(x,y) = P^{(X,Y)}[(-\infty, x] \times (-\infty, y]].$$

- A partir de F, on obtient

$$\begin{aligned} P^{(X,Y)}[(a_1, b_1] \times (a_2, b_2]] &= P[a_1 < X \leq b_1, a_2 < Y \leq b_2] \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2), \end{aligned}$$

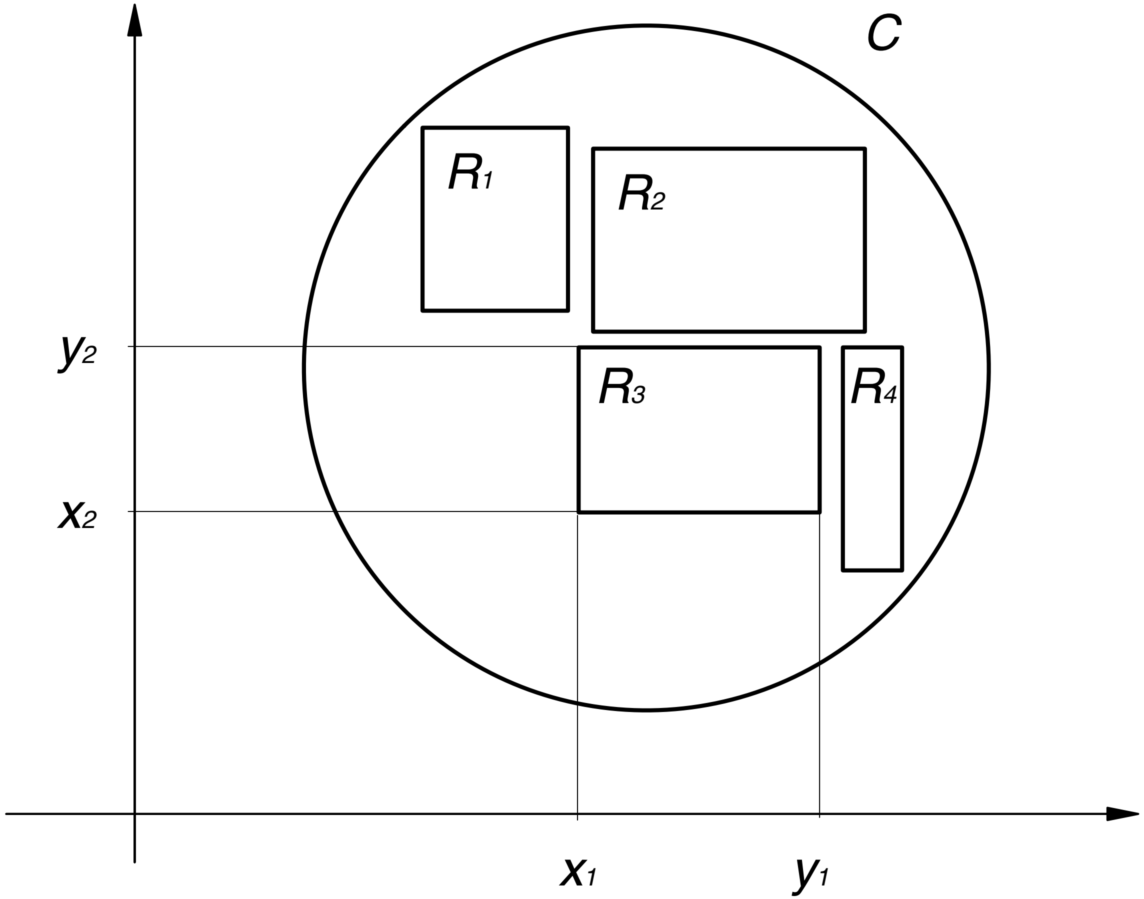
ce qui permet d'obtenir $P^{(X,Y)}[B]$ pour n'importe quel $B \in \mathcal{B}^2$ (puisque tout B est $\cup/\cap$ d'une collection dénombrable de tels rectangles $(a_1, b_1] \times (a_2, b_2]$).

Distribution jointe, fonction de répartition



$$P[a_1 < X \leq b_1, a_2 < Y \leq b_2] = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

Distribution jointe, fonction de répartition



On a $C = R_1 \cup R_2 \cup R_3 \cup \ldots$, ce qui permet d'obtenir $P[(X, Y) \in C]$
$= P[(X, Y) \in R_1] + P[(X, Y) \in R_2] + P[(X, Y) \in R_3] + \ldots$ à partir de F

Distribution jointe, fonction de répartition

F fournit l'information nécessaire pour obtenir $P[(X, Y) \in B] \; \forall B$.
Néanmoins, le lien entre F et ces probabilités n'est pas très explicite.
Comment alors calculer en pratique ces probabilités ?

Nous traiterons séparément les v.a. discrets et continus.

Définition

(X, Y) est discret $\Leftrightarrow$ l'ensemble des valeurs possibles $\{(X(\omega), Y(\omega)) : \omega \in \Omega\}$ est fini ou infini dénombrable.

Définition

(X, Y) est continu $\Leftrightarrow$ (plus tard : ch.3-p.25)

Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Distribution jointe et distributions marginales : le cas discret

Définition

(X, Y) est discret $\Leftrightarrow$ l'ensemble des valeurs possibles $\{(X(\omega), Y(\omega)) : \omega \in \Omega\}$ est fini ou infini dénombrable.

$E =$ lancer de deux dés (distinguables)
$\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
$X =$ somme des résultats de chaque dé
$Y =$ différence des résultats de chaque dé

Autrement dit,

$$\begin{array}{rccc} (X, Y) : & \Omega & \to & \mathbb{R}^2 \\ & (i,j) & \mapsto & (i+j, |i-j|). \end{array}$$

Puisque $X(\Omega) = \{2, 3, \ldots, 12\}$ et $Y(\Omega) = \{0, 1, \ldots, 5\}$ sont finis, (X, Y) est discret.

Distribution jointe et distributions marginales : le cas discret

Supposons que (X, Y) est discret.
Notons x_i, $i \in \mathcal{I}$, les valeurs possibles de X.
Notons y_j, $j \in \mathcal{J}$, les valeurs possibles de Y.
$\Rightarrow$ L'ensemble des valeurs possibles de $(X, Y) \subset \{(x_i, y_j) : i \in \mathcal{I}, j \in \mathcal{J}\}$.

La distribution jointe de (X, Y) est complètement caractérisée par le tableau des valeurs possibles et des probabilités correspondantes $p_{ij} = P[(X, Y) = (x_i, y_j)]$

	x_1	x_2	$\ldots$	(x_k)
y_1	p_{11}	p_{21}	$\ldots$	(p_{k1})
y_2	p_{12}	p_{22}	$\ldots$	(p_{k2})
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
(y_ℓ)	$(p_{1\ell})$	$(p_{2\ell})$	$\ldots$	$(p_{k\ell})$

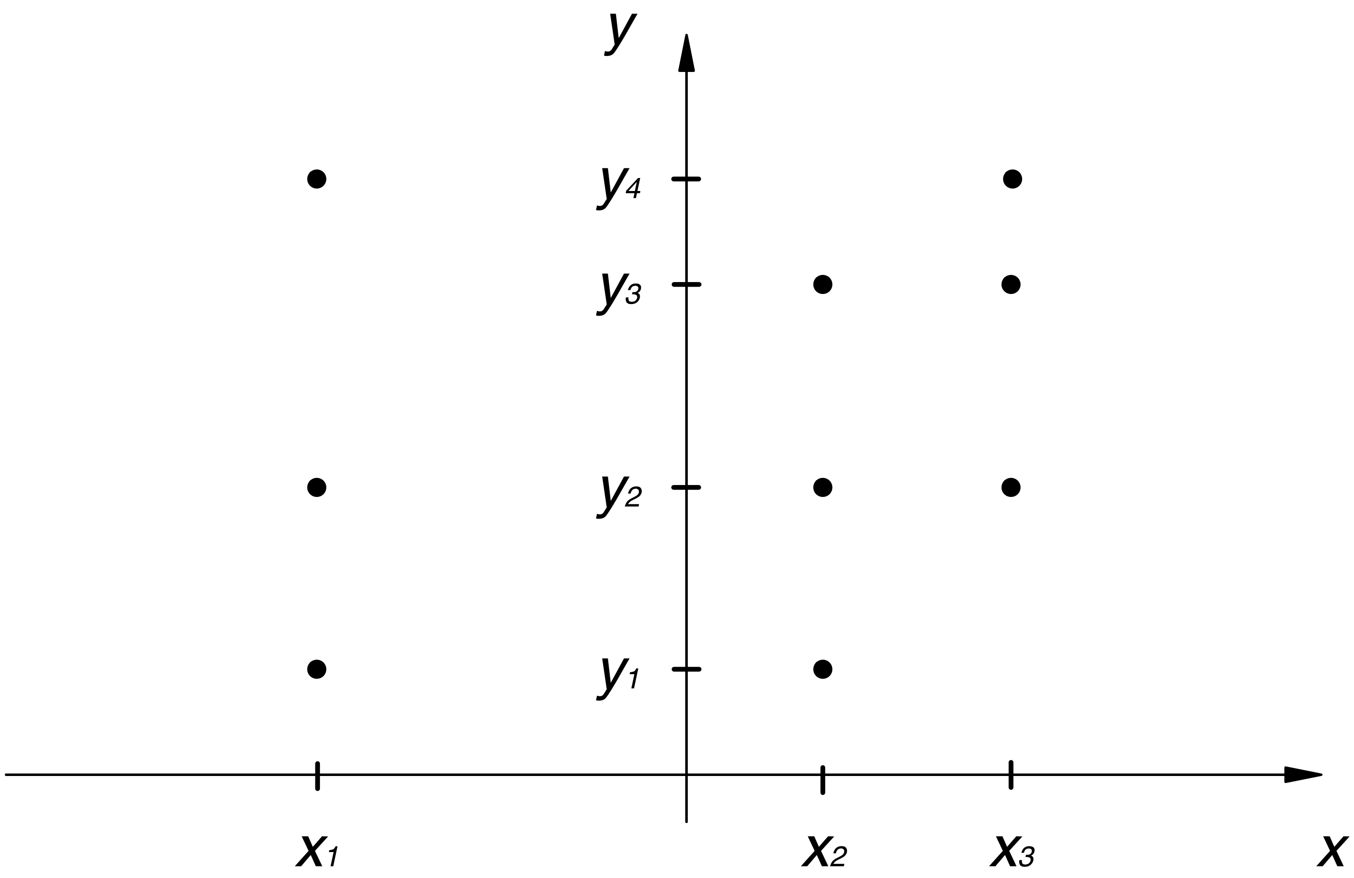
On a en effet

$$P[(X, Y) \in B] = \sum_{i,j:(x_i,y_j)\in B} P[(X, Y) = (x_i, y_j)] \quad \forall B \in \mathcal{B}^2.$$

Distribution jointe et distributions marginales : le cas discret

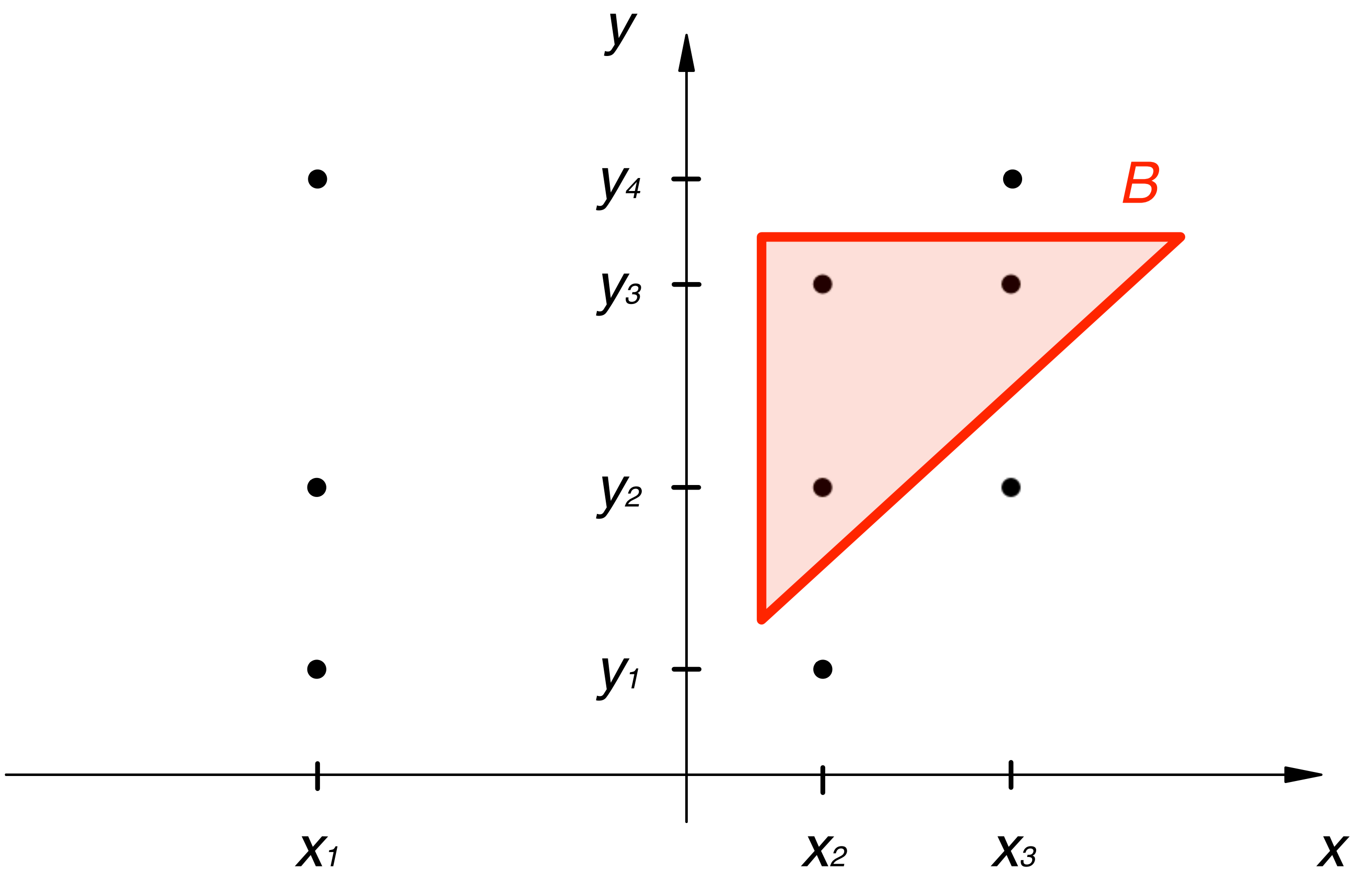
$E =$ lancer de deux dés (distinguables)
$\Omega = \{(1,1),(1,2),\dots,(1,6),(2,1),\dots,(6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
$X =$ somme des résultats de chaque dé
$Y =$ différence des résultats de chaque dé

	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
1		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$	
2			$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		
3				$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$			
4					$\frac{2}{36}$		$\frac{2}{36}$				
5						$\frac{2}{36}$					



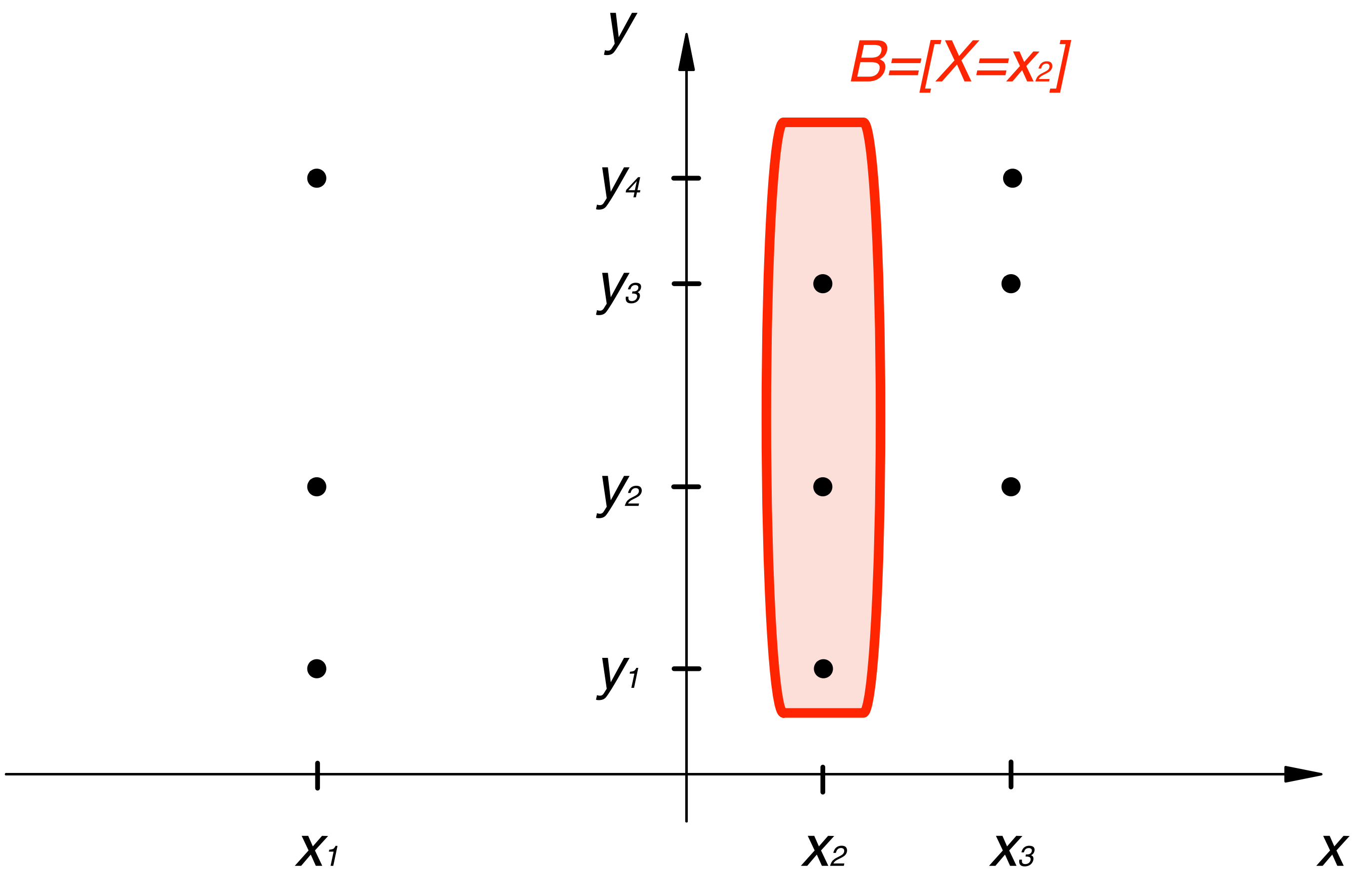
Collection des valeurs possibles pour (X, Y)

Distribution jointe et distributions marginales : le cas discret



$$P[(X,Y) \in B] = P[(X,Y) = (x_2, y_2)] + P[(X,Y) = (x_2, y_3)] + P[(X,Y) = (x_3, y_3)]$$

Distribution jointe et distributions marginales : le cas discret



$$P[X = x_2] = P[(X, Y) \in B] = P[(X, Y) = (x_2, y_1)] + \ldots + P[(X, Y) = (x_2, y_4)]$$

(cas particulier important)

Distribution jointe et distributions marginales : le cas discret

Ceci explique comment calculer la distribution de X à partir de la distribution jointe :

$$p_{i\bullet} := P[X = x_i] = \sum_{j \in \mathcal{J}} P[(X, Y) = (x_i, y_j)] = \sum_{j \in \mathcal{J}} p_{ij}$$

valeurs possibles	x_1	x_2	$\ldots$	(x_k)
probabilités	$p_{1\bullet}$	$p_{2\bullet}$	$\ldots$	$(p_{k\bullet})$

On parlera de distribution marginale ($><$ distribution jointe).

Cette distribution marginale est celle d'une variable aléatoire ($><$ vecteur aléatoire).
$\rightsquigarrow$ On peut calculer $P[X \in B]$, l'espérance de X, la variance de X, etc. Par exemple,

$$\mathrm{E}[X] = \sum_{i \in \mathcal{I}} x_i p_{i\bullet}$$

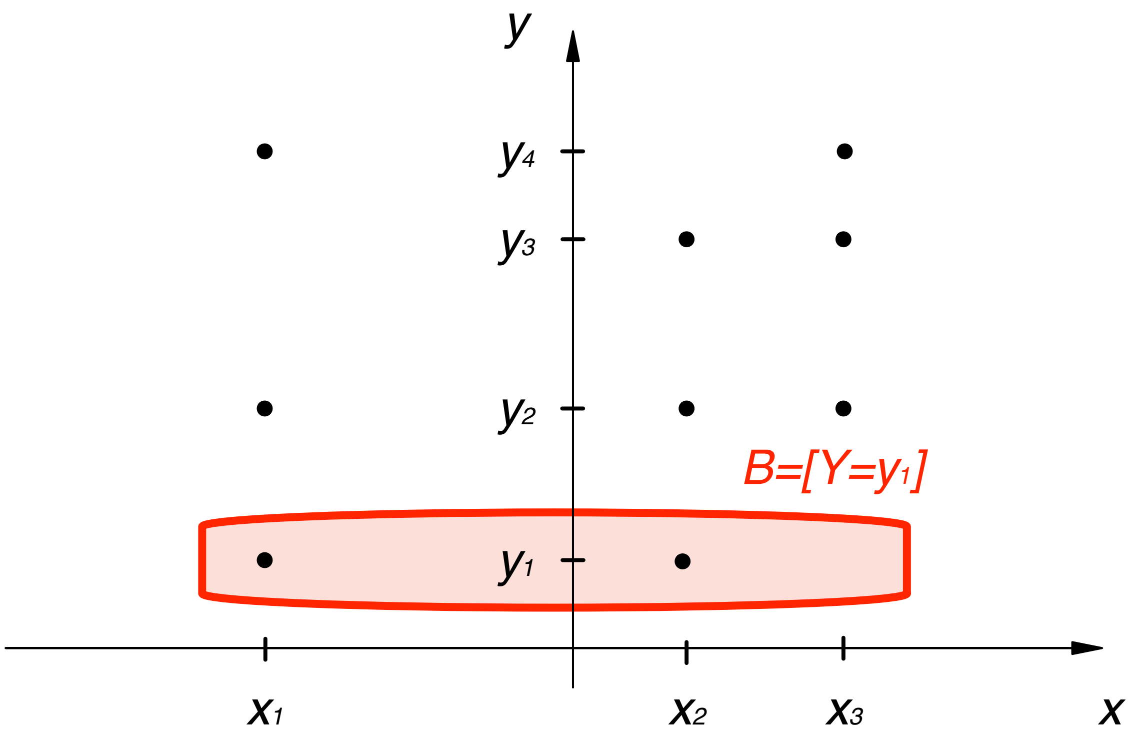
$$\mathrm{Var}[X] = \begin{cases} \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_{i \in \mathcal{I}} (x_i - \mathrm{E}[X])^2 p_{i\bullet} & \text{(pour l'interprétation)} \\ \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \sum_{i \in \mathcal{I}} (x_i)^2 p_{i\bullet} - (\mathrm{E}[X])^2 & \text{(pour le calcul)} \end{cases}$$

Distribution jointe et distributions marginales : le cas discret

De même, la distribution marginale de Y est donnée par

$$p_{\bullet j} := P[Y = y_j] = \sum_{i \in \mathcal{I}} P[(X, Y) = (x_i, y_j)] = \sum_{i \in \mathcal{I}} p_{ij}$$

valeurs possibles	y_1	y_2	$\ldots$	(y_ℓ)
probabilités	$p_{\bullet 1}$	$p_{\bullet 2}$	$\ldots$	$(p_{\bullet \ell})$



Distribution jointe et distributions marginales : le cas discret

On reporte souvent ces distributions marginales dans le tableau de la distribution jointe (auquel on peut aussi ajouter les espérances et variances marginales) :

	x_1	x_2	$\ldots$	(x_k)			
y_1	p_{11}	p_{21}	$\ldots$	(p_{k1})	$p_{\bullet 1}$	$\mathrm{E}[Y]$	$\mathrm{Var}[Y]$
y_2	p_{12}	p_{22}	$\ldots$	(p_{k2})	$p_{\bullet 2}$		
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$		
(y_ℓ)	$(p_{1\ell})$	$(p_{2\ell})$	$\ldots$	$(p_{k\ell})$	$(p_{\bullet \ell})$		
	$p_{1\bullet}$	$p_{2\bullet}$	$\ldots$	$(p_{k\bullet})$	1		
	$\mathrm{E}[X]$						
	$\mathrm{Var}[X]$						

Ce n'est pas parce qu'on s'intéresse à (X, Y) qu'on ne s'intéresse pas aux distributions marginales !

