Vecteurs aléatoires

Diana Nurbakova

Contents

Dans la vie réelle, on va souvent s'intéresser aux plusieures variables aléatoires qui décrivent un phénomène en question. D'abord, nous allons étudier le cas d'un couple de variables aléatoires réelles, autrement dit deux variables aléatoires réelles. Ensuite, on va génaliser pour le cas de n variables aléatoires.

Couple de variables aléatoires réelles

Variables discrètes

Chat ou chien ?

600 personnes d'âge différent ont été interrogées sur leur préférence des chats ou des chiens. Les résultats sont présentés dans le tableau ci-dessous :

<18 ans	80	125	
18-30 ans	110	90	
>30 ans	95	100	
			600

On peut vérifier que :

$$80 + 110 + 95 + 125 + 90 + 100 = 600$$

Qu'est-ce qu'on peut dire sur la probabilité de préférer les chats et avoir un certain âge ?

Il s'agit donc de deux variables aléatoires : l'âge de personne et la préférence des chats ou des chiens.

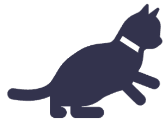
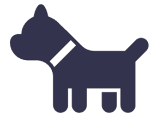
Tout d'abord, notons qu'on peut calculer la somme par colonnes et par lignes :

<18 ans	80	125	205	
18-30 ans	110	90	200	600
>30 ans	95	100	195	
	285	315	600	
	600			

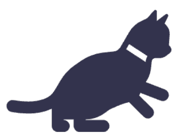
Ainsi, la somme de la 1ère colonne montre combien de personnes préfèrent les chats tout âge confondu (sans prendre en compte l'âge). La somme de la 2ème ligne montre combien de personnes sont dans la tranche d'âge 18-30 sans prendre en compte leur préférence pour les chats ou les chiens.

Notons que pour l'instant, il ne s'agit pas de probabilités.

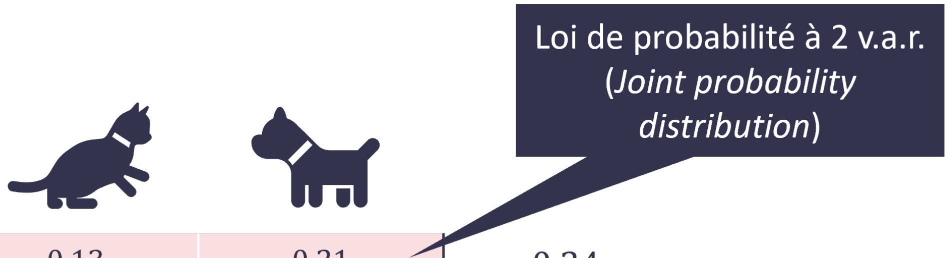
Divisons tout par le nombre total des participants, i.e. 600, afin d'avoir les valeurs qui ressemblent plus aux probabilités :

			
<18 ans	$\frac{80}{600} = 0.13$	$\frac{125}{600} = 0.21$	$\frac{205}{600} = 0.34$
18-30 ans	$\frac{110}{600} = 0.18$	$\frac{90}{600} = 0.15$	$\frac{200}{600} = 0.33$
>30 ans	$\frac{95}{600} = 0.16$	$\frac{100}{600} = 0.17$	$\frac{195}{600} = 0.33$
	$\frac{285}{600} = 0.47$	$\frac{315}{600} = 0.53$	$\frac{600}{600} = 1$

ou en ré-écrivant :

				
<18 ans	0.13	0.21	0.34	
18-30 ans	0.18	0.15	0.33	1
>30 ans	0.16	0.17	0.33	
	0.47	0.53	1	
	1			

Le centre de la table représente maintenant la fonction de masse du couple de variables :



<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

Notons que toutes les valeurs sont non-négatives et leur somme est 1.

D'une manière plus formelle :

Soit (X, Y) un couple de v.a.r. discrèt défini sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs finis ou dénombrables de $\mathbb{R}_{XY} \subset \mathbb{R}_X \times \mathbb{R}_Y = \{(x_i, y_j) | x_i \in \mathbb{R}_X, y_j \in \mathbb{R}_Y\}$.
La fonction de masse du couple de v.a.r. (X, Y) (en. *joint probability mass function*) est définie comme :

$$\mathbb{P}_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$$

où :

- $\sum_{(x_i, y_j) \in \mathbb{R}_{XY}} \mathbb{P}_{XY}(x_i, y_j) = 1$
- $\forall (x_i, y_j) \in \mathbb{R}_{XY} : \mathbb{P}_{XY}(x_i, y_j) \geq 0$

Remarque : $\mathbb{R}_{XY}$ est souvent défini comme $\mathbb{R}_{XY} = \mathbb{R}_X \times \mathbb{R}_Y$. Notons que dans ce cas-là, pour certaines paires (x_i, y_j), la probabilité $\mathbb{P}_{XY}(x_i, y_j)$ peut être égale à 0.
Remarque : lorsqu'il s'agit des variables aléatoires discrètes, souvent on va considérer $\mathbb{N}^2$ comme l'espace des valeurs.

Remarque : les notations suivantes peuvent aussi être utilisées : $\mathbb{P}(X = x \text{ et } Y = y)$ et $\mathbb{P}(X = x \cap Y = y)$ (l'intersection).

On appelle **fonction de répartition** (en. *joint cumulative distribution function*) du couple de v.a.r. (X, Y), l'application $F_{XY} : \mathbb{R}^2 \to \mathbb{R}$

$$F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \ \forall (x_i, y_j) \in \mathbb{R}_{XY}$$

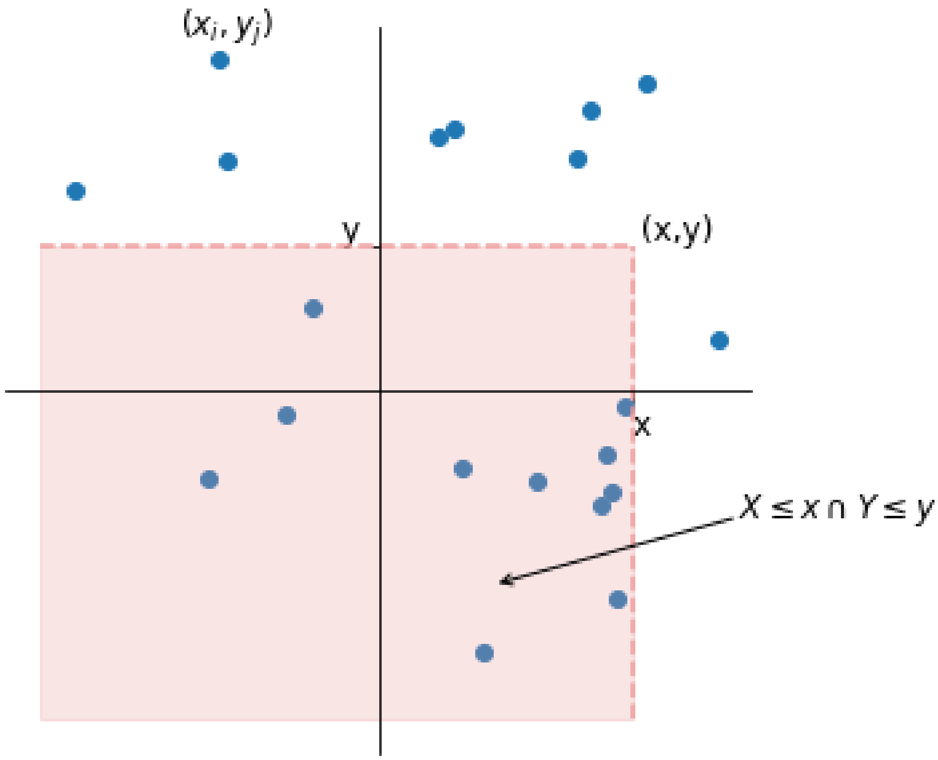
où $0 \leq F_{XY}(x, y) \leq 1$.
Autrement dit :

- $\lim_{x \to -\infty y \to -\infty} F_{XY}(x, y) = 0$
- $\lim_{x \to -\infty} F_{XY}(x, y) = \lim_{y \to -\infty} F_{XY}(x, y) = 0$
- $\lim_{x \to +\infty y \to +\infty} F_{XY}(x, y) = 1$

La fonction de répartition jointe a également des propriétés suivantes :

- $F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)$ (*fonction de répartition marginale*, en. *marginal CDF*)
- $F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y)$ (*fonction de répartition marginale*, en. *marginal CDF*)
- $\mathbb{P}(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$
- si X et Y sont indépendantes (voir plus bas), alors : $F_{XY}(x, y) = F_X(x) F_Y(y)$

De point de vue graphique, la fonction de répartition du couple de v.a.r. (X, Y) correspond à la probabilité que (X, Y) appartienne à une région bornée par x et y :



Notons que l'évènement $X = x$ peut être écrit comme $\{(x_i, y_j) : x_i = x, y_j \in \mathbb{R}_Y\}$. D'une manière similaire, l'évènement $Y = y$ peut être défini comme $\{(x_i, y_j) : x_i \in \mathbb{R}_X, y_j = y\}$. Alors :

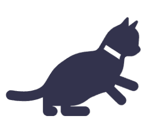
$$\mathbb{P}_{XY}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}((X = x) \cap (Y = y))$$

.

A présent, on peut aussi parler de la *loi marginale* qui doit son nom au fait qu'elle soit donnée sur *les marges*.

Pour chaque tranche d'âge, regardons quelle est la probabilité qu'une personne y appartienne sans tenir compte de ses préférences par rapport aux chats ou chiens :

Loi marginale (*marginal probability distribution*)

			
<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

Ainsi, la valeur 0.34 correspond à la probabilité qu'une personne soit de < 18 ans et peut être obtenue comme la somme des valeurs de la même ligne, i.e. $0.13 + 0.21 = 0.34$.

Notons que la somme des valeurs de toute cette colonne est égale à 1 :

$$0.34 + 0.33 + 0.33 = 1$$

Pareil, pour chaque préférence des chats ou des chiens, on peut calculer la probabilité qu'une personne l'ait sans tenir compte de son âge :

<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

Loi marginale (*marginal probability distribution*)

Vérifions que $0.47 + 0.53 = 1$ et notons que les éléments de cette ligne peuvent être obtenus comme la somme des éléments des colonnes correspondantes, e.g. : $0.13 + 0.18 + 0.16 = 0.47$.

Introduisons maintenant la définition formelle :

On appelle **la loi marginale** (en. *margin probability* ou *simple probability*) du couple (X, Y) de v.a.r. discrètes :

$$\forall x \in \mathbb{R}_X : \ \mathbb{P}_X(x) = \sum_{y_j \in \mathbb{R}_Y} \mathbb{P}_{XY}(x_i, y_j)$$

$$\forall y \in \mathbb{R}_Y : \ \mathbb{P}_Y(y) = \sum_{x_i \in \mathbb{R}_X} \mathbb{P}_{XY}(x_i, y_j)$$

Dans le R :

```r
# matrice de probabilite
p <- matrix(c(0.13, 0.21, 0.18, 0.15, 0.16, 0.17), nrow = 3, ncol = 2, byrow = TRUE); p
```

```
##      [,1] [,2]
## [1,] 0.13 0.21
## [2,] 0.18 0.15
## [3,] 0.16 0.17
```

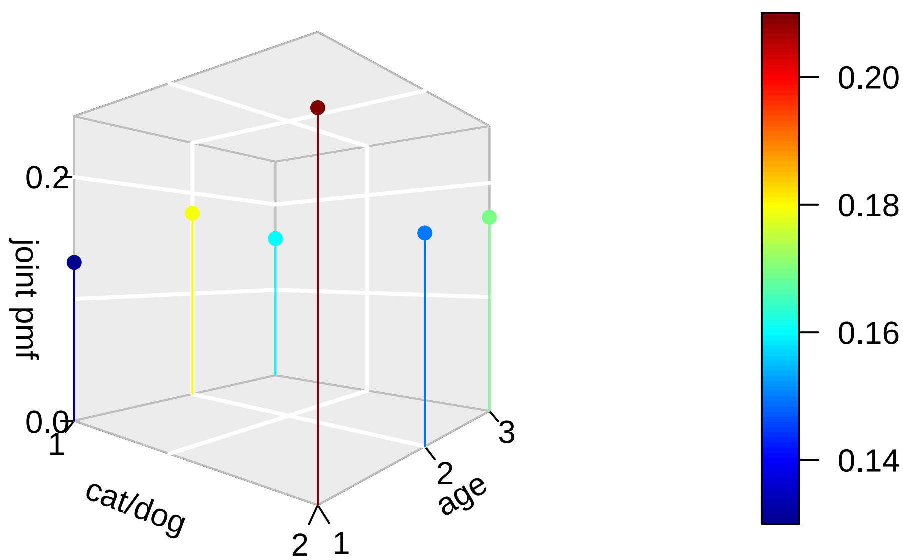
```r
# distribution marginale
# axe 1 (lignes)
marginTable(p, 1)
```

```
## [1] 0.34 0.33 0.33
```

```r
# axe 2 (colonnes)
marginTable(p, 2)
```

```
## [1] 0.47 0.53
```

```r
# plot
scatter3D(x=c(1, 2, 1, 2, 1, 2), y=c(1, 1, 2, 2, 3, 3),
          z=c(0.13, 0.21, 0.18, 0.15, 0.16, 0.17),
          phi = 0, bty = "g",  type = "h", zlim=c(0, 0.25),
          ticktype = "detailed", pch = 19, cex = 0.9, nticks = c(2,3),
          xlab = "cat/dog", ylab = "age", zlab = "joint pmf")
```



Il est aussi possible de définir *la fonction de répartition marginale* :

Soit (X, Y) un couple de v.a.r. de fonction de répartition $F_{XY}(x, y)$.
On appelle **fonctions de répartition marginales** (en. *marginal CDFs*) de X et Y les fonctions définis comme suit :

$$\forall x \in \mathbb{R}_X : F_X(x) = F_{XY}(x, \infty) = \lim_{y \to \infty} F_{XY}(x, y)$$

$$\forall y \in \mathbb{R}_Y : F_Y(y) = F_{XY}(\infty, y) = \lim_{x \to \infty} F_{XY}(x, y)$$

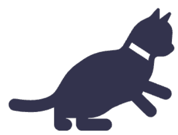
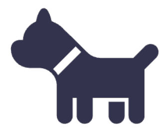
Soit (X, Y) un couple de v.a.r. Soit $x_1 \leq x_2$, $y_1 \leq y_2$, $x_1, x_2, y_1, y_2 \in \mathbb{R}$.
Alors:

$$\mathbb{P}(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$$

En se servant de ces données, il est possible de répondre aux nombreuses questions. Voici quelques exemples :

Quelle est la probabilité qu'une personne préfère des chiens ?

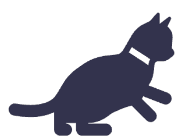
Dans ce cas-là, l'âge de la personne n'est pas pris en compte. On suffit donc de considérer la loi marginale :

			
<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

Ainsi, $\mathbb{P}(\heartsuit\text{chiens}) = 0.53$

Quelle est la probabilité qu'une personne préfère les chats ET soit de la tranche d'âge 18-30 ?

Dans ce cas-là, on considère une seule valeur qui se situe sur l'intersection de la colonne *Chats* et la ligne *18-30 ans* :

			
<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

Donc, $\mathbb{P}(\heartsuit\text{chats} \cap \text{18-30 ans}) = 0.18$

Maintenant, changeons un peu la question :

Quelle est la probabilité qu'une personne préfère les chats OU soit de la tranche d'âge 18-30 ?

