24 - Probabilités sur un univers fini

Jeremy Daniel

Introduction - Le paradoxe de Bertrand

D'après Joseph Bertrand, *Calcul des probabilités*, 1889.

On considère un cercle $\mathcal{C}$ de rayon 1 et on en choisit *au hasard* une corde [1]. Quelle est la probabilité que la corde soit plus longue que $\sqrt{3}$, la longueur du côté d'un triangle équilatéral inscrit dans $\mathcal{C}$.

- **Alice :** On choisit d'abord un premier point A au hasard sur $\mathcal{C}$. Le triangle équilatéral inscrit dans $\mathcal{C}$ ayant A pour sommet délimite $\mathcal{C}$ en trois partie de même longueur d'arc. Quand on choisit un point B au hasard sur $\mathcal{C}$, la longueur AB est plus grande que $\sqrt{3}$ si, et seulement si B est dans celle des trois zones la plus éloignées de A.
  On trouve donc 1 **chance sur** 3.
- **Bob :** On peut obtenir n'importe quelle corde de la façon suivante. D'abord, on choisit au hasard un rayon sur le cercle. Puis on prend au hasard un point sur ce rayon et on trace la perpendiculaire au rayon en ce point. Le segment de cette perpendiculaire inclus dans l'intérieur de $\mathcal{C}$ est une corde et toute corde se décrit de cette façon.
  Si d est la distance du point pris sur le rayon au centre du cercle, la longueur de la corde est $2\sqrt{1-d^2}$. Elle est supérieure à $\sqrt{3}$ si, et seulement si $d \leq 1/2$.
  On trouve donc 1 **chance sur 2.**
- **Ève :** On choisit un point P au hasard dans le disque intérieur à $\mathcal{C}$ et on trace, comme précédemment, la perpendiculaire en P au rayon du cercle passant P. Cette perpendiculaire définit une corde du cercle.
  Cette corde est de longueur supérieure à $\sqrt{3}$ si, et seulement si la distance du point au centre du cercle est inférieure à 1/2. Or, le disque de rayon 1 a une aire 4 fois plus grande que celui de rayon 1/2.
  On trouve donc 1 **chance sur 4.**

Remarque 0.1
Les trois calculs sont corrects. Le problème est que la notion de *hasard* n'est pas suffisamment précisée pour qu'on puisse donner une unique réponse. En des termes modernes, on n'a pas spécifié quelle était la mesure de probabilité sur l'espace des cordes.

1. On rappelle qu'une corde d'un cercle est un segment de droite joignant deux points du cercle.

Remarque 0.2
Ce paradoxe se place sur un univers infini non dénombrable (l'espace des cordes d'un cercle). On ne peut donc pas le formaliser avec le programme de MPSI/MP. Mais on pourrait donner des paradoxes analogues avec des univers finis ou dénombrables.

1 Espaces probabilisés finis

1.1 Dictionnaire

Définition 1.1 (Univers fini, évènement, issue)
On utilise un vocabulaire probabiliste spécifique pour les notions de théorie des ensembles.

- Un univers fini Ω est un ensemble fini.
- Un évènement A de Ω est une partie de Ω.
- Une issue de Ω est un élément de Ω.

Définition 1.2 (Vocabulaire sur les évènements)
On fixe un univers fini Ω.

- L'évènement certain est la partie Ω.
- L'évènement impossible est la partie $\emptyset$.
- Les évènements élémentaires sont les singletons $\{\omega\}$, où $\omega \in \Omega$ est une issue.
- Si A est un évènement, l'évènement contraire de A est le complémentaire de A dans Ω, noté $\overline{A}$.
- Si A et B sont deux évènements, on appelle A *et* B (resp. A *ou* B) l'évènement $A \cap B$ (resp. $A \cup B$).
- Deux évènements A et B sont incompatibles si A *et* B est l'évènement impossible.

Définition 1.3 (Système complet d'évènements)
Un système complet d'évènements est une famille finie $(A_1, \ldots, A_p)$ d'évènements de Ω telle que les A_i sont deux à deux incompatibles et telle que $\Omega = \bigcup_{i=1}^{p} A_i$.

Remarque 1.4
On ne demande pas, dans la définition d'un système complet d'évènements, que les A_i soient chacun non vides. Quand cette condition supplémentaire est vérifiée, la famille $(A_1, \ldots, A_p)$ définit une partition $\{A_1, \ldots, A_p\}$ de Ω.