Distribution jointe et distributions marginales : le cas discret

E = lancer de deux dés (distinguables)
$\Omega = \{(1,1),(1,2),\dots,(1,6),(2,1),\dots,(6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
X = somme des résultats de chaque dé
Y = différence des résultats de chaque dé

	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
1		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$	
2			$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		
3				$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$			
4					$\frac{2}{36}$		$\frac{2}{36}$				
5						$\frac{2}{36}$					

Distribution jointe et distributions marginales : le cas discret

E = lancer de deux dés (distinguables)
$\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
X = somme des résultats de chaque dé
Y = différence des résultats de chaque dé

	2	3	4	5	6	7	8	9	10	11	12			
0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$	$\frac{6}{36}$		
1		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{10}{36}$	$\mathrm{E}[Y]$	$\mathrm{Var}[Y]$
2			$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$			$\frac{8}{36}$	$= \frac{35}{18}$	$= \frac{665}{324}$
3				$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$				$\frac{6}{36}$	≈ 1.94	≈ 2.05
4					$\frac{2}{36}$		$\frac{2}{36}$					$\frac{4}{36}$		
5						$\frac{2}{36}$						$\frac{2}{36}$		
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1		

$\mathrm{E}[X] = 7$

$\mathrm{Var}[X] = \frac{35}{6} \approx 5.83$

Distribution jointe et distributions marginales : le cas continu

Définition

(X, Y) *est* continu $\Leftrightarrow$ *il existe* $f : \mathbb{R}^2 \to \mathbb{R}^+$ *telle que*

$$P[a_1 < X \leq b_1, a_2 < Y \leq b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y)\, dy\, dx.$$

La fonction f est appelée fonction de densité (de probabilité) de (X, Y).

En particulier, on a

$$F(s, t) = P[-\infty < X \leq s, -\infty < Y \leq t] = \int_{-\infty}^{s} \int_{-\infty}^{t} f(x, y)\, dy\, dx.$$

De façon similaire aux variables aléatoires, on peut montrer que

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

de sorte que f fournit une autre description équivalente de $P^{(X,Y)}$ ($P^{(X,Y)} \Leftrightarrow F \Leftrightarrow f$)

Distribution jointe et distributions marginales : le cas continu

Plus généralement que dans la définition ci-dessus,

$$P[(X, Y) \in B] = \iint_B f(x, y)\, dy\, dx$$

pour tout borélien B.

Propriétés de f :

- $f(x, y) \geq 0 \;\; \forall (x, y) \in \mathbb{R}^2$
- $\iint_{\mathbb{R}^2} f(x, y)\, dy\, dx = 1$

Toute fonction f satisfaisant ces propriétés est en fait la densité d'un v.a. continu.

Distribution jointe et distributions marginales : le cas continu

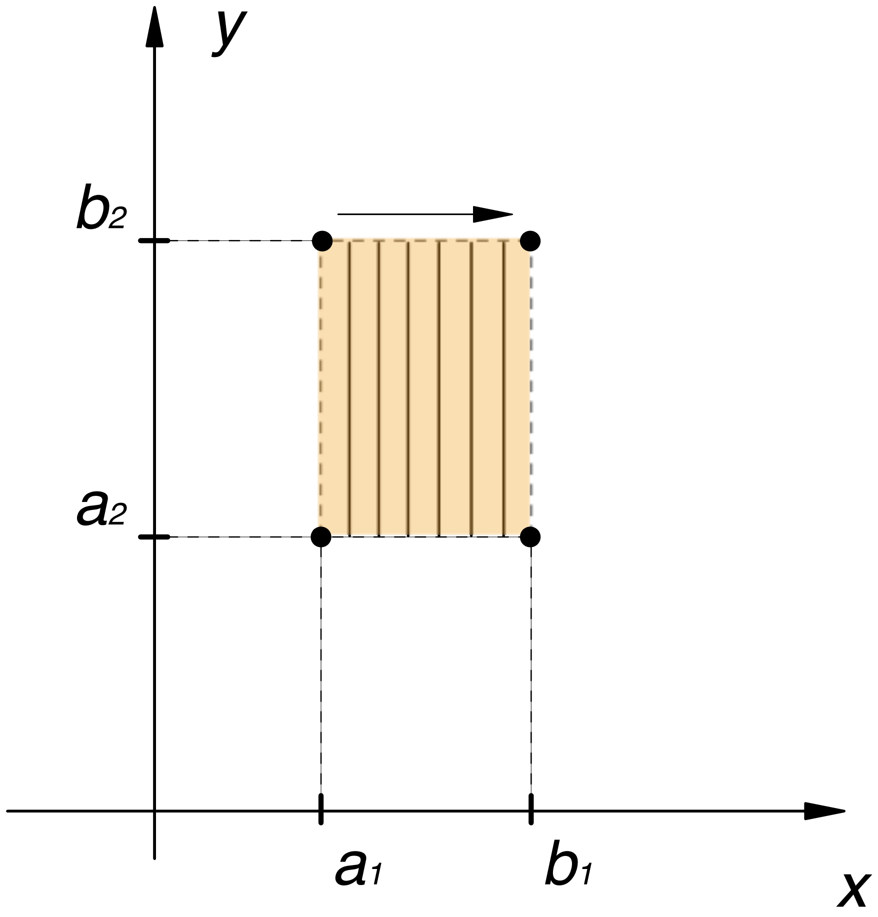
En pratique, l'évaluation de cette "intégrale double" se fait de façon séquentielle (en traitant d'abord une variable comme constante) :

$$
\begin{aligned}
P[a_1 < X \leq b_1, a_2 < Y \leq b_2] &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x,y)\, dy\, dx \\
&= \int_{a_1}^{b_1} \underbrace{\left(\int_{a_2}^{b_2} f(x,y)\, dy \right)}_{\text{une fonction de } x} dx
\end{aligned}
$$

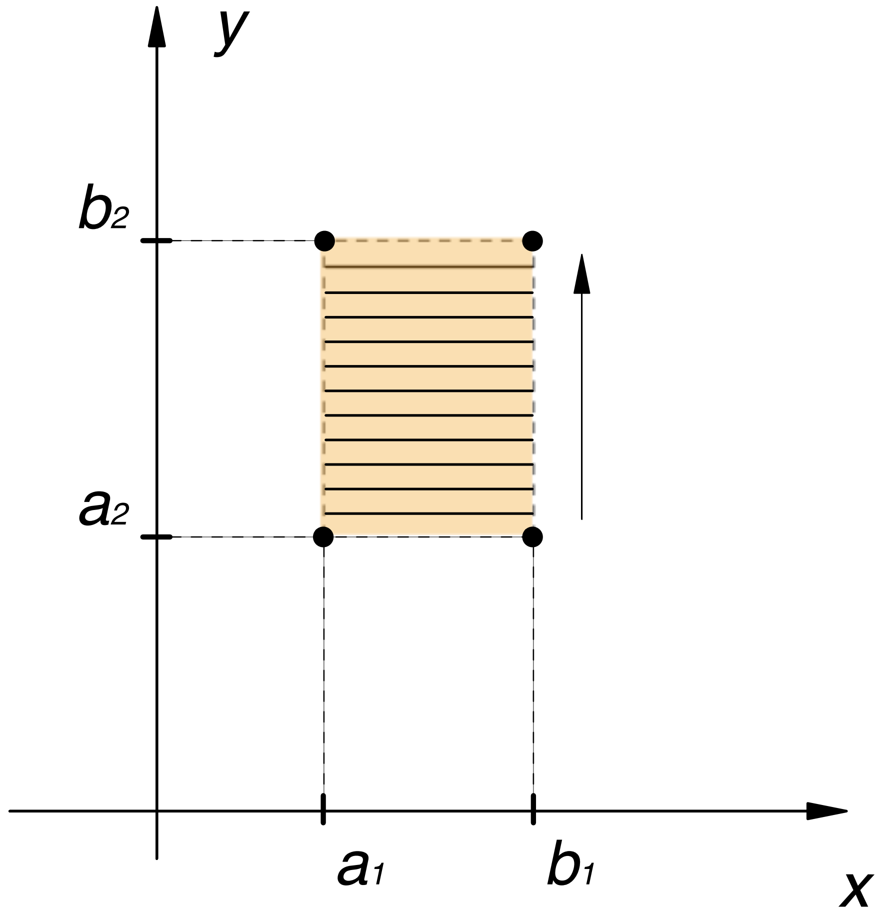
ou, de manière équivalente (le résultat est le même !),

$$
\begin{aligned}
P[a_1 < X \leq b_1, a_2 < Y \leq b_2] &= \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x,y)\, dx\, dy \\
&= \int_{a_2}^{b_2} \underbrace{\left(\int_{a_1}^{b_1} f(x,y)\, dx \right)}_{\text{une fonction de } y} dy
\end{aligned}
$$

Distribution jointe et distributions marginales : le cas continu



$$\int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} \dots \, dy \right) dx$$



$$\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} \dots \, dx \right) dy$$

Distribution jointe et distributions marginales : le cas continu

Exercice

Si $f(x, y) = f_1(x) f_2(y)$ pour tout x, y, alors

$$\begin{aligned}
\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y)\, dx\, dy &= \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_1(x) f_2(y)\, dx\, dy \\
&= \left(\int_{a_1}^{b_1} f_1(x)\, dx \right) \left(\int_{a_2}^{b_2} f_2(y)\, dy \right).
\end{aligned}$$

Distribution jointe et distributions marginales : le cas continu

Exemple :

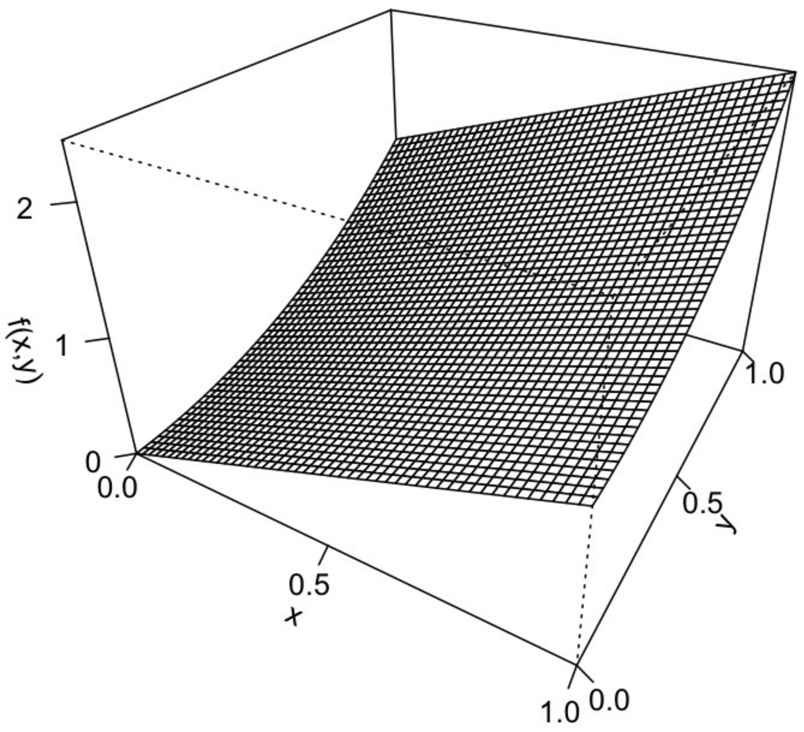
Une chaîne de restauration rapide vend des hamburgers selon deux modalités :

- un comptoir standard
- un drive-in

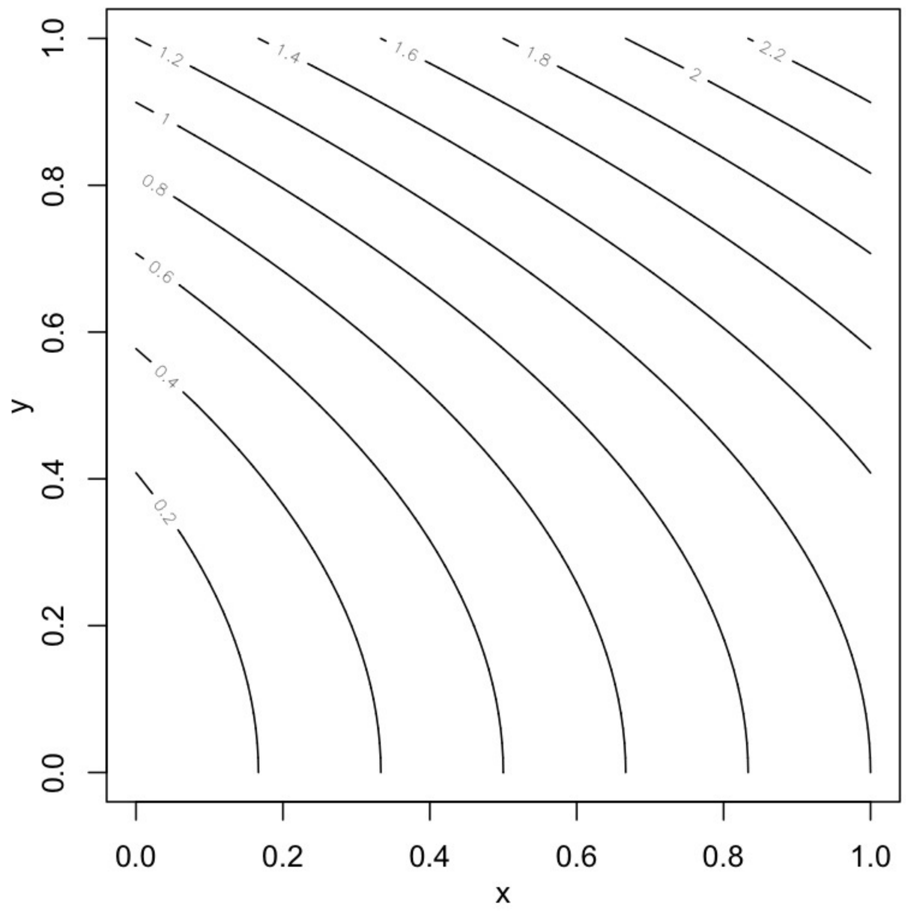
Soit X la proportion du temps où le comptoir standard est occupé le jeudi.
Soit Y la proportion du temps où le drive-in est occupé le jeudi.

Supposons que (X, Y) admette la fonction de densité

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{sinon.} \end{cases}$$



Graphe de $f(x, y)$



Courbes de niveau de $f(x, y)$

Distribution jointe et distributions marginales : le cas continu

Le manager est satisfait si le comptoir standard est occupé 50% du temps au moins et le drive-in 25% du temps au moins : la probabilité, $P[\frac{1}{2} \leq X \leq 1, \frac{1}{4} \leq Y \leq 1]$, que ceci arrive vaut

$$\int_{\frac{1}{2}}^{1} \int_{\frac{1}{4}}^{1} f(x,y)\, dy\, dx = \int_{\frac{1}{2}}^{1} \left(\int_{\frac{1}{4}}^{1} \frac{6}{5}(x+y^2)\, dy \right) dx = \int_{\frac{1}{2}}^{1} \left[\frac{6}{5}\left(xy + \frac{y^3}{3} \right) \right]_{y=\frac{1}{4}}^{y=1} dx$$

$$= \int_{\frac{1}{2}}^{1} \left(\frac{9x}{10} + \frac{63}{160} \right) dx = \left[\frac{9x^2}{20} + \frac{63x}{160} \right]_{\frac{1}{2}}^{1} = \frac{171}{320},$$

ou encore

$$\int_{\frac{1}{4}}^{1} \int_{\frac{1}{2}}^{1} f(x,y)\, dx\, dy = \int_{\frac{1}{4}}^{1} \left(\int_{\frac{1}{2}}^{1} \frac{6}{5}(x+y^2)\, dx \right) dy = \int_{\frac{1}{4}}^{1} \left[\frac{6}{5}\left(\frac{x^2}{2} + y^2 x \right) \right]_{x=\frac{1}{2}}^{x=1} dy$$

$$= \int_{\frac{1}{4}}^{1} \left(\frac{9}{20} + \frac{3y^2}{5} \right) dy = \left[\frac{9y}{20} + \frac{y^3}{5} \right]_{\frac{1}{4}}^{1} = \frac{171}{320}.$$

Distribution jointe et distributions marginales : le cas continu

Exercice :

En procédant de la même façon, montrer que la fonction de répartition

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\, dy\, dx$$

est donnée par

$$F(x, y) = \begin{cases} 0 & \text{si } x < 0 \text{ ou } y < 0 \\ \frac{1}{5}xy(3x + 2y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ \frac{1}{5}x(3x + 2) & \text{si } x \in [0, 1] \text{ et } y > 1 \\ \frac{1}{5}y(3 + 2y^2) & \text{si } x > 1 \text{ et } y \in [0, 1] \\ 1 & \text{sinon.} \end{cases}$$

Distribution jointe et distributions marginales : le cas continu

Par analogie avec les formules discrètes

$$P[X = x_i] = \sum_{j \in \mathcal{J}} P[(X, Y) = (x_i, y_j)]$$

$$P[Y = y_j] = \sum_{i \in \mathcal{I}} P[(X, Y) = (x_i, y_j)],$$

on a

$$f^X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$f^Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Ceci montre comment extraire, dans le cas continu, les distributions marginales de la distribution jointe.

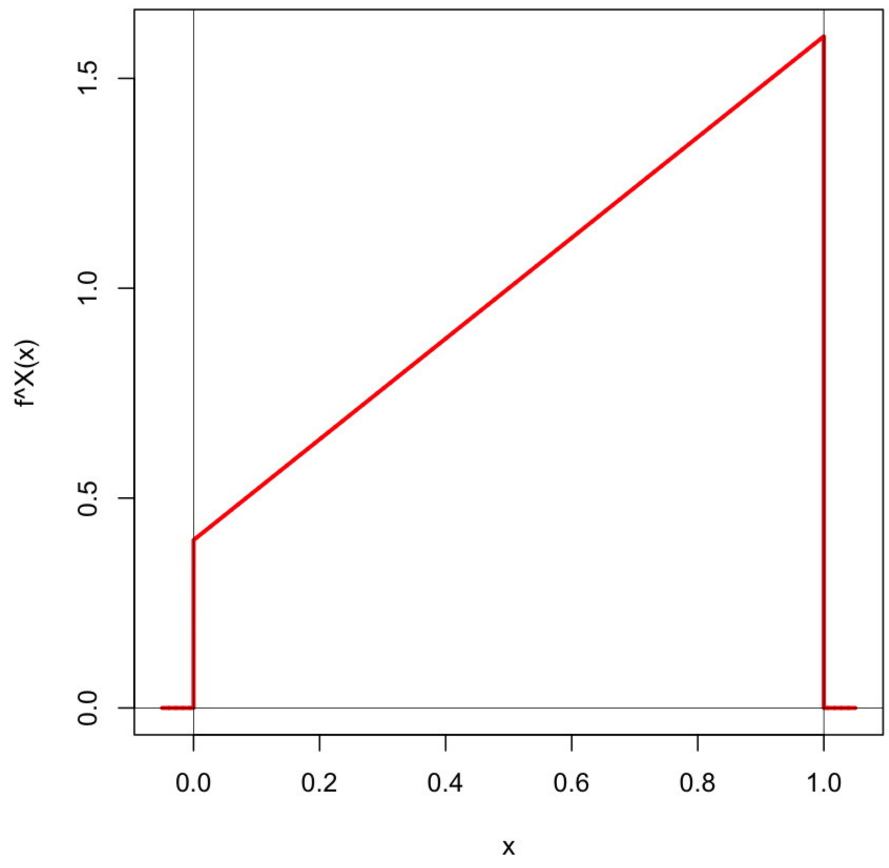
Dans le cas du fast food, ceci donne

$$f^X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{6}{5}(x+y^2)\,dy = \left[\frac{6xy}{5} + \frac{2y^3}{5}\right]_0^1 = \frac{6x}{5} + \frac{2}{5},\ 0 \le x \le 1,$$

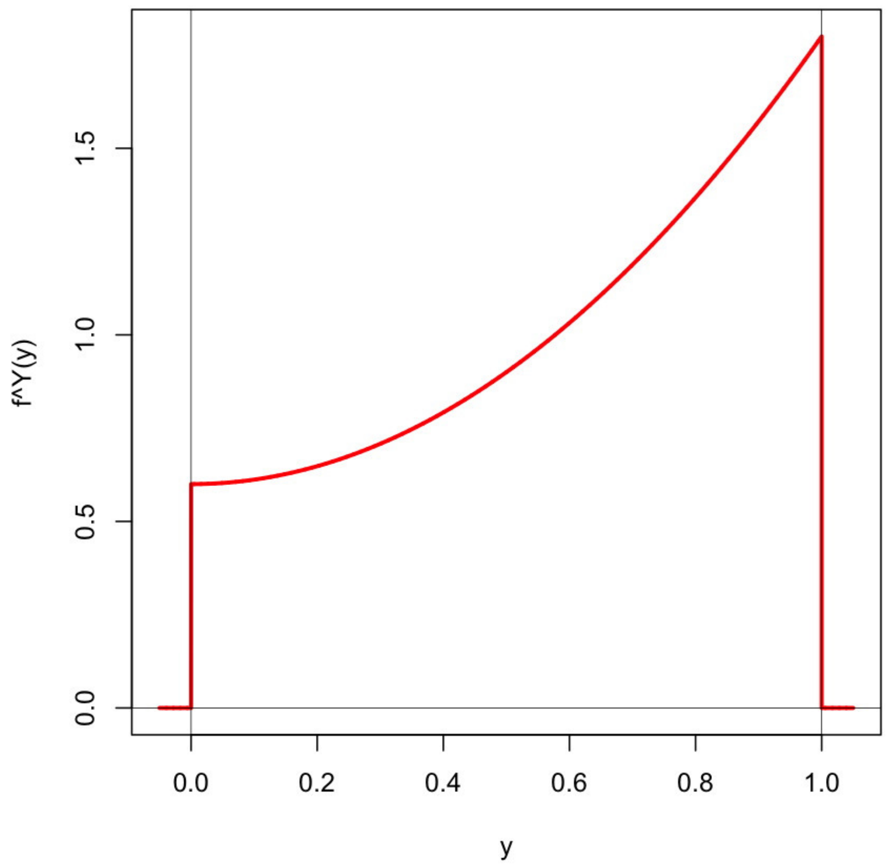
$$f^Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 \frac{6}{5}(x+y^2)\,dx = \left[\frac{3x^2}{5} + \frac{6xy^2}{5}\right]_0^1 = \frac{3}{5} + \frac{6y^2}{5},\ 0 \le y \le 1$$

(et ces fonctions prennent la valeur zéro ailleurs).

Distribution jointe et distributions marginales : le cas continu



Densité marginale de X



Densité marginale de Y

Distribution jointe et distributions marginales : le cas continu

Sur la base des distributions marginales, on peut calculer des probabilités du type $P[X \in B]$ et $P[Y \in B]$, les espérances de X et Y, leur variance, etc.

Par exemple,

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f^X(x)\, dx$$

$$\mathrm{Var}[X] = \begin{cases} \mathrm{E}[(X - \mathrm{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f^X(x)\, dx & \text{(pour l'interprétation)} \\ \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f^X(x)\, dx - (\mathrm{E}[X])^2 & \text{(pour le calcul)} \end{cases}$$

A titre d'illustration, on a par exemple

$$P\Big[X \geq \frac{1}{2}\Big] = P\Big[X \in \Big[\frac{1}{2}, 1\Big]\Big] = \int_{\frac{1}{2}}^{1} f^X(x)\, dx = \int_{\frac{1}{2}}^{1} \Big(\frac{6x}{5} + \frac{2}{5}\Big)\, dx = \ldots = \frac{13}{20}$$

$$P\Big[Y \geq \frac{1}{4}\Big] = P\Big[Y \in \Big[\frac{1}{4}, 1\Big]\Big] = \int_{\frac{1}{4}}^{1} f^Y(y)\, dy = \int_{\frac{1}{4}}^{1} \Big(\frac{3}{5} + \frac{6y^2}{5}\Big)\, dy = \ldots = \frac{27}{32}$$

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f^X(x)\, dx = \int_0^1 x\Big(\frac{6x}{5} + \frac{2}{5}\Big)\, dx = \ldots = \frac{3}{5}$$

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y\, f^Y(y)\, dy = \int_0^1 y\Big(\frac{3}{5} + \frac{6y^2}{5}\Big)\, dy = \ldots = \frac{3}{5}$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f^X(x)\, dx - \Big(\frac{3}{5}\Big)^2 = \ldots = \frac{11}{150} \approx 0.073$$

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = \int_{-\infty}^{\infty} y^2 f^Y(y)\, dy - \Big(\frac{3}{5}\Big)^2 = \ldots = \frac{2}{25} = 0.080$$

On a montré comment calculer des probabilités via la formule

$$P[(X,Y) \in B] = \iint_B f(x,y)\, dy\, dx$$

pour des zones rectangulaires du type $B = (a_1, b_1] \times (a_2, b_2]$.