Il s'agit maintenant de toutes les possibilités pour les chats et toutes les possibilités pour 18-30 ans :

<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

Donc, $\mathbb{P}(\heartsuit\text{chats} \cup \text{18-30 ans}) = 0.13 + 0.18 + 0.16 + 0.15 = 0.62$

Notons qu'on ne prend en compte la valeur 0.18 qui correspond à *préférer les chats* et *être de 18-30 ans* qu'une fois.

On peut également se souvenir de la formule suivante :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Ainsi, on peut prendre des valeurs de probabilités marginales et soustraire la valeur de l'intersection comme suit :

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)}$$

<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

$\mathbb{P}(\heartsuit\text{chats} \cup \text{18-30 ans}) = 0.47 + 0.33 - 0.18 = 0.8 - 0.18 = 0.62$

Ceci nous permet de parler de *la loi conditionnelle*.

Loi conditionnelle

Quelle est la probabilité qu'une personne préfère les chiens sachant qu'elle est de 18-30 ans ?

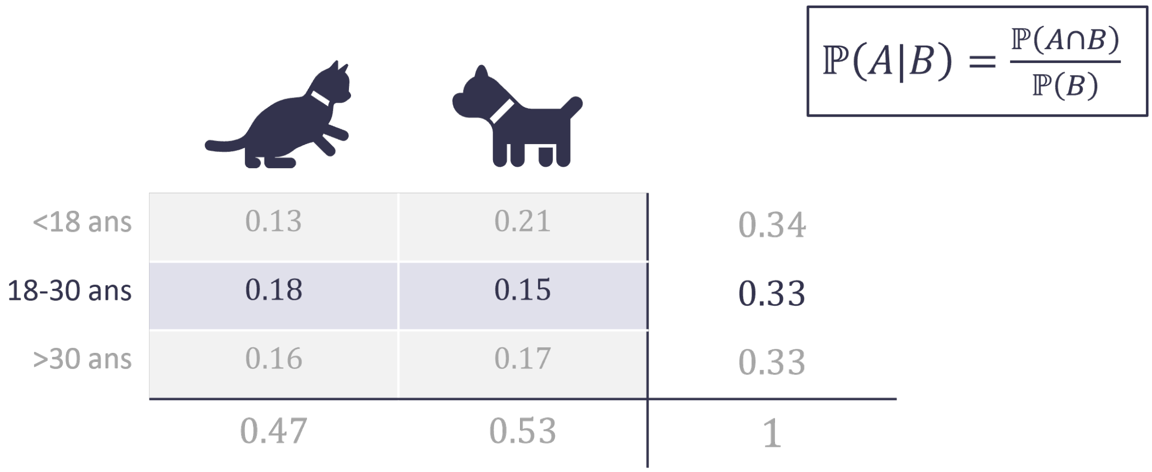
Afin de répondre à cette question, on va s'intéresser surtout à la ligne qui correspond à 18-30 ans.

Pour rappel la probabilité conditionnelle de l'évènement A sachant B ($\mathbb{P}(B) \neq 0$) est donnée par :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \text{ où } \mathbb{P}(B) \neq 0$$

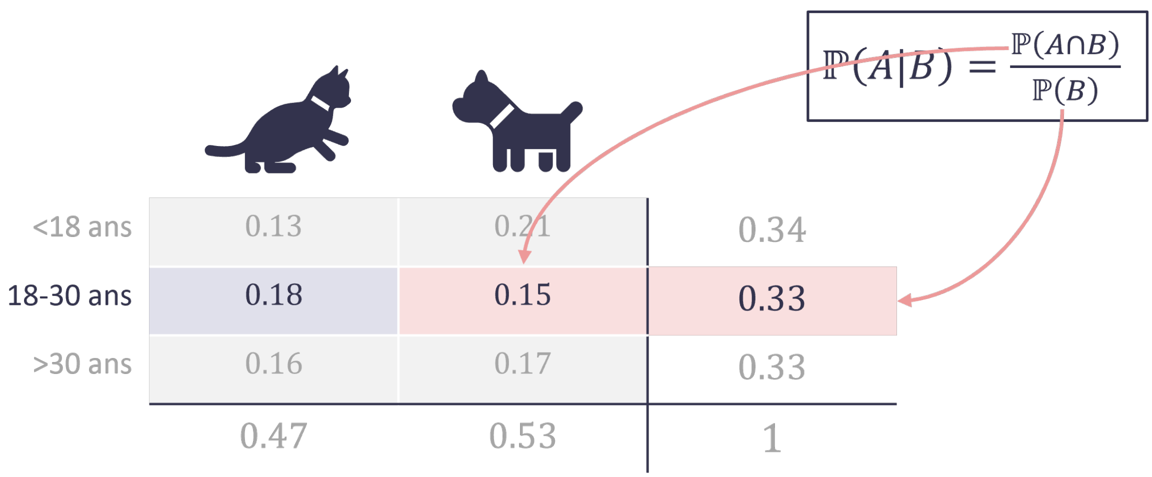
Dans notre cas :

- $A = \heartsuit chien$
- $B =$ 18-30 ans



L'intersection de A et B correspond à $A \cap B = \heartsuit chien$ ET $18\text{-}30\ ans$. Donc la probabilité $\mathbb{P}(A \cap B) = 0.15$.

La probabilité $\mathbb{P}(B)$ correspond à la probabilité d'avoir l'âge 18-30 ans sans tenir compte de préférences chats/chiens, donc de la probabilité marginale, c'est-à-dire $\mathbb{P}(B) = 0.33$.



Donc, $\mathbb{P}(\heartsuit chien | 18\text{-}30\ ans) = \frac{\mathbb{P}(\heartsuit chien \cap 18\text{-}30\ ans)}{\mathbb{P}(18\text{-}30\ ans)} = \frac{0.15}{0.33} \approx 0.45$

Il est alors possible de rajouter une ligne dans la table qui contient les probabilités conditionnelles de la préférence des chats / chiens sachant la tranche d'âge. Voici un example :

<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
$\mathbb{P}(animal \mid 18{-}30\ ans)$	0.55	0.45	1
TOTAL :	0.47	0.53	1

Loi conditionnelle (*conditional probability distribution*)

Notons que la somme des éléments de cette nouvelle ligne est égale à 1.

D’une manière similaire, il est possible de répondre aux questions comme :

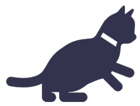
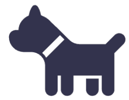
> Quelle est la probabilité qu’une personne ait un certain âge sachant qu’elle préfère les chats ?

Dans ce cas-là, on va s’intéresser à la colonne *chat* :

<18 ans	0.13	0.21	0.34
18-30 ans	0.18	0.15	0.33
>30 ans	0.16	0.17	0.33
	0.47	0.53	1

On obtient ainsi une nouvelle colonne de la probabilité conditionnelle :

Loi conditionnelle (*conditional probability distribution*)

			$\mathbb{P}(age \mid chat)$	
<18 ans	0.13	0.21	$\frac{0.13}{0.47} = 0.28$	0.34
18-30 ans	0.18	0.15	$\frac{0.18}{0.47} = 0.38$	0.33
>30 ans	0.16	0.17	$\frac{0.16}{0.47} = 0.34$	0.33
	0.47	0.53	1	1

Définissons maintenant *la fonction de masse conditionnelle* et *la fonction de répartition conditionnelle* d'une manière plus formelle :

Soit (X, Y) un couple de v.a.r. discrètes à valeurs dans $\mathbb{R}_{XY}$.
On appelle **fonction de masse conditionnelle** (en. *conditional PMF* of X given $Y = y_j$) de X sachant $Y = y_j$, l'application p telle que :

$$\forall x_i \in \mathbb{R}_X,\ p:\ x_i \rightarrow \mathbb{P}_{X|Y}(x_i, y_j) = \mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{\mathbb{P}_{XY}(x_i, y_j)}{\mathbb{P}_Y(y_j)}$$

La loi conditionnelle de X sachant $Y = A$ est donc la loi définie par cette fonction de masse.
On appelle **fonction de répartition conditionnelle** (en. *conditional CDF* of X given $Y = y_j$) de X sachant $Y = y_j$ l'application $F_X^{[Y=y_j]}$ de $\mathbb{R}$ dans $[0, 1]$ définie pour tout $x \in \mathbb{R}$ par :

$$F_X^{[Y=y_j]}(x) = F_{X|Y=y_j}(x) = \mathbb{P}(X \leq x | Y = y_j) = \frac{\mathbb{P}(X \leq x, Y = y_j)}{\mathbb{P}(Y = y_j)}$$

Remarque : il est possible de définir d'une manière plus générale la loi conditionnelle de X sachant n'importe quel évènement A :

$$\forall x_i \in \mathbb{R}_X,\ \mathbb{P}_{X|A}(x_i) = \mathbb{P}(X = x_i | A) = \frac{\mathbb{P}(X = x_i \text{ ET } A)}{\mathbb{P}(A)}$$

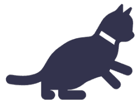
La fonction de répartition de X sachant A est donc donnée par :

$$F_{X|A}(x) = \mathbb{P}(X \leq x | A)$$

Indépendance

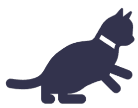
La question qui se pose maintenant est de dire si la préférence des chats ou des chiens et l'âge d'une personne sont indépendants. Si les variables X et Y sont indépendantes, alors savoir la valeur de Y n'apporte pas d'information sur X. De ce point de vue, on peut dire que si A et B sont indépendants, alors $\mathbb{P}(A|B) = \mathbb{P}(A)$. Ainsi la condition sur B n'a pas d'influence sur A.

$$\boxed{\mathbb{P}(A|B) = \mathbb{P}(A) \;\Leftarrow A \text{ et } B \textit{ indépendants}}$$

			$\mathbb{P}(age\|\,chat)$	
<18 ans	0.13	0.21	0.28	0.34
18-30 ans	0.18	0.15	0.38	0.33
>30 ans	0.16	0.17	0.34	0.33
	0.47	0.53	1	1

On peut donc comparer les valeurs de la colonne avec la loi conditionnelle et la colonne de la loi marginale :

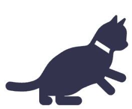
$$\boxed{\mathbb{P}(A|B) = \mathbb{P}(A) \;\Leftarrow A \text{ et } B \textit{ indépendants}}$$

			$\mathbb{P}(age\|\,chat)$	
<18 ans	0.13	0.21	0.28	0.34
18-30 ans	0.18	0.15	0.38	0.33
>30 ans	0.16	0.17	0.34	0.33
	0.47	0.53	1	1

Pour illustration, considérons la valeur de la probabilité d'être dans la tranche d'âge $< 18\ ans$ si vous préférez les chats $\mathbb{P}(< 18\ ans|\heartsuit chat) = 0.28$. Mais la probabilité de juste être dans cette tranche d'âge est $\mathbb{P}(< 18\ ans) = 0.34$. Si les variables étaient indépendantes, alors ces valeurs seraient égales, ce qui n'est pas le cas $0.28 \neq 0.34$. On peut faire donc une conclusion que les variables NE SONT PAS INDEPENDANTES.

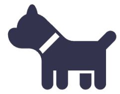
Pour la meilleure vision, rajoutons une colonne avec la loi conditionnelle de la tranche d'âge sachant que la personne préfère les chiens :

$$\boxed{\mathbb{P}(A|B) = \mathbb{P}(A) \;\Leftarrow A \text{ et } B \textit{ indépendants}}$$

			$\mathbb{P}(age\|\,chat)$	$\mathbb{P}(age\|\,chien)$	
<18 ans	0.13	0.21	0.28	0.4	0.34
18-30 ans	0.18	0.15	0.38	0.28	0.33
>30 ans	0.16	0.17	0.34	0.32	0.33
	0.47	0.53	1	1	1

On va donc comparer les valeurs :

$$\boxed{\mathbb{P}(A|B) = \mathbb{P}(A) \Leftarrow A \text{ et } B \textit{ indépendants}}$$

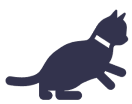
			$\mathbb{P}(age\vert\ chat)$	$\mathbb{P}(age\vert\ chien)$	
<18 ans	0.13	0.21	0.28	0.4	0.34
18-30 ans	0.18	0.15	0.38	0.28	0.33
>30 ans	0.16	0.17	0.34	0.32	0.33
	0.47	0.53	1	1	1

Une autre façon d'aborder le sujet d'indépendance est via l'intersection des évènements, i.e. dans le cas où les évènements A et B sont indépendants, on a :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

Dans notre cas :

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B) \Leftrightarrow A \text{ et } B \textit{ indépendants}}$$

			$\mathbb{P}(age\vert\ chat)$	$\mathbb{P}(age\vert\ chien)$	
<18 ans	0.13	0.21	0.28	0.4	0.34
18-30 ans	0.18	0.15	0.38	0.28	0.33
>30 ans	0.16	0.17	0.34	0.32	0.33
	0.47	0.53	1	1	1

D'un côté, nous avons :

$$\mathbb{P}(\text{18-30 ans } \cap \heartsuit chat) = 0.18$$

De l'autre côte :

- $\mathbb{P}(\text{18-30 ans}) = 0.33$
- $\mathbb{P}(\heartsuit chat) = 0.47$

Donc:

$$\mathbb{P}(\text{18-30 ans}) \times \mathbb{P}(\heartsuit chat) = 0.33 \times 0.47 = 0.16$$

On peut voir que :

$$0.18 \neq 0.16$$

donc les variables ne sont pas strictement indépendantes. Cependant, on peut noter que la différence dans ce cas-là n'est pas si grande.