Définition 1.5 (Variable aléatoire)
Soit Ω un univers fini, soit E un ensemble. Une variable aléatoire sur Ω, à valeurs dans E, est une application $X : \Omega \to E$.

Remarque 1.6
Si ω est une issue quelconque, on pense à la valeur $X(\omega)$ comme une variable dans E,

dépendant de l'aléa ω (donc du hasard).

Remarque 1.7
Presque toujours, E est une partie de $\mathbb{R}$ et on parle de variable aléatoire réelle. Si E est une partie de $\mathbb{Z}$, on parle de variable aléatoire entière.

Notation 1.8
Soit X une variable aléatoire définie sur Ω, à valeurs dans E. Soient A une partie de E, x un élément de E.

- On note $(X \in A)$ l'évènement $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\} \subset \Omega$.
- On note $(X = x)$ l'évènement $X^{-1}\big(\{x\}\big) = \{\omega \in \Omega \mid X(\omega) = x\} \subset \Omega$.

Notation 1.9
Si de plus $E \subset \mathbb{R}$, on note $(X < x)$ l'évènement $X^{-1}\big(]-\infty, x[\big) = \{\omega \in \Omega \mid X(\omega) < x\} \subset \Omega$. De même, pour les autres inégalités.

1.2 Probabilité

Définition 1.10 (Mesure de probabilité)
Soit Ω un univers fini. Une mesure de probabilité (ou simplement probabilité) est une fonction $P : \mathcal{P}(\Omega) \to [0, 1]$ telle que :

- Si A et B sont des évènements incompatibles, alors $P(A \cup B) = P(A) + P(B)$ (*propriété d'additivité*)
- $P(\Omega) = 1$ (*propriété de normalisation*)

Définition 1.11 (Espace probabilisé fini)
Un espace probabilisé fini (Ω, P) est la donnée d'un univers fini Ω et d'une mesure de probabilité P sur Ω.

Exemple 1.12 (Probabilité uniforme)
Si Ω est un univers fini non vide, on le munit d'une mesure de probabilité P définie par :

$$\forall A \subset \Omega, P(A) = \frac{|A|}{|\Omega|}.$$

C'est la probabilité uniforme sur Ω. Pour toute issue ω, $P\big(\{\omega\}\big) = \dfrac{1}{|\Omega|}$.

Exemple 1.13
On lance un dé à 6 faces et on demande la probabilité d'obtenir un score pair. On modélise cela par l'univers $\Omega = [\![1, 6]\!]$, muni de la probabilité uniforme. L'évènement correspondant à l'obtention d'un score pair est $A = \{2, 4, 6\}$. Il est de cardinal 3 donc la probabilité d'obtenir un score pair est $P(A) = 3/6 = 1/2$.

Exemple 1.14
On lance un dé à 6 faces deux fois d'affilée et on demande la probabilité que la somme des deux scores soit 8. On modélise l'expérience par l'univers $\Omega = [\![1, 6]\!]^2$, muni de la probabilité uniforme. En effet, en l'absence d'autres hypothèses (dé truqué, non-indépendance des lancers...), on considère que c'est cette notion de hasard qui est sous-entendue par l'énoncé. L'évènement A : *somme des scores égale à* 8 est de cardinal 5, car

$$A = \{(2, 6); (3, 5); (4, 4); (5, 3); (6, 2)\}.$$

On a donc $P(A) = 5/6^2 = 5/36$.

Attention !
Dans l'exemple précédent, on pourrait être tenté de modéliser l'expérience par l'ensemble des sommes possibles et prendre $\Omega = [\![2, 12]\!]$. C'est possible mais déterminer la probabilité *naturelle* à mettre sur cet espace n'est pas aisé. De façon générale, le choix de l'univers ne dépend que de l'expérience (*on lance deux dés*), pas de la mesure qu'on fait de cette expérience (*on regarde la somme des deux scores*).

Remarque 1.15
La description précise de l'univers Ω et de la mesure de probabilité P est rarement pertinente. Le plus important est de bien définir les évènements étudiés et de traduire les hypothèses de l'énoncé en termes de probabilités.