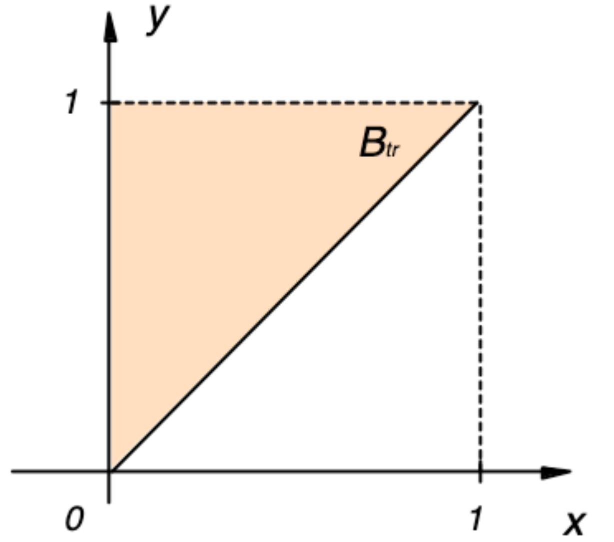
Mais les applications conduisent souvent à des zones non rectangulaires.

Par exemple, le manager pourrait s'intéresser à la probabilité que le comptoir standard soit moins utilisé que le drive-in.

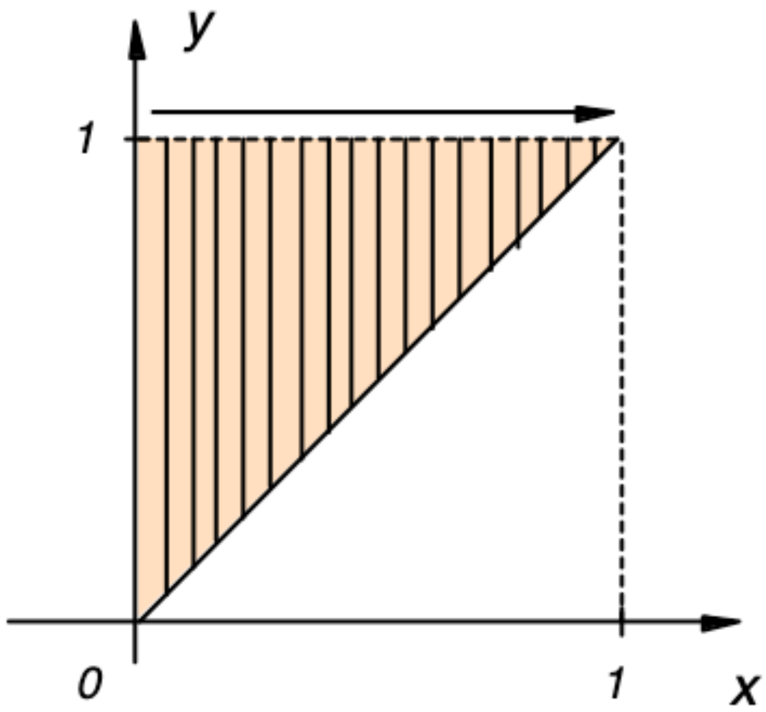
Ceci correspond à

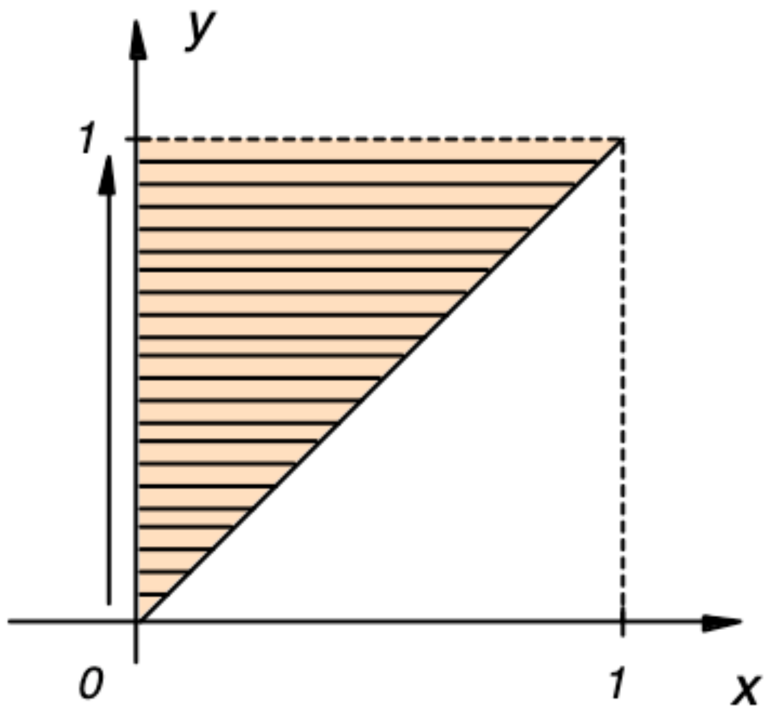
$$P[X \leq Y] = P[(X,Y) \in B_{\mathrm{tr}}],$$

où $B_{\mathrm{tr}} = \{(x,y) \in [0,1] \times [0,1] : x \leq y\}$.



Distribution jointe et distributions marginales : le cas continu





$$P[X \leq Y] = P[(X, Y) \in B_{\text{tr}}]$$
$$= \int_0^1 \left(\int_x^1 f(x, y)\, dy \right) dx = \ldots = \frac{1}{2}$$

$$P[X \leq Y] = P[(X, Y) \in B_{\text{tr}}]$$
$$= \int_0^1 \left(\int_0^y f(x, y)\, dx \right) dy = \ldots = \frac{1}{2}$$

Distribution jointe et distributions marginales : le cas continu

Parfois, c'est la structure même du v.a. qui est non rectangulaire...

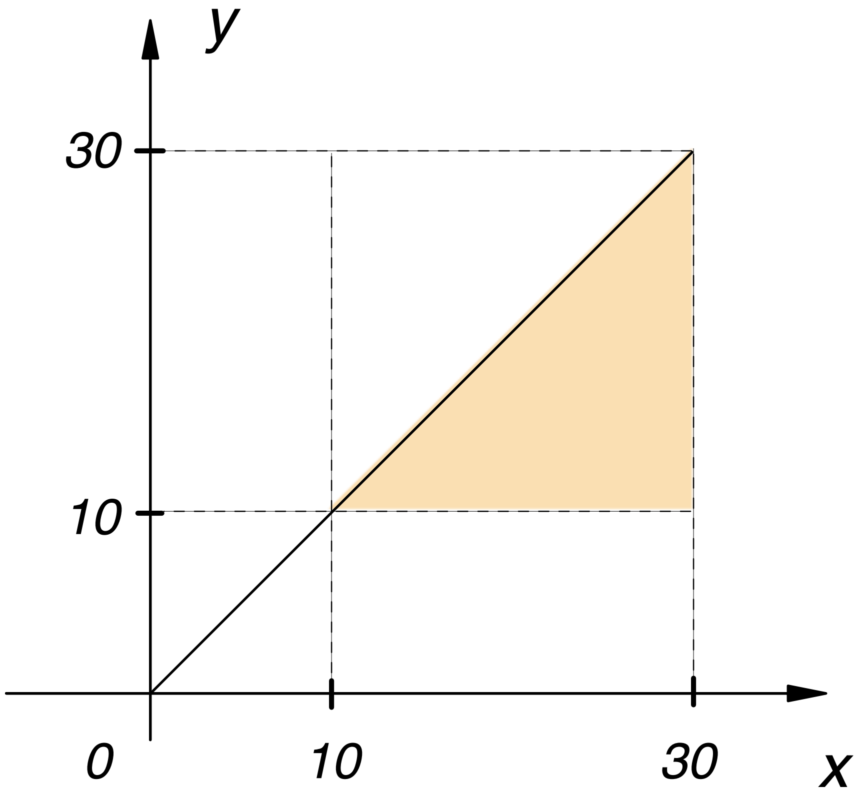
Exemple :

Soit X les revenus annuels nets d'un ménage belge (en milliers d'€).
Soit Y les dépenses annuelles de ce ménage (en milliers d'€).

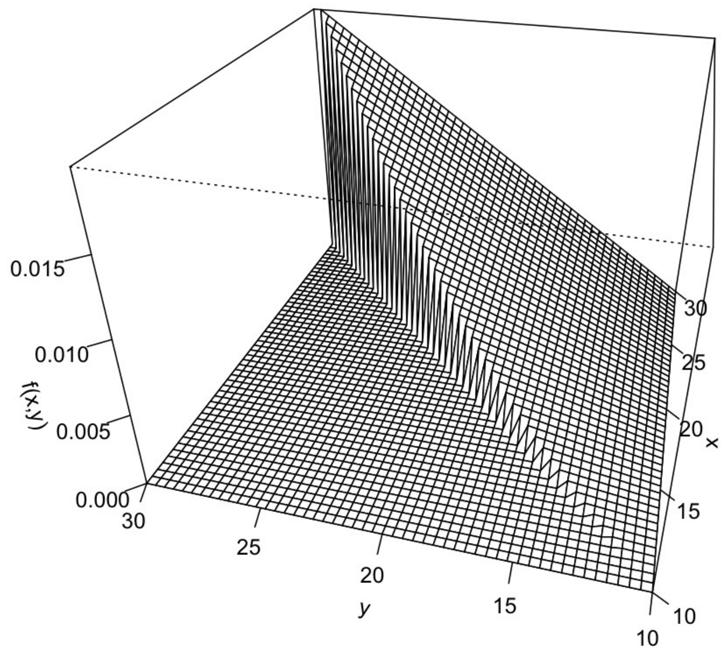
Supposons que (X, Y) admette la fonction de densité

$$f(x, y) = \begin{cases} \dfrac{1}{20000}(x-10)(y-10) & \text{si } 10 < y < x < 30 \\ 0 & \text{sinon.} \end{cases}$$

Distribution jointe et distributions marginales : le cas continu



La zone colorée est celle
où $f(x, y)$ est non nulle



Graphe de $f(x, y)$

Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Distributions conditionnelles

Au chapitre 1, nous avons vu que connaître la réalisation d'un événement permettait en général d'affiner la probabilité de réalisation d'un autre événement.

Ici, on considère une situation où on connaît la valeur qu'a prise l'une des variables aléatoires et on veut savoir si cette information permet d'affiner

- la probabilité que l'autre variable prenne sa valeur dans une région donnée (et par suite la distribution de l'autre variable), ou
- l'espérance ou la variance de l'autre variable.

Au contraire des distributions marginales, les distributions conditionnelles permettent d'appréhender le lien entre X et Y.

Distributions conditionnelles : le cas discret

Soit (X, Y) un v.a. discret.
Notons encore x_i, $i \in \mathcal{I}$, les valeurs possibles de X, et y_j, $j \in \mathcal{J}$, les valeurs possibles de Y. Supposons que la distribution jointe de (X, Y) est donnée par

	x_1	x_2	$\ldots$	(x_k)
y_1	p_{11}	p_{21}	$\ldots$	(p_{k1})
y_2	p_{12}	p_{22}	$\ldots$	(p_{k2})
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
(y_ℓ)	$(p_{1\ell})$	$(p_{2\ell})$	$\ldots$	$(p_{k\ell})$

où $p_{ij} = P[X = x_i, Y = y_j]$.

Rappelons que

$$p_{i\bullet} := P[X = x_i] = \sum_{j \in \mathcal{J}} p_{ij},$$

$$p_{\bullet j} := P[Y = y_j] = \sum_{i \in \mathcal{I}} p_{ij}.$$

Distributions conditionnelles : le cas discret

Pour chaque x_i, la distribution conditionnelle de $Y|[X = x_i]$ est donnée par

$$P[Y = y_j|X = x_i] = \frac{P[X = x_i, Y = y_j]}{P[X = x_i]} = \frac{p_{ij}}{p_{i\bullet}}, \qquad j = 1, 2, \ldots (, \ell)$$

valeurs possibles	y_1	y_2	$\ldots$	(y_ℓ)
probabilités	$\frac{p_{i1}}{p_{i\bullet}}$	$\frac{p_{i2}}{p_{i\bullet}}$	$\ldots$	$(\frac{p_{i\ell}}{p_{i\bullet}})$

Comme pour toute distribution discrète univariée, on peut en calculer l'espérance et la variance (qui seront dites ici "conditionnelles") :

$$\mathrm{E}[Y|X = x_i] = \sum_{j \in \mathcal{J}} y_j \frac{p_{ij}}{p_{i\bullet}}$$

$$\mathrm{Var}[Y|X = x_i] = \begin{cases} \sum_{j \in \mathcal{J}} (y_j - \mathrm{E}[Y|X = x_i])^2 \frac{p_{ij}}{p_{i\bullet}} & \text{(pour l'interprétation)} \\ \sum_{j \in \mathcal{J}} (y_j)^2 \frac{p_{ij}}{p_{i\bullet}} - (\mathrm{E}[Y|X = x_i])^2 & \text{(pour le calcul)} \end{cases}$$

Distributions conditionnelles : le cas discret

E = lancer de deux dés (distinguables)
$\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
X = somme des résultats de chaque dé
Y = différence des résultats de chaque dé

Distributions conditionnelles de Y sachant les diverses valeurs de X :

	2	3	4	5	6	7	8	9	10	11	12
0	1		$\frac{1}{3}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{3}$		1
1		1		$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$		1	
2			$\frac{2}{3}$		$\frac{2}{5}$		$\frac{2}{5}$		$\frac{2}{3}$		
3				$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$			
4					$\frac{2}{5}$		$\frac{2}{5}$				
5						$\frac{1}{3}$					
	1	1	1	1	1	1	1	1	1	1	1

Sachant $[X = 3] = \{(1,2),(2,1)\}$, le résultat de E est encore aléatoire, mais plus Y !

Distributions conditionnelles : le cas discret

Distributions conditionnelles de Y sachant les diverses valeurs de X, avec les espérances et variances correspondantes :

	2	3	4	5	6	7	8	9	10	11	12
0	1		$\frac{1}{3}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{3}$		1
1		1		$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$		1	
2			$\frac{2}{3}$		$\frac{2}{5}$		$\frac{2}{5}$		$\frac{2}{3}$		
3				$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$			
4					$\frac{2}{5}$		$\frac{2}{5}$				
5						$\frac{1}{3}$					
	1	1	1	1	1	1	1	1	1	1	1
E	0	1	1.33	2	2.4	3	2.4	2	1.33	1	0
Var	0	0	0.89	1	2.24	2.67	2.24	1	0.89	0	0

Les fonctions $x_i \mapsto \mathrm{E}[Y|X = x_i]$ et $x_i \mapsto \mathrm{Var}[Y|X = x_i]$ portent souvent une information importante sur le lien entre X et Y.

Distributions conditionnelles : le cas discret

Jusqu'ici, on a supposé que X prenait une certaine valeur fixée x_i.

Si on rend à X son caractère aléatoire, on obtient deux nouvelles v.a. :

- la v.a. "moyenne conditionnelle $\mathrm{E}[Y|X]$", de distribution

valeurs possibles	$\mathrm{E}[Y\|X = x_1]$	$\mathrm{E}[Y\|X = x_2]$	$\ldots$	$(\mathrm{E}[Y\|X = x_k])$
probabilités	$p_{1\bullet} = P[X = x_1]$	$p_{2\bullet} = P[X = x_2]$	$\ldots$	$(p_{k\bullet} = P[X = x_k])$

- la v.a. "variance conditionnelle $\mathrm{Var}[Y|X]$", de distribution

valeurs possibles	$\mathrm{Var}[Y\|X = x_1]$	$\mathrm{Var}[Y\|X = x_2]$	$\ldots$	$(\mathrm{Var}[Y\|X = x_k])$
probabilités	$p_{1\bullet} = P[X = x_1]$	$p_{2\bullet} = P[X = x_2]$	$\ldots$	$(p_{k\bullet} = P[X = x_k])$

Un miracle :

Théorème

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

Preuve :

$$\begin{aligned}
\mathrm{E}[\mathrm{E}[Y|X]] &= \sum_{i\in\mathcal{I}} \mathrm{E}[Y|X = x_i]\, p_{i\bullet} && \text{(définitions de } \mathrm{E}[\cdot] \text{ et de } \mathrm{E}[Y|X]\text{)} \\
&= \sum_{i\in\mathcal{I}} \Big(\sum_{j\in\mathcal{J}} y_j\, \frac{p_{ij}}{p_{i\bullet}} \Big)\, p_{i\bullet} && \text{(expression de } \mathrm{E}[Y|X = x_i]\text{)} \\
&= \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} y_j\, p_{ij} \\
&= \sum_{j\in\mathcal{J}} y_j \Big(\sum_{i\in\mathcal{I}} p_{ij} \Big) = \sum_{j\in\mathcal{J}} y_j\, p_{\bullet j} = \mathrm{E}[Y].
\end{aligned}$$

$\square$

Distributions conditionnelles : le cas discret

Distributions conditionnelles de Y sachant les diverses valeurs de X, avec les espérances et variances correspondantes (+ illustration du miracle) :

	2	3	4	5	6	7	8	9	10	11	12
0	1		$\frac{1}{3}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{3}$		1
1		1		$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$		1	
2			$\frac{2}{3}$		$\frac{2}{5}$		$\frac{2}{5}$		$\frac{2}{3}$		
3				$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$			
4					$\frac{2}{5}$		$\frac{2}{5}$				
5						$\frac{1}{3}$					
	1	1	1	1	1	1	1	1	1	1	1
$\mathrm{E}[Y \mid X = x_i]$	0	1	1.33	2	2.4	3	2.4	2	1.33	1	0
$\mathrm{Var}[Y \mid X = x_i]$	0	0	0.89	1	2.24	2.67	2.24	1	0.89	0	0
$p_{i\bullet}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
	$\mathrm{E}[\mathrm{E}[Y \mid X]] = 0 \times \frac{1}{36} + 1 \times \frac{2}{36} + \ldots = \frac{35}{18} = \mathrm{E}[Y]$ (ch.3-p.24)										

Distributions conditionnelles : le cas discret

On a
$$\mathrm{E}[\mathrm{E}[Y|X]] = \mathrm{E}[Y].$$

Par contre, il est en général faux que

$$\mathrm{E}[\mathrm{Var}[Y|X]] = \mathrm{Var}[Y].$$

Mais on a le résultat suivant.

Théorème

$\mathrm{E}[\mathrm{Var}[Y|X]] = \mathrm{Var}[Y] - \mathrm{Var}[\mathrm{E}[Y|X]].$

Ce théorème sera établi à la page 97 de ce chapitre.

Distributions conditionnelles : le cas discret

On traite de la même façon les distributions conditionnelles de X sachant Y...

Pour chaque y_j, la distribution conditionnelle de $X|[Y = y_j]$ est donnée par

valeurs possibles	x_1	x_2	$\dots$	(x_k)
probabilités	$\frac{p_{1j}}{p_{\bullet j}}$	$\frac{p_{2j}}{p_{\bullet j}}$	$\dots$	$(\frac{p_{kj}}{p_{\bullet j}})$

où $P[X = x_i | Y = y_j] = \frac{P[X=x_i, Y=y_j]}{P[Y=y_j]} = \frac{p_{ij}}{p_{\bullet j}}$, $i = 1, 2, \dots (, k)$. Et on peut définir

$$\mathrm{E}[X|Y = y_j] = \sum_{i \in \mathcal{I}} x_i \frac{p_{ij}}{p_{\bullet j}}$$

$$\mathrm{Var}[X|Y = y_j] = \sum_{i \in \mathcal{I}} (x_i - \mathrm{E}[X|Y = y_j])^2 \frac{p_{ij}}{p_{\bullet j}}$$

$\dots$

Distributions conditionnelles : le cas discret

Distributions conditionnelles de X sachant les diverses valeurs de Y, avec les espérances et variances correspondantes :

	2	3	4	5	6	7	8	9	10	11	12		E	Var
0	$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$	1	7	11.67
1		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{5}$		1	7	8
2			$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$			1	7	5
3				$\frac{1}{3}$		$\frac{1}{3}$		$\frac{1}{3}$				1	7	2.67
4					$\frac{1}{2}$		$\frac{1}{2}$					1	7	1
5						1						1	7	0

(on a encore que $\mathrm{E}[\mathrm{E}[X|Y]] = \mathrm{E}[X]$ et $\mathrm{E}[\mathrm{Var}[X|Y]] = \mathrm{Var}[X] - \mathrm{Var}[\mathrm{E}[X|Y]]$).

Distributions conditionnelles : le cas continu

Par analogie avec le cas discret où

$$P[Y = y_j | X = x_i] = \frac{P[X = x_i, Y = y_j]}{P[X = x_i]} \quad \text{et} \quad P[X = x_i | Y = y_j] = \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]},$$

les densités conditionnelles de $Y|[X = x]$ et de $X|[Y = y]$ sont

$$f^{Y|[X=x]}(y) = \frac{f(x, y)}{f^X(x)} \quad \text{et} \quad f^{X|[Y=y]}(x) = \frac{f(x, y)}{f^Y(y)}.$$

Il s'agit encore de densités de variables aléatoires ($><$vecteurs aléatoires) : on a

$$P[a \leq Y \leq b | X = x] = \int_a^b f^{Y|[X=x]}(y)\, dy,$$

et on définit

$$\mathrm{E}[Y|X = x] = \int_{-\infty}^{\infty} y\, f^{Y|[X=x]}(y)\, dy$$

$$\mathrm{Var}[Y|X = x] = \begin{cases} \int_{-\infty}^{\infty} (y - \mathrm{E}[Y|X = x])^2\, f^{Y|[X=x]}(y)\, dy & \text{(pour l'interprétation)} \\ \int_{-\infty}^{\infty} y^2 f^{Y|[X=x]}(y)\, dy - (\mathrm{E}[Y|X = x])^2 & \text{(pour le calcul)} \end{cases}$$

Exemple :

Une chaîne de restauration rapide vend des hamburgers selon deux modalités :

- un comptoir standard
- un drive-in

Soit X la proportion du temps où le comptoir standard est occupé le jeudi.
Soit Y la proportion du temps où le drive-in est occupé le jeudi.