D'une manière formelle :

Deux v.a.r. X et Y sont dites **indépendantes** (en. *independent*) si

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

Ce qui est équivalent à la condition suivante :

$$\forall x \in \mathbb{R}_X, \forall y \in \mathbb{R}_Y,\ F_{XY}(x, y) = F_X(x) \times F_Y(y)$$

Soit X et Y deux v.a.r. discrètes. X et Y sont indépendants si et seulement si :

$$\forall x \in \mathbb{R}_X, \forall y \in \mathbb{R}_Y,\ \mathbb{P}_{XY}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}_X(x) \times \mathbb{P}_Y(y)$$

Soit X et Y deux v.a.r. discrètes indépendantes. Alors :

$$\mathbb{P}_{X|Y}(x_i, y_j) = \mathbb{P}(X = x_i|Y = y_j) = \frac{\mathbb{P}_{XY}(x_i, y_j)}{\mathbb{P}_Y(y_j)} = \frac{\mathbb{P}_X(x_i) \times \mathbb{P}_Y(y_j)}{\mathbb{P}_Y(y_j)} = \mathbb{P}_X(x_i)$$

Variables continues

Dans le cas univarié, nous savons que les différences au niveau de formules entre les v.a.r. discrètes et continues se manifestent principalement d'une manière suivante [3], [2] :

- les sommes sont remplacées par des intégrals
- au lieu de la fonction de masse (cas discrèt) on utilise la fonction de densité (cas continu)

Soit X et Y deux v.a.r. continues. Les variables aléatoires X et Y sont appelées **absolument continues** (en. *jointly continuous random variables*) s'il existe une fonction $f_{XY}:\ \mathbb{R}^2 \to \mathbb{R}$ non-négative telle que pour tout ensemble $A \in \mathbb{R}^2$ on a :

$$\forall A \in \mathbb{R}^2,\ \mathbb{P}\left((X, Y) \in A\right) = \iint\limits_{(x,y)\in A} f_{XY}(x, y)dxdy$$

La fonction $f_{XY}(x, y)$ est appelée **la fonction de densité de probabilité jointe** ou **loi jointe** (en. *joint probability density function* ou *joint PDF*) de v.a.r. X et Y.
La fonction de densité de probabilité jointe a ainsi les propriétés suivantes :

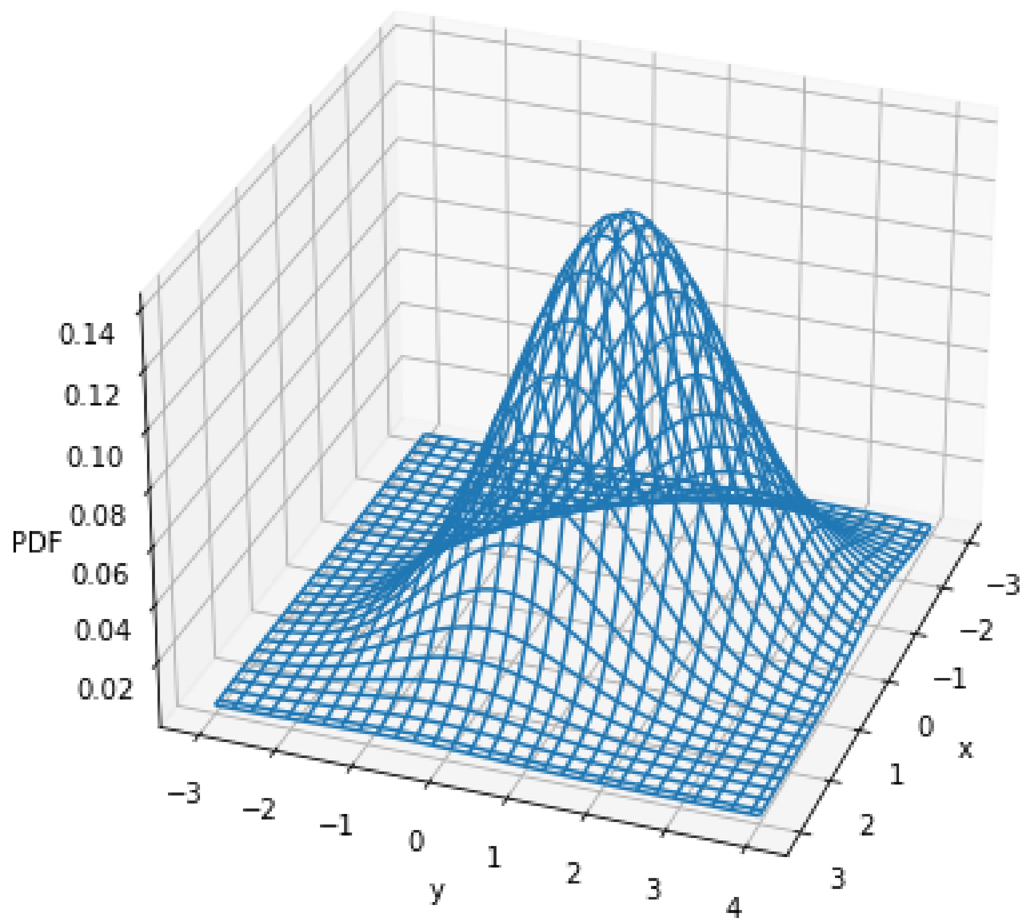
1. $f_{XY}(x, y) \geq 0$ (*non-négativité*)
2. $\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f_{XY}(x, y)dxdy = 1$

La densité de probabilité jointe $f_{XY}(x, y)$ de v.a.r. X et Y de la fonction de répartition jointe $F_{XY}(x, y)$ peut être obtenu par une double dérivation de $F_{XY}(x, y)$:

$$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$$

Remarque : comme dans le cas univarié, la fonction de densité jointe $f_{XY}(x, y)$ peut avoir des valeurs supérieures à 1, car il s'agit de la densité et **pas** de la probabilité.

Voici un exemple du graphique de la distribution normale bivariée :



Pour rappel, dans le cas univarié, nous disions que la probabilité était reflétée par l'aire sous la courbe de la fonction de densité $f_X(x)$. D'où, la probabilité dans un point concret est égal à 0.

Nous allons développer cette idée pour le cas d'un couple de v.a.r. continues.

Dans le cas de deux v.a.r. continues X et Y de la fonction de densité de la probabilité jointe $f_{XY}(x, y)$, la probabilité est donnée par le volume sous la courbe, car selon la définition $\forall A \in \mathbb{R}^2,\ \mathbb{P}\left((X, Y) \in A\right) = \iint\limits_{(x,y)\in A} f_{XY}(x, y)dxdy$.

La fonction de densité de la probabilité jointe peut ainsi être vue comme une mesure de la probabilité par unité de surface [3].

Ainsi, dans le cas univarié, pour un petit $\Delta > 0$, on peut présenter la fonction de densité comme suit :

$$f_X(x) = \lim_{\Delta\to 0} \frac{\mathbb{P}(x < X \leq x + \Delta)}{\Delta}$$

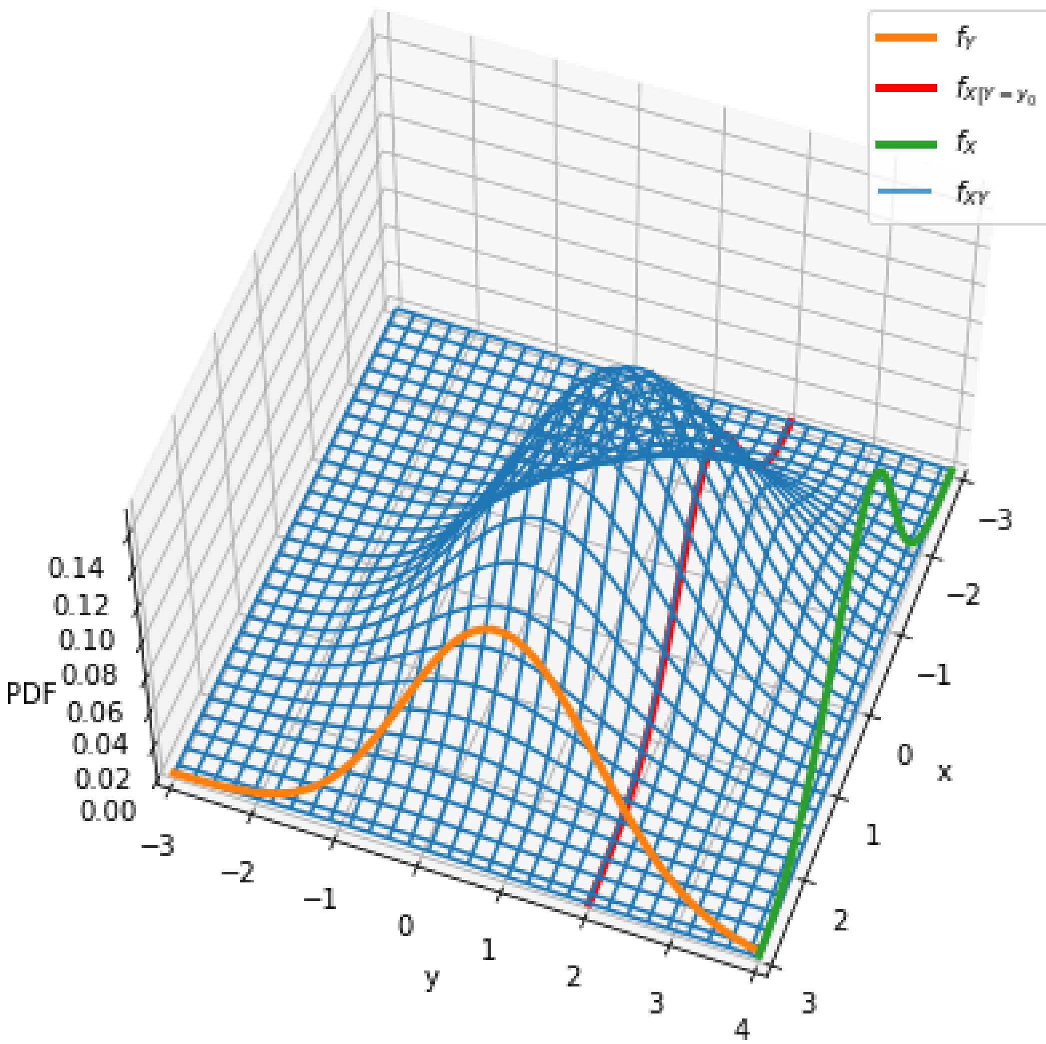
Dans le cas du couple (X, Y), la fonction de densité $f_{XY}(x, y)$ peut être interprétée comme la probabilité que le couple (X, Y) a ses valeurs dans un petit rectangle de largeur δ_x et hauteur δ_y autour du point (x, y), c.à.d. :

$$\mathbb{P}(x < X \leq x + \delta_x,\ y < Y \leq y + \delta_y) \approx f_{XY}(x, y)\delta_x\delta_y$$

Soit (X, Y) un couple de v.a.r. absolument continue de densité $f_{XY}(x, y)$. Les **lois marginales** de v.a.r. X et Y sont données par les **densités marginales** (en. *marginal density*) suivantes :

$$\forall x \in \mathbb{R}_X, \quad f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$$

$$\forall y \in \mathbb{R}_Y, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$



La **fonction de répartition jointe** (en. *joint cumulative distribution function* ou *joint CDF*) du couple des v.a.r. absolument continues a la même définition générale que dans le cas discret :

$$F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

Cepandant, il est possible de définir cette fonction de répartition via la représentation intégrale suivante :

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(t_1, t_2) dt_2 dt_1$$

On peut en définir les **fonctions de répartition marginales** de X et Y comme suit :

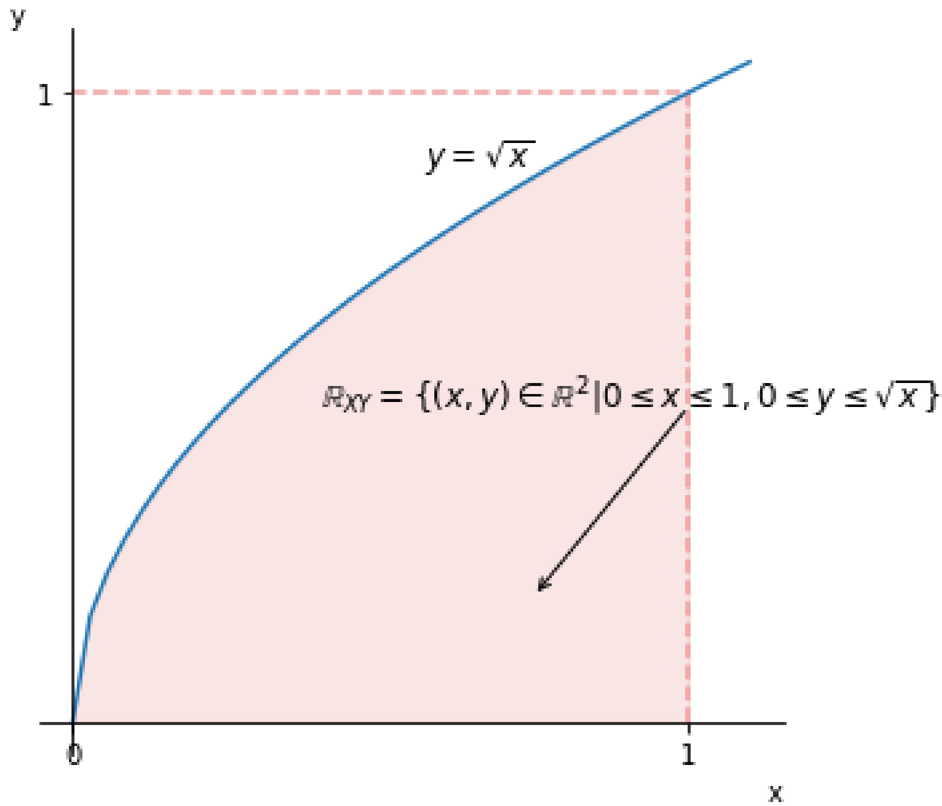
$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{XY}(t_1, t_2) dt_2 dt_1, \forall x \in \mathbb{R}_X$$

$$F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{+\infty} f_{XY}(t_1, t_2) dt_1 dt_2, \forall y \in \mathbb{R}_Y$$

Exemple Soit (X, Y) un couple de v.a.r. absolument continues à la fonction de densité jointe $f_{XY}(x, y)$:

$$f_{XY}(x, y) = \begin{cases} 10xy & \text{si } 0 \leq x \leq 1, 0 \leq y \leq \sqrt{x} \\ 0 & \text{sinon} \end{cases}$$

Tout d'abord illustrons sur le graphique la région correspondant à la condiiton que la fonction de densité jointe est non nulle si $0 \leq x \leq 1, 0 \leq y \leq \sqrt{x}$:



En dehors de cette région $\mathbb{R}_{XY}$ la fonction de densité jointe est 0.

Quelles sont les fonctions de densité marginales de X et Y ?

Notons qu'à partir de $0 \le x \le 1$, $0 \le y \le \sqrt{x}$, on peut définir que $\mathbb{R}_X = [0, 1]$, $\mathbb{R}_Y = [0, \sqrt{x}]$. Lorsque x est limité par 1, on peut considérer que y est limité par $\sqrt{1} = 1$ également.

Afin de trouver la fonction de densité marginale de X, notée $f_X(x)$, il suffit de se concentrer sur l'intervalle $[0, \sqrt{x}]$ par raport à y car en dehors de cet intervalle la fonction de densité jointe est égale à 0, c.à.d. :

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy = \int_{0}^{\sqrt{x}} 10xydy = 10x\frac{y^2}{2}\bigg|_0^{\sqrt{x}} = 5x\left((\sqrt{x})^2 - 0\right) = 5x^2$$

Donc :

$$f_X(x) = \begin{cases} 5x^2, \text{ si } 0 \le x \le 1 \\ 0, \text{ sinon} \end{cases}$$

D'une manière similaire, trouvons $f_Y(y)$, la fonction de densité marginale de Y. Notons qu'à partir de $y = \sqrt{x}$, nous avons $x = y^2$. Alors pour $0 \le y \le \sqrt{x}$:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy = \int_{y^2}^{1} 10xydx = 10y\frac{x^2}{2}\bigg|_{y^2}^{1} = 5y\left(1 - (y^2)^2\right) = 5y(1 - y^4)$$

Donc :

$$f_Y(y) = \begin{cases} 5y(1 - y^4), \text{ si } 0 \le y \le \sqrt{x} \\ 0, \text{ sinon} \end{cases}$$

Quelle est la fonction de répartition jointe ?

