

Rappels de probabilités

univers fini

Sylvain Pelletier

LMSC - PSI

Définitions

- Une probabilité p sur l'univers fini Ω est une application $p : P(\Omega) \rightarrow [0, 1]$, qui vérifie : $p(\Omega) = 1$, et si $A \cap B = \emptyset$, alors : $p(A \cup B) = p(A) + p(B)$.
- On obtient alors facilement :
 $p(\bar{A}) = 1 - p(A)$ $p(\emptyset) = 0$ $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
- Les événements sont les éléments de $p(\Omega)$, i.e. les parties de Ω . Une probabilité = une mesure des événements.
- Un système complet d'événements est une liste d'événements qui forment une partition de Ω :

$$(A_i)_{i=1 \dots n} \text{ est un SCE si } \Omega = \bigcup_{i=1}^n A_i \text{ union disjointe}$$

Probabilité définie par la mesure des singletons

- Une probabilité p sur un ensemble fini est entièrement déterminée par les $\text{Card}(\Omega)$ valeurs donnant la probabilité des singletons

$p(\{w_i\})_{i \in [1, n]}$ pour $w_i \in \Omega$:

$$\forall A \in P(\Omega), P(A) = \sum_{i \in [1, n] \text{ tq } w_i \in A} p(\{w_i\})$$

- On a de plus :

$$\sum_{i=1}^n p(\{w_i\}) = 1 \text{ avec } n = \text{Card}(\Omega).$$

- Utile lorsque la probabilité dépend d'un paramètre à déterminer.
- Exemple : un dé où $p(D = i)$ est proportionnel à i .
- Permet de définir la probabilité uniforme comme l'unique probabilité telle que chaque singleton a la probabilité $\frac{1}{n}$.

Probabilités conditionnelles

La probabilité de B sachant A est

$$p(B|A) = p_A(B) = \frac{p(A \cap B)}{p(A)} \text{ avec } p(A) \neq 0.$$

On peut écrire :

$$p(A \cap B) = p(A)p_A(B).$$

- La probabilité conditionnelle sachant A est une probabilité et vérifie donc toutes les propriétés d'une probabilité.
Par exemple : $p_A(\bar{B}) = 1 - p_A(B)$.
- Souvent A est chronologiquement avant B si ce n'est pas le cas, c'est la formule de Bayes !

Probabilités composées

(A_i) une suite d'événements vérifiant : $p(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$ alors :

$$p(A_1 \cap A_2 \cap \dots \cap A_n) = p(A_1)p_{A_1}(A_2)p_{A_1 \cap A_2}(A_3) \dots p_{A_1 \cap A_2 \cap \dots \cap A_{n-1}}(A_n)$$

- ▶ L'hypothèse est souvent vérifiée sous la forme : « la modélisation assure que ... »
- ▶ Souvent A_i concerne chronologiquement l'étape i .
- ▶ Utile pour considérer la probabilités d'une suite de tirage.

Exemple : loi de la première blanche dans un tirage sans remise.

Probabilités totales

Soit B un événement, et (A_i) un SCE. On a alors :

$$p(B) = \sum_{i=1}^m p(A_i \cap B) = \sum_{i=1}^m p_{A_i}(B)p(A_i).$$

- ▶ Sans doute la formule la plus utile du cours de probabilité
- ▶ On découpe l'événement B selon le SCE (A_i) .
Souvent B concerne la deuxième étape, et (A_i) est l'ensemble des résultats possibles d'une première étape.
- ▶ Si on veut déterminer la loi de Y à partir de la loi du couple (X, Y) , on écrit :

$$\forall j \in Y(\Omega), p(Y = j) = \sum_{k=1}^n p(Y = j \cap X = k).$$

Formule de Bayes

- ▶ Soit (A_i) un système complet d'événements et B un autre événement, alors :

$$\forall j \in \llbracket 1, m \rrbracket, p_B(A_j) = \frac{p_{A_j}(B)p(A_j)}{p(B)} = \frac{p_{A_j}(B)p(A_j)}{\sum_{i=1}^m p_{A_i}(B)p(A_i)}$$

- ▶ En particulier soit A un événement avec $0 < p(A) < 1$ et B un autre événement, alors

$$p_B(A) = \frac{p_A(B)p(A)}{p(B)} = \frac{p_A(B)p(A)}{p_A(B)p(A) + p_{\bar{A}}(B)p(\bar{A})}$$

- ▶ Formule des causes : on constate l'événement B de la deuxième étape et on se demande quel est la probabilité que la première étape ait donné A .