Supposons que (X, Y) admette la fonction de densité

$$f(x,y) = \begin{cases} \dfrac{6}{5}(x+y^2) & \text{si } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{sinon.} \end{cases}$$

Dans le cas du fast food, ceci donne en particulier (pour chaque $x \in [0, 1]$)

$$f^{Y|[X=x]}(y) = \frac{f(x,y)}{f^X(x)} = \begin{cases} \dfrac{3(x+y^2)}{3x+1} & \text{si } y \in [0,1] \\ 0 & \text{sinon,} \end{cases}$$

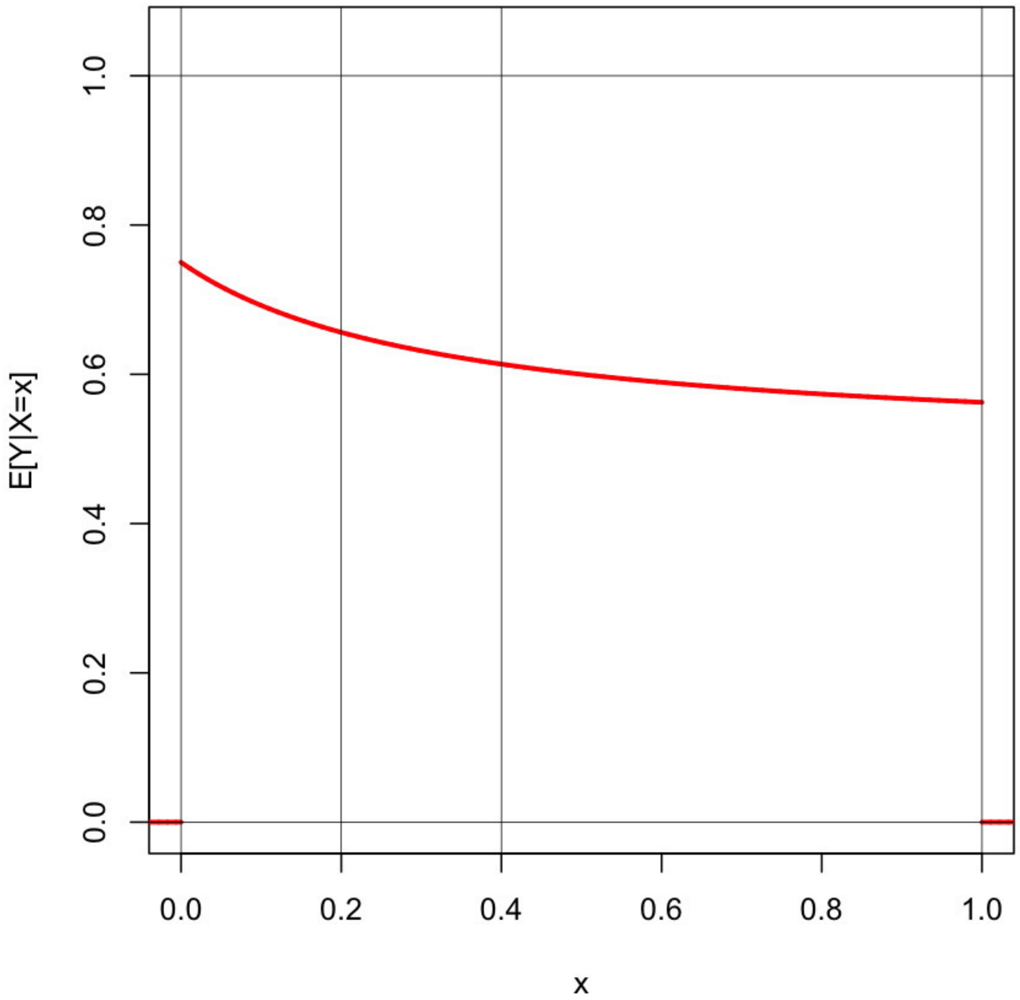
ce qui livre

$$\mathrm{E}[Y|X=x] = \int_{-\infty}^{\infty} y\, f^{Y|[X=x]}(y)\, dy = \int_0^1 y\, \frac{3(x+y^2)}{3x+1}\, dy = \ldots = \frac{6x+3}{12x+4}$$

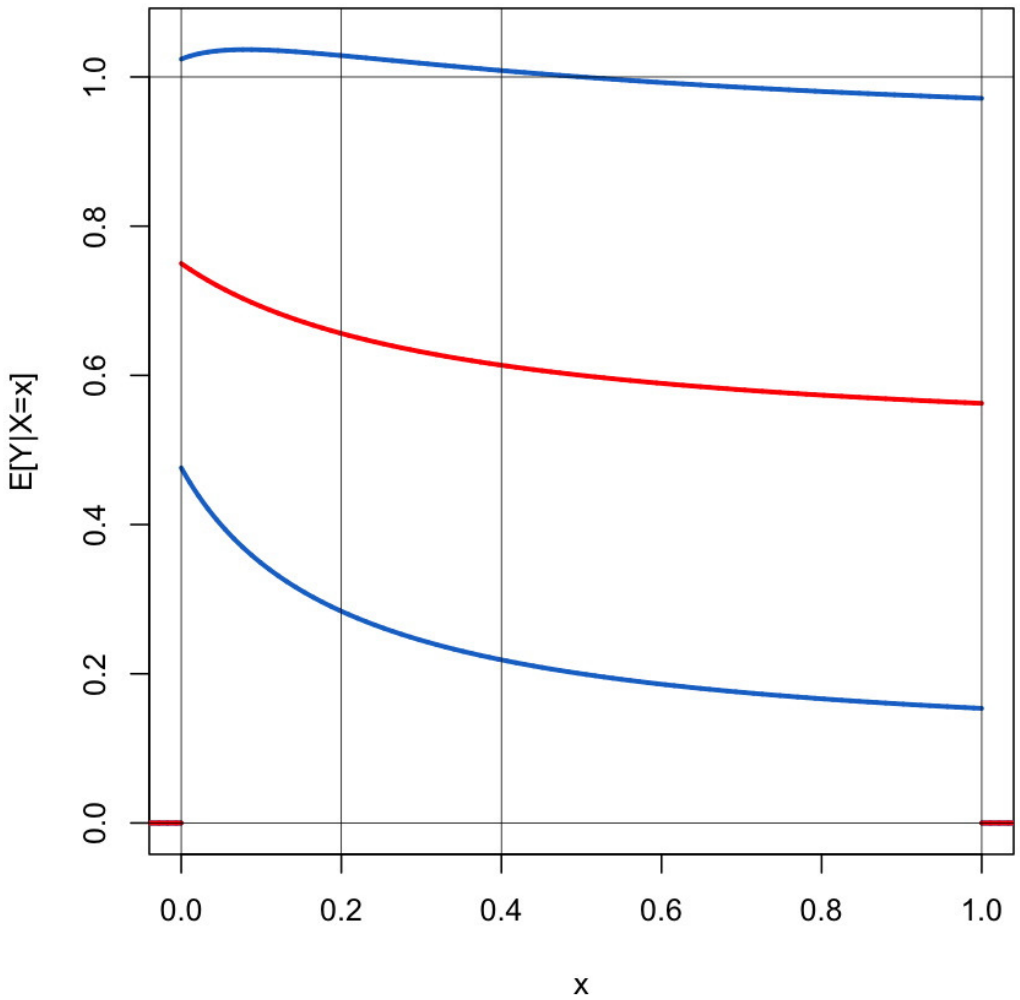
et

$$\mathrm{Var}[Y|X=x] = \int_{-\infty}^{\infty} y^2\, f^{Y|[X=x]}(y)\, dy - (\mathrm{E}[Y|X=x])^2$$

$$= \int_0^1 y^2\, \frac{3(x+y^2)}{3x+1}\, dy - \left(\frac{6x+3}{12x+4}\right)^2 = \ldots = \frac{60x^2+44x+3}{80(3x+1)^2}.$$



Au plus le comptoir standard est occupé, au moins le drive-in l'est (en moyenne)



Graphes de $\mathrm{E}[Y|X = x]$ et $g_{\pm}(x) = \mathrm{E}[Y|X = x] \pm \sqrt{2}\sqrt{\mathrm{Var}[Y|X = x]}$
$\forall x$, on a $P[g_{-}(x) \leq Y \leq g_{+}(x)|X = x] \geq \frac{1}{2}$ (Tchebychev)

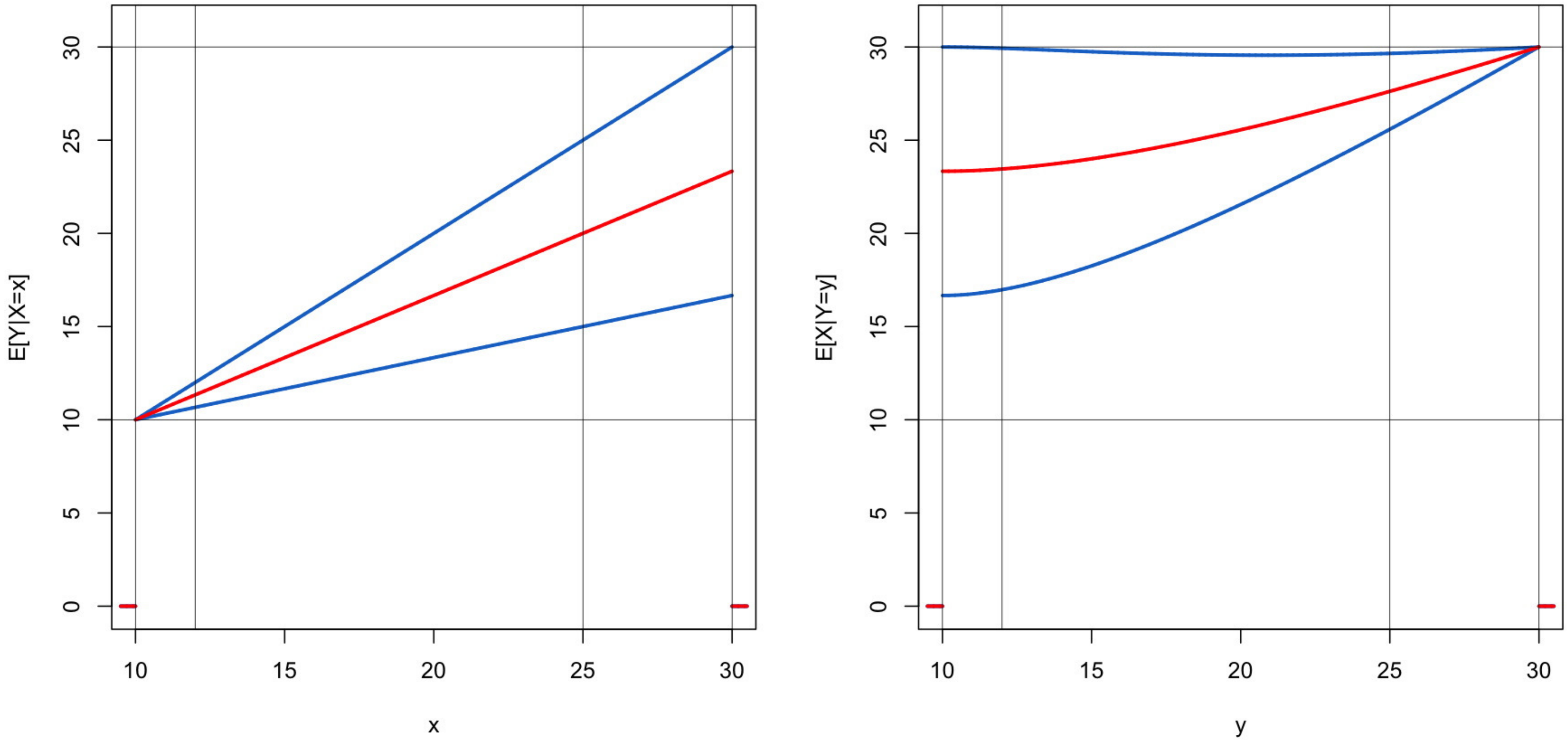
Notre second exemple dans le cas continu :

Soit X les revenus annuels nets d'un ménage belge (en milliers d'€).
Soit Y les dépenses annuelles de ce ménage (en milliers d'€).

Supposons que (X, Y) admette la fonction de densité

$$f(x,y) = \begin{cases} \frac{1}{20000}(x-10)(y-10) & \text{si } 10 < y < x < 30 \\ 0 & \text{sinon.} \end{cases}$$

Distributions conditionnelles : le cas continu



Gauche : plus on gagne, plus on dépense en moyenne (mais l'incertitude ↗)

Droite : plus on dépense, plus on gagne en moyenne (et l'incertitude ↘)

Distributions conditionnelles : le cas continu

Dans le cas continu, on peut définir

- la v.a. "moyenne conditionnelle $\mathrm{E}[Y|X]$" comme la variable aléatoire prenant la valeur $\mathrm{E}[Y|X=x]$ avec densité $f^X(x)$, et
- la v.a. "variance conditionnelle $\mathrm{Var}[Y|X]$" comme la variable aléatoire prenant la valeur $\mathrm{Var}[Y|X=x]$ avec densité $f^X(x)$.

Alors, comme dans le cas discret, on a le résultat suivant.

Théorème

(i) $\mathrm{E}[\mathrm{E}[Y|X]] = \mathrm{E}[Y]$
(ii) $\mathrm{E}[\mathrm{Var}[Y|X]] = \mathrm{Var}[Y] - \mathrm{Var}[\mathrm{E}[Y|X]]$

Exercice : adapter la preuve de (i) au cas continu.

Distributions conditionnelles : le cas continu

A titre d'illustration : dans le cas du fast food, on a

$$
\begin{aligned}
\mathrm{E}[\mathrm{E}[Y|X]] &= \int_{-\infty}^{\infty} \mathrm{E}[Y|X=x]\, f^X(x)\, dx \\
&= \int_0^1 \left(\frac{6x+3}{12x+4}\right)\left(\frac{2}{5}(3x+1)\right) dx \\
&= \int_0^1 \frac{1}{10}(6x+3)\, dx \\
&= \left[\frac{1}{10}(3x^2+3x)\right]_0^1 \\
&= \frac{3}{5},
\end{aligned}
$$

ce qui coïncide bien avec $\mathrm{E}[Y]$ (voir ch.3-p.38).

Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Indépendance

Un cas particulier de "lien" entre X et Y est la situation d'indépendance.

Définition

X et Y sont indépendantes (notation : $X \perp\!\!\!\perp Y$)

$\Leftrightarrow$ *Pour tout $B_1, B_2 \in \mathcal{B}$, $[X \in B_1] \perp\!\!\!\perp [Y \in B_2]$ (au sens du chapitre 1)*

$\Leftrightarrow$ *Pour tout $B_1, B_2 \in \mathcal{B}$, $P[X \in B_1, Y \in B_2] = P[X \in B_1]P[Y \in B_2]$*

Si ces probabilités sont non nulles, ceci se réécrit $P[Y \in B_2 | X \in B_1] = P[Y \in B_2]$ ou $P[X \in B_1 | Y \in B_2] = P[X \in B_1]$ (ce qui fournit l'intuition d'indépendance)

Pour les dés :

$0 = P[X = 12, Y = 5] \neq P[X = 12]P[Y = 5] = \frac{1}{36} \times \frac{2}{36}$.

Pour le fast food :

$\frac{171}{320} = P[\frac{1}{2} \leq X \leq 1, \frac{1}{4} \leq Y \leq 1] \neq P[\frac{1}{2} \leq X \leq 1]P[\frac{1}{4} \leq Y \leq 1] = \frac{13}{20} \times \frac{27}{32}$.

Dans les deux cas, X et Y ne sont donc pas indépendantes.

Indépendance

Le résultat suivant présente des caractérisations de l'indépendance.

Théorème

$X \perp\!\!\!\perp Y$

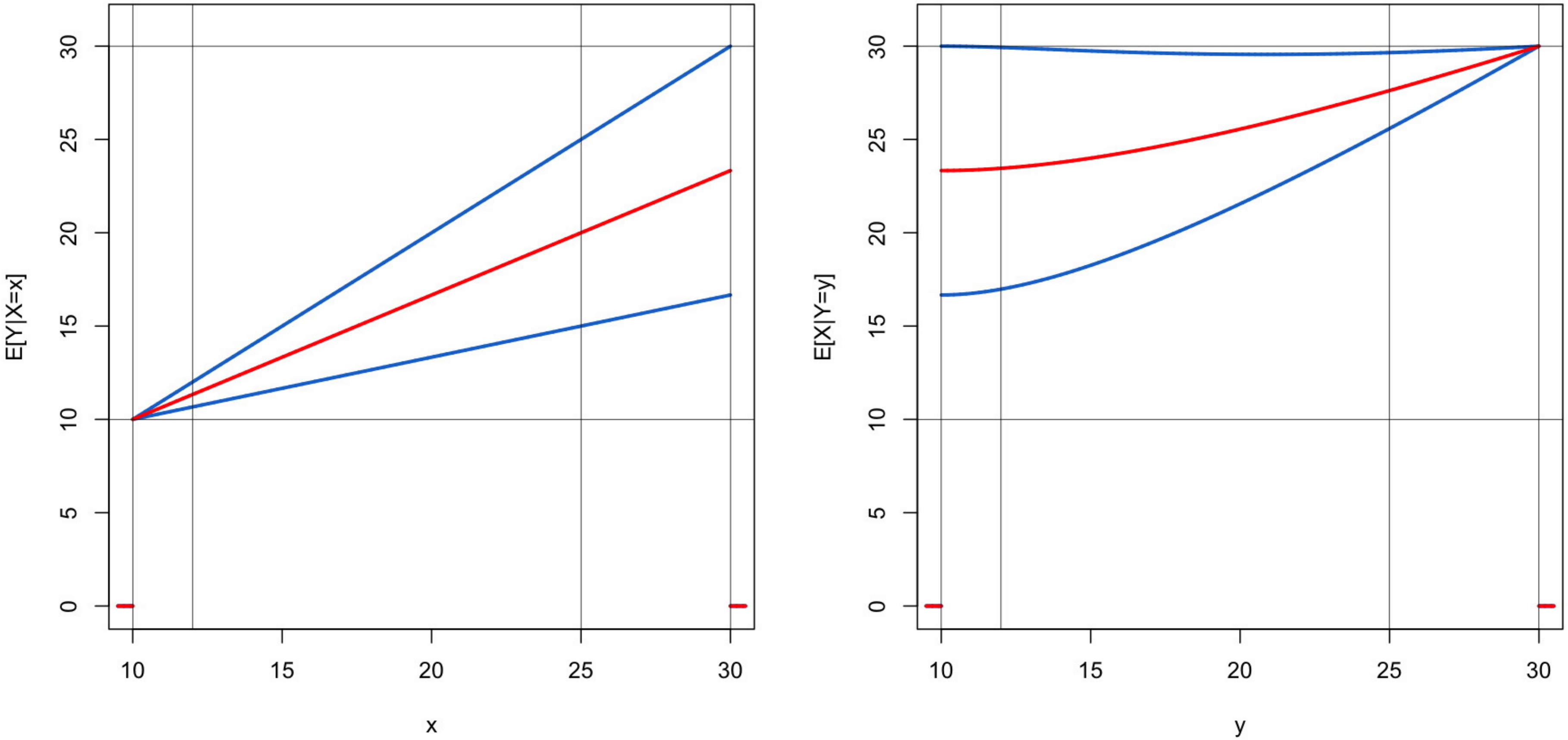
$\Leftrightarrow \forall x, y,\ F(x,y) = F^X(x)F^Y(y)$, *où* $F^X(x) = P[X \leq x]$ *et* $F^Y(y) = P[Y \leq y]$

$\Leftrightarrow \forall i, j,\ P[X = x_i, Y = y_j] = P[X = x_i]P[Y = y_j]$ *(cas discret)*
$\forall x, y,\ f(x,y) = f^X(x)f^Y(y)$ *(cas continu)*

$\Leftrightarrow \forall i, j,\ P[Y = y_j | X = x_i] = P[Y = y_j]$ *(cas discret)*
$\forall x, y,\ f^{Y|[X=x]}(y) = f^Y(y)$ *(cas continu)*

$\Leftrightarrow \forall i, j,\ P[X = x_i | Y = y_j] = P[X = x_i]$ *(cas discret)*
$\forall x, y,\ f^{X|[Y=y]}(x) = f^X(x)$ *(cas continu)*

On a que $X \perp\!\!\!\perp Y$ exactement quand les distributions marginales et conditionnelles coïncident (ce qui est intuitivement clair !), une situation où, en particulier, les fonctions d'espérances conditionnelles $x \mapsto \mathrm{E}[Y|X = x]$ et $y \mapsto \mathrm{E}[X|Y = y]$ (comme celles de variances conditionnelles) sont constantes.



X (les revenus) et Y (les dépenses) ne sont donc pas indépendantes...

Indépendance

E = lancer de deux dés (distinguables)
$\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
X = résultat du 1er dé
Y = résultat du 2nd dé

	1	2	3	4	5	6	
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$\forall i,j \; p_{ij} = p_{i\bullet}p_{\bullet j} \Rightarrow X \perp\!\!\!\perp Y$$

Cette indépendance se traduit aussi par le fait que, pour chaque x_i, la distribution conditionnelle de $Y|[X = x_i]$ coïncide avec la distribution marginale de Y

	1	2	3	4	5	6
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
4	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
5	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
6	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	1	1	1	1	1	1

y_j	$p_{\bullet j}$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$
	1

Indépendance

En haut de l'avenue Héger, on observe

- le temps X (en min) jusqu'à la prochaine arrivée d'un bus 71 (direction ville)
- le temps Y (en min) jusqu'à la prochaine arrivée d'un tram 8 (direction ville)

Supposons qu'un 71 passe exactement toutes les 10 minutes, qu'un 8 passe exactement toutes les 8 minutes, et que (X, Y) admette la densité

$$f(x,y) = \begin{cases} \frac{1}{80} & \text{si } (x,y) \in [0,10] \times [0,8] \\ 0 & \text{sinon.} \end{cases}$$

On vérifie alors aisément que

$$f^X(x) = \begin{cases} \frac{1}{10} & \text{si } x \in [0,10] \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f^Y(y) = \begin{cases} \frac{1}{8} & \text{si } y \in [0,8] \\ 0 & \text{sinon,} \end{cases}$$

ce qui implique que $f(x,y) = f^X(x) f^Y(y) \ \forall x, y$. On a donc $X \perp\!\!\!\perp Y$.

Indépendance

Puisqu'on a défini au chapitre 1 l'indépendance mutuelle de plus de deux événements, on peut adopter les extensions suivantes de la définition en page 63.

Définition

$X_1, X_2, \ldots, X_k$ sont mutuellement indépendantes
$\Leftrightarrow$ Pour tout $B_1, B_2, \ldots, B_k \in \mathcal{B}$, $[X_1 \in B_1], [X_2 \in B_2], \ldots, [X_k \in B_k]$ sont mutuellement indépendants (au sens du chapitre 1)

Définition

$X_1, X_2, \ldots$ sont mutuellement indépendantes
$\Leftrightarrow$ Pour tout $B_1, B_2, \ldots \in \mathcal{B}$, $[X_1 \in B_1], [X_2 \in B_2], \ldots$ sont mutuellement indépendants (au sens du chapitre 1)

Ceci sera important pour l'inférence statistique, qui supposera (dans ce cours) que les observations sont des réalisations de v.a. indépendantes (pas de couples dans les sondages, ni de "séries chronologiques" !)

Indépendance

Exemple :

X_1 = nombre d'accidents de voiture par jour à Ixelles
X_2 = nombre d'accidents de voiture par jour à Etterbeek
X_3 = nombre d'accidents de voiture par jour à Uccle

L'ensemble des valeurs possibles de (X_1, X_2, X_3) est $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$.

Si la distribution jointe est telle que

$$P[X_1 = k_1, X_2 = k_2, X_3 = k_3] = e^{-\lambda_1-\lambda_2-\lambda_3} \frac{\lambda_1^{k_1} \lambda_2^{k_2} \lambda_3^{k_3}}{(k_1!)(k_2!)(k_3!)}, \quad \lambda_1, \lambda_2, \lambda_3 > 0,$$

on peut vérifier que X_1, X_2 et X_3 sont mutuellement indépendantes.

Remarques :
- $X_i \sim \mathrm{Poi}(\lambda_i)$, $i = 1, 2, 3$.
- Si $\lambda_1 = \lambda_2 = \lambda_3$, X_1, X_2 et X_3 sont i.i.d. (indépendantes et identiquement distribuées).

Dans la suite, nous considérerons des espérances du type $\mathrm{E}[g(X,Y)]$, qui font intervenir à la fois X et Y.

Définition

(i) Pour (X,Y) discret, de distribution $(x_i, y_j, p_{ij} = P[X = x_i, Y = y_j])$, $i \in \mathcal{I}$, $j \in \mathcal{J}$,

$$\mathrm{E}[g(X,Y)] = \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} g(x_i, y_j) p_{ij}.$$

(ii) Pour (X,Y) continu, de fonction de densité $f(x,y)$,

$$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dy\, dx.$$

On peut aussi définir des versions conditionnelles de telles espérances :

Définition

(i) Pour (X, Y) *discret, de distribution* $(x_i, y_j, p_{ij} = P[X = x_i, Y = y_j])$*,* $i \in \mathcal{I}, j \in \mathcal{J}$*,*

$$\mathrm{E}[g(X,Y)|X = x_i] = \sum_{j \in \mathcal{J}} g(x_i, y_j) P[Y = y_j | X = x_i] = \sum_{j \in \mathcal{J}} g(x_i, y_j) \frac{p_{ij}}{p_{i\bullet}}.$$

(ii) Pour (X, Y) *continu, de fonction de densité* $f(x, y)$*,*

$$\begin{aligned} \mathrm{E}[g(X,Y)|X = x] &= \int_{-\infty}^{\infty} g(x,y)\, f^{Y|[X=x]}(y)\, dy \\ &= \int_{-\infty}^{\infty} g(x,y)\, \frac{f(x,y)}{f^X(x)}\, dy. \end{aligned}$$

(iii) Dans les deux cas, on définit la variable aléatoire $\mathrm{E}[g(X,Y)|X]$ *à travers la relation* $\mathrm{E}[g(X,Y)|X](\omega) = \mathrm{E}[g(X,Y)|X = X(\omega)]$*.*

Alors $\mathrm{E}[\mathrm{E}[g(X,Y)|X]] = \mathrm{E}[g(X,Y)]$ et $\mathrm{E}[h(X)g(X,Y)|X] = h(X)\mathrm{E}[g(X,Y)|X]$

Théorème

$X \perp\!\!\!\perp Y \Rightarrow \mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$

Preuve : dans le cas continu (le cas discret est laissé comme exercice), on a

$$\begin{aligned}
\mathrm{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\, dy\, dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f^X(x) f^Y(y)\, dy\, dx \\
&= \left(\int_{-\infty}^{\infty} x\, f^X(x)\, dx\right)\left(\int_{-\infty}^{\infty} y\, f^Y(y)\, dy\right) = \mathrm{E}[X]\,\mathrm{E}[Y],
\end{aligned}$$

où on a utilisé la remarque de ch.3-p.29. $\square$

Remarque : l'hypothèse d'indépendance ne peut être retirée
(si on pouvait, on aurait $\mathrm{Var}[X] = \mathrm{E}[X\,X] - (\mathrm{E}[X])(\mathrm{E}[X]) = 0$ pour toute v.a. X !)

Indépendance

Si $X \perp\!\!\!\perp Y$, alors $h(X) \perp\!\!\!\perp g(Y)$ $\forall h, g$ (exercice).
Ceci permet de montrer le résultat suivant.

Théorème

$X \perp\!\!\!\perp Y \Rightarrow M_{X+Y}(t) = M_X(t)M_Y(t)$

Preuve : $X \perp\!\!\!\perp Y \Rightarrow e^{tX} \perp\!\!\!\perp e^{tY}$. Donc

$$M_{X+Y}(t) = \mathrm{E}[e^{t(X+Y)}] = \mathrm{E}[e^{tX}e^{tY}] = \mathrm{E}[e^{tX}]\mathrm{E}[e^{tY}] = M_X(t)M_Y(t),$$

par le théorème de la page précédente. $\square$

Ceci permet de prouver certaines propriétés d'additivité.

Théorème

Supposons que $X \perp\!\!\!\perp Y$. Alors
(i) $X \sim \mathrm{Bin}(n_1, p)$ *et* $Y \sim \mathrm{Bin}(n_2, p) \Rightarrow X + Y \sim \mathrm{Bin}(n_1 + n_2, p)$.
(ii) $X \sim \mathrm{Poi}(\lambda_1)$ *et* $Y \sim \mathrm{Poi}(\lambda_2) \Rightarrow X + Y \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$.
(iii) $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ *et* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2) \Rightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
(iv) $X \sim \chi^2_{k_1}$ *et* $Y \sim \chi^2_{k_2} \Rightarrow X + Y \sim \chi^2_{k_1+k_2}$.