Notons que :

- si $x < 0$ ou $y < 0$, alors $F_{XY}(x, y) = 0$
- si $x \geq 1$ et $y \geq \sqrt{x}$, alors $F_{XY}(x, y) = 1$

Afin de trouver la fonction de répartition jointe pour $x > 0, y > 0$, il suffit d'intégrer la fonction de densité jointe :

$$F_{XY}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u, v) du dv = \int_{0}^{y} \int_{0}^{x} f_{XY}(u, v) du dv = \int_{0}^{\min(y, \sqrt{x})} \int_{0}^{\min(x, 1)} 10uv du dv$$

Pour $0 \leq x \leq 1, 0 \leq y \leq \sqrt{x}$, on obtient :

$$F_{XY}(x, y) = \int_{0}^{y} \int_{0}^{x} 10uv du dv = \int_{0}^{y} 10v \frac{u^2}{2}\bigg|_{0}^{x} dv = \int_{0}^{y} 5vx^2 dv = 5x^2 \frac{v^2}{2}\bigg|_{0}^{y} = \frac{5}{2}(xy)^2$$

Pour $0 \leq x \leq 1, y \geq \sqrt{x}$, on obtient :

$$F_{XY}(x, y) = F_{XY}(x, \sqrt{x}) = \frac{5}{2}(x\sqrt{x})^2 = \frac{5}{2}x^3$$

Pour $0 \leq y \leq \sqrt{x}, x \geq 1$, on obtient :

$$F_{XY}(x, y) = F_{XY}(1, y) = \frac{5}{2}(1y)^2 = \frac{5}{2}y^2$$

Ainsi :

$$F_{XY}(x, y) = \begin{cases} 0 & x < 0 \text{ ou } y < 0 \\ \frac{5}{2}(xy)^2 & 0 \leq x \leq 1, 0 \leq y \leq \sqrt{x} \\ \frac{5}{2}x^3 & 0 \leq x \leq 1, y \geq \sqrt{x} \\ \frac{5}{2}y^2 & 0 \leq y \leq \sqrt{x}, x \geq 1 \\ 1 & x \geq 1 \text{ et } y \geq \sqrt{x} \end{cases}$$

Loi conditionnelle

Comme dans le cas d'un couple de v.a.r. discrètes, la formule de base de la probabilité conditionnelle est la suivante :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \text{ où } \mathbb{P}(B) > 0$$

Nous pouvons donc définir la fonction de densité conditionne et la fonction de répartition conditionnelle comme suit :

Soit (X, Y) un couple de v.a.r. de densité $f_{XY}(x, y)$ et $f_Y(y)$ la densité de Y.
La **fonction de densité conditionnelle** (en. *conditional PDF*) de X sachant $Y = y$ où $f_Y(Y = y) \neq 0$:

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

La probabilité conditionnelle de $X \in A$ sachant $Y = y$:

$$\mathbb{P}(X \in A|Y = y) = \int_A f_{X|Y}(x, y)dx$$

La **fonction de répartition conditionnelle** (en. *conditional CDF*) de X sachant $Y = y$:

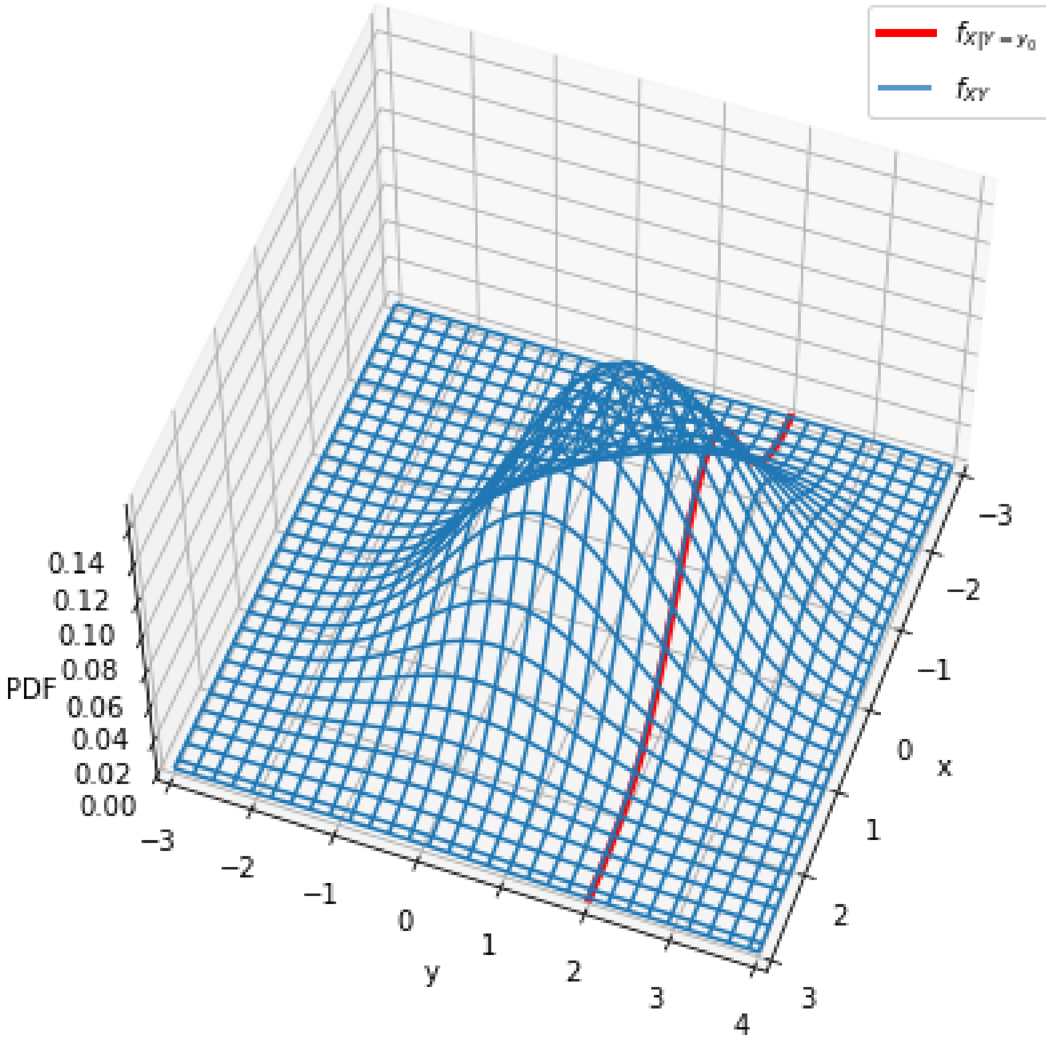
$$F_{X|Y}(x|y) = \mathbb{P}(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(x|y)dx$$

Soit A un évènement défini comme $a < X < b$ (où il est possible que $a = -\infty$ ou $b = +\infty$), alors :

$$F_{X|A}(x) = \begin{cases} 1 & x > b \\ \frac{F_X(x)-F_X(a)}{F_X(b)-F_X(a)} & a \leq x < b \\ 0 & x < a \end{cases}$$

et

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(A)} & a \leq x < b \\ 0 & \text{sinon} \end{cases}$$



Reprenons l'exemple précédent où le couple de v.a.r. absolument continues est caractérisé par la fonction de densité jointe suivante :

$$f_{XY}(x,y) = \begin{cases} 10xy, \text{ si } 0 \leq x \leq 1, 0 \leq y \leq \sqrt{x} \\ 0, \text{ sinon} \end{cases}$$

Trouvez la densité conditionnelle $f_{X|Y}(x|y)$

Selon la définition :

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

Précédement, nous avons trouvé la desité marginale de Y :

$$f_Y(y) = \begin{cases} 5y(1-y^4), \text{ si } 0 \leq y \leq \sqrt{x} \\ 0, \text{ sinon} \end{cases}$$

Notons que $y \leq \sqrt{x}$ peut être ré-écrit comme $y^2 \leq x$. De l'autre côté on sait que la densité jointe est non nulle si $0 \leq x \leq 1, 0 \leq y \leq \sqrt{x}$. Donc x est limité d'un côté par $y^2 \leq x$ et de l'autre côté par $x \leq 1$.

On obtient donc :

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \begin{cases} \frac{10xy}{5y(1-y^4)} y^2 \leq x \leq 1 \\ 0 \qquad \text{sinon} \end{cases} = \begin{cases} \frac{2x}{1-y^4} y^2 \leq x \leq 1 \\ 0 \quad \text{sinon} \end{cases}$$

Indépendance

Comme donné dans le cas discrèt, deux v.a.r. X et Y sont dites indépendantes si

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

De même façon cette définition peut être ré-écrit comme suit pour $\forall x \in \mathbb{R}, \forall y \in \mathbb{R}$:

$$F_{XY}(x,y) = F_X(x) \times F_Y(y)$$

Regardons maintenant ce que cette définition implique dans le cas continue.

Soit (X, Y) un couple de v.a.r. absolument continu à la densité $f_{XY}(x,y)$. Soit $f_X(x)$ et $f_Y(y)$ les fonctions de denstié marginale de X et Y respectivement. X et Y sont indépendantes si et seulement si :

$$\forall s \in \mathbb{R}_X, \forall t \in \mathbb{R}_Y,\ f_{XY}(s,t) = f_X(s) \times f_Y(t)$$

Reprenons notre exemple :

Est-ce que les v.a.r. X et Y sont indépendantes ?

Vérifions la condition $f_{XY}(s,t) = f_X(x) \times f_Y(y)$.

D'un côté, nous avons :

$$f_{XY}(x,y) = \begin{cases} 10xy, \text{ si } 0 \leq x \leq 1, 0 \leq y \leq \sqrt{x} \\ 0, \text{ sinon} \end{cases}$$

De l'autre côté :

$$f_X(x) = \begin{cases} 5x^2, \text{ si } 0 \leq x \leq 1 \\ 0, \text{ sinon} \end{cases}$$

$$f_Y(y) = \begin{cases} 5y(1-y^4), \text{ si } 0 \leq y \leq \sqrt{x} \\ 0, \text{ sinon} \end{cases}$$

Considérons le cas $0 \leq x \leq 1$ et $0 \leq y \leq \sqrt{x}$:

$$f_X(x) \times f_Y(y) = 5x^2 \times 5y(1-y^4) = 25x^2 y(1-y^4) \neq f_{XY}(xy)$$

Donc, X et Y **ne sont pas** indépendantes.

Moments d'un couple de v.a.r.

Espérance

On appelle **espérance** (en. *expectation*) du couple de v.a.r. (X, Y), notée $\mathbb{E}(X, Y)$, l'élément de $\mathbb{R}^2$ défini comme suit :

$$\mathbb{E}(X, Y) = (\mathbb{E}(X), \mathbb{E}(Y))$$

L'espérance peut être calculé de la façon suivante.
(cas discret)
Soit $h : \mathbb{R}^2 \to \mathbb{R}$ une fonction bornée et continu par morceaux. L'espérance de la v.a.r. $Z = h(X, Y)$ est donnée par :

$$\mathbb{E}[Z] = \sum_{(i,j)\in\mathbb{R}_{XY}} h(i, j) \times \mathbb{P}(X = i, Y = j)$$

Notons que $\mathbb{R}_{XY}$ est un sous-ensemble fini ou dénombrable de $\mathbb{N}^2$ dans lequel se trouvent les valeurs de (X, Y).
(cas continu)
Soit (X, Y) a pour densité la fonction $f_{XY}(x, y)$. Soit $h : \mathbb{R}^2 \to \mathbb{R}$ une fonction bornée et continu par morceaux. L'espérance de la v.a.r. $Z = h(X, Y)$ est donnée par :

$$\mathbb{E}[Z] = \int_{\mathbb{R}^2} h(u, v) \times f_{XY}(u, v) du dv$$

lorsque cette intégrale existe.
L'espérance de la fonction de deux variable a comme propriétés :

- *linéarité* (en. *linearity*) : $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

Espérance conditionnelle

Il est tout à fait possible de calculer *l'espérance conditionnelle* de X sachant que l'évènement A s'est produit ou sachant la valeur de la variable aléatoire $Y = y_j$. L'espérance conditionnelle ressemble à l'espérance simple sauf que la fonction de masse / la densité sont remplacés par leur analogues conditionnelles [2].

Soit (X, Y) un couple de v.a.r. discrètes. Soit A un évènement.
L'**espérance conditionnelle** (en. *conditional expectation*) de X :

1. sachant A est définie par :
$$\mathbb{E}[X|A] = \sum_{x_i\in\mathbb{R}_X} x_i \mathbb{P}_{X|A}(x_i)$$
2. sachant $Y = y_j$ est définie par :
$$\mathbb{E}[X|Y = y_j] = \sum_{x_i\in\mathbb{R}_X} x_i \mathbb{P}_{X|Y}(x_i|y_j)$$

Dans le cas continu, l'*espérance conditionnelle* de X sachant $Y = y$ est donnée par :

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

Notons que la valeur de l'espérance conditionnelle change en fonction de la valeur $Y = y$.

La loi de l'espérance totale (en. *Law of Total Expectation* ou *Law of Iterated Expectation*) :
(cas discret)
Soit (X, Y) un couple de v.a.r. discrètes. L'espérance totale de X peut être calculée comme suit :

$$\mathbb{E}X = \sum_{y_j \in \mathbb{R}_Y} \mathbb{E}[X|Y = y_j]\mathbb{P}_Y(y_j)$$

(cas continu)
Soit (X, Y) un couple de v.a.r. absolument continu. L'espérance totale de X peut être calculée comme suit :

$$\mathbb{E}X = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y]f_Y(y)dy = \mathbb{E}\left[\mathbb{E}[X|Y]\right]$$

Quelle est l'espérance $\mathbb{E}[X|Y = y]$ pour $0 \leq y \leq \sqrt{x}$?

Pour rappel, la fonction de densité jointe est donnée par :

$$f_{XY}(x, y) = \begin{cases} 10xy, \text{ si } 0 \leq x \leq 1, 0 \leq y \leq \sqrt{x} \\ 0, \text{ sinon} \end{cases}$$

La densité conditionnelle est :

$$f_{X|Y}(x|y) = \begin{cases} \frac{2x}{1-y^4} y^2 \leq x \leq 1 \\ 0 \quad \text{sinon} \end{cases}$$

Notons que $y \leq \sqrt{x}$ est équivalent à $y^2 \leq x$ pour $x \geq 0$. C'est-à-dire que dans notre cas on peut considérer $y^2 \leq x \leq 1$.

Selon la formule, l'espérance conditionnelle $\mathbb{E}[X|Y = y]$ pour $0 \leq y \leq 1$ peut être calculée comme suit :

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx = \int_{y^2}^{1} x\frac{2x}{1-y^4}dx = \int_{y^2}^{1} \frac{2x^2}{1-y^4}dx =$$

$$= \frac{2}{1-y^4}\frac{x^3}{3}\bigg|_{y^2}^{1} = \frac{2}{3(1-y^4)}(1-y^6) = \frac{2(1-y^6)}{3(1-y^4)}$$

Variance conditionnelle

(cas discret)
Soit (X, Y) un couple de v.a.r. discrètes. Notons par $\mu_{X|Y}(y) = \mathbb{E}[X|Y = y]$. **La variance conditionnelle** (en. *conditional variance*) de X sachant $Y = y$, notée $Var(X|Y = y)$, est définie par :

$$Var(X|Y = y) = \mathbb{E}\left[(X - \mu_{X|Y}(y))^2 | Y = y\right] = \sum_{x_i \in \mathbb{R}_X} (x_i - \mu_{X|Y}(y))^2 \, \mathbb{P}_{X|Y}(x_i) = \mathbb{E}[X^2|Y = y] - \mu_{X|Y}(y)^2$$

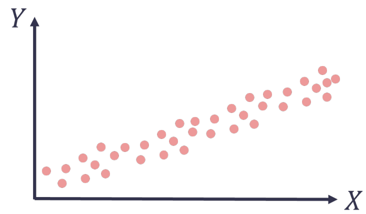
Notons que comme dans le cas de l'espérance conditionnelle, la variance conditionnelle est une fonction de la v.a.r. Y car sa valeur dépend de la valeur de $Y = y$.