Proposition 1.16 (Probabilité de l'union)
Soient A et B deux évènements de Ω. Alors, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proposition 1.17 (Probabilités dans un système complet d'évènements)
Soit $(A_1, \ldots, A_p)$ un système complet d'évènements de Ω. Alors, $\displaystyle\sum_{i=1}^{p} P(A_i) = 1$.
En particulier, si A est un évènement quelconque, alors $P(\overline{A}) = 1 - P(A)$.

Proposition 1.18 (Croissance)
Soient $A \subset B$ deux évènements. Alors, $P(A) \leq P(B)$.

Théorème 1.19 (Caractérisation d'une mesure de probabilité)
Écrivons $\Omega = \{\omega_1, \ldots, \omega_n\}$ et notons $p_i = P(\{\omega_i\})$ les probabilités des évènements élémentaires. Alors $\displaystyle\sum_{i=1}^{n} p_i = 1$.
Réciproquement, considérons $\Omega = \{\omega_1, \ldots, \omega_n\}$ un univers fini. Étant donnée une famille de réels $(p_1, \ldots, p_n)$ telle que
- $\forall i \in [\![1, n]\!] : p_i \in [0, 1]$,

– $\sum_{i=1}^{n} p_i = 1$,

il existe une unique mesure de probabilité P sur Ω telle que $\forall i \in [\![1, n]\!] : P(\{\omega_i\}) = p_i$.

1.3 Loi d'une variable aléatoire

Proposition 1.20
Soit X une variable aléatoire sur l'espace probabilisé fini (X, Ω), à valeurs dans E. Pour tout évènement A de l'univers fini $X(\Omega)$, on note $P_X(A) = P(X \in A)$. Alors, $(X(\Omega), P_X)$ est un espace probabilisé fini.

Définition 1.21 (Loi d'une variable aléatoire)
La mesure de probabilité P_X est la loi de X.

Remarque 1.22
Formellement, la loi de X dépend de l'ensemble $X(\Omega)$. Il peut cependant arriver que $P_X(\omega) = 0$ pour certaines issues ω de $X(\Omega)$. Considérons par exemple l'univers $\Omega = [\![1, 6]\!]^2$, modélisant les lancers successifs de 2 dés et considérons $X : (a, b) \in \Omega \mapsto a + b$, qui correspond à la somme des deux scores. Si les issues $(i, 6)$ de Ω ont probabilité 0 (parce que le deuxième dé est truqué et ne renvoie jamais 6), alors $X(\Omega) = [\![2, 12]\!]$, et pourtant $P_X(12) = 0$. En pratique, du point de vue des probabilités, tout se passe comme si X prenait des valeurs dans $[\![2, 11]\!]$. La détermination précise de $X(\Omega)$ n'est donc pas fondamentale ; seul compte le support de X, défini comme l'ensemble des $x \in E$ tels que $P(X = x) > 0$.

Remarque 1.23
Donner la loi d'une variable aléatoire X revient à expliciter l'ensemble $X(\Omega)$ (ou le support de X) et à donner les probabilités $P_X(x) = P(X = x)$, pour tout x dans $X(\Omega)$.

Exercice 1.24
On lance successivement deux dés à 6 faces (indépendants, non truqués). Donner la loi de la variable aléatoire donnant la somme des deux scores.

Définition 1.25 (Image d'une variable aléatoire)
Soit X une variable aléatoire sur Ω, à valeurs dans E. Soit $f : E \to F$. L'image de la variable aléatoire X par f, notée $f(X)$, est la variable aléatoire $f \circ X$.

Remarque 1.26
De nouveau, on tâchera de ne pas se laisser abuser par le vocabulaire. L'image d'une variable aléatoire correspond en fait à une composée de fonctions. Mais le vocabulaire est cohérent avec l'idée qu'une variable aléatoire correspond à un élément de E, soumis à l'aléa.

Proposition 1.27 (Loi de $f(X)$)

La loi de la variable aléatoire $f(X)$ est déterminée par la loi de X. On a :

$$\forall y \in F, P\big(f(X) = y\big) = \sum_{x \in f^{-1}(y)} P(X = x).$$

1.4 Lois usuelles

Définition 1.28 (Loi uniforme)
Soit E un ensemble fini de cardinal n. Une variable aléatoire $X : (\Omega, P) \to E$ suit la loi uniforme sur E si pour tout x dans E, $P(X = x) = \dfrac{1}{n}$.

Notation 1.29
On note alors $X \sim \mathcal{U}(E)$.