Preuve : (i) Si $X \sim \mathrm{Bin}(n, p)$, alors $M_X(t) = (1 - p + pe^t)^n$ (ch.2-p.94). Donc,

$$M_{X+Y}(t) = M_X(t)M_Y(t) = (1-p+pe^t)^{n_1}(1-p+pe^t)^{n_2} = (1 - p + pe^t)^{n_1+n_2} = M_Z(t),$$

où $Z \sim \mathrm{Bin}(n_1 + n_2, p)$. Puisque $X + Y$ et Z ont la même fonction génératrice des moments, elles ont la même distribution (ch.2-p.95).

(ii)–(iv) Exercices ! $\square$

Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Motivation

Considérons un portefeuille boursier valant

$$Z = X + Y,$$

où X est la valeur de l'actif A_1 et Y est la valeur de l'actif A_2.

- La valeur attendue pour Z est

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

  qu'on peut évaluer sur la seule base des distributions marginales de X et de Y (pas besoin de la distribution jointe).

- Qu'en est-il du risque de Z, qui peut être mesuré par $\mathrm{Var}[Z]$?

Motivation

Les propriétés de l'espérance mathématique fournissent

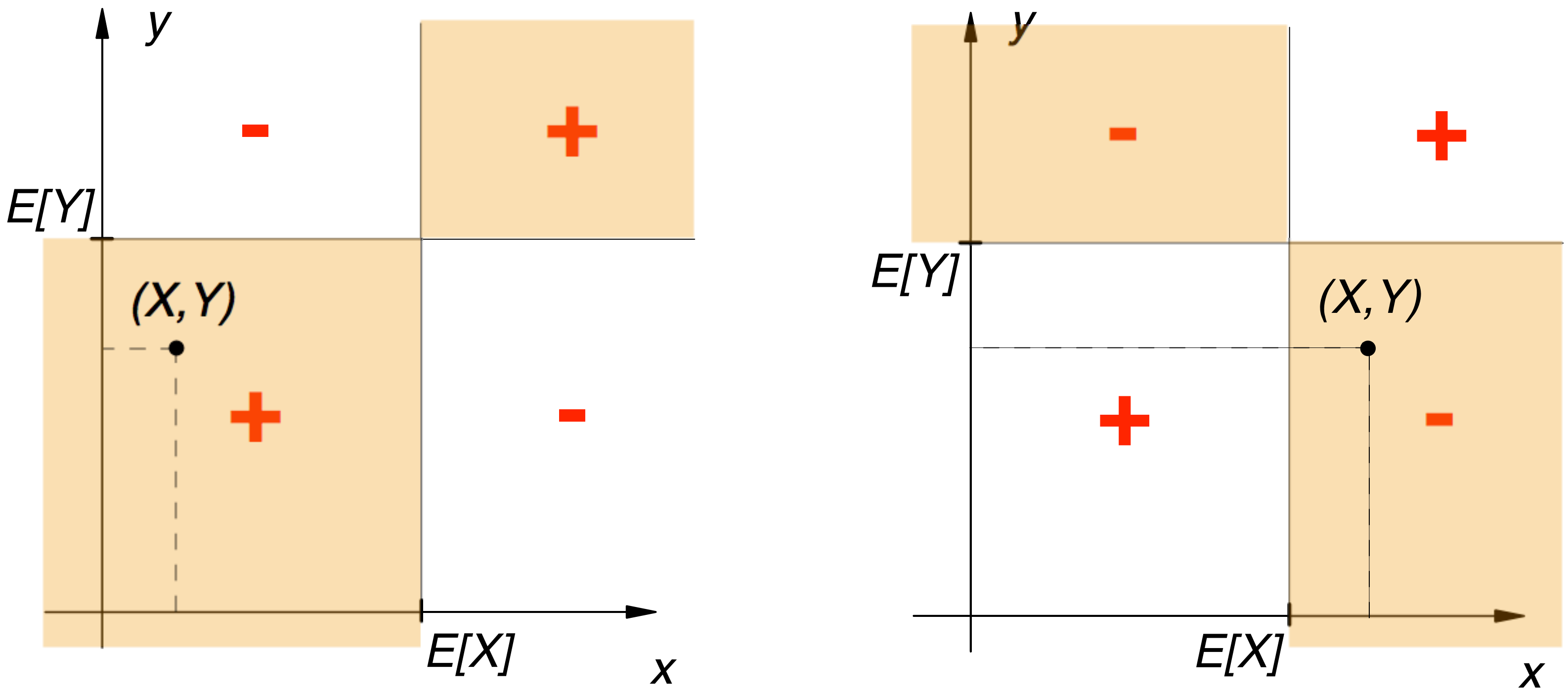
$$\begin{aligned}
\mathrm{Var}[Z] &= \mathrm{E}[(Z-\mathrm{E}[Z])^2] = \mathrm{E}[(X+Y-\underbrace{\mathrm{E}[X+Y]}_{\mathrm{E}[X]+\mathrm{E}[Y]})^2] \\
&= \mathrm{E}[\{(X-\mathrm{E}[X])+(Y-\mathrm{E}[Y])\}^2] \\
&= \mathrm{E}[(X-\mathrm{E}[X])^2]+\mathrm{E}[(Y-\mathrm{E}[Y])^2]+\mathrm{E}[2(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\,\mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])].
\end{aligned}$$

Le risque de $Z = X + Y$ n'est donc pas égal à la somme des risques de X et Y.
Il peut être plus grand ou plus petit en fonction du signe de la quantité suivante.

Définition

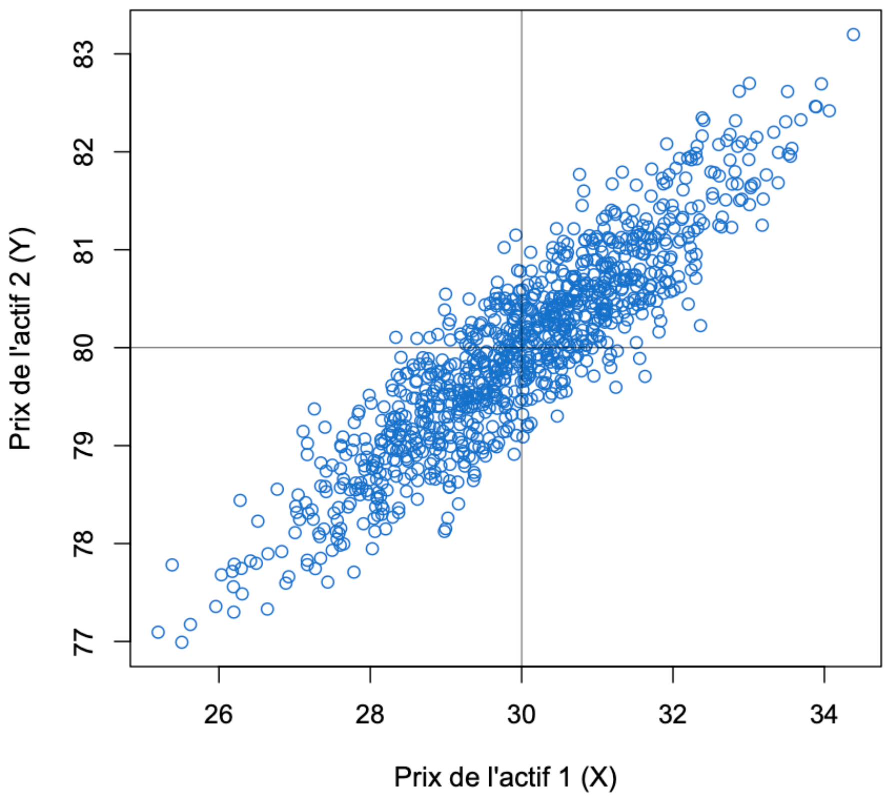
La covariance entre X et Y est $\mathrm{Cov}[X, Y] = \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])]$

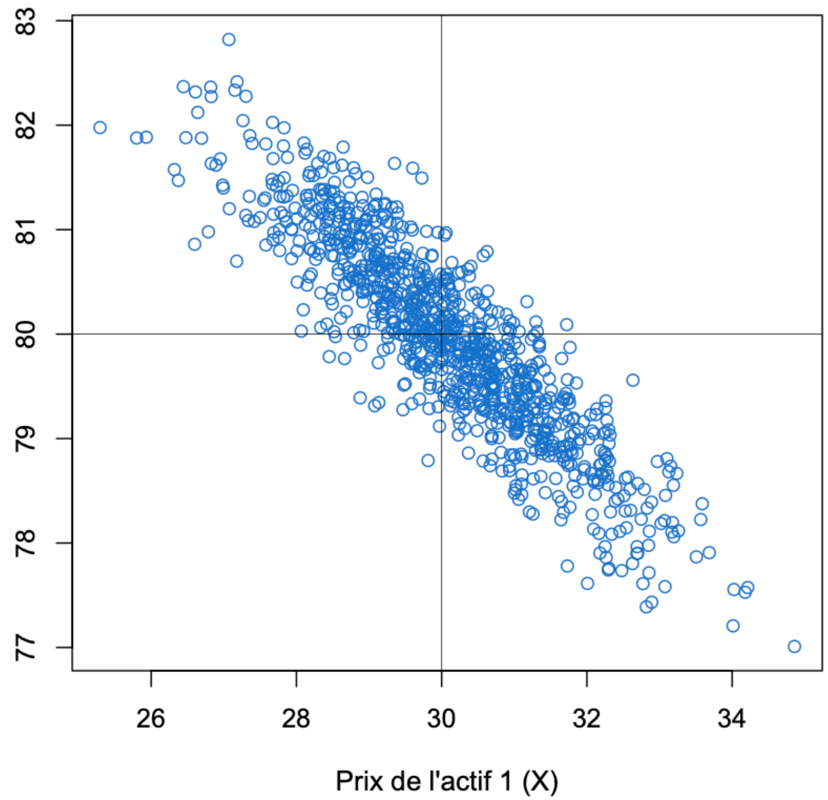
Comment interpréter la covariance $\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$?



Si (X, Y) se réalise plus souvent dans les zones "+" (ou de façon plus extrême), $\mathrm{Cov}[X, Y]$ sera positif, et inversément.

$\rightsquigarrow$ $\mathrm{Cov}[X, Y]$ est une mesure de dépendance linéaire





Dépendance positive ($\mathrm{Cov}[X, Y] > 0$)

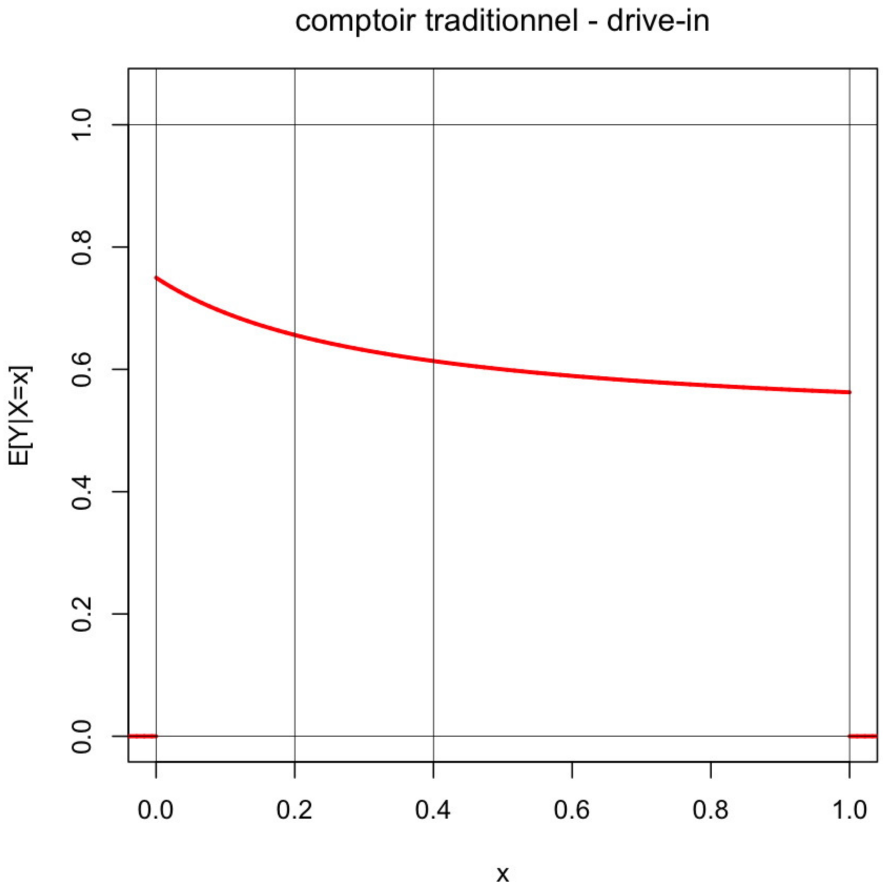
$\mathrm{Var}[X + Y] > \mathrm{Var}[X] + \mathrm{Var}[Y]$

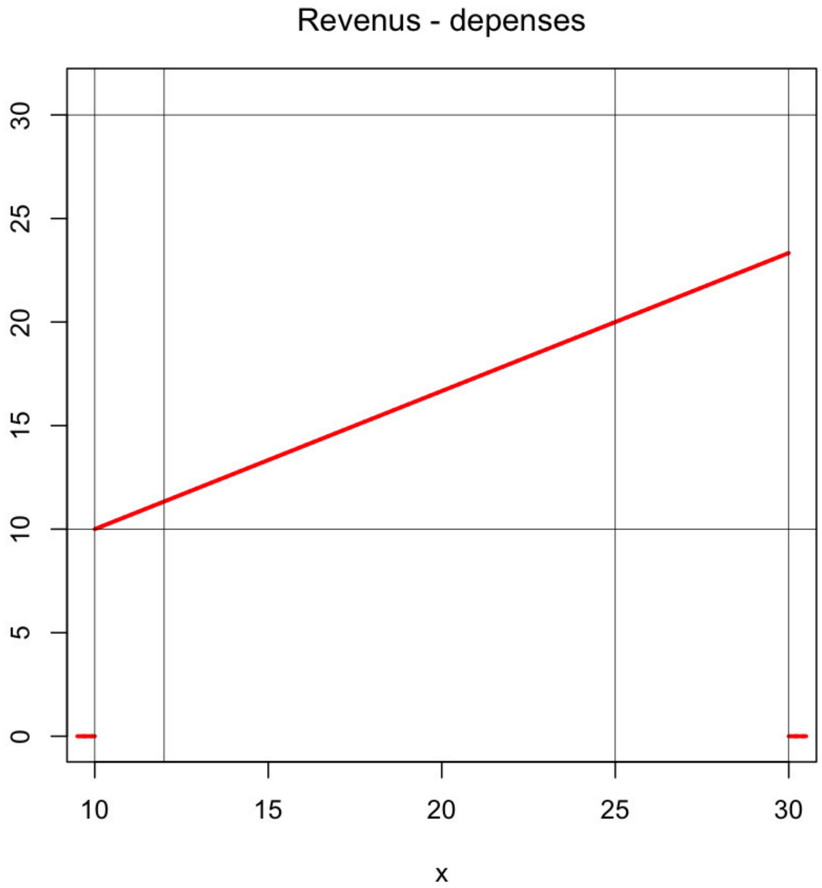
$X =$ ING et $Y =$ BNP ?

Dépendance négative ($\mathrm{Cov}[X, Y] < 0$)

$\mathrm{Var}[X + Y] < \mathrm{Var}[X] + \mathrm{Var}[Y]$

$X =$ ING et $Y =$ Colruyt ?





Dépendance négative

$\mathrm{Cov}[X, Y] = -0.01$

Dépendance positive

$\mathrm{Cov}[X, Y] \approx 7.11$

Covariance

Définition

La covariance entre X et Y est $\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$

Théorème

(i) $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X, Y]$
(ii) $\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$
(iii) Si $X \perp\!\!\!\perp Y$, $\mathrm{Cov}[X, Y] = 0$
(iv) Si $X \perp\!\!\!\perp Y$, $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

Preuve : (i) ceci a été prouvé en ch.3-p.77. (ii) On a

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

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

(iii) Le résultat découle directement de (ii) et du théorème en ch.3-p.73. (iv) Ceci découle directement de (i) et (iii). $\square$

Covariance

L'expression $\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$, qui conduit aux formules

$$\mathrm{Cov}[X, Y] = \begin{cases} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} (x_i - \mathrm{E}[X])(y_j - \mathrm{E}[Y]) p_{ij} & \text{(cas discret)} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathrm{E}[X])(y - \mathrm{E}[Y])\, f(x, y)\, dy\, dx & \text{(cas continu),} \end{cases}$$

a permis ci-dessus l'interprétation de la covariance.

L'expression $\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$, qui mène à

$$\mathrm{Cov}[X, Y] = \begin{cases} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} x_i y_j p_{ij} - \mathrm{E}[X]\mathrm{E}[Y] & \text{(cas discret)} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dy\, dx - \mathrm{E}[X]\mathrm{E}[Y] & \text{(cas continu),} \end{cases}$$

est souvent plus pratique pour l'évaluation de la covariance.

La situation est donc similaire à celle de la variance.

Covariance

$E =$ lancer de deux dés (distinguables)
$\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$ ($\rightsquigarrow \mathcal{A} = \mathcal{P}(\Omega)$)
$X =$ somme des résultats de chaque dé
$Y =$ différence des résultats de chaque dé

	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
1		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$	
2			$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		
3				$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$			
4					$\frac{2}{36}$		$\frac{2}{36}$				
5						$\frac{2}{36}$					

On a $\mathrm{Cov}[X, Y] = 0$ (exercice), ce qui n'est pas si surprenant

Covariance

Dans l'exemple précédent, $\mathrm{Cov}[X, Y] = 0$ mais X et Y ne sont pas indépendants.

On en conclut que

- L'implication $X \perp\!\!\!\perp Y \Rightarrow \mathrm{Cov}[X, Y] = 0$ est un théorème.
- L'implication $X \perp\!\!\!\perp Y \Leftarrow \mathrm{Cov}[X, Y] = 0$ n'est pas vraie en général.

La covariance n'est donc pas une mesure parfaite de dépendance
(c'est une mesure de dépendance <u>linéaire</u>)

Covariance

Propriétés supplémentaires de la covariance :

Théorème

Soient X, X_1, X_2, Y, Y_1, Y_2 *des variables aléatoires et* $c, c_1, c_2 \in \mathbb{R}$. *Alors*
(i) $\mathrm{Cov}[c_1 X_1 + c_2 X_2, Y] = c_1 \mathrm{Cov}[X_1, Y] + c_2 \mathrm{Cov}[X_2, Y]$
(ii) $\mathrm{Cov}[X, c_1 Y_1 + c_2 Y_2] = c_1 \mathrm{Cov}[X, Y_1] + c_2 \mathrm{Cov}[X, Y_2]$
(iii) $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$
(iv) $\mathrm{Cov}[X, c] = 0$
(v) $\mathrm{Cov}[X, X] = \mathrm{Var}[X]$
(vi) $|\mathrm{Cov}[X, Y]| \leq \sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}$, *et l'égalité a lieu si et seulement si* $Y = \alpha + \beta X$ *p.s. ou* $X = \alpha + \beta Y$ *p.s. pour certains* $\alpha, \beta \in \mathbb{R}$.

Exercice : prouver (vi) en utilisant l'inégalité de Cauchy-Schwarz (ch.2-p.43).

Corrélation

La corrélation est une version normalisée de la covariance, dont on peut non seulement interpréter le signe mais aussi la valeur absolue.

Définition

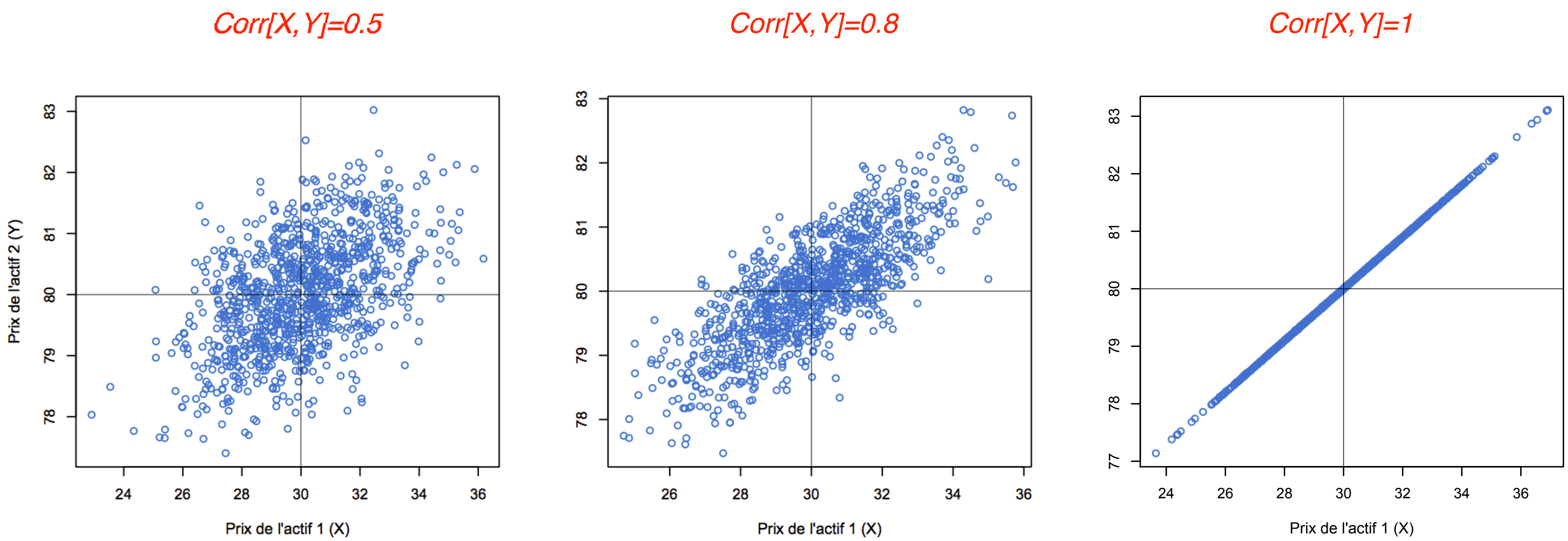
La corrélation entre X et Y est $\mathrm{Corr}[X, Y] = \dfrac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}}$

Théorème

(i) $\mathrm{Corr}[X, Y]$ *et* $\mathrm{Cov}[X, Y]$ *ont le même signe.*
(ii) $|\mathrm{Corr}[X, Y]| \leq 1$*, et l'égalité a lieu* $\Leftrightarrow \exists \alpha \in \mathbb{R}, \beta \in \mathbb{R}_0$ *tels que* $Y = \alpha + \beta X$ *p.s.*

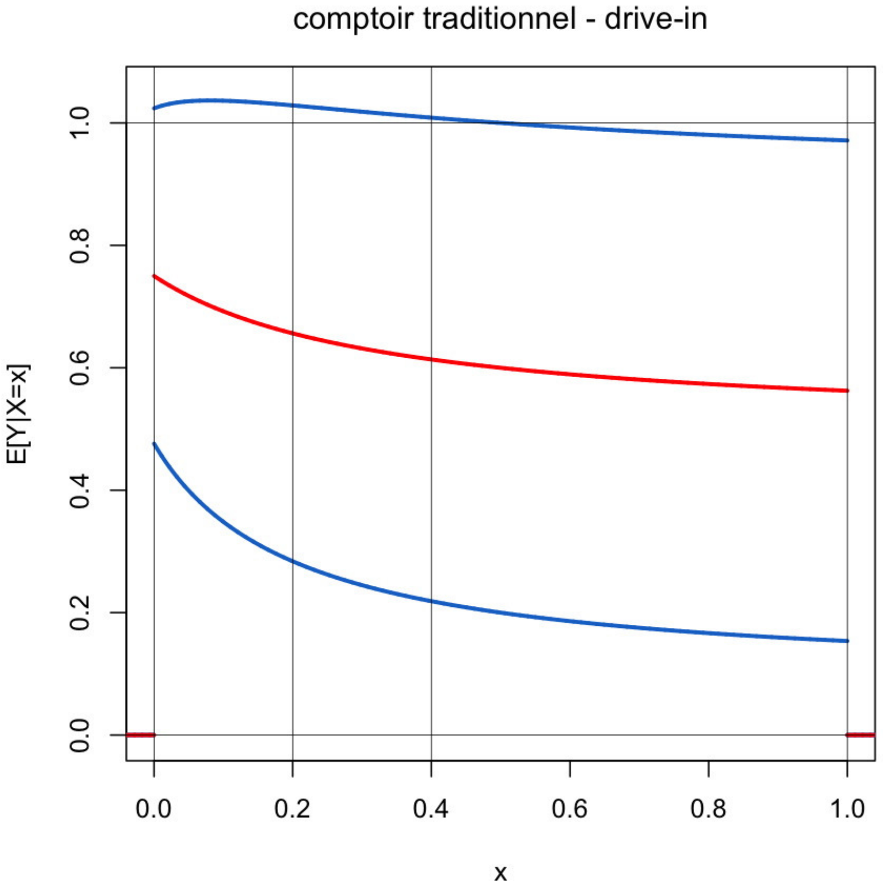
Au plus $|\mathrm{Corr}[X, Y]|$ est proche de 1, au plus la dépendance linéaire est forte