La loi de la variance totale (en. *Law of Total Variance*) : Soit (X, Y) un couple de v.a.r. discrètes. La variance totale de X peut être calculée comme suit :

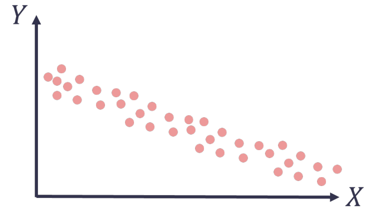
$$Var(X) = \mathbb{E}[Var(X|Y)] + Var\left(\mathbb{E}[X|Y]\right)$$

Covariance

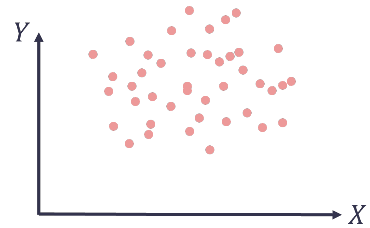
On peut s'intérroger sur la relation des variations des valeurs de deux variables aléatoires X et Y.



$Cov(X, Y) > 0$
Les plus grandes valeurs de Y correspondent aux plus grandes valeurs de X



$Cov(X, Y) < 0$
Les plus grandes valeurs de Y correspondent aux plus petites valeurs de X



$Cov(X, Y) \approx 0$
Pas de relation évidente

Ainsi on introduit la notion de la *covariance* de deux v.a.r.

Soit X et Y deux v.a.r. Si $\mathbb{E}X$ et $\mathbb{E}Y$ existent, la **covariance** (en. *covariance*) entre X et Y, notée $Cov(X, Y)$ ou σ_{XY}, est définie par :

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

Propriétés :

- $Cov(X, X) = \mathbb{E}[(X - \mathbb{E}X)(X - \mathbb{E}X)] = \mathbb{E}[XX] - \mathbb{E}X \times \mathbb{E}X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = Var(X)$

Notons que $Cov(X, X) \geq 0$. Si $Cov(X, X) = 0$, alors $X = const$ presque sûrement.

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(aX_1 + bY_1, X_2) = a\ Cov(X_1, X_2) + b\ Cov(Y_1, X_2),\ \forall a, b \in \mathbb{R}$
- $Cov(X_1, aX_2 + bY_2) = a\ Cov(X_1, X_2) + b\ Cov(X_1, Y_2),\ \forall a, b \in \mathbb{R}$
- $Cov(X, a) = 0$
- $Cov(aX, bY) = ab\ Cov(X, Y)$
- $Cov(X + c, Y) = Cov(X, Y)$

La covariance entre X et Y reflète le comportement d'une v.a.r. par rapport à l'autre. Elle traduit leurs variations conjointes (les écarts conjoints par rapport à leurs espérances respectives) mais elle reste assez difficile à interpréter.

Cependant, on considère que :

- si $Cov(X, Y) > 0$, alors X et Y varient plutôt dans le même sens
- si $Cov(X, Y) < 0$, alors X et Y varient plutôt dans le sens contraire

Exemple :

Prenons un exemple proposé par [2].

Soit X une v.a.r. continue de la loi uniforme sur $[1, 2]$, c.à.d. $X \sim \mathcal{U}([1, 2])$. Soit Y une v.a.r. qui sous condition $X = x$ suit la loi exponentielle avec le paramètre $\lambda = x$, c.à.d. $\mathcal{E}(\lambda = x)$. Trouvez la covariance $Cov(X, Y)$.

Solution :

Pour rappel :

	$f_X(x) \neq 0$	EX	$Var(X)$
loi uniforme, $\mathcal{U}([a, b]),\ a < b$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
loi exponentielle, $\mathcal{E}(\lambda),\ \lambda > 0$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Afin de trouver la covariance, on peut se servir de la formule :

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

Lorsque $X \sim \mathcal{U}([1, 2])$, son espérance est donnée par :

$$\mathbb{E}X = \frac{a+b}{2} = \frac{1+2}{2} = \frac{3}{2}$$

Concernant la v.a.r. Y, nous savons sa distribution conditionnelle. Donc, pour trouver l'espérance de Y, on peut utiliser la loi de l'espérance totale, selon laquelle :

$$\mathbb{E}Y = \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_X(x) dx = \mathbb{E}[\mathbb{E}[Y|X]]$$

Lorsque Y sachant $X = x$ suit la loi exponentielle avec le paramètre $\lambda = x$, autrement dit $Y|X \sim \mathcal{E}(X)$, alors :

$$\mathbb{E}[Y|X] = \frac{1}{\lambda} = \frac{1}{X}$$

Maintenant, il suffit de trouver :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}\left[\frac{1}{X}\right]$$

Il s'agit donc de l'espérance d'une fontion de X. Pour rappel :

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx$$

Lorsque $X \sim \mathcal{U}([1, 2])$, on va s'intéresser qu'à l'intervalle $[1, 2]$. Ainsi :

$$\mathbb{E}\left[\frac{1}{X}\right] = \int_1^2 \frac{1}{x} \times \frac{1}{2-1} dx = \int_1^2 \frac{1}{x} dx = \ln x \Big|_1^2 = \ln 2 - \ln 1 = \ln 2$$

A présent, trouvons $\mathbb{E}[XY]$.

En s'appuyant sur la loi de l'espérance totale, on peut ré-écrire l'expression pour $\mathbb{E}[XY]$ comme suit :

$$\mathbb{E}[XY] = \mathbb{E}\left[\mathbb{E}[XY|X]\right]$$

Notons que $\mathbb{E}[X|X = x] = x$. Alors :

$$\mathbb{E}[XY] = \mathbb{E}\left[\mathbb{E}[XY|X]\right] = \mathbb{E}\left[X\mathbb{E}[Y|X]\right]$$

Nous savons que $\mathbb{E}[Y|X] = \frac{1}{X}$. Donc :

$$\mathbb{E}[XY] = \mathbb{E}\left[\mathbb{E}[XY|X]\right] = \mathbb{E}\left[X\mathbb{E}[Y|X]\right] = \mathbb{E}\left[X\frac{1}{X}\right] = \mathbb{E}[1] = 1$$

Maintenant rassemblons tout :

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y = 1 - \frac{3}{2}\ln 2$$

Lien avec l'indépendance

Soit X et Y deux v.a.r. indépendantes. Alors, on a :

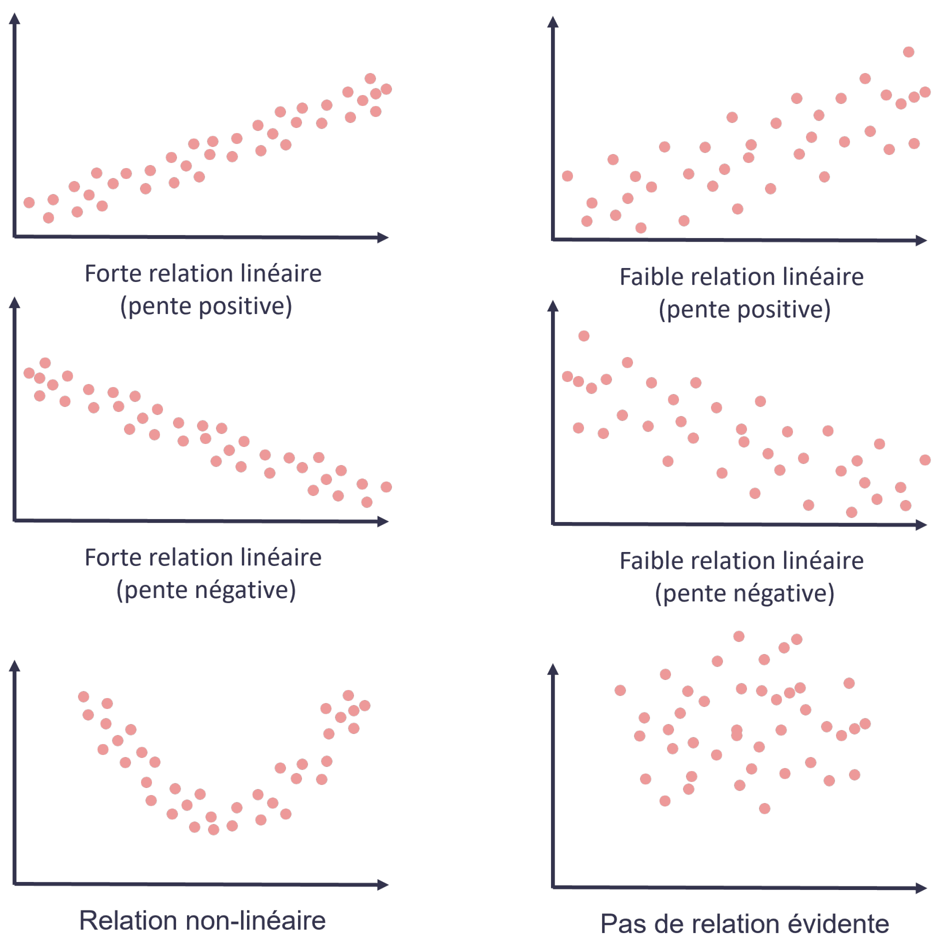
1. $\mathbb{E}[XY] = \mathbb{E}X \times \mathbb{E}Y$ et $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \times \mathbb{E}[h(Y)]$
2. $\mathbb{E}[X|Y] = \mathbb{E}[X]$ et $\mathbb{E}[g(X)|Y] = \mathbb{E}[g(X)]$
3. $Var(X+Y) = Var(X) + Var(Y)$
4. $Cov(X,Y) = 0$ car $\mathbb{E}[XY] = \mathbb{E}X \times \mathbb{E}Y$

Soit X et Y deux v.a.r., et h et g deux fonctions de $\mathbb{R}$ dans $\mathbb{R}$.
Si X et Y sont indépendantes, alors les deux v.a.r. $g(X)$ et $h(Y)$ sont indépendantes et :

$$\mathbb{E}\left(g(X)h(Y)\right) = \mathbb{E}(g(X)) \times \mathbb{E}(h(Y))$$

Coefficient de corrélation

Il existe une forme normalisée de la covariance, appelée *coefficient de correlation*, qui sert de l'indicateur de la relation linéaire entre deux variables.



Calculons la covariance de version standartisée des v.a.r. X et Y, c.à.d. :

Soit X et Y deux v.a.r. à l'espérance $\mathbb{E}X$ et $\mathbb{E}Y$, respectivement, et à la variance σ_X^2 et σ_Y^2, respectivement.

Soit U et V les versions standartisées de X et Y :

$$U = \frac{X - \mathbb{E}X}{\sigma_X}$$

$$V = \frac{Y - \mathbb{E}Y}{\sigma_Y}$$

Trouvons maintenant la covariance de U et V :

$$Cov(U, V) = Cov\left(\frac{X - \mathbb{E}X}{\sigma_X}, \frac{Y - \mathbb{E}Y}{\sigma_Y}\right)$$

Notons que selon les propriétés de la covariance : $Cov(X + c, Y) = Cov(X, Y)$. Les valeurs $\mathbb{E}X$ et $\mathbb{E}Y$ sont des constantes, donc :

$$Cov(U, V) = Cov\left(\frac{X - \mathbb{E}X}{\sigma_X}, \frac{Y - \mathbb{E}Y}{\sigma_Y}\right) = Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

Selon les propriétés de la covariance : $Cov(aX, bY) = ab\ Cov(X, Y)$. Alors :

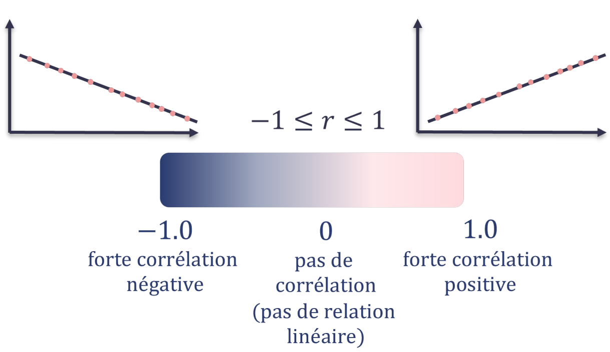
$$Cov(U, V) = Cov\left(\frac{X - \mathbb{E}X}{\sigma_X}, \frac{Y - \mathbb{E}Y}{\sigma_Y}\right) = \frac{1}{\sigma_X \sigma_Y} Cov\,(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

On appelle **coefficient de corrélation** (en. *correlation coefficient*) de X et Y, noté ρ_{XY}, $\rho(X, Y)$ ou r_{XY}, le réel défini par :

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Propriétés :

- $|\rho_{XY}| \leq 1$
- si X et Y sont indépendantes, alors $\rho_{XY} = 0$ (la réciproque n'est pas forcément vraie)
- $\exists (a, b) \in \mathbb{R}^2 :\ Y = aX + b \Rightarrow |\rho_{XY}| = 1$. Le coefficient de corrélation indique la relation linéaire entre les variables X et Y.
- si $\rho_{XY} = 0$, alors $Var(X + Y) = Var(X) + Var(Y)$



Mais attention : **la corrélation N'implique PAS la causalité**.

Fonction caractéristique et fonction génératrice d'un couple de v.a.r.

Reprenons les définitions de [1].

On appelle **fonction caractéristique** (en. *characteristic function* ou *CF*) du couple de v.a.r. (X, Y) la fonction ϕ_{XY} définie de $\mathbb{R}^2$ dans $\mathbb{C}$ par :

$$\phi_{XY}(t_1, t_2) = \mathbb{E}\left[e^{i(t_1 X + t_2 Y)}\right]$$

On appelle **fonction génératrice** (en. *generating function*) du couple de v.a.r. (X, Y) la fonction $G_{XY} : \mathbb{R}^2 \to \mathbb{R}$ définie par :

$$G_{XY}(s_1, s_2) = \mathbb{E}\left[s_1^X s_2^Y\right]$$

Il est possible de formuler l'indépendance de deux variables aléatoires via des fonctions caractéristiques et des fonctions génératrices [1].

1. Soit (X, Y) un couple de v.a.r. de fonction caractéristique ϕ_{XY}. Soit ϕ_X et ϕ_Y les fonctions caractéristiques de X et Y respectivement.

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

$$\forall s, t \in \mathbb{R},\ \phi_{XY}(s, t) = \phi_X(s) \times \phi_Y(t)$$

2. Soit (X, Y) un couple de v.a.r. discrètes de fonction génératrice G_{XY}. Soit G_X et G_Y les fonctions génératrices de X et Y respectivement.

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

$$\forall s, t \in [-1, 1],\ G_{XY}(s, t) = G_X(s) \times G_Y(t)$$

Somme de deux variables aléatoires réelles indépendantes

Jusqu'à présent nous nous sommes intéressés à deux v.a.r. et nous avons commencé à voir quelques examples des fonctions de deux v.a.r.

(cas discret)
Soit X et Y deux v.a.r. discrètes. Soit $Z = g(X, Y)$ où $g : \mathbb{R}^2 \to \mathbb{R}$. Alors :

$$\mathbb{P}_Z(z) = \mathbb{P}_Z(g(X, Y) = z) = \sum_{(x_i, y_j) \in A} \mathbb{P}_{XY}(x_i, y_j)$$

où $A = \{(x_i, y_j) \in \mathbb{R}_{XY} :\ g(x_i, y_j) = z\}$
(cas continu)
Dans le cas continu, la fonction de répartition de $Z = g(X, Y)$ peut être défini comme :

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(g(X, Y) \leq z) = \iint\limits_D f_{XY}(x, y) dxdy$$

où $D = \{(x, y) \mid g(x, y) < z\}$

S'il s'agit de trouvez l'espérance de $Z = g(X, Y)$, on peut se servir de l'expression suivante :

(cas discret) Soit X et Y deux v.a.r. discrètes. Soit $Z = g(X, Y)$ où $g : \mathbb{R}^2 \to \mathbb{R}$. Alors, l'espérance de Z est donnée par (en. *Law of the unconscious statistician (LOTUS)*) :

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \sum_{(x_i, y_j) \in \mathbb{R}_{XY}} g(x_i, y_j)\mathbb{P}_{XY}(x_i, y_j)$$

(cas continu) Soit (X, Y) un couple de v.a.r. absolument continu à la fonction de densité $f_{XY}(x, y)$. Soit $Z = g(X, Y)$ où $g : \mathbb{R}^2 \to \mathbb{R}$. Alors, l'espérance de Z est donnée par :

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dxdy$$

Notons que dans le cas où X et Y sont idépendantes, $f_{XY}(x, y) = f_X(x) \times f_Y(y)$.