Exemple fondamental

- ▶ On a un dé et 6 urnes, l'urne numéro i contient i B et $6 - i$ N. Si $D = i$ on tire plusieurs fois dans l'urne i avec remise.

- ▶ $p(B_1)$? Expérience aléatoire en deux étapes : formule des probabilités totales sur la première étape!
- ▶ B_1 et B_2 sont-ils indépendants? NON! Même si les tirages se font avec remise, le résultat du première tirage donne de l'information sur le numéro du dé et donc sur le deuxième tirage
- ▶ Sachant B_1 quelle est la probabilité que le dé a donné i ? C'est la formule de Bayes!

Bien comprendre la différence entre indépendance et indépendance conditionnelle.

Loi d'une somme de VARs

- ▶ On considère $X \sim \mathcal{U}(n)$ et $Y \sim \mathcal{U}(n)$ indépendantes. Loi de $Z = X + Y$?

Bien retenir la méthode

- ▶ Les probabilités totales.
- ▶ Exprimer $(X + Y = k \cap X = l)$ par $(X = l \cap Y = k - l)$
- ▶ Identifier les valeurs de l telles que $(X = l \cap Y = k - l)$ est impossible. Les enlever de la somme.
- ▶ Conditionner si besoin pour se ramener à $p(X = l)p_{X=l}(Y = k - l)$
- ▶ Simplifier la somme.

Variables aléatoires

- ▶ Une VAR est une **fonction** $X : \Omega \rightarrow \mathbb{R}$.
- ▶ L'univers image $X(\Omega)$ est l'**ensemble des valeurs possibles** de X .
- ▶ La famille d'événements $((X = k))_{k \in X(\Omega)}$ forme un SCE.

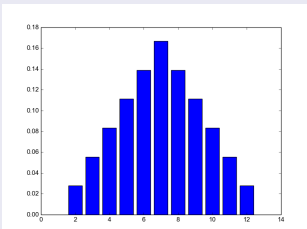
- ▶ Mesure le **gain d'une expérience**. Exemple valeur d'un dé.
- ▶ On utilise des **variables caractéristiques** (de bernoulli) pour certains événements. Exemple tirage B/N, $X_i = 1$ si B_i . Alors $\sum_{i=1}^n X_i$ est le nombre de blanches.
- ▶ Autre exemple : N jetons, on en tire n au hasard. $X_i = 1$ si on tire le jeton i . Alors :

$$\sum_{i=1}^N X_i = n \quad \text{et} \quad \sum_{i=1}^N iX_i \text{ est la valeur des jetons tirés.}$$

Lois et fonction de répartition

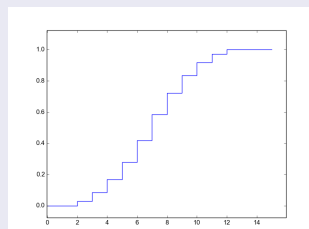
Loi de X :

$$f_X : \begin{cases} X(\Omega) & \rightarrow [0, 1] \\ x & \mapsto p(X = x) \end{cases}$$



Fonction de répartition

$$F_X : \begin{cases} \mathbb{R} & \rightarrow [0, 1] \\ x & \mapsto p(X \leq x) \end{cases}$$



La fonction de répartition détermine la loi. Exemple : maximum / minimum.

Espérance

C'est le **gain moyen**

$$E(X) = \sum_{i=1}^n x_i p(X = x_i) = \sum_{x \in X(\Omega)} x p(X = x).$$

- ▶ L'espérance est linéaire :
 $\forall (a, b) \in \mathbb{R}^2, E(aX + b) = aE(X) + b$ et $E(X + Y) = E(X) + E(Y)$

- ▶ Soit U une fonction de $\mathbb{R}$ dans $\mathbb{R}$

$$E(U(X)) = \sum_{i=1}^n U(x_i) p(X = x_i)$$

- ▶ Utile pour avoir l'**espérance d'une composée sans calculer sa loi** : la loi de $U(X)$ est dure à calculer mais son espérance se calcule facilement.