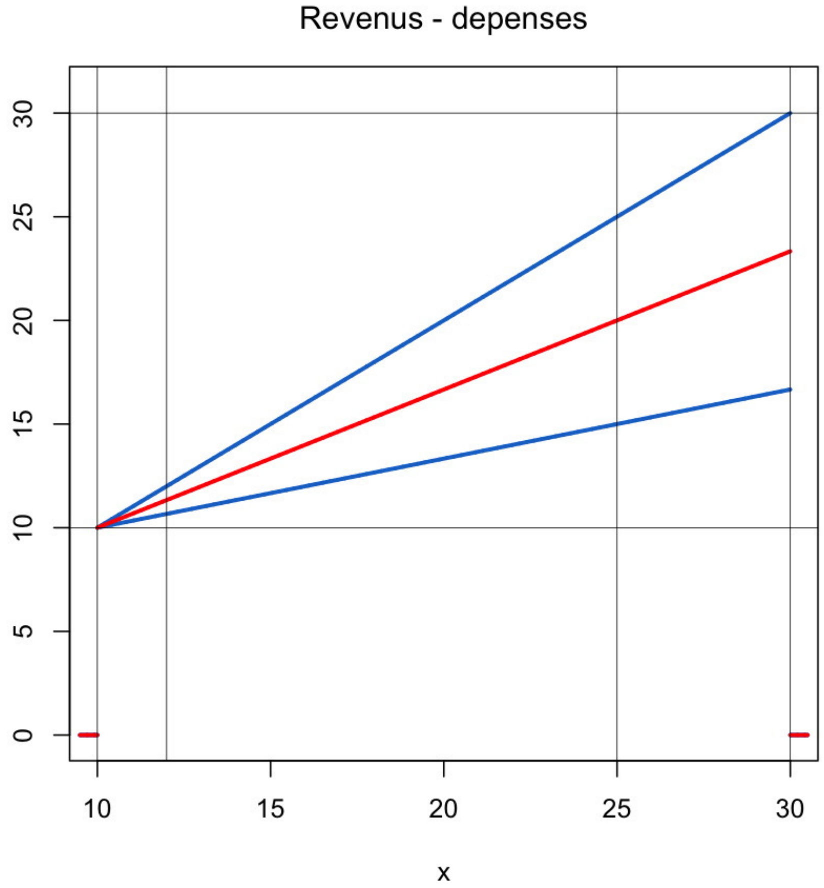
Corrélation



Au plus $|\mathrm{Corr}[X, Y]|$ est proche de 1, au plus la dépendance linéaire est forte

Corrélation





$\mathrm{Corr}[X, Y] \approx -0.13$

$\mathrm{Corr}[X, Y] \approx 0.49$

Corrélation

Théorème

Soient X, Y, X_1, X_2, Y_1, Y_2 *des variables aléatoires et* $c_1, c_2, d_1, d_2 \in \mathbb{R}$. *Alors*

(i) $\mathrm{Corr}[c_1 X + d_1, c_2 Y + d_2] = \mathrm{S}(c_1)\mathrm{S}(c_2)\mathrm{Corr}[X, Y]$*, où* $\mathrm{S}(a)$ *est le signe de a*

(ii) $\mathrm{Corr}[X, Y] = \mathrm{Corr}[Y, X]$

(iii) $\mathrm{Corr}[X, c] = 0$

(iv) $\mathrm{Corr}[X, X] = 1$

(v) $X \perp\!\!\!\perp Y \Rightarrow \mathrm{Corr}[X, Y] = 0$

Matrice de variance-covariance

L'opérateur d'espérance $\mathrm{E}[\cdot]$ peut être appliqué à un vecteur ou à une matrice, auquel cas on prend l'espérance composante par composante.

A un v.a. $Z = \binom{X}{Y}$, il est alors classique d'associer

- son vecteur moyen

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

- sa matrice de variance-covariance

$$\Sigma_Z = \mathrm{Var}[Z] = \mathrm{E}[(Z - \mathrm{E}[Z])(Z - \mathrm{E}[Z])^T] = \mathrm{E}\left[\begin{pmatrix} X - \mathrm{E}[X] \\ Y - \mathrm{E}[Y] \end{pmatrix}\begin{pmatrix} X - \mathrm{E}[X] & Y - \mathrm{E}[Y] \end{pmatrix}\right]$$

$$= \mathrm{E}\left[\begin{pmatrix} (X - \mathrm{E}[X])^2 & (X - \mathrm{E}[X])(Y - \mathrm{E}[Y]) \\ (Y - \mathrm{E}[Y])(X - \mathrm{E}[X]) & (Y - \mathrm{E}[Y])^2 \end{pmatrix}\right] = \begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}[X, Y] \\ \mathrm{Cov}[X, Y] & \mathrm{Var}[Y] \end{pmatrix}$$

Matrice de variance-covariance

Soit A une matrice 2×2, $b = \binom{b_1}{b_2}$ et $\lambda \in \mathbb{R}$. On vérifie facilement que

- $\mu_{AZ+b} = A\mu_Z + b$
- $\Sigma_{AZ+b} = A\Sigma_Z A^T$

et

- $\mathrm{E}[b^T Z + \lambda] = b^T \mu_Z + \lambda$
- $\mathrm{Var}[b^T Z + \lambda] = b^T \Sigma_Z (b^T)^T = b^T \Sigma_Z b.$

En particulier, $b^T \Sigma_Z b = \mathrm{Var}[b^T Z + \lambda] \geq 0.$
$\rightsquigarrow$ Σ_Z est semi-définie positive (et bien entendu symétrique).

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions *k*-variées

Motivation

Soient X et Y deux v.a.
Supposons que X est observé, mais pas Y.
Il est commun de vouloir "prédire" Y sur la base de X.

Exemples :

$X =$ revenus annuels d'un ménage belge (en milliers d'€)
$Y =$ dépenses annuelles d'un ménage belge (en milliers d'€)

$X =$ la proportion du temps où le comptoir standard du fast food est occupé
$Y =$ la proportion du temps où le drive-in est occupé

$X =$ mesure du stress à un examen (en pourcentage du maximum)
$Y =$ note sur 20 obtenue à cet examen

La prédiction requiert d'identifier une fonction de régression m telle que $Y \approx m(X)$.
$\rightsquigarrow$ Le prédicteur de Y sera alors simplement $\hat{Y} = m(X)$.

Régression générale

Puisque $\mathrm{E}[(U - V)^2] = 0$ implique que $U = V$ p.s. (voir (v), ch.2-p.43), on peut lire $\mathrm{E}[(U - V)^2]$ comme une distance entre les v.a. U et V. La définition suivante vise donc à rendre l'approximation $Y \approx m(X)$ aussi bonne que possible.

Définition

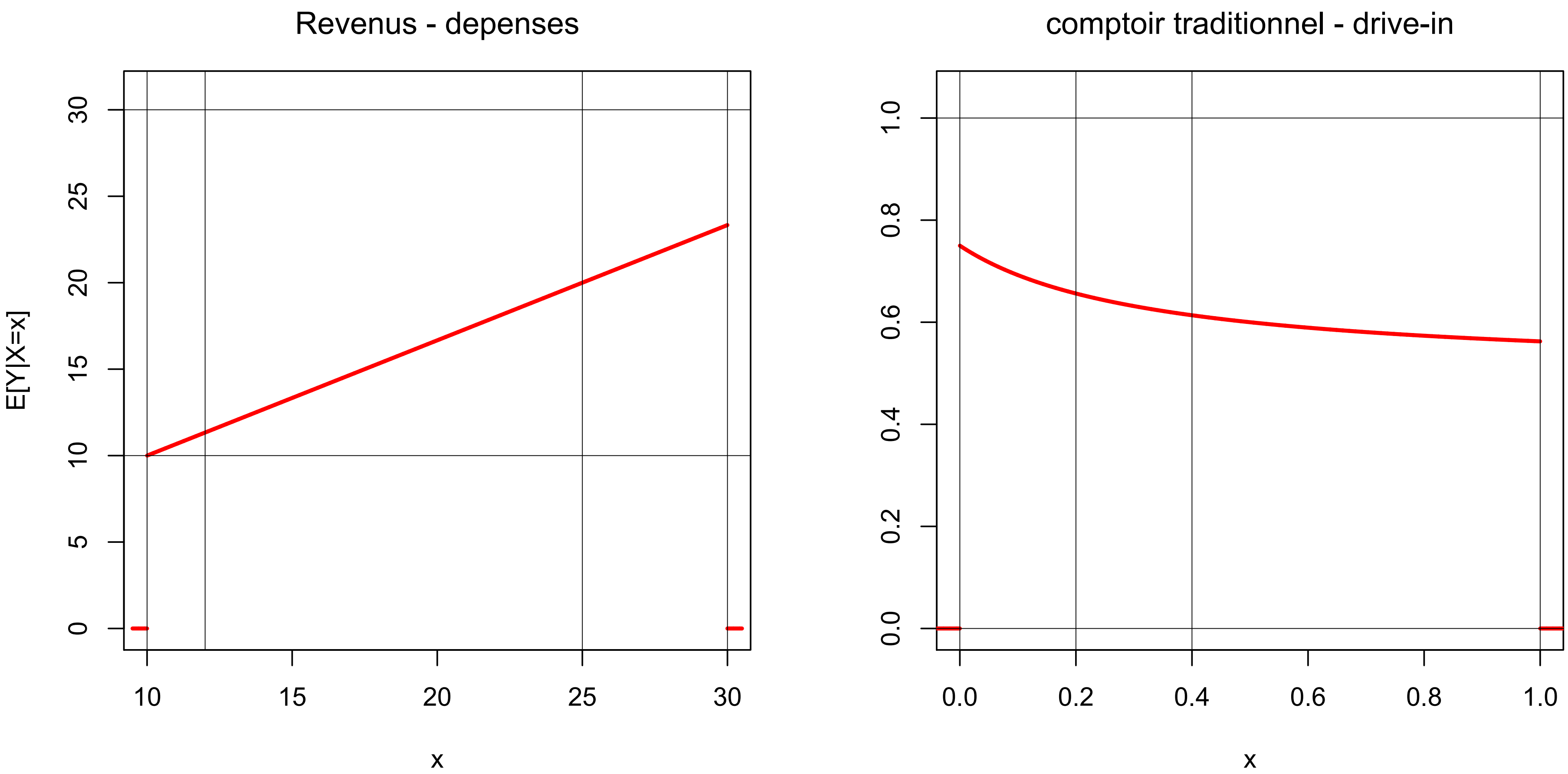
La fonction de régression m_{reg} est celle qui minimise l'erreur quadratique moyenne $\mathrm{E}[(Y - m(X))^2]$.

Le prédicteur de Y qui en résulte est $\hat{Y} = m_{\mathrm{reg}}(X)$.
Sa qualité est mesurée par l'erreur de prédiction $\mathrm{E}[(Y - m_{\mathrm{reg}}(X))^2]$.

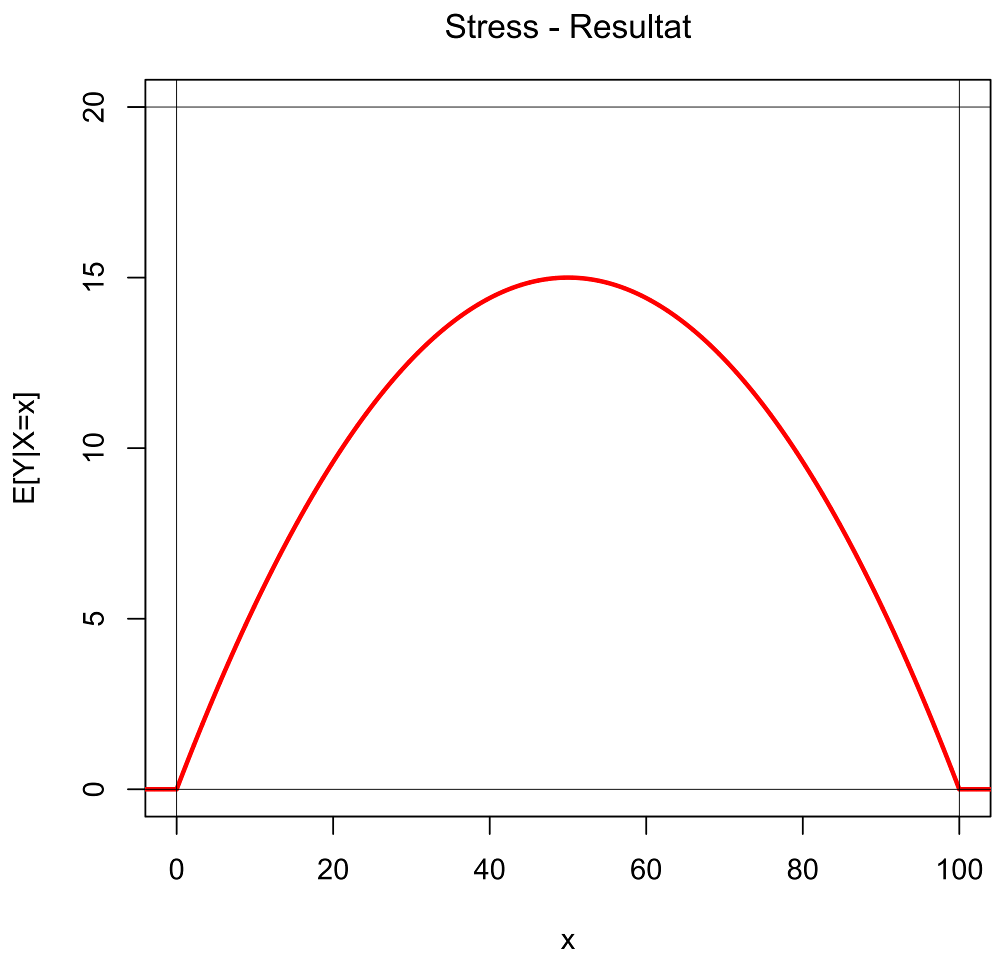
Théorème

(i) $m_{\mathrm{reg}}(x) = \mathrm{E}[Y|X = x]$ *pour tout x*
(ii) $\mathrm{E}[(Y - m_{\mathrm{reg}}(X))^2] = \mathrm{E}[\mathrm{Var}[Y|X]]$

Régression générale



Graphes de $m_{\mathrm{reg}}(x) = \mathrm{E}[Y|X = x]$



Graphe de $m_{\mathrm{reg}}(x) = \mathrm{E}[Y|X = x]$

Régression générale

Preuve : (i) pour toute fonction m, on a

$$\mathrm{E}[(Y-m(X))^2|X]$$

$$= \mathrm{E}[\{(Y-\mathrm{E}[Y|X]) + (\mathrm{E}[Y|X] - m(X))\}^2|X]$$

$$= \mathrm{E}[(Y-\mathrm{E}[Y|X])^2|X] + \mathrm{E}[(\mathrm{E}[Y|X]-m(X))^2|X] + 2\mathrm{E}[(Y-\mathrm{E}[Y|X])(\mathrm{E}[Y|X]-m(X))|X]$$

$$= \mathrm{Var}[Y|X] + (\mathrm{E}[Y|X]-m(X))^2 + 2(\mathrm{E}[Y|X]-m(X))\underbrace{\mathrm{E}[(Y-\mathrm{E}[Y|X])|X]}_{\mathrm{E}[Y|X]-\mathrm{E}[Y|X]=0}$$

$$= \mathrm{Var}[Y|X] + (\mathrm{E}[Y|X]-m(X))^2.$$

En prenant l'espérance, on obtient

$$\mathrm{E}[(Y-m(X))^2] = \mathrm{E}[\mathrm{Var}[Y|X]] + \mathrm{E}[(\mathrm{E}[Y|X]-m(X))^2]. \qquad (*)$$

Pour minimiser ceci, il faut donc prendre $m(X) = \mathrm{E}[Y|X]$.

(ii) Au vu de (*), la valeur minimale de $\mathrm{E}[(Y-m(X))^2])$ est donc $\mathrm{E}[\mathrm{Var}[Y|X]]$. $\square$

Régression générale

Preuve du théorème énoncé en ch.3-p.51 :

Ci-dessus, on a prouvé (*) pour une fonction m quelconque.
Si on prend $m(x) = \mathrm{E}[Y]$, on obtient

$$\mathrm{E}[(Y - \mathrm{E}[Y])^2] = \mathrm{E}[\mathrm{Var}[Y|X]] + \mathrm{E}[(\mathrm{E}[Y|X] - \mathrm{E}[Y])^2],$$

ce qui fournit

$$\begin{aligned}\mathrm{Var}[Y] = \mathrm{E}[(Y - \mathrm{E}[Y])^2] &= \mathrm{E}[\mathrm{Var}[Y|X]] + \mathrm{E}[(\mathrm{E}[Y|X] - \underbrace{\mathrm{E}[Y]}_{=\mathrm{E}[\mathrm{E}[Y|X]]})^2] \\ &= \mathrm{E}[\mathrm{Var}[Y|X]] + \mathrm{Var}[\mathrm{E}[Y|X]]. \qquad \square\end{aligned}$$

Régression linéaire

Parfois, la relation entre X et Y est linéaire ou presque linéaire.

Exemple : revenus - dépenses
Exemple : le fast food

On gagne alors à se restreindre à des fonctions m de la forme $x \mapsto m(x) = \alpha x + \beta$, car cela fournit un modèle simple et interprétable pour la relation entre X et Y.

Définition

La fonction de régression linéaire m_{reglin} est la fonction $x \mapsto m(x) = \alpha x + \beta$ qui minimise l'erreur quadratique moyenne $\mathrm{E}[(Y - m(X))^2]$

La qualité de la prédiction sera alors mesurée par

$$\mathrm{E}[(Y - m_{\mathrm{reglin}}(X))^2] \quad \Big(\geq \mathrm{E}[(Y - m_{\mathrm{reg}}(X))^2] \Big).$$

Régression linéaire

Notons $\sigma_{XY} = \mathrm{Cov}[X, Y]$ et $\rho_{XY} = \mathrm{Corr}[X, Y]$.

Théorème

(i) $m_{\mathrm{reglin}}(x) = \alpha_{XY} x + \beta_{XY}$, *où*

$$\alpha_{XY} = \rho_{XY} \frac{\sigma_Y}{\sigma_X} \quad \textit{et} \quad \beta_{XY} = \mu_Y - \alpha_{XY} \mu_X.$$

(ii) $\mathrm{E}[(Y - m_{\mathrm{reglin}}(X))^2] = (1 - \rho_{XY}^2)\sigma_Y^2$.

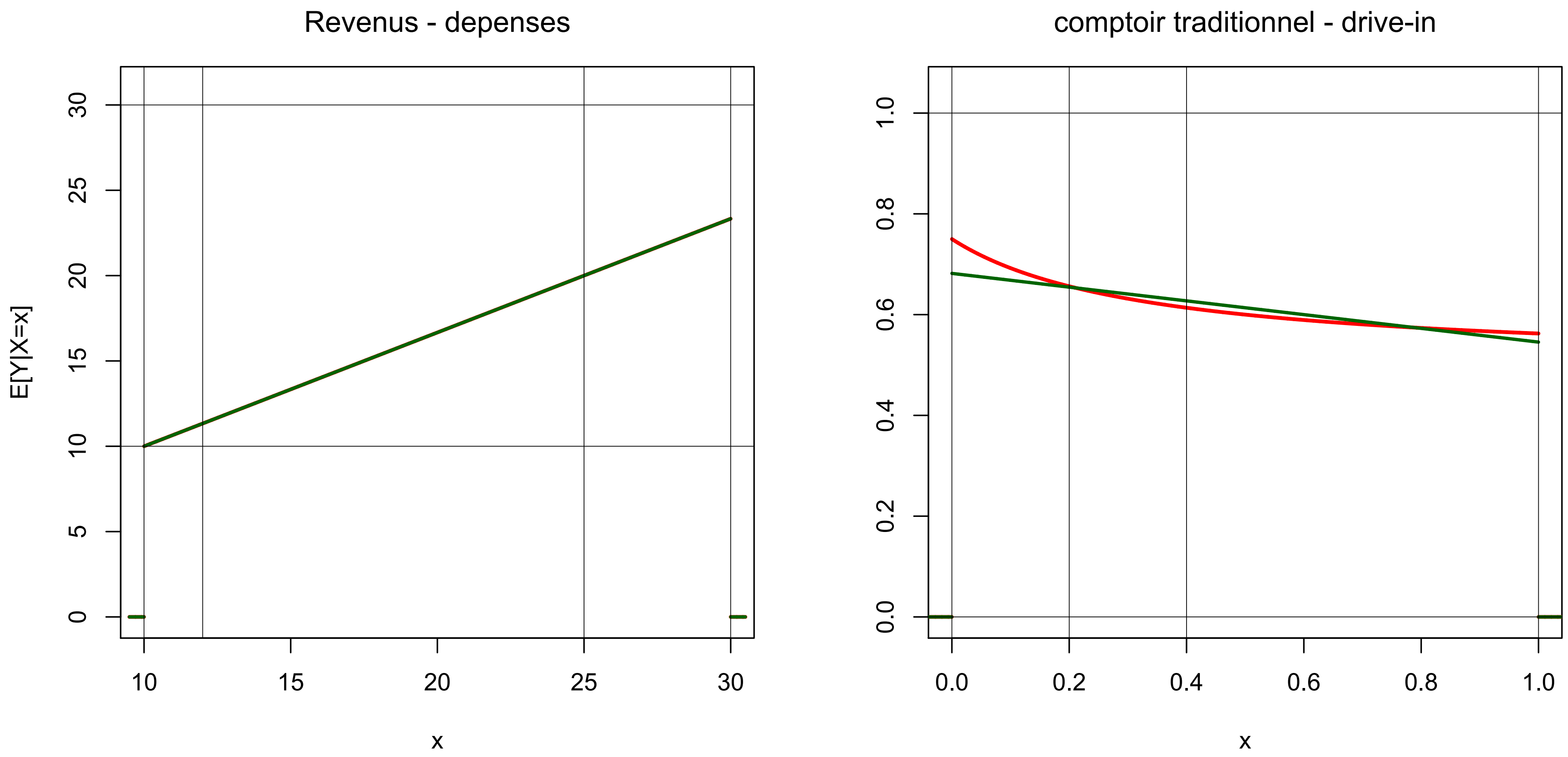
L'équation $y = m_{\mathrm{reglin}}(x) = \alpha_{XY} x + \beta_{XY}$ de la droite de régression se réécrit donc $(y - \mu_Y) = \alpha_{XY}(x - \mu_X)$. La droite est de pente α_{XY} et passe par le point (μ_X, μ_Y).

Le signe de la pente est le signe de ρ_{XY}.

L'erreur de prédiction est décroissante en $|\rho_{XY}|$.

- $|\rho_{XY}| = 0$ fournit donc l'erreur de prévision maximale
- $|\rho_{XY}| = 1$ fournit donc l'erreur de prévision minimale (en fait nulle !)