Dans le cas du couple de v.a.r. absolument continu (X, Y), on peut formuler un théorème suivant [2].

Soit (X, Y) un couple de v.a.r. absolument continu. Soit l'application $g : \mathbb{R}^2 \to \mathbb{R}$ une fonction admettant des dérivées partielles continues.
Soit (Z, W) un couple de v.a.r. défini par :

$$(Z, W) = g(X, Y) = (g_1(X, Y), g_2(X, Y))$$

Soit $h = g^{-1}$, i.e. :

$$(X, Y) = h(Z, W) = (h_1(Z, W), h_2(Z, W))$$

Alors (Z, W) est un couple de v.a.r. absolument continu à la fonction de densité jointe $f_{ZW}(z, w)$ défini par :

$$\forall (z, w) \in \mathbb{R}_{ZW}, \; f_{ZW}(z, w) = f_{XY}(h_1(z, w), h_2(z, w))|J|$$

où J est le Jacobian de h donné par :

$$J = \det \begin{bmatrix} \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \\ \frac{\partial h_2}{\partial z} & \frac{\partial h_2}{\partial w} \end{bmatrix} = \frac{\partial h_1}{\partial z} \cdot \frac{\partial h_2}{\partial w} - \frac{\partial h_2}{\partial z} \cdot \frac{\partial h_1}{\partial w}$$

En reprenant le déroulement de [2], considérons deux v.a.r. X et Y à la fonction de densité jointe $f_{XY}(x, y)$. Définissons une nouvelle v.a.r. Z comme suit : $Z = X + Y$.

Soit $Z = X + Y$. Quelle est sa fonction de densité $f_Z(z)$?

Afin de pouvoir appliquer le théorème précédent, il nous faut deux v.a.r. Z et W. On peut définir $W = X$. Dans ce cas-là, la fonction g assure les transformations suivantes :

$$\begin{cases} z = x + y \\ w = x \end{cases}$$

Trouvons maintenant la fonction inverse h pour x et y :

$$\begin{cases} x = w \\ y = z - w \end{cases}$$

Le Jacobian est donc donné par :

$$J = \det \begin{bmatrix} \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \\ \frac{\partial h_2}{\partial z} & \frac{\partial h_2}{\partial w} \end{bmatrix} = \det \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} = 0 \times (-1) - 1 \times 1 = -1$$

Donc :

$$|J| = |-1| = 1$$

Alors :

$$f_{ZW}(z, w) = f_{XY}(h_1(z, w), h_2(z, w))|J| = f_{XY}(w, z - w) \times 1$$

Selon l'énoncé, nous cherchons la fonction de densité $f_Z(z)$. Donc il s'agit de la fonction de densité marginale, c.à.d. :

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(w, z - w)dw$$

Notons que si X et Y sont indépendant, alors $f_{XY}(x, y) = f_X(x)f_Y(y)$. Dans ce cas-là :

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(w, z - w)dw = \int_{-\infty}^{\infty} f_X(w)f_Y(z - w)dw$$

Cet intégral est appelé la **convolution** ou **le produit de convolution** (en. *convolution*) de f_X et f_y. La notation suivante est utilisée :

$$f_Z(z) = f_X(z)*f_Y(z) = \int_{-\infty}^{\infty} f_X(w)f_Y(z - w)dw = \int_{-\infty}^{\infty} f_Y(w)f_X(z - w)dw$$

Soit X et Y deux v.a.r. indépendantes. Soit $Z = X + Y$. La loi de Z est obtenue en effectuant le **produit de convolution** (en. *convolution*) des lois de X et Y comme suit :
(cas discret)

$$\forall k \in \mathbb{N},\ \mathbb{P}(Z = k) = \sum_{i \in \mathbb{N}} \mathbb{P}(X = k - i) \times \mathbb{P}(Y = i)$$

(cas continu)
Soit $f_X(x)$ et $f_Y(y)$ les fonctions de densité de X et Y respectives. On obtient :

$$\forall z \in \mathbb{R},\ f_Z(z) = \int_{\mathbb{R}} f_X(x)f_Y(z - x)dx$$

Considérons un exemple suivant :

Soit X et Y deux v.a.r. discrètes.

X est définie par :

x	-1	1
$\mathbb{P}(X = x)$	$1/3$	$2/3$

Y est définie par :

y	-2	0	2
$\mathbb{P}(Y = y)$	$2/6$	$1/6$	$3/6$

Quelle est la probabilité que $X + Y$ soit égale à 1 ?

Regardons les cas où la somme $X + Y$ va être égale à 1 :

$\downarrow X \; Y \rightarrow$	-2	0	2
-1	-3	0	**1**
1	-1	**1**	3

Les deux possibilités sont :

- $X = -1$ et $Y = 2$
- $X = 1$ et $Y = 0$

Alors :

$$\mathbb{P}_{X+Y}(1) = \mathbb{P}_X(X = -1) \times \mathbb{P}_Y(Y = 2) + \mathbb{P}_X(X = 1) \times \mathbb{P}_Y(Y = 0) =$$
$$= \frac{1}{3} \times \frac{3}{6} + \frac{2}{3} \times \frac{1}{6} = \frac{1}{6} + \frac{1}{9} = \frac{5}{18}$$

Soient X, Y deux v.a.r. indépendantes de fonctions caractéristiques ϕ_X et ϕ_Y. Soit $Z = X + Y$. Alors la fonction caractéristique de Z, notée ϕ_Z est donnée par :

$$\phi_Z(t) = \phi_X(t) \times \phi_Y(t), \ \forall t \in \mathbb{R}$$

Soient X, Y deux v.a.r. indépendantes de fonctions génératrice G_X et G_Y. Soit $Z = X + Y$. Alors :

$$G_Z(s) = G_X(s) \times G_Y(s), \ \forall s \in [-1, 1]$$

Quelques cas particuliers

Soit $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Soit X et Y indépendantes. Alors :

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

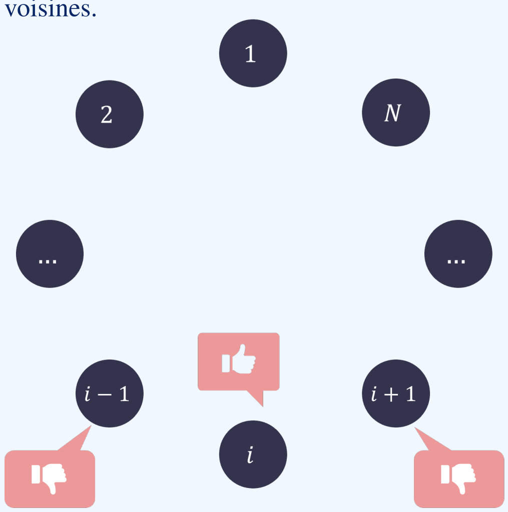
Soit $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$. Soit X et Y indépendantes. Alors :

$$X + Y \sim \mathcal{P}(\lambda + \mu)$$

Vecteurs aléatoires

Combien de gagnants ?

$N > 5$ annonces publicitaires sont affichées en carrousel. Chaque annonce peut être aimé ou pas aimé par utilisateur avec la probabilité 0.5 (comme dans le cas d'un lacement d'une pièce). Une prime est associée à une publicité si la préférence de l'utilisateur (aimé ou pas aimé) est différente de celles faite par rapport à ces voisines.



$\Rightarrow i^{\text{ème}}$ élément gagne une prime

$\Rightarrow i^{\text{ème}}$ élément ne gagne pas de prime

Combien de primes en moyenne vont être attribuées ?

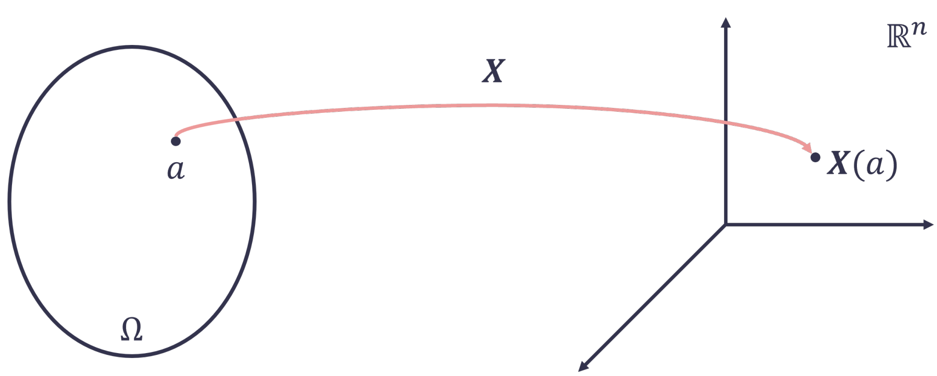
A présent, nous allons étendre les notions vues précédement pour le cas de n variables.

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix}$$

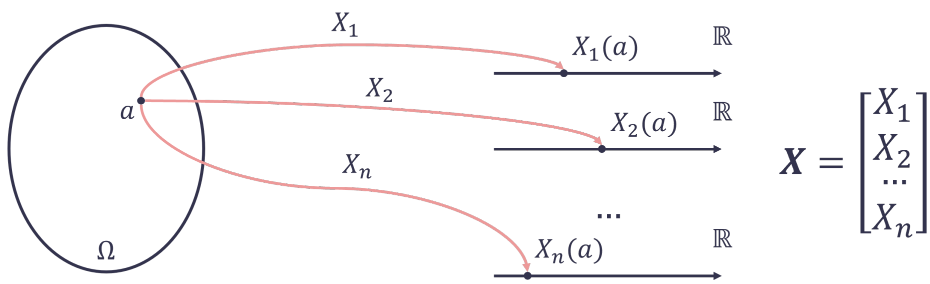
Remarque : en terme de notation, pour les vecteurs on utilise souvent les lettres majuscules en gras, e.g. **X**, **Y**, **Z**. Une autre façon de noter est $\underset{\sim}{X}$.

Graphiquement on peut représenter un vecteur aléatoire **X** de deux façons suivantes :

1. une application de Ω dans $\mathbb{R}^n$



2. un vecteur de n variables aléatoires



Loi

Soit $X_1, X_2, ...X_n$ n v.a.r. définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.
(cas discret) Soit X_i, $i = 1, ..., n$ v.a.r. discrètes.
La **fonction de masse jointe** (en. *joint probability mass function* ou *joint PMF*) est donnée par :

$$\mathbb{P}_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$$

La fonction de répartition jointe (en. *joint CDF*) de n v.a.r. $X_1, ..., X_n$ est définie comme :

$$F_{X_1 X_2 ... X_n}(x_1, x_2, ..., x_n) = \mathbb{P}([X_1 \le x_1] \cap [X_2 \le x_2] \cap \ldots \cap [X_n \le x_n])$$

(cas continu) Soit $X_1, X_2, ...X_n$ absolument continues.
La **fonction de densité jointe** (en. *joint PDF*) $f_{X_1 X_2 ... X_n}(x_1, x_2, ..., x_n)$ est une fonction telle que pour tout ensemble $A \subset \mathbb{R}^n$ la probabilité des v.a.r. $X_1, X_2, ..., X_n$ de se retrouver dans A est donnée par l'intégrale de cette fonction sur l'ensemble A, i.e. :

$$\forall A \in \mathbb{R}^n,\ \mathbb{P}\left((X_1, X_2, ..., X_n) \in A\right) = \int \ldots \int_A \ldots \int f_{X_1 X_2 ... X_n}(t_1, t_2, ..., t_n) dt_1 dt_2 ... dt_n$$

La **fonction de répartition jointe** (en. *joint CDF*) de $X_1, ..., X_n$ est donnée par :

$$F_{X_1 X_2 ... X_n}(x_1, x_2, ..., x_n) = \mathbb{P}(X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n) =$$
$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_n} f_{X_1 X_2 ... X_n}(t_1, t_2, ..., t_n) dt_1 dt_2 ... dt_n$$

Notons que :
- $\lim\limits_{x \to -\infty} F_{\mathbf{X}}(x, ..., x) = 0$
- $\lim\limits_{x \to +\infty} F_{\mathbf{X}}(x, ..., x) = 1$

Considérons un exemple suivant :

Soit X, Y et Z trois v.a.r. absolument continues de la fonction de densité jointe :

$$f_{XYZ}(x, y, z) = \begin{cases} c(3x + 2y + z) 0 \le x \le 1,\ 0 \le y \le 1,\ 0 \le z \le 1 \\ 0 \qquad\qquad\qquad sinon \end{cases}$$

où c est une constante.

Trouvons c.

Solution :

Lorsqu'il s'agit de la fonction de densité jointe, le suivant est vraie :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)dxdydz = 1$$

En se servant de cette expression, on peut trouver c. Notons qu'en dehors de $0 \leq x \leq 1,\ 0 \leq y \leq 1,\ 0 \leq z \leq 1$ la fonction de densité est égale à 0. Alors :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)dxdydz = \int_0^1\int_0^1\int_0^1 c(3x+2y+z)dxdydz = 1$$

$$\int_0^1\int_0^1\int_0^1 c(3x+2y+z)dxdydz = \int_0^1\int_0^1 c\left(\frac{3x^2}{2}+2yx+zx\right)\Big|_0^1 dydz = \int_0^1\int_0^1 c\left(\frac{3}{2}+2y+z\right)dydz =$$

$$= \int_0^1 c\left(\frac{3}{2}y+\frac{2y^2}{2}+zy\right)\Big|_0^1 dz = \int_0^1 c\left(\frac{3}{2}y+y^2+zy\right)\Big|_0^1 dz = \int_0^1 c\left(\frac{3}{2}+1+z\right)dz =$$

$$= \int_0^1 c\left(\frac{5}{2}+z\right)dz = c\left(\frac{5}{2}z+\frac{z^2}{2}\right)\Big|_0^1 = c\left(\frac{5}{2}+\frac{1}{2}\right) = 3c$$

Alors :

$$3c = 1$$
$$c = \frac{1}{3}$$

Donc, on peut substituer c par sa valeur dans f_{XYZ} :

$$f_{XYZ}(x,y,z) = \begin{cases} \frac{1}{3}(3x+2y+z) 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ 0 \leq z \leq 1 \\ 0 \qquad\qquad sinon \end{cases}$$

Comme dans le cas bivarié, il est également possible de définir la loi marginale des variables aléatoires.

> Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un vecteur aléatoire de $\mathbb{R}^n$. On appelle k^{ime} **loi marginale** (en. *marginal distribution*) $k \in \{1, \ldots, n\}$ du $\mathbf{X}$ la loi de la v.a.r. X_k.