Variance

Mesure de l'écart quadratique moyen à la moyenne

$$V(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

L'écart-type est : $\sigma(X) = \sqrt{V(X)}$.

$$\forall (a, b) \in \mathbb{R}^2, V(aX + b) = a^2 V(X) \text{ et } \sigma(aX + b) = |a| \sigma(X). \\ V(X + Y) = V(X) + 2\text{cov}(X, Y) + V(Y).$$

► Inégalité de Bienaymé-Tchebychev

$$\forall A \in \mathbb{R}^+, p(|X - m| \geq A) \leq \frac{\sigma^2}{A^2}$$

donne la probabilité qu'une variable s'éloigne de sa moyenne.

Lois usuelles

Variables	Univers images	Loi	Modèle	E	V
Certaine	$\{a\}$	$P(X = a) = 1$	issue certaine	a	0
$\mathcal{U}([1, n])$ Uniforme sur $[1, n]$	$[1, n]$	$P(X = k) = \frac{1}{n}$	issues équiprobables, lancer de dé	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$ (*)
$\mathcal{B}(1, p)$ Bernoulli	$\{0, 1\}$	$P(X = 1) = p,$ $P(X = 0) = q = 1 - p$	Choix binaires, succès ou échec	p	pq
$\mathcal{B}(n, p)$ Binomial	$[0, n]$	$P(X = k) = \binom{n}{k} p^k q^{n-k}$	Nombre de succès dans un schéma de Bernoulli, tirage avec remise	np	npq

Couples de VAR

► C'est la donnée de deux VAR (X, Y) . L'univers image est $(X, Y)(\Omega) = X(\Omega) \times Y(\Omega)$.

► Loi conjointe

$$\begin{cases} X(\Omega) \times Y(\Omega) \rightarrow [0, 1] \\ (x, y) \mapsto p([X = x] \cap [Y = y]) \end{cases}$$

► Toujours commencer par identifier les valeurs tels que $p(X = x \cap Y = y) = 0$.

► Vocabulaire : lois marginales (= loi de X seul), lois conditionnelles.

► La loi du couple détermine les lois marginales par la formule des probabilités totales.

Théorème de transfert

Soit U une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$, alors :

$$\begin{aligned} E(U(X, Y)) &= \sum_{i=1}^n \sum_{j=1}^m U(x_i, y_j) p([X = x_i] \cap [Y = y_j]) \\ &= \sum_{(x, y) \in X(\Omega) \times Y(\Omega)} U(x, y) p([X = x] \cap [Y = y]) \end{aligned}$$

► On peut ainsi calculer l'espérance de la composée sans connaître la loi de $U(X, Y)$.

Covariance

- Mesure l'influence d'une variable sur une autre

$$\text{cov}(X, Y) = E([X - E(X)][Y - E(Y)]) = E(XY) - E(X)E(Y).$$

- Interprétation du signe.
- Propriétés :

$$\text{cov}(X, Y) = \text{cov}(Y, X)$$

$$\text{cov}(\lambda X + X', Y) = \lambda \text{cov}(X, Y) + \text{cov}(X', Y)$$

$$\text{cov}(X, Y + \mu Y') = \text{cov}(X, Y) + \mu \text{cov}(X, Y')$$

$$\text{cov}(X, X) = V(X) \geq 0$$

- Si les variables aléatoires sont indépendantes, alors la covariance est nulle, mais la réciproque est fautive !

Variance d'une somme

Avec la définition, on constate :

$$V(X + Y) = V(X) + 2\text{cov}(X, Y) + V(Y)$$

Deux utilisations

- on connaît $\text{cov}(X, Y)$ on en déduit la variance de la somme.
 - on connaît $X + Y$ qui suit une loi usuelle, on en déduit la covariance.
-
- on lance n boules au hasard qui tombent dans 3 boîtes. On note X_i nombre de boules dans la boîte i .
 - Loi de X_i ?
 - $\text{cov}(X_1, X_2)$?