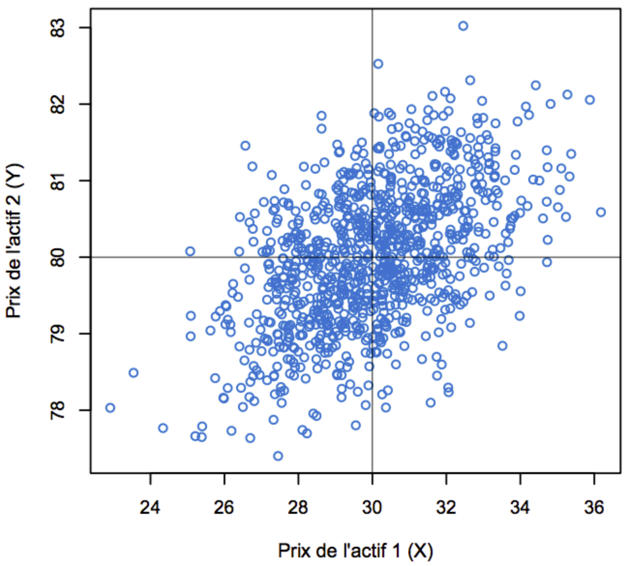
Régression linéaire



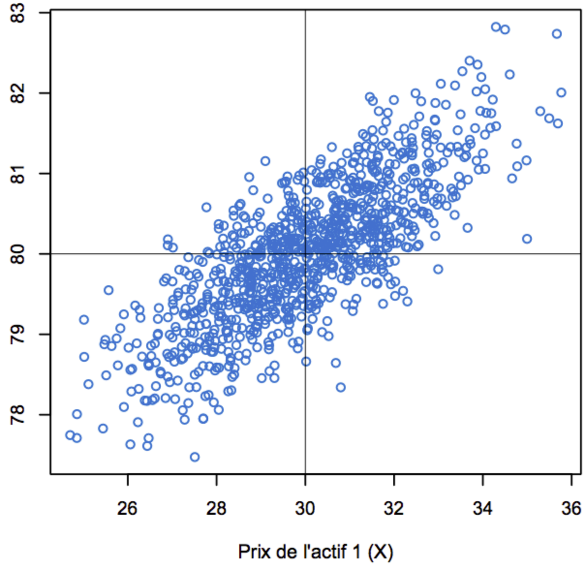
Graphes de $m_{\mathrm{reg}}(x) = \mathrm{E}[Y|X = x]$ et de $m_{\mathrm{reglin}}(x) = \alpha_{XY}x + \beta_{XY}$

Corrélation

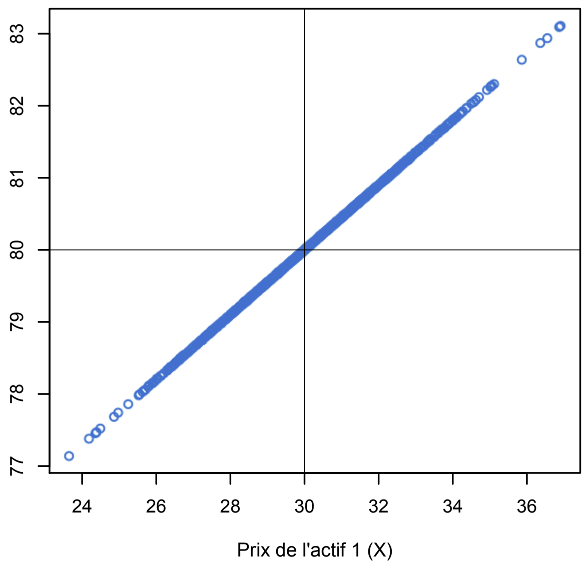
Corr[X,Y]=0.5



Corr[X,Y]=0.8



Corr[X,Y]=1



Au plus $|\mathrm{Corr}[X,Y]|$ est proche de 1, au plus la dépendance linéaire est forte

Régression linéaire

Preuve : (i) il s'agit de trouver les minima $(\alpha_{XY}, \beta_{XY})$ de la fonction

$$(\alpha, \beta) \mapsto h(\alpha, \beta) = \mathrm{E}[(Y - \alpha X - \beta)^2]$$

$$= \mathrm{E}[\{(Y - \mu_Y) - \alpha\,(X - \mu_X) + (\mu_Y - \alpha\mu_X - \beta)\}^2]$$

$$= \underbrace{\mathrm{E}[(Y - \mu_Y)^2]}_{\sigma_Y^2} + \alpha^2\,\underbrace{\mathrm{E}[(X - \mu_X)^2]}_{\sigma_X^2} + (\mu_Y - \alpha\mu_X - \beta)^2 - 2\alpha\,\underbrace{\mathrm{E}[(X - \mu_X)(Y - \mu_Y)]}_{\sigma_{XY}} + 0 + 0.$$

Ces minima se trouvent parmi les solutions du système

$$\begin{cases} \dfrac{\partial h}{\partial \alpha}(\alpha, \beta) = 0 \\[2ex] \dfrac{\partial h}{\partial \beta}(\alpha, \beta) = 0 \end{cases}$$

(voir le cours de math de bloc 2)

Régression linéaire

Puisque $h(\alpha,\beta) = \sigma_Y^2 + \alpha^2\sigma_X^2 + (\mu_Y - \alpha\mu_X - \beta)^2 - 2\alpha\sigma_{XY}$, ce système se réécrit

$$\begin{cases} 2\alpha\sigma_X^2 + 2(\mu_Y - \alpha\mu_X - \beta) \times (-\mu_X) - 2\sigma_{XY} = 0 \\ 2(\mu_Y - \alpha\mu_X - \beta) \times (-1) = 0, \end{cases}$$

et admet pour unique solution

$$\begin{cases} \alpha = \dfrac{\sigma_{XY}}{\sigma_X^2} = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y} \times \dfrac{\sigma_Y}{\sigma_X} = \rho_{XY}\dfrac{\sigma_Y}{\sigma_X} & \stackrel{!}{=} \alpha_{XY} \\ \beta = \mu_Y - \alpha\mu_X & \stackrel{!}{=} \beta_{XY}. \end{cases}$$

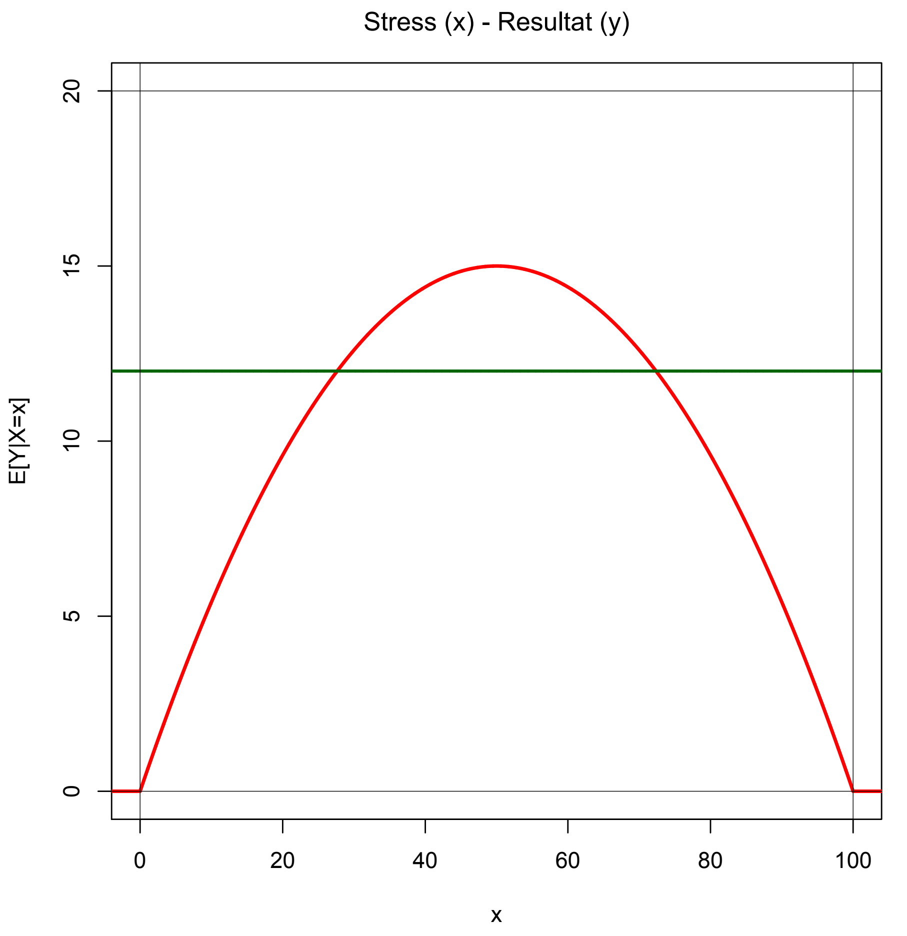
On montre facilement que $(\alpha_{XY}, \beta_{XY})$ est bien un minimum global.

(ii) La valeur minimale associée est alors

$$\begin{aligned} h(\alpha_{XY},\beta_{XY}) &= \sigma_Y^2 + \alpha_{XY}^2\sigma_X^2 + (\mu_Y - \alpha_{XY}\mu_X - \beta_{XY})^2 - 2\alpha_{XY}\rho_{XY}\sigma_X\sigma_Y \\ &= \sigma_Y^2 + \rho_{XY}^2\sigma_Y^2 + 0^2 - 2\rho_{XY}^2\sigma_Y^2 \\ &= (1-\rho_{XY}^2)\sigma_Y^2. \end{aligned}$$

$\square$

Un tel modèle linéaire est parfois à éviter !



Graphes de $m_{\mathrm{reg}}(x) = \mathrm{E}[Y|X = x]$ et $m_{\mathrm{reglin}}(x) = \alpha_{XY} x + \beta_{XY}$

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Le v.a. $Z = (X, Y)$ est de loi normale bivariée de paramètres $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ si Z admet la fonction de densité

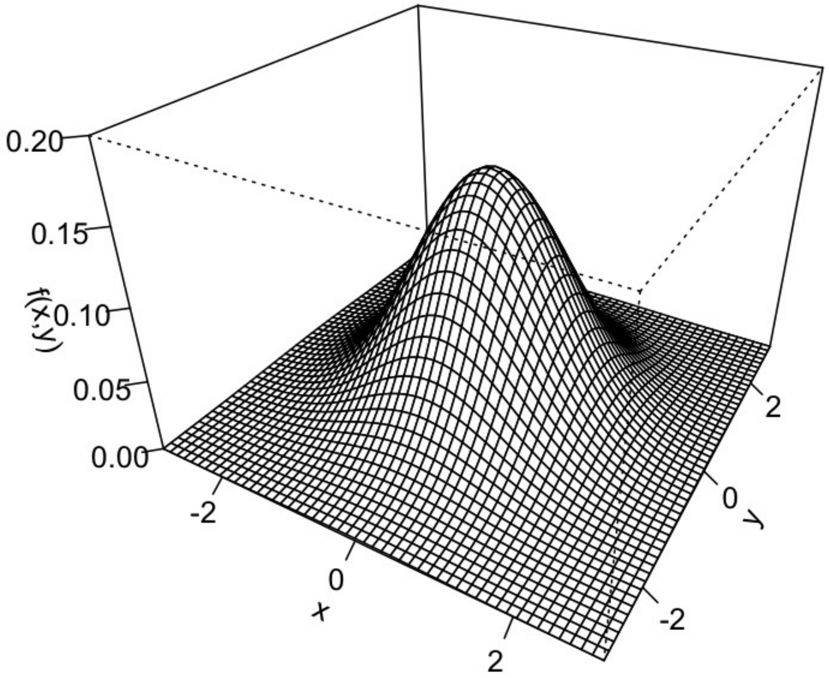
$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)}{\sigma_1}\frac{(y-\mu_2)}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}.$$

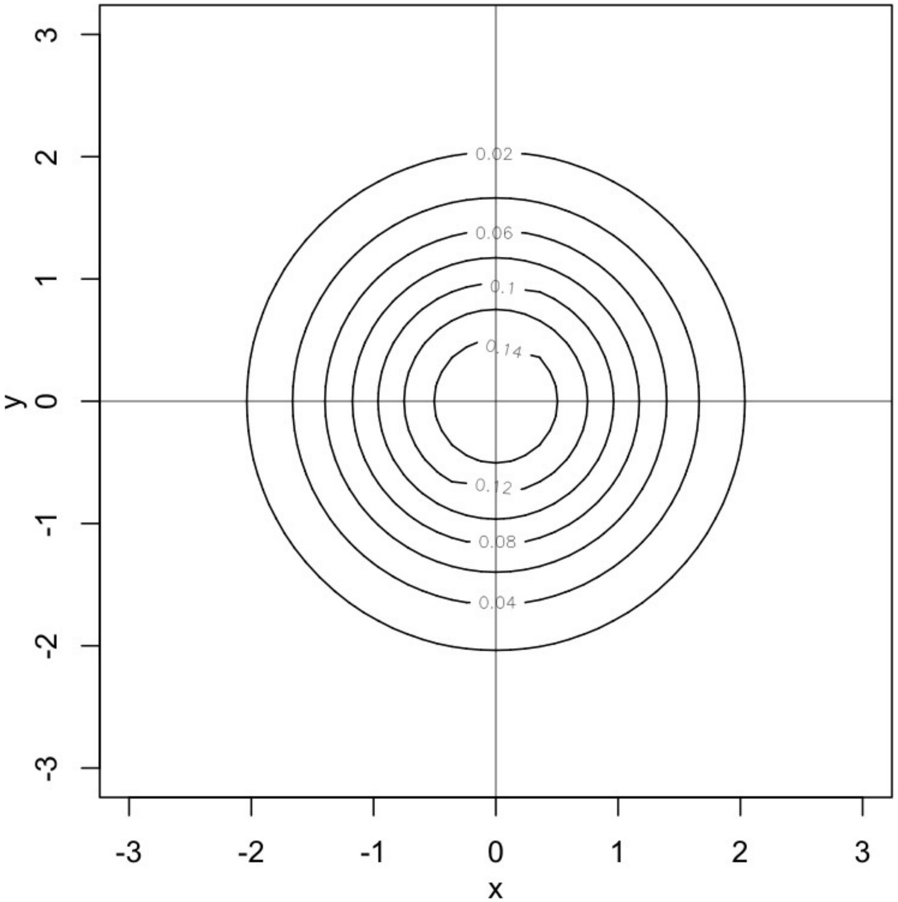
Ici, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_0^+$, et $\rho \in (-1, 1)$.

Pour rappel, la densité permet de calculer des probabilités via

$$P[(X,Y) \in B] = \iint_B f(x,y)\, dy\, dx \qquad \forall B \in \mathcal{B}^2.$$

Loi normale bivariée standard

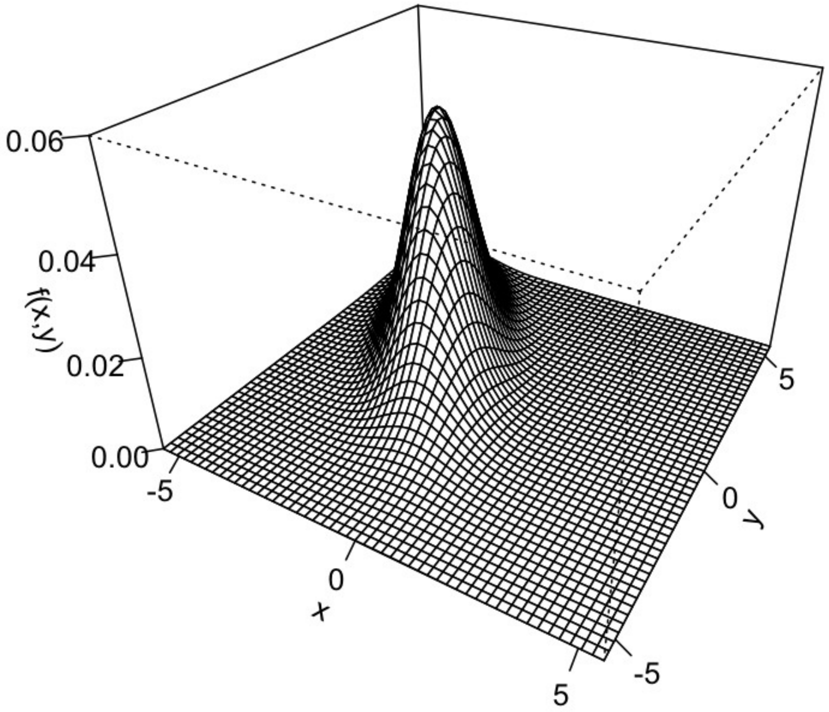




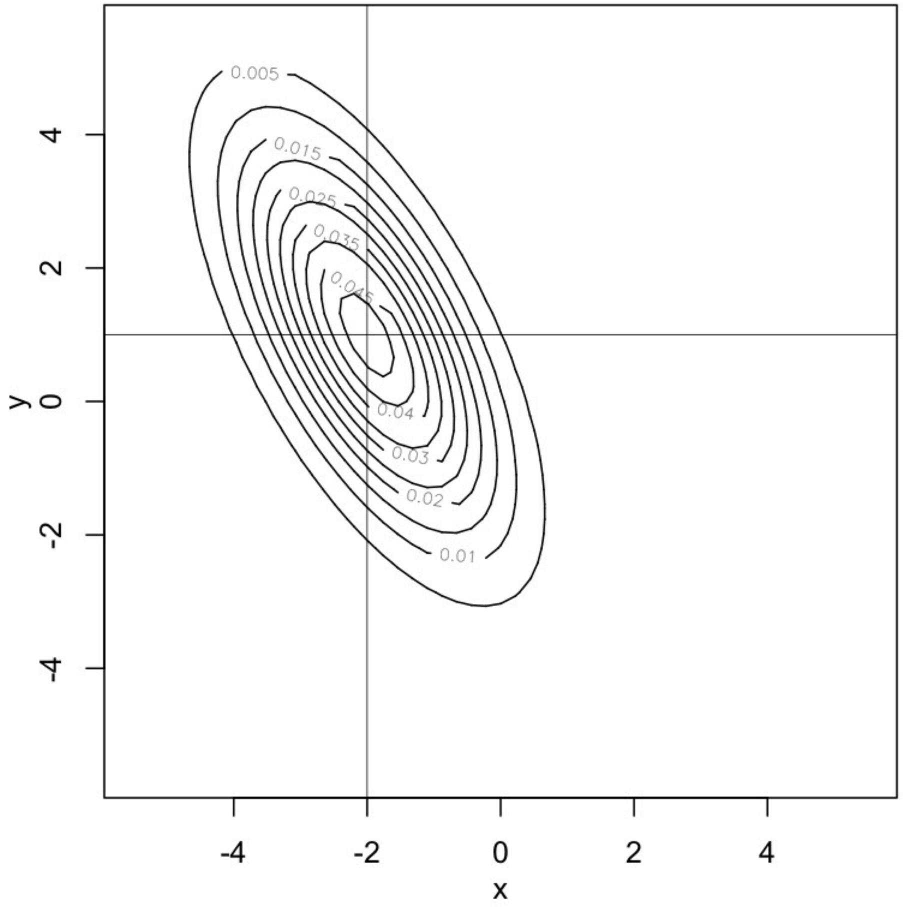
Densité de la loi normale bivariée avec $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) = (0, 0, 1, 1, 0)$

Courbes de niveau de cette densité

Loi normale bivariée



Densité de la loi normale bivariée avec
$(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) = (-2, 1, 1.5, 3.5, -0.65)$



Courbes de niveau de cette densité

Lois normales bivariées

Pour interpréter les paramètres $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$, nous allons déterminer les distributions marginales et conditionnelles de cette loi.

Nous aurons besoin de l'identité $f(x, y) = h_1(x) h_2(x, y)$, où

$$x \mapsto h_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}$$

est la fonction de densité de la loi $\mathcal{N}(\mu_1, \sigma_1^2)$, et où

$$y \mapsto h_2(x, y) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}(y-[\mu_2+\rho\sigma_2\sigma_1^{-1}(x-\mu_1)])^2}$$

est, $\forall x$, la densité de la loi $\mathcal{N}(\mu_2 + \rho\sigma_2\sigma_1^{-1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$.

(Etablir cette égalité est un exercice calculatoire facile).

Théorème

Soit $Z = (X, Y)$ de loi normale bivariée de paramètres $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. Alors
(i) $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$
(ii) $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Les distributions marginales d'une loi normale bivariée sont normales (univariées).
On a $\mu_1 = \mathrm{E}[X]$, $\mu_2 = \mathrm{E}[Y]$, $\sigma_1^2 = \mathrm{Var}[X]$ et $\sigma_2^2 = \mathrm{Var}[Y]$.

Preuve : (i) l'identité $f(x, y) = h_1(x)h_2(x, y)$ livre

$$f^X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = h_1(x) \underbrace{\int_{-\infty}^{\infty} h_2(x, y)\, dy}_{=1} = h_1(x).$$

Puisque h_1 est la densité de la loi $\mathcal{N}(\mu_1, \sigma_1^2)$, ceci établit le résultat.

(ii) La preuve est similaire, mais requiert une autre identité pour f (exercice). $\square$

Théorème

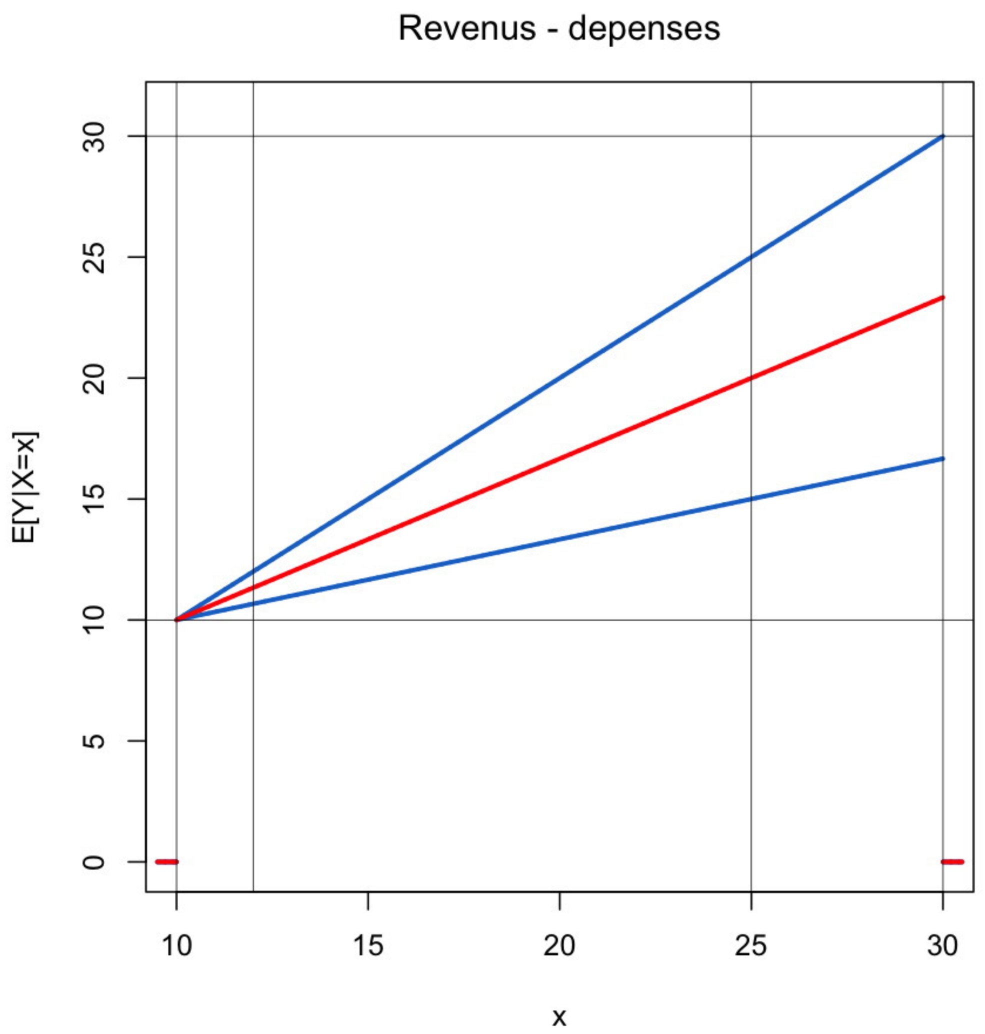
Soit $Z = (X, Y)$ de loi normale bivariée de paramètres $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. Alors
(i) $Y|[X = x] \sim \mathcal{N}(\mu_2 + \rho\sigma_2\sigma_1^{-1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$
(ii) $X|[Y = y] \sim \mathcal{N}(\mu_1 + \rho\sigma_1\sigma_2^{-1}(y - \mu_2), \sigma_1^2(1 - \rho^2))$.

Preuve : (i) le théorème précédent livre

$$f^{Y|[X=x]}(y) = \frac{f(x, y)}{f^X(x)} = \frac{h_1(x)h_2(x, y)}{h_1(x)} = h_2(x, y).$$

Puisque $y \mapsto h_2(x, y)$ est, $\forall x$, la densité de la loi $\mathcal{N}(\mu_2 + \rho\sigma_2\sigma_1^{-1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$, ceci établit le résultat. (ii) La preuve, qui fait usage de l'identité requise pour montrer la partie (ii) du théorème précédent, est entièrement similaire. $\square$

$m_{\text{reg}}(x) = \mathrm{E}[Y|X = x] = \mu_2 + \rho\sigma_2\sigma_1^{-1}(x - \mu_1)$ est une fonction "linéaire" de x.
$\mathrm{Var}[Y|X = x] = \sigma_2^2(1 - \rho^2)$ ne dépend pas de x (on parle d'homoscédasticité).



$\mathrm{E}[Y|X=x]$ n'est pas linéaire $\forall x \in \mathbb{R}$ et $\mathrm{Var}[Y|X=x]$ dépend de x (hétéroscédasticité)

$\Rightarrow (X,Y)$ n'est pas de loi normale bivariée

Lois normales bivariées

Ceci nous permet d'interpréter le paramètre ρ.