La densité marginale de X_i peut être obtenue par l'intégration de tous les autres $X_j,\ j \neq i$, i.e. :

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{X_1X_2\ldots X_n}(x_1, x_2, \ldots, x_n)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n$$

$$F_{X_i}(x) = \mathbb{P}(X_i \leq x) = \mathbb{P}(X_1 \in \mathbb{R}, \ldots, X_{i-1} \in \mathbb{R}, X_i \leq x, X_{i+1} \in \mathbb{R}, \ldots, X_n \in \mathbb{R}) =$$

$$= \lim_{y \to +\infty} F_X(\underbrace{y, \ldots, y}_{i-1\ lments}, \overbrace{x}^{i^{me}}, \underbrace{y, \ldots, y}_{partir\ de\ (i+1)^{me}})$$

Dans le cas continu :

$$F_{X_i}(x) = \int_{-\infty}^{x} f_{X_i}(t_1, \ldots, t_n)dt_i$$

Reprenons notre exemple et trouvons la densité marginale de X.

Comme nous avons noté précédement, la fonction de densité jointe est égale à 0 en dehors de $0 \leq x \leq 1,\ 0 \leq y \leq 1,\ 0 \leq z \leq 1$. Alors pour $0 \leq x \leq 1$, la densité marginale peut être calculée comme suit :

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)dydz = \int_0^1\int_0^1 \frac{1}{3}(3x+2y+z)dydz = \int_0^1 \frac{1}{3}(3xy+2\frac{y^2}{2}+zy)\Big|_0^1 dz =$$

$$= \int_0^1 \frac{1}{3}(3x+1+z)dz = \frac{1}{3}\left(3xz+z+\frac{z^2}{2}\right)\Big|_0^1 = \frac{1}{3}\left(3x+1+\frac{1}{2}\right) = \frac{1}{3}\left(3x+\frac{3}{2}\right) = x+\frac{1}{2}$$

Moments de vecteur aléatoire

Espérance

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un vecteur aléatoire défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Soit $\mathbb{E}[X_i] \in \mathbb{R}, \forall i = 1, \ldots, n$ l'espérance de X_i. L'espérance de $\mathbf{X}$ est définie comme :

$$\mathbb{E}\mathbf{X} = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]) \in \mathbb{R}^n$$

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un vecteur aléatoire défini sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.
Soit $Z = h(X_1, \ldots, X_n)$ où $h : \mathbb{R}^n \to \mathbb{R}$ une fonction bornée et continue par morceaux.
(cas discret) Soit $\mathbf{X}$ a des valeurs dans $D = \mathbf{X}(\Omega) \subset \mathbb{R}^n$ sous ensemble fini ou dénombrable.
Alors, l'espérance $\mathbb{E}Z$ est donnée par :

$$\mathbb{E}Z = \sum_{k=(k_1,\ldots,k_n)\in D} h(k_1, \ldots, k_n)\mathbb{P}(X_1 = k_1, \ldots, X_n = k_n)$$

(cas continu) Soit $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ la fonction de densité de $\mathbf{X}$.
Alors, l'espérance $\mathbb{E}Z$ est donnée par :

$$\mathbb{E}Z = \int_{\mathbb{R}^n} h(t_1, \ldots, t_n) f_{\mathbf{X}}(t_1, \ldots, t_n) dt_1 \ldots dt_n$$

lorsque cette intégrale existe.

Soient $\mathbf{X}$ et $\mathbf{Y}$ deux vecteurs aléatoires de $\mathbb{R}^n$. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée d'ordre n à coefficients réels.
On a :

1. $\mathbb{E}[A\mathbf{X}] = A\,\mathbb{E}\mathbf{X}$
2. $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}\mathbf{X} + \mathbb{E}\mathbf{Y}$

Matrice de covariance

Soit $\mathbf{X} = (X_1, \ldots, X_n)^T$ un vecteur aléatoire dans $\mathbb{R}^n$. La **matrice de covariance** (en. *covariance matrix*), notée $\Sigma_{\mathbf{X}}$ ou $K_{\mathbf{XX}}$, est une matrice carrée symétrique telle que :

$$K_{\mathbf{XX}} = \Sigma_{\mathbf{X}} = \Big(Cov(X_i, X_j)\Big)_{i,j=1,\ldots,n}$$

Propriétés :

1. *symétrique*, i.e. : $Cov(X_i, X_j) = Cov(X_j, X_i),\ \forall i, j = 1, \ldots, n$
2. *semi-définie positive*, i.e. : pour $\forall \mathbf{x} = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^T \in \mathbb{R}^n$ on a $\mathbf{x}^T M \mathbf{x} \geq 0$

Soit $Z = X_1 + \ldots + X_n$. Alors **la variance** peut être calculée comme suit :

$$Var(Z) = Cov\left(\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

Moment de l'ordre n

Soit X une v.a.r. On appelle **moment de l'ordre** n (en. n^{th} *moment*) de X la valeur $\mathbb{E}[X^n]$.
On appelle **moment central de l'ordre** n (en. n^{th} *central moment*) de X la valeur $\mathbb{E}[(X - \mathbb{E}X)^n]$.

Ainsi, l'espérance de X est le moment de l'ordre $n = 1$, la variance de X est le moment central de l'ordre $n = 2$.

Indépendance

Pour rappel, l'intuition derrière la notion de l'indépendance des évènements est la suivante. Les connaissances qu'on a sur un évènement (une variable aléatoire) n'ont pas d'influence sur la probabilité des évènements qui restent.

Supposons qu'on a des évènements $A_1, A_2, \ldots, A_n$. Le fait que tous les évènements soient indépendants implique que :

$$\mathbb{P}\left(A_5 \cap \overline{A_6}\right) = \mathbb{P}\left(\underbrace{A_5 \cap \overline{A_6}}_{vnement\ d'intrt} \ \middle|\ \underbrace{\overline{A_1} \cup A_2 \cup \left(A_3 \cap \overline{A_4}\right) \cup A_7}_{ce\ qu'on\ sait}\right)$$

Notons que les indices des évènements (dans notre cas 5 et 6) formant l'évènement d'intérêt ($A_5 \cap \overline{A_6}$) sont différents des indices des évènements sur la réalisations dequels nous avons l'information (A_1, A_2, A_3, A_4, A_7).

Les évènements $A_1, A_2, \ldots, A_n$ sont dits **mutuellement indépendants** (en. *independent*), si pour tous indices distincts $\forall i, j, \ldots, m : i \neq j \neq \ldots \neq m$ et pour tout nombre d'évènements choisis :

$$\mathbb{P}(A_i \cap A_j\ \cap\ \ldots\ \cap\ A_m) = \mathbb{P}(A_i) \times \mathbb{P}(A_j) \times \ldots \times \mathbb{P}(A_m)$$

Les v.a.r. $X_1, X_2, \ldots, X_n$ sont dites **mutuellement indépendantes** (en. *mutually independent*) si $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$, les évènements $[X_i \leq x_i], i = 1, \ldots, n$ sont *mutuellement indépendants*.

Si $X_1, \ldots, X_n$ sont mutuellement indépendantes, alors :

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Les v.a.r. $X_1, X_2, \ldots, X_n$ sont dites **indépendantes et identiquement distribuées (i.i.d.)** (en. *independent and identically distributed (i.i.d.)*) si $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$, les v.a.r. X_i, $i = 1, \ldots, n$ sont mutuellement indépendantes et ont la même fonction de répartition :

$$F_{X_1}(x) = F_{X_2}(x) = \ldots = F_{X_n}(x),\ \forall x \in \mathbb{R}$$

Soit $X_1, \ldots, X_n$ n v.a.r. i.i.d., alors :

$$\mathbb{E}[X_1 \ldots X_n] = (indpendance) = \mathbb{E}X_1 \times \mathbb{E}X_2 \times \ldots \times \mathbb{E}X_n =$$
$$= (identiquement\ distribues) = \mathbb{E}X_1 \times \mathbb{E}X_1 \times \ldots \times \mathbb{E}X_1 = (\mathbb{E}X_1)^n$$

Exemple

Prenons l'exemple de publicité carrousel inspiré par [2].

Soit X le nombre de primes attribuées. Il s'agit de trouver $\mathbb{E}X$.

On numérote les annonces de 1 à N.

Soit X_i une v.a.r. qui sert de l'indicateur si une prime est attribuée à l'annonce i, i.e. :

$$X_i = \begin{cases} 1\text{prime est attribuée à l'annonce } i \\ 0 sinon \end{cases}$$

Alors :

$$X = X_1 + X_2 + \ldots + X_n$$

Prenons X_i. La prime est attribuée dans deux cas :

1. $A = (\text{👎}_{i-1} \cap \text{👍}_i \cap \text{👎}_{i+1})$
2. $B = (\text{👍}_{i-1} \cap \text{👎}_i \cap \text{👍}_{i+1})$

Notons que les évènements (préférences) sont indépendantes. Alors :

1. $\mathbb{P}(\text{👎}_{i-1} \cap \text{👍}_i \cap \text{👎}_{i+1}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$
2. $\mathbb{P}(\text{👍}_{i-1} \cap \text{👎}_i \cap \text{👍}_{i+1}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

Ainsi :

$\mathbb{P}(X_i = 1) = \mathbb{P}(\text{👎}_{i-1} \cap \text{👍}_i \cap \text{👎}_{i+1}) + \mathbb{P}(\text{👍}_{i-1} \cap \text{👎}_i \cap \text{👍}_{i+1}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$

Donc :

$$\mathbb{E}X_i = 1 \times \mathbb{P}(X_i = 1) + 0 \times \mathbb{P}(X_i = 0) = 1 \times \frac{1}{4} + 0 \times \frac{3}{4} = \frac{1}{4}$$

Ceci est vrai pour tout $i = 1, \ldots, N$.

Ainsi, on peut trouver $\mathbb{E}X$ comme suit :

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_N = N \times \mathbb{E}X_i = N \times \frac{1}{4} = \frac{N}{4}$$

Quelle est la variance de X ?

Selon la formule :

$$Var(X) = \sum_{i=1}^{N} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j) = \sum_{i=1}^{N} Var(X_i) + \sum_{i\neq j} Cov(X_i, X_j) =$$

$$= \sum_{i=1}^{N} Var(X_i) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i\neq j}}^{N} Cov(X_i, X_j)$$

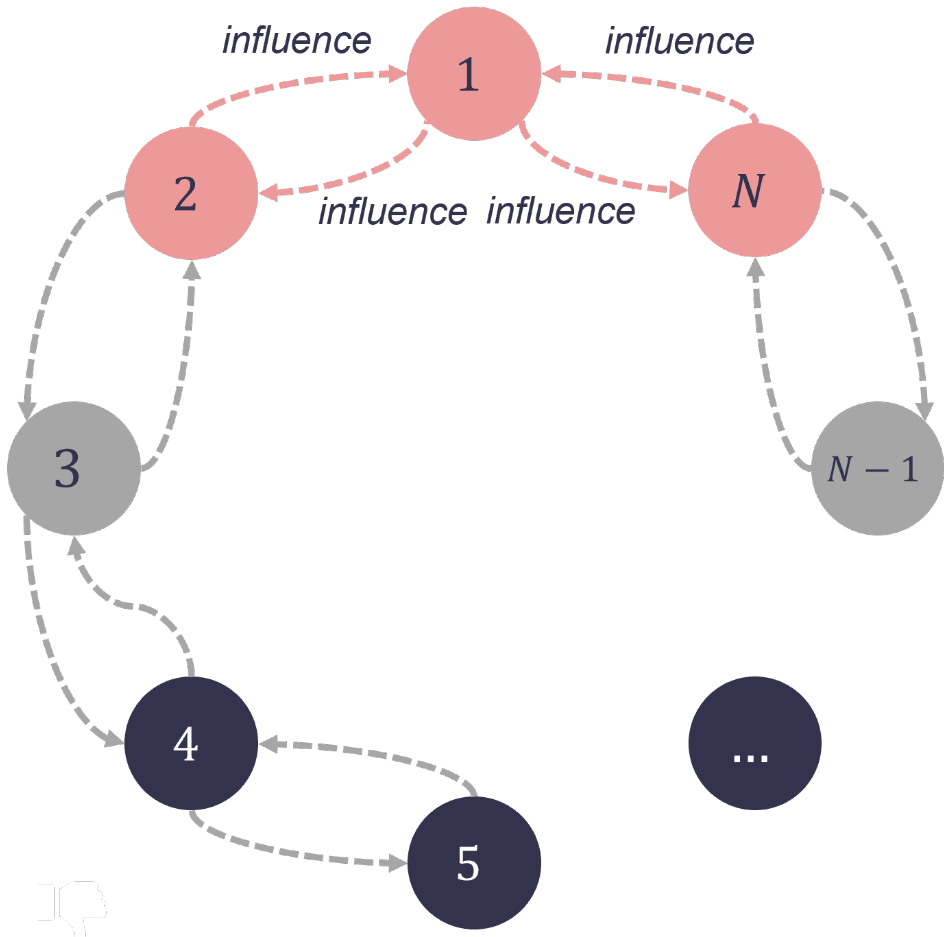
Considérons X_i. Comme noté précédement, il y a deux issues possibles (1 et 0), dont le succès ($X_i = 1$) se réalise avec la probabilité $\frac{1}{4}$. On peut donc dire que X_i suit la loi de Bernoulli $X_i \sim \mathcal{B}\left(\frac{1}{4}\right)$ pour tout $i = 1, \ldots, N$. Donc :

$$Var(X_i) = p(1-p) = \frac{1}{4} \times \left(1 - \frac{1}{4}\right) = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$$

Donc : $\sum_{i=1}^{N} Var(X_i) = N \times \frac{3}{16} = \frac{3N}{16}$

Explorons l'indépendance des X_i.

Prenons $i = 1$ comme exemple. Afin d'évaluer la valuer de X_i, nous regardons les 2 voisins, i.e. avec les indices $i-1$ et $i+1$. Si $i = 1$, alors $i+1 = 2$ et $i-1 = N$:



Etant donné que X_1 dépend des issues des éléments aux indices 2 et N qui se situe à la distance de 2, on peut conclure que :

- X_1 et X_2 sont dépendantes
- X_1 et X_3 sont dépendantes
- X_1 et X_4 sont **indépendantes**
- X_1 et X_2 sont **indépendantes**
- ...
- X_1 et X_{N-2} sont **indépendantes**
- X_1 et X_{N-1} sont dépendantes
- X_1 et X_N sont dépendantes

On peut généraliser cette conclusion d'une manière suivante : X_i et X_j sont indépendantes si $2 < |i - j| < N - 2$.