Variance d'une somme

Pour deux variables : $V(X_1 + X_2) = V(X_1) + 2\text{cov}(X_1, X_2) + V(X_2)$

Pour trois variables :

$$V(X_1 + X_2 + X_3) = V(X_1) + V(X_2) + V(X_3) + 2\text{cov}(X_1, X_2) + 2\text{cov}(X_1, X_3) + 2\text{cov}(X_2, X_3)$$

Pour n variables :

$$\begin{aligned} V\left(\sum_{i=1}^n V(X_i)\right) &= \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \\ &= \sum_{i=1}^n V(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j) \\ &= \sum_{(i,j)} \text{cov}(X_i, X_j) \end{aligned}$$

Variance d'une somme

$$\begin{aligned} V\left(\sum_{i=1}^n V(X_i)\right) &= \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \\ &= \sum_{i=1}^n V(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j) \\ &= \sum_{(i,j)} \text{cov}(X_i, X_j) \end{aligned}$$

- Il faut avoir compris les trois expressions de cette formule (les double produit). En particulier le nombre de termes. Toujours écrire dans un tableau les valeurs de $\text{cov}(X_i, X_j)$.
- Très souvent la somme double se simplifie (beaucoup de termes sont nuls, $\text{cov}(X_i, X_j)$ constant, $\text{cov}(X_i, X_j)$ est nul si j est « loin » de i).

- ▶ Un **vecteur aléatoire** est n VARs : $(X_1, \dots, X_n)$.
- ▶ $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$ dans tous les cas.
- ▶ Si les VARs (X_i) sont indépendantes :

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

Dans le cas général, c'est la **variance d'une somme** :

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{i < j} \text{cov}(X_i, X_j)$$

Coefficient de corrélation linéaire

Soit X et Y non constantes. On appelle **coefficient de corrélation**, le nombre :

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

On a :

$$\rho(X, Y) \in [-1, 1], \quad \text{ie} \quad |\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y).$$

L'égalité a lieu si et seulement si $\exists (a, b) \in \mathbb{R}, Y = aX + b$. (sauf pour des valeurs de probabilités nulles).

Démonstration à connaître (lien avec les produits scalaires).

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j)$$

- ▶ Espérance et variance d'une **VAR binomiale** (= cas de variables indépendantes).
- ▶ Espérance et variance du **nbr de B dans un tirage sans remise** (=cas sans indépendance).

On considère

$$P : t \mapsto V(tX + Y) = t^2 V(X) + 2t \text{cov}(X, Y) + V(Y)$$

- ▶ P est un polynôme de degré 2.
- ▶ $\forall t \in \mathbb{R}, P(t) \geq 0$,
- ▶ donc $\Delta \leq 0$, ce qui donne : $|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y)$

- ▶ Si $\Delta = 0$, alors $\exists t_0 \in \mathbb{R}, V(t_0 X + Y) = 0$
- ▶ et donc $t_0 X + Y$ est une constante.
- ▶ Réciproque à rédiger.

Indépendance de deux événements

A et B deux événements

- Indépendance de deux événements :

$$p(A \cap B) = p(A)p(B).$$

- L'indépendance dépend de la probabilité. Deux événements peuvent être indépendants pour une probabilité et pas pour une autre.
- Voir l'exemple des 6 urnes et du dé : B_1 et B_2 ne sont pas indépendants pour la proba p , mais sont indépendants pour la probabilité conditionnelle.
- C'est souvent une conséquence de la modélisation, mais si c'est le sujet de la question on démontre.
- Parfois c'est contre intuitif.

Indépendance des événements

A et B deux événements

- Indépendance de deux événements :

$$p(A \cap B) = p(A)p(B).$$

$(A_i)_{i \in \llbracket 1, n \rrbracket}$ n événements.

- Indépendance deux à deux de n événements :

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j \implies A_i \text{ et } A_j \text{ indépendants}$$

$$\text{ie } p(A_i \cap A_j) = p(A_i)p(A_j)$$

- Indépendance mutuelle de n événements :

$$\forall J \subset \llbracket 1, n \rrbracket, J \neq \emptyset \implies p\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} p(A_j)$$

Indépendance d'une famille finie d'événements

$(A_i)_{i \in \llbracket 1, n \rrbracket}$ n événements.