Puisque $E[Y|X] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$, on a

$$\mathrm{Cov}[X,Y] \stackrel{\mathrm{def}}{=} E[(X - E[X])(Y - E[Y])] = E[(X - \mu_1)(Y - \mu_2)] = E[E[(X - \mu_1)(Y - \mu_2)|X]]$$

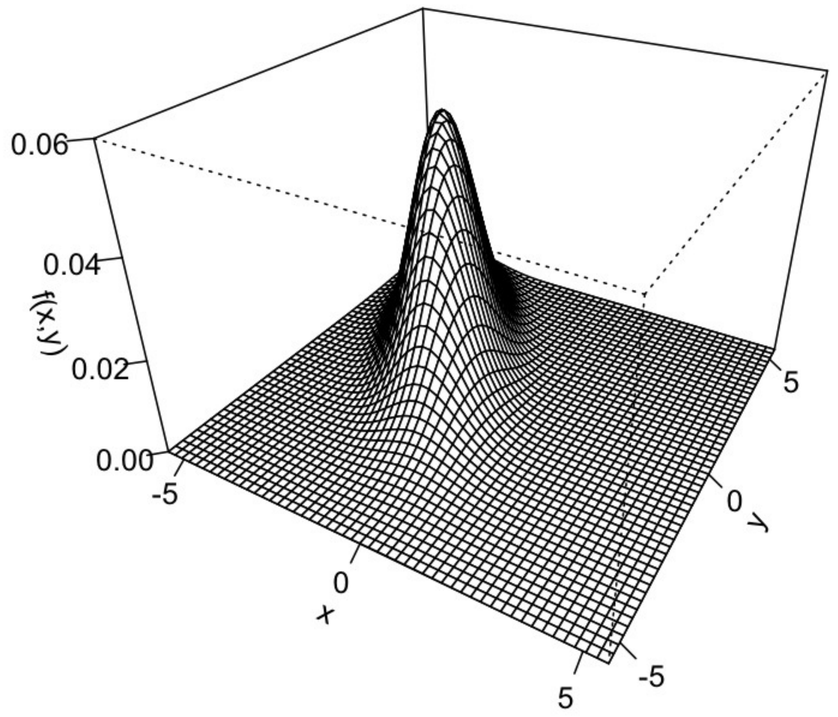
$$= E[(X - \mu_1)E[(Y - \mu_2)|X]] = E[(X - \mu_1)\underbrace{(E[Y|X] - \mu_2)}_{=\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)}] = \rho\frac{\sigma_2}{\sigma_1}\underbrace{\mathrm{Var}[X]}_{=\sigma_1^2} = \rho\sigma_1\sigma_2.$$

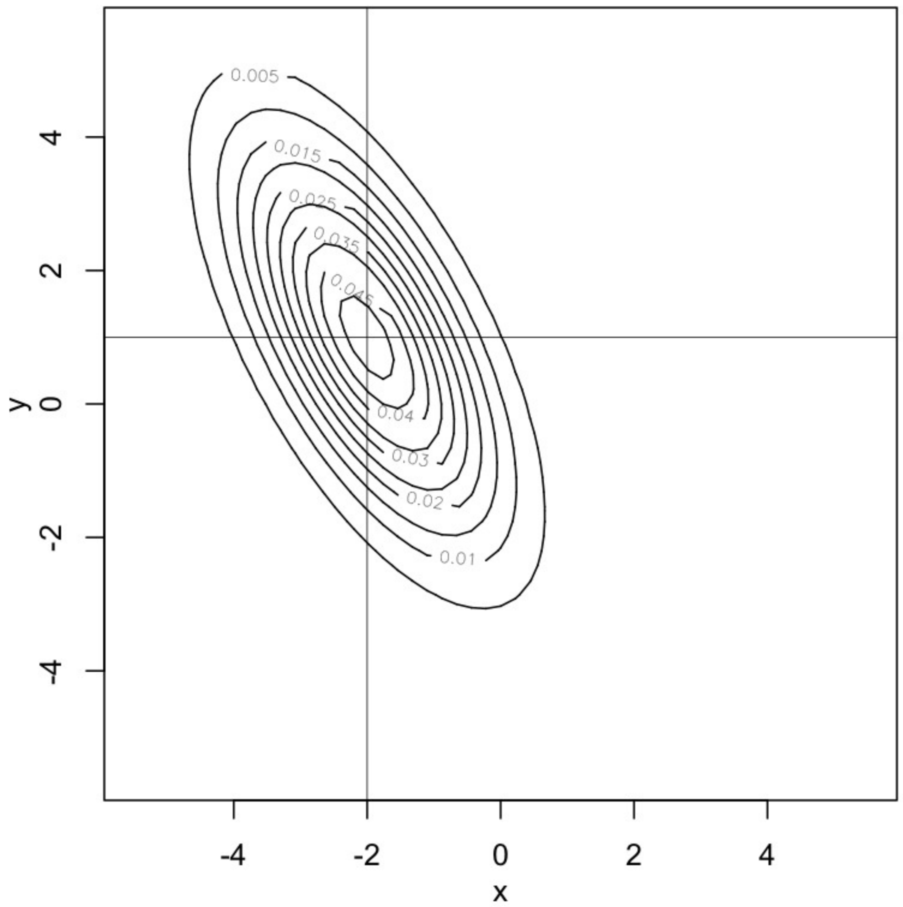
Ceci livre

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho.$$

On conclut donc que $\rho = \mathrm{Corr}[X,Y]$

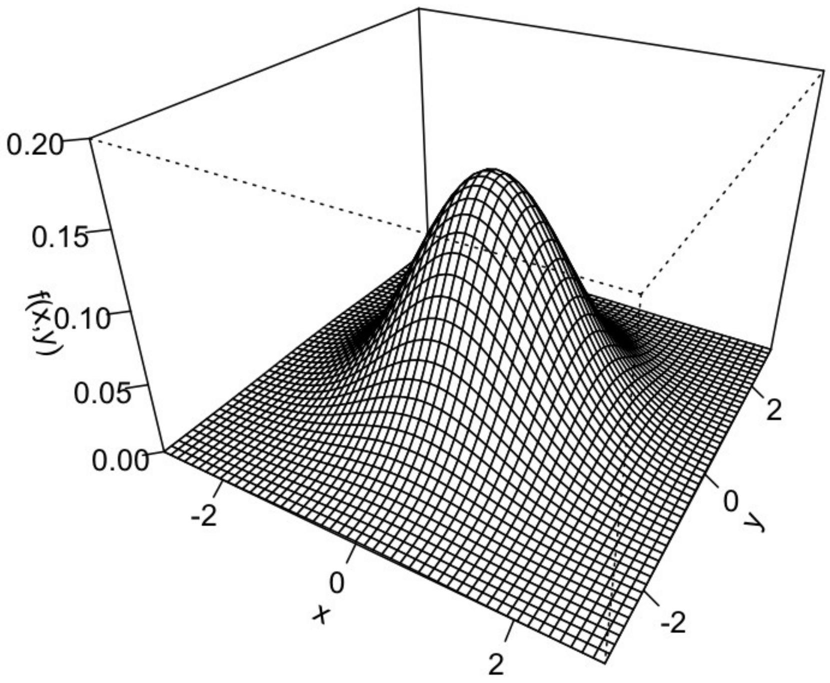
Lois normales bivariées

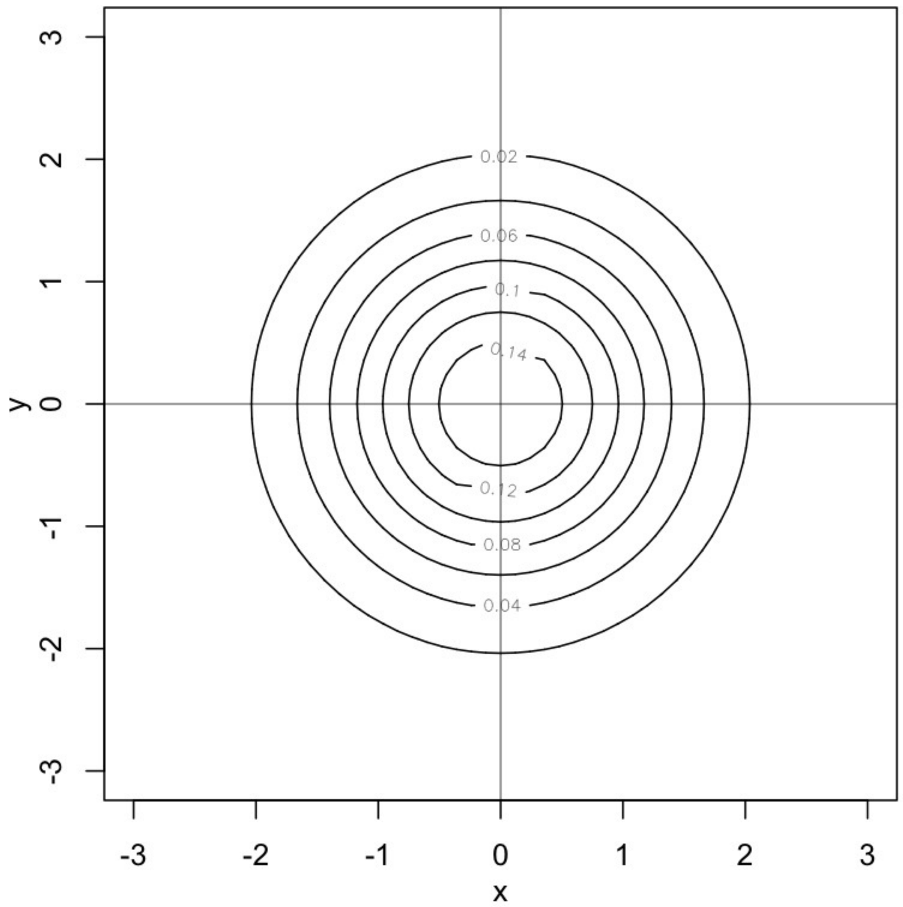




$\mu_1 = \mathrm{E}[X] = -2, \mu_2 = \mathrm{E}[Y] = 1, \sigma_1^2 = \mathrm{Var}[X] = 1.5, \sigma_2^2 = \mathrm{Var}[Y] = 3.5, \rho = \mathrm{Corr}[X, Y] = -0.65$

Lois normales bivariées





$$\mu_1 = \mathrm{E}[X] = 0, \mu_2 = \mathrm{E}[Y] = 0, \sigma_1^2 = \mathrm{Var}[X] = 1, \sigma_2^2 = \mathrm{Var}[Y] = 1, \rho = \mathrm{Corr}[X, Y] = 0$$

En utilisant l'interprétation des paramètres et le fait que $\mathrm{E}[Y|X=x] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$, on obtient

$$
\begin{aligned}
m_{\text{reglin}}(x) &= \rho_{XY}\frac{\sigma_Y}{\sigma_X}x + \mu_Y - \rho_{XY}\frac{\sigma_Y}{\sigma_X}\mu_X \\
&= \rho\frac{\sigma_2}{\sigma_1}x + \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 \\
&= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \\
&= \mathrm{E}[Y|X=x] \\
&= m_{\text{reg}}(x).
\end{aligned}
$$

Ceci confirme que, pour la loi normale bivariée, la fonction de régression m_{reg} coïncide avec la fonction de régression linéaire m_{reglin}.

Théorème

Soit (X,Y) de loi normale bivariée.
Alors $X \perp\!\!\!\perp Y \Leftrightarrow \mathrm{Cov}[X,Y] = 0$

Preuve : ($\Rightarrow$) Cette implication est toujours vraie (voir ch.3-p.81).
($\Leftarrow$) Si $\mathrm{Cov}[X,Y] = 0$, alors $\rho = \mathrm{Corr}[X,Y] = 0$. Donc la densité de (X,Y) est

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}}\right)\left(\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\frac{(y-\mu_2)^2}{\sigma_2^2}}\right) = f^X(x) f^Y(y) \end{aligned}$$

pour tout x, y, ce qui implique que $X \perp\!\!\!\perp Y$. $\square$

Nous avons vu précédemment que l'implication ($\Leftarrow$) n'est pas vraie en général

Nous terminons la présentation de ces lois par leur notation matricielle.

Désignons le vecteur moyen et la matrice de variance-covariance par

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

On vérifiera alors facilement que

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\quad \times e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)}{\sigma_1}\frac{(y-\mu_2)}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]} \\ &= \frac{1}{2\pi}\frac{1}{\sqrt{\det\Sigma}}\, e^{-\frac{1}{2}\left[\left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu\right)^T \Sigma^{-1} \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu\right)\right]}. \end{aligned}$$

Plan du chapitre 3

3 Vecteurs aléatoires

- Définition, distribution jointe et fonction de répartition
- Distribution jointe et distributions marginales
- Distributions conditionnelles
- Indépendance
- Covariance, corrélation, et matrice de variance-covariance
- Courbes de régression
- Lois normales bivariées
- Distributions k-variées

Motivation

Jusqu'ici, nous nous sommes restreints aux vecteurs aléatoires bivariés.

Mais on doit souvent considérer plus de deux variables aléatoires simultanément, que ce soit

- pour étudier un portefeuille boursier composé de plus de deux actifs,
- pour prédire une variable sur la base de plusieurs autres variables (le résultat à l'examen sur la base du stress, du nombre d'heures d'étude, etc.),
- ...

Sans rentrer dans autant de détails que dans le cas bivarié, nous allons donc considérer des v.a. k-variés

$$X = (X_1, \ldots, X_k)^T,$$

où chaque X_ℓ est une variable aléatoire. On définit la fonction de répartition comme $F(x_1, \ldots, x_k) = P[X_1 \leq x_1, \ldots, X_k \leq x_k]$ pour tout $(x_1, \ldots, x_k)$.

Le cas discret

Dans le cas discret, chaque X_ℓ n'a qu'un nombre fini ou infini dénombrable de valeurs possibles : $x_i^{(\ell)}$, $i \in \mathcal{I}^{(\ell)}$.

La distribution de X est encore déterminée par la collection de toutes les valeurs possibles de $X = (X_1, \ldots, X_k)^T$ accompagnée des probabilités correspondantes $p_{i_1 \ldots i_k} := P[X_1 = x_{i_1}^{(1)}, \ldots, X_k = x_{i_k}^{(k)}]$.

On détermine la probabilité que X se réalise dans un borélien $B \subset \mathbb{R}^k$ via

$$P[X \in B] = \sum_{(i_1,\ldots,i_k):(x_{i_1}^{(1)},\ldots,x_{i_k}^{(k)}) \in B} p_{i_1 \ldots i_k}$$

et on calcule des espérances selon

$$\mathrm{E}[g(X_1, \ldots, X_k)] = \sum_{(i_1,\ldots,i_k)} g(x_{i_1}^{(1)}, \ldots, x_{i_k}^{(k)})\, p_{i_1 \ldots i_k}.$$

Le cas discret

Pour des espérances du type $\mathrm{E}[g(X_\ell)]$, on utilisera plutôt

$$\mathrm{E}[g(X_\ell)] = \sum_{i_\ell} g(x_{i_\ell}^{(\ell)})\, p_{i_\ell}^{(\ell)}, \quad \text{où } p_{i_\ell}^{(\ell)} = P[X_\ell = x_{i_\ell}^{(\ell)}] = \sum_{(i_1,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_k)} p_{i_1\ldots i_k},$$

fondée sur la distribution marginale de X_ℓ (valeurs possibles $x_{i_\ell}^{(\ell)}$, de probabilités respectives $p_{i_\ell}^{(\ell)}$).

De même, des espérances du type $\mathrm{E}[g(X_\ell, X_m)]$ peuvent être calculées via

$$\mathrm{E}[g(X_\ell, X_m)] = \sum_{i_\ell, i_m} g(x_{i_\ell}^{(\ell)}, x_{i_m}^{(m)})\, p_{i_\ell, i_m}^{(\ell),(m)},$$

où

$$p_{i_\ell, i_m}^{(\ell),(m)} = P[X_\ell = x_{i_\ell}^{(\ell)}, X_m = x_{i_m}^{(m)}] = \sum_{(i_1,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_{m-1},i_{m+1},\ldots,i_k)} p_{i_1\ldots i_k}$$

caractérise la distribution jointe (bivariée) de (X_ℓ, X_m).

Le cas discret

On peut aussi considérer des distributions marginales de dimension supérieure. C'est en fait nécessaire pour calculer les distributions conditionnelles puisque le dénominateur de

$$P[X_\ell = x_{i_\ell}^{(\ell)} | X_1 = x_{i_1}^{(1)}, \ldots, X_{\ell-1} = x_{i_{\ell-1}}^{(\ell-1)}, X_{\ell+1} = x_{i_{\ell+1}}^{(\ell+1)}, \ldots, X_k = x_{i_k}^{(k)}]$$

$$= \frac{P[X_1 = x_{i_1}^{(1)}, \ldots, X_k = x_{i_k}^{(k)}]}{P[X_1 = x_{i_1}^{(1)}, \ldots, X_{\ell-1} = x_{i_{\ell-1}}^{(\ell-1)}, X_{\ell+1} = x_{i_{\ell+1}}^{(\ell+1)}, \ldots, X_k = x_{i_k}^{(k)}]}$$

est associé à une distribution marginale $(k-1)$-variée (on peut encore définir des espérances et variances conditionnelles sur base de ces distributions conditionnelles).

Remarque : on peut vérifier que $X_1, \ldots, X_k$ sont mutuellement indépendantes si et seulement si $p_{i_1 \ldots i_k} = p_{i_1}^{(1)} \ldots p_{i_k}^{(k)}$ pour tout $i_1, \ldots, i_k$.

On utilisera ceci dans la partie "inférence statistique", où les $k(= n)$ observations seront supposées être des variables aléatoires mutuellement indépendantes.

Une distribution multivariée discrète particulière

Soient $k, n \in \mathbb{N}_0$ et $p_1, \ldots, p_k \in (0,1)$ tels que $\sum_{\ell=1}^{k} p_\ell = 1$.
Soit E une expérience aléatoire à k résultats possibles, avec $P[\text{résultat } \ell] = p_\ell$.

Alors $X = (X_1, \ldots, X_k)^T$ est de distribution multinomiale de paramètres $n, p_1, \ldots, p_k$ (notation : $X \sim \mathrm{Multin}(n, p_1, \ldots, p_k)$) si X_ℓ, $\ell = 1, \ldots, k$, compte le nombre de résultats ℓ dans une suite de n répétitions indépendantes de E.

Les valeurs possibles sont tous les $(n_1, \ldots, n_k)$ tels que $\sum_{\ell=1}^{k} n_\ell = n$.
Les probabilités correspondantes sont

$$P[X_1 = n_1, \ldots, X_k = n_k] = \frac{n!}{(n_1!) \ldots (n_k!)} \, p_1^{n_1} \ldots p_k^{n_k}.$$

Clairement, $X_\ell \sim \mathrm{Bin}(n, p_\ell)$.
On a donc $\mathrm{E}[X_\ell] = np_\ell$ et $\mathrm{Var}[X_\ell] = np_\ell(1 - p_\ell)$.

Il est aussi clair que, pour $\ell \neq m$, on a $X_\ell + X_m \sim \mathrm{Bin}(n, p_\ell + p_m)$.
On a donc $\mathrm{Var}[X_\ell + X_m] = n(p_\ell + p_m)(1 - (p_\ell + p_m))$.

Puisque $\mathrm{Var}[X_\ell + X_m] = \mathrm{Var}[X_\ell] + \mathrm{Var}[X_m] + 2\mathrm{Cov}[X_\ell, X_m]$, on obtient donc (exercice)

$$\mathrm{Cov}[X_\ell, X_m] = -np_\ell p_m < 0.$$

Les X_ℓ ne sont donc pas indépendantes. Ce n'est pas étonnant (pourquoi ?)
Le signe de la covariance entre X_ℓ et X_m n'est pas étonnant non plus (pourquoi ?)

Un exemple de loi multinomiale :

Au 1er tour de la dernière élection présidentielle française, on interroge n personnes sur leurs intentions de vote parmi les k candidats. En notant X_ℓ le nombre de sondés déclarant vouloir voter pour le candidat ℓ,

$$(X_1, \ldots, X_k)^T \sim \mathrm{Multin}(n, p_1, \ldots, p_k),$$

où p_ℓ est la proportion des Français en faveur du candidat ℓ.

Le cas continu

Pour décrire le cas continu, nous adoptons volontairement le même schéma de présentation que pour le cas discret, dans le but de mettre en évidence les analogies fortes entre les deux types de formules.

Le v.a. X est continu si il existe une fonction $f : \mathbb{R}^k \to \mathbb{R}^+$ telle que

$$P[X \in B] = \int_B f(x_1, \ldots, x_k)\, dx_k \ldots dx_1$$

pour tout borélien $B \subset \mathbb{R}^k$. Cette fonction f, encore appelée fonction de densité de probabilité, est liée à F à travers la relation

$$f(x_1, \ldots, x_k) := \frac{\partial^k F}{\partial x_1 \ldots \partial x_k}(x_1, \ldots, x_k).$$

On calcule alors des espérances selon

$$\mathrm{E}[g(X_1, \ldots, X_k)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_k)\, f(x_1, \ldots, x_k)\, dx_k \ldots dx_1.$$

Le cas continu

Pour des espérances du type $\mathrm{E}[g(X_\ell)]$, on utilisera plutôt

$$\mathrm{E}[g(X_\ell)] = \int_{-\infty}^{\infty} g(x_\ell)\, f^{X_\ell}(x_\ell)\, dx_\ell,$$

où

$$f^{X_\ell}(x_\ell) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_k)\, dx_k \dots dx_{\ell+1} dx_{\ell-1} \dots dx_1$$

est la densité marginale de X_ℓ.

De même, des espérances du type $\mathrm{E}[g(X_\ell, X_m)]$ peuvent être calculées via

$$\mathrm{E}[g(X_\ell, X_m)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_\ell, x_m)\, f^{(X_\ell, X_m)}(x_\ell, x_m)\, dx_m dx_\ell,$$

où

$$f^{(X_\ell, X_m)}(x_\ell, x_m) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_k)\, dx_k \dots dx_{m+1} dx_{m-1} \dots dx_{\ell+1} dx_{\ell-1} \dots dx_1$$

est la densité marginale de (X_ℓ, X_m).

Le cas continu

On peut ici aussi considérer des distributions marginales de dimension supérieure, comme c'est le cas (au dénominateur) dans les densités conditionnelles

$$f^{X_\ell|[X_1=x_1,\ldots,X_{\ell-1}=x_{\ell-1},X_{\ell+1}=x_{\ell+1},\ldots,X_k=x_k]}(x_\ell)$$

$$= \frac{f(x_1,\ldots,x_k)}{f^{(X_1,\ldots,X_{\ell-1},X_{\ell+1},\ldots,X_k)}(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)}$$

(on peut encore définir des espérances et variances conditionnelles sur base de ces densités conditionnelles).

Remarque : on peut vérifier que $X_1,\ldots,X_k$ sont mutuellement indépendantes si et seulement si $f(x_1,\ldots,x_k) = f^{X_1}(x_1)\ldots f^{X_k}(x_k)$ $\forall x_1,\ldots,x_k$.

Pour la même raison que dans le cas discret, ceci sera souvent utilisé dans la partie "inférence statistique" du cours.

Vecteur moyen et matrice de variance-covariance

Par analogie avec le cas bivarié, nous définissons le vecteur moyen et la matrice de variance-covariance du v.a. $X = (X_1, \ldots, X_k)^T$ respectivement comme le vecteur

$$\mathrm{E}[X] = \begin{pmatrix} \mathrm{E}[X_1] \\ \mathrm{E}[X_2] \\ \vdots \\ \mathrm{E}[X_k] \end{pmatrix}$$

et la matrice (symétrique et semi-définie positive, comme pour $k = 2$)

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{E}[(X - \mathrm{E}[X])(X - \mathrm{E}[X])^T] \\ &= \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \ldots & \mathrm{Cov}[X_1, X_k] \\ \mathrm{Cov}[X_2, X_1] & \mathrm{Var}[X_2] & \ldots & \mathrm{Cov}[X_2, X_k] \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}[X_k, X_1] & \mathrm{Cov}[X_k, X_2] & \ldots & \mathrm{Var}[X_k] \end{pmatrix}. \end{aligned}$$

Il est facile de vérifier que, pour toute matrice $(r \times k)$ A et pour tout vecteur $b \in \mathbb{R}^r$, on a $\mathrm{E}[AX + b] = A\mathrm{E}[X] + b$ et $\mathrm{Var}[AX + b] = A\mathrm{Var}[X]A^T$.

La loi normale k-variée

Une distribution multivariée continue particulière

Soient $\mu \in \mathbb{R}^k$ et Σ une matrice $(k \times k)$ symétrique et définie positive.
Alors $X = (X_1, \ldots, X_k)^T$ est de loi normale k-variée de paramètres μ et Σ (notation : $X \sim \mathcal{N}_k(\mu, \Sigma)$) si X admet la densité

$$f(x) = \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} \frac{1}{\sqrt{\det \Sigma}}\, e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},$$

où $x = (x_1, \ldots, x_k)^T$.

Ceci généralise donc

- la loi normale univariée ($k = 1$; voir ch.2-p.71) et
- la loi normale bivariée ($k = 2$; voir ch.3-p.117).

La loi normale k-variée

On peut montrer que

(i) $\mu = \mathrm{E}[X]$ et $\Sigma = \mathrm{Var}[X]$.

(ii) Pour toute matrice $(r \times k)$ A et pour tout vecteur $b \in \mathbb{R}^r$,

$$AX + b \sim \mathcal{N}_r(A\mu + b, A\Sigma A^T).$$

En particulier, si on prend $A = (0, \ldots, 0, 1, 0, \ldots, 0)$ (avec le "1" en position ℓ) et $b = 0$, on obtient que $X_\ell \sim \mathcal{N}_1(\mu_\ell, \Sigma_{\ell\ell})$.

(iii) Si Σ est une matrice diagonale, $X_1, \ldots, X_k$ sont mutuellement indépendantes.

Remarque : ce qu'on appellera "loi normale k-variée standard" est le cas particulier obtenu pour $\mu = 0$ et $\Sigma = I_k$ (la matrice identité de dimension k).