Pourquoi la notion d'indépendance est utile dans notre cas ? C'est parce que la covariance de deux v.a.r. indépendantes est égale à 0, i.e. :

$$Cov(X_i, X_j) = 0, \ \forall i, j : 2 < |i - j| < N - 2$$

Illustrons cela pour $N = 6$. Dans ce cas $N - 2 = 4$. La matrice de covariance est donnée par :

$$K = \begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) & \mathbf{0} & Cov(X_1,X_5) & Cov(X_1,X_6) \\ Cov(X_2,X_1) & Var(X_2) & Cov(X_2,X_3) & Cov(X_2,X_4) & \mathbf{0} & Cov(X_2,X_6) \\ Cov(X_3,X_1) & Cov(X_3,X_2) & Var(X_3) & Cov(X_3,X_4) & Cov(X_3,X_5) & \mathbf{0} \\ \mathbf{0} & Cov(X_4,X_2) & Cov(X_4,X_3) & Var(X_4) & Cov(X_4,X_5) & Cov(X_4,X_6) \\ Cov(X_5,X_1) & \mathbf{0} & Cov(X_5,X_3) & Cov(X_5,X_4) & Var(X_5) & Cov(X_5,X_6) \\ Cov(X_6,X_1) & Cov(X_6,X_2) & \mathbf{0} & Cov(X_6,X_4) & Cov(X_6,X_5) & Var(X_6) \end{pmatrix}$$

Dans le cas de $N = 7$, la matrice de covariance est donnée par :

$$\begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) & \mathbf{0} & \mathbf{0} & Cov(X_1,X_6) & Cov(X_1,X_7) \\ Cov(X_2,X_1) & Var(X_2) & Cov(X_2,X_3) & Cov(X_2,X_4) & \mathbf{0} & \mathbf{0} & Cov(X_2,X_7) \\ Cov(X_3,X_1) & Cov(X_3,X_2) & Var(X_3) & Cov(X_3,X_4) & Cov(X_3,X_5) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Cov(X_4,X_2) & Cov(X_4,X_3) & Var(X_4) & Cov(X_4,X_5) & Cov(X_4,X_6) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & Cov(X_5,X_3) & Cov(X_5,X_4) & Var(X_5) & Cov(X_5,X_6) & Cov(X_5,X_7) \\ Cov(X_6,X_1) & \mathbf{0} & \mathbf{0} & Cov(X_6,X_4) & Cov(X_6,X_5) & Var(X_6) & Cov(X_6,X_7) \\ Cov(X_7,X_1) & Cov(X_7,X_2) & \mathbf{0} & \mathbf{0} & Cov(X_7,X_5) & Cov(X_7,X_6) & Var(X_7) \end{pmatrix}$$

On peut également noter la symétrie suivante :

$Cov(X_1, X_2) = Cov(X_2, X_3)... = Cov(X_{i-1}, X_i) = Cov(X_i, X_{i+1})... = Cov(X_{N-1}, X_N) = Cov(X_N, X_1)$

$Cov(X_1, X_3) = Cov(X_2, X_4)... = Cov(X_{i-1}, X_{i+1}) = Cov(X_i, X_{i+2})... = Cov(X_{N-1}, X_1) = Cov(X_N, X_2)$

Pour rappel, la covariance de X_i et X_j est donnée par :

$$Cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}X_i \mathbb{E}X_j$$

Précédement, nous avons trouvé que $\mathbb{E}X_i = \frac{1}{4}, \ \forall i = 1, ..., N$.

Notons que pour chaque $i = 1, ..., N$, il faut trouver $Cov(X_i, X_{i+1})$ et $Cov(X_i, X_{i+2})$

$$\mathbb{E}[X_i X_{i+1}] = \mathbb{P}(X_i = 1, X_{i+1} = 1)$$

Regardons sur l'exemple de $i = 1$, i.e. X_1 et X_2.

Pour que $X_1 = 1$ et $X_2 = 1$, les deux cas sont possibles :

1. $A =$ (👎$_N$ ∩ 👍$_1$ ∩ 👎$_2$ ∩ 👍$_3$)
2. $B =$ (👍$_N$ ∩ 👎$_1$ ∩ 👍$_2$ ∩ 👎$_3$)

Dont les probabilités sont données par :

1. $\mathbb{P}(A) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
2. $\mathbb{P}(B) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

Donc :

$$\mathbb{E}[X_i X_{i+1}] = \mathbb{P}(X_i = 1, X_{i+1} = 1) = \mathbb{P}(A) + \mathbb{P}(B) = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}$$

Alors :

$$Cov(X_i, X_{i+1}) = \mathbb{E}[X_i X_{i+1}] - \mathbb{E}X_i \mathbb{E}X_{i+1} = \frac{1}{8} - \frac{1}{4} \times \frac{1}{4} = \frac{1}{8} - \frac{1}{16} = \frac{1}{16}$$

Maintenant trouvons $\mathbb{E}[X_i X_{i+2}] = \mathbb{P}(X_i = 1, X_{i+2} = 1)$

Reprenons l'exemple de $i = 1$. Deux options sont possibles :

1. $A = ($👎$_N \cap$ 👍$_1 \cap$ 👎$_2 \cap$ 👍$_3 \cap$ 👎$_4$ $)$
2. $B = ($ 👍$_N \cap$ 👎$_1 \cap$ 👍$_2 \cap$ 👎$_3 \cap$ 👍$_4$ $)$

Les probabilités de ces deux évènements sont données par :

1. $\mathbb{P}(A) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{32}$
2. $\mathbb{P}(B) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{32}$

Donc :

$$\mathbb{E}[X_i X_{i+2}] = \mathbb{P}(X_i = 1, X_{i+2} = 1) = \mathbb{P}(A) + \mathbb{P}(B) = \frac{1}{32} + \frac{1}{32} = \frac{1}{16}$$

Donc :

$$Cov(X_i, X_{i+2}) = \mathbb{E}[X_i X_{i+2}] - \mathbb{E}X_i \times \mathbb{E}X_{i+2} = \frac{1}{16} - \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} - \frac{1}{16} = 0$$

Ainsi, nous pouvons trouver $\sum_{\substack{i=1 \\ i \neq j}}^{N} Cov(X_i, X_j)$ utilisé dans la formule de $Var(X)$:

$$\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} Cov(X_i, X_j) = 2N \times Cov(X_1, X_2) + 2N \times Cov(X_1, X_3) = 2N \times \frac{1}{16} + 2N \times 0 = \frac{2N}{16}$$

La formule finale est donc :

$$Var(X) = \frac{3N}{16} + \frac{2N}{16} + 0 = \frac{5N}{16}$$

Fonction caractéristique et fonction génératrice

Dans le cas bivariée, nous avons défini la fonction de densité de $Z = X + Y$ dans le cas où X et Y étaient indépendantes. Pour cela, nous avons utilisé la notion du produit de convolution, i.e. :

$$f_Z(z) = f_X(z) * f_Y(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

Pour le cas de n v.a.r. mutuellement indépendantes $X_1, X_2, \dots, X_n$, on pourrait se servir du même principe et obtenir la fonction de densité de $Z = X_1 + \dots + X_n$ comme suit :

$$f_Z(z) = f_{X_1}(z) * f_{X_2}(z) \ * \ \dots \ * \ f_{X_n}(z)$$

Cependant, cette approche devient assez coûteuse en terme de calcul. Ainsi, la fonction caractéristique et la fonction génératrice sont à privilégier dans ce cas-là.

Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire dans $\mathbb{R}^n$. La fonction $\phi_{\mathbf{X}}$ définie pour $\forall \tau = (\tau_1, \dots, \tau_n) \in \mathbb{R}^n$ est appellée **fonction caractéristique** (en. *characteristic function*) de $\mathbf{X}$ qui est donnée par :

$$\phi_{\mathbf{X}}(\tau) = \mathbb{E}\left(\exp\left(i \sum_{i=1}^{n} \tau_k X_k\right)\right) = \mathbb{E}\left[e^{i(\tau.\mathbf{X})}\right]$$

où $\tau.\mathbf{X} \in \mathbb{R}$ est le produit scalaire entre deux vecteurs.

Soit $\mathbf{X}$ un vecteur aléatoire de fonction caractéristique $\phi_{\mathbf{X}}(\tau)$. Pour toute matrice $\forall A \in \mathcal{M}_n(\mathbb{R})$ et vecteur $\forall \mathbf{B} \in \mathbb{R}^n$, la fonction caractéristique du vecteur aléatoire $\mathbf{Y} = A\mathbf{X} + \mathbf{B}$ est donnée par :

$$\phi_{\mathbf{Y}}(\tau) = e^{i(\tau.\mathbf{B})}\phi_{\mathbf{X}}(\tau A)$$

où $\tau A \in \mathbb{R}^n$ est le produit matriciel.

Soit $\mathbf{X} = (X_1, ..., X_n)$ un vecteur aléatoire discret dans $\mathbb{R}^n$. On appelle **fonction génératrice** (en. *generating function*) de $\mathbf{X}$, notée $G_{\mathbf{X}}(s)$, la fonction définie pour $\forall s = (s_1, ..., s_n) \in [0, 1]^n$ comme suit :

$$G_{\mathbf{X}}(s) = \mathbb{E}\left[s_1^{X_1} ... s_n^{X_n}\right]$$

Soit X une v.a.r. On appelle **fonction génératrice des moments** (en. *moment generating function* ou *MGF*) de X, notée $M_X(s)$, la fonction définie comme :

$$M_X(s) = \mathbb{E}\left[e^{sX}\right]$$

$M_X(s)$ existe s'il existe une constante $a > 0, a \in \mathbb{R}$ telle que $M_X(s)$ est finie $\forall s \in [-a, a]$.

En se servant de séries de Taylor pour $e^x, \forall x \in \mathbb{R}$:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

On peut obtenir que :

$$M_X(s) = \mathbb{E}\left[e^{sX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(sX)^k}{k!}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{s^k X^k}{k!}\right] = \sum_{k=0}^{\infty} \mathbb{E}[X^k]\frac{s^k}{k!}$$

Soit X une v.a.r. de fonction génératrice des moments $M_X(s)$. On obtient :

$$M_X(s) = \mathbb{E}\left[e^{sX}\right] = \sum_{k=0}^{\infty} \mathbb{E}[X^k]\frac{s^k}{k!}$$

où

$$\mathbb{E}[X^k] = \frac{d^k}{ds^k} M_X(s)|_{s=0}$$

Soit $X_1, X_2, ..., X_n$ n v.a.r. mutuellement indépendantes de fonction génératrice des moments $M_{X_i}, \ \forall i = 1, ..., n$.
Soit $Z = X_1 + ... + X_n$.
Alors la fonction génératrice des moments $M_Z(s)$ est donnée par :

$$M_Z(s) = M_{X_1}(s) \times M_{X_2}(s) \times ... \times M_{X_n}(s)$$

Quelques distributions importantes

Loi multinomiale (en. *multinomial distribution*), notée $\mathcal{M}(n, p_1, ..., p_k)$ où $n \in \mathbb{N}^*$ et $p_i \in]0, 1[\ \forall i \in \{1, ..., k\}$ est la généralisation de la loi binomiale, i.e. :

$$\mathcal{B}(n, p) = \mathcal{M}(n, p, 1 - p)$$

La fonction de masse de $\mathcal{M}(n, p_1, ..., p_k)$ est définie par :

$$\mathbb{P}(X_1 = \eta_1, ..., X_n = \eta_n) = \frac{n!}{\eta_1! \times ... \times \eta_k!} p_1^{\eta_1} \times ... \times p_k^{\eta_k}, \ \forall \eta = (\eta_1, ..., \eta_k) \in \mathbb{N}^k$$

où :

- $\sum_{i=1}^{k} \eta_i = n$
- $\sum_{i=1}^{k} p_i = 1$

La **loi normale n-dimensionnelle** ou **loi normale multidimensionnelle** (en. *multivariate normal distribution* ou *multivariate Gaussian distribution* ou *joint normal distribution*), notée $\mathcal{N}(\mu, \Sigma)$ où $\mu = (\mu_1, ..., \mu_n) \in \mathbb{R}^n$ et Σ est une matrice de covariance (carrée d'ordre n, symétrique définie positive), est la généralisation de la loi normale $\mathcal{N}(m, \sigma^2)$ où $\mu = (m)$ et $\Sigma = (\sigma^2)$

$$f_{\mathbf{X}}(x_1, ..., x_n) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right)$$

où

- $(\mathbf{x} - \mu)^T$ désigne le vecteur colonne composé de $x_i - \mu_i, \ \forall i \in \{1, ..., n\}$
- Σ^{-1} désigne l'inverse de la matrice Σ

La valeur $\sqrt{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}$ est appelé **la distance de Mahalanobis** (en. *Mahalanobis distance*) entre le point $\mathbf{x}$ et l'espérance μ.

Somme de n v.a.r.

Soit $X_1, ..., X_n$ n v.a.r. mutuellement indépendantes.
Soit $S = \sum_{i=1}^{n} X_i$.
(cas discret)
Soit $X_i, i = 1, ..., n$ v.a.r. discrètes de fonction génératrice G_{X_i}.
Alors, la fonction génératrice $G_S(s)$ de S pour $\forall s \in [0, 1]$ est donnée par :

$$G_S(s) = \prod_{i=1}^{n} G_{X_i}(s)$$

(cas continu)
Soit φ_{X_i} une fonction caractéristique de $X_i, i = 1, ..., n$.
Alors, la fonction caractéristique de S pour $\forall t \in \mathbb{R}$ est donnée par :

$$\varphi_S(t) = \prod_{i=1}^{n} \varphi_{X_i}(t)$$

Soit $X_1, X_2, \ldots, X_n$ n v.a.r. mutuellement indépendantes de même loi de Bernoulli de paramètre p, c.à.d. $\forall i = 1, \ldots, n,\ X_i \sim \mathcal{B}(p)$. Alors, la v.a.r. définie comme la somme $\sum_{i=1}^n X_i$ suit une loi binomiale de paramètre (n, p).

Soit $X_1, X_2, \ldots, X_n$ n v.a.r. mutuellement indépendantes, dont chaque $X_k,\ \forall k \in \{1, \ldots, n\}$ suit une loi normale $\mathcal{N}(\mu_k, \sigma_k^2)$.
Alors, la v.a.r. définie comme la somme $\sum_{i=1}^n X_i$ suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, où :

- $\mu = \sum_{k=1}^{n} \mu_k$
- $\sigma^2 = \sum_{k=1}^{n} \sigma_k^2$

Si $\forall k = \{1, \ldots, n\}$, $X_k \sim \mathcal{N}(0, 1)$, alors $T = \sum_{k=1}^{n} X_k^2$ suit la loi Chi-Carré à n degré de liberté, notée χ_n^2 ou $\chi^2(n)$.

Identité de Wald ou **formule de Wald** (en. *Wald's equation* ou *Wald's identity* ou *Wald's lemma*) :
Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé discret.
Soit N une v.a.r. discrète définie sur Ω à valeurs dans $\mathbb{N}^*$ de fonction génératrice G_N.
Soit $X_1, \ldots, X_N$ des v.a.r. i.i.d. définies sur Ω de fonction génératrice commune G_X.
Soit S une v.a.r. définie comme :

$$S : \omega \in \Omega \rightarrow S(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega)$$

Alors l'espérance de S est donnée par :

$$\mathbb{E}[S] = \mathbb{E}[X_i] \times \mathbb{E}[N]$$

S admet pour fonction génératrice :
$\forall s \in [0, 1],\ G_S(s) = G_N(G_X(s))$

La variance de S est donnée par :

$$Var(S) = \mathbb{E}[N] \times Var(X_i) + Var(N) \times (\mathbb{E}[X_i])^2$$

Liens utiles

1. Jeremy Orloff, and Jonathan Bloom. *18.05 Introduction to Probability and Statistics*. Spring 2014. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.
2. Basic probability: Joint, marginal and conditional probability | Independence by zedstatistics
3. Joint PDFs by MIT OpenCourseWare
4. Mean of sum and difference of random variables by Khan Academy
5. Variance of sum and difference of random variables by Khan Academy

References

[1] Stéphane Balac and Olivier Mazet. *Introduction aux Probabilités*. URL: https://perso.univ-rennes1.fr/stephane.balac/publis/polypbs.pdf (visited on 05/13/2021).
[2] Hossein Pishro-Nik. *Introduction to Probability, Statistics, and Random Processes*. English. Blue Bell, PA: Kappa Research, LLC, Aug. 2014. ISBN: 9780990637202. URL: https://www.probabilitycourse.com/.

[3] John Tsitsiklis and Patrick Jaillet. *RES.6-012 Introduction to Probability*. MIT OpenCourseWare. 2018. URL: https://ocw.mit.edu/resources/res-6-012-introduction-to-probability-spring-2018/ (visited on 11/02/2021).