- Indépendance deux à deux :

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j \implies A_i \text{ et } A_j \text{ indépendants}$$

- Indépendance mutuelle :

$$\forall J \subset \llbracket 1, n \rrbracket, J \neq \emptyset \implies p\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} p(A_j)$$

- Ce n'est pas la même chose
indépendance mutuelle $\implies$ indépendance deux à deux
- Quand on ne précise pas, c'est l'indépendance mutuelle.
- C'est souvent une conséquence de la modélisation,
- Parfois c'est contre intuitif.

Indépendance de deux VARs

X et Y deux VARs

- Indépendance de deux VARs :

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), p(X = x \cap Y = y) = p(X = x)p(Y = y).$$

- Cela signifie en particulier que les lois marginales donnent la loi conjointe : chaque "case tu tableau" est le produit des "cases du bords".
- Pour le démontrer, il faut traiter toutes les cases.
- Si on doit démontrer que deux VARs ne sont pas indépendantes, il suffit de trouver un contre exemple.

Indépendance de VARs

X et Y deux VARs

► **Indépendance de deux VARs :**

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), p(X = x \cap Y = y) = p(X = x)p(Y = y).$$

$(X_i)_{i=1\dots n}$ un vecteur de n VARs

► Les variables $(X_i)_{i=1\dots n}$ sont **deux à deux indépendantes** si :

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j \implies X_i \text{ et } X_j \text{ sont indépendantes}$$

► Les variables $(X_i)_{i=1\dots n}$ sont **mutuellement indépendantes** si :

$$\forall (x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega),$$

$$p((X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n))$$

$$= p(X_1 = x_1)p(X_2 = x_2) \dots p(X_n = x_n)$$

Indépendance de VARs

$(X_i)_{i=1\dots n}$ un vecteur de n VARs

► Les variables $(X_i)_{i=1\dots n}$ sont **deux à deux indépendantes** si :

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j \implies X_i \text{ et } X_j \text{ sont indépendantes}$$

En conséquence $\text{cov}(X_i, X_j) = 0$.

► Les variables $(X_i)_{i=1\dots n}$ sont **mutuellement indépendantes** si :

$$\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega),$$

$$p((X_1 = x_1) \cap \dots \cap (X_n = x_n)) = p(X_1 = x_1) \dots p(X_n = x_n)$$

► Ce n'est pas la même chose

(mutuellement indépendante $\implies$ indépendantes 2 à 2)

► Dès que l'on a **indépendance deux à deux**, la variance de la somme est la somme des variances.

Les variables $(X_i)_{i=1\dots n}$ sont **mutuellement indépendantes** si :

$$\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega),$$

$$p((X_1 = x_1) \cap \dots \cap (X_n = x_n)) = p(X_1 = x_1) \dots p(X_n = x_n)$$

► Si on ne précise pas, **indépendantes** signifie mutuellement indépendantes

► Souvent c'est la modélisation qui donne l'indépendance mutuelle. On le démontre très rarement.

► suite de VAR i.i.d. = indépendantes et identiquement distribuées.

► Si $(X_i)_{i=1\dots n}$ sont mutuellement indépendantes.

► si la VAR Y est construite avec certaines variables (X_i) et la VAR Z est construite avec d'autres variables (X_i) (aucune variable ne sert à la fois pour Y et Z)

► alors Y et Z sont indépendantes.

Exemple d'exercice de synthèse

► Pour $i \in \llbracket 1, n+1 \rrbracket$, on considère $X_i \sim \mathcal{B}(p)$ des VAR indépendantes

► pour $i \in \llbracket 1, n \rrbracket$, on note $Y_i = (X_i - X_{i+1})^2$.

► Loi de Y_i ? $Y_i = 1$ si $X_i \neq X_{i+1}$, on obtient $Y_i \sim \mathcal{B}(2pq)$

► $Z = \sum_{i=1}^n Y_i$, interpréter Z . C'est le nombre de changement.

► Donner $E(Z)$. L'espérance est linéaire, il n'y a pas de difficulté.

► Donner $V(Z)$. C'est la variance d'une somme, on doit d'abord calculer $\text{cov}(Y_i, Y_j)$.

Si $j > i+1$, on a Y_i et Y_j indépendantes.

On écrit le tableau des covariances et il ne reste plus qu'à faire la somme.