Exemple 1.30
Quand on lance un dé, la variable aléatoire donnant son score suit une loi uniforme sur $[\![1, 6]\!]$.

Définition 1.31 (Loi de Bernoulli)
Soit $p \in [0, 1]$. Une variable aléatoire X à valeurs dans $\{0, 1\}$ suit la loi de Bernoulli de paramètre p si $P(X = 0) = 1 - p$ et $P(X = 1) = p$.

Notation 1.32
On note alors $X \sim \mathcal{B}(p)$.

Remarque 1.33
Si X est à valeurs dans $\{0, 1\}$, elle suit une loi de Bernoulli de paramètre $p = P(X = 1)$.

Exemple 1.34
On lance une pièce qui a une probabilité p de tomber sur Pile et $1-p$ sur Face. La variable aléatoire X valant 1 si la pièce tombe sur Pile et 0 sinon suit une loi de Bernoulli de paramètre p.

Définition 1.35 (Loi de Rademacher)
Une variable X à valeurs dans $\{-1, 1\}$ suit la loi de Rademacher si

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Remarque 1.36
X suit la loi de Rademacher ssi $\dfrac{X+1}{2}$ suit la loi binomiale de paramètre $\dfrac{1}{2}$.

Définition 1.37 (Loi binomiale)
Soient n un entier naturel non nul, soit $p \in [0,1]$. Une variable aléatoire X à valeurs dans $[\![0,n]\!]$ suit la loi binomiale de paramètres n et p si

$$\forall k \in [\![0,n]\!] : P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Notation 1.38
On note alors $X \sim \mathcal{B}(n,p)$.

Exercice 1.39
On reprend l'exemple d'une pièce tombant sur Pile avec probabilité p. On la lance n fois. Quelle est la loi de la variable aléatoire donnant le nombre de Pile obtenus ?

2 Conditionnement et indépendance

2.1 Probabilités conditionnelles

Définition 2.1 (Probabilité de B sachant A)
Soit A un évènement de probabilité non nulle, soit B un évènement. La probabilité de B sachant A, notée $P_A(B)$ ou $P(B \mid A)$, est définie par

$$P_A(B) = \frac{P(A \cap B)}{P(A)}.$$

Attention !
Il n'existe pas d'évènement $(B \mid A)$! La notation $P(B \mid A)$ est donc à prendre dans sa globalité ; ce n'est pas *a priori* la probabilité $P(C)$, pour un certain évènement C.

Exemple 2.2
Considérons le lancer d'un dé à 6 faces, modélisé comme précédemment. Notons A l'évènement : *faire un score pair* et B l'évènement : *faire un score divisble par* 3.
L'évèment $A \cap B$ correspond à faire un 6. Donc $P(A \cap B) = 1/6$, tandis que $P(A) = 1/2$ et $P(B) = 1/3$. On a donc $P_A(B) = (1/6)/(1/2) = 1/3$ et $P_B(A) = (1/6)/(1/3) = 1/2$.

Remarque 2.3
Intuitivement, la probabilité qu'a un évènement de se réaliser est le nombre moyen de fois que cet évènement va se réaliser, divisé par le nombre d'essais, quand le nombre d'essais tend vers l'infini.
Avec cette approche *fréquentiste*[2], la probabilité conditionnelle $P_A(B)$ s'obtient en ne testant la réalisation de l'évènement B que dans les cas où A se réalise. Or, après N essais, A se réalise environ $N \times P(A)$ fois. Et quand A se réalise, B se réalise ssi $A \cap B$ se réalise, ce qui arrive environ $N \times P(A \cap B)$. On en déduit la définition donnée.

2. qu'on justifiera à la fin du chapitre

Proposition 2.4 (P_A est une probabilité sur Ω)
Soit A un évènement tel que $P(A) > 0$. Alors l'application $P_A : \mathcal{P}(\Omega) \to [0,1]$, définie par

$$P_A : B \mapsto P_A(B)$$

est une mesure de probabilité sur Ω.

Proposition 2.5 (Formule des probabilités composées)
Soient $A_1, \dots, A_n$ des évènements tels que $P(A_1 \cap \dots \cap A_n) > 0$. Alors

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \times P_{A_1}(A_2) \times P_{A_1 \cap A_2}(A_3) \times \dots \times P_{A_1 \cap \dots \cap A_{n-1}} P(A_n).$$

Proposition 2.6 (Formule des probabilités totales)
Soit $(A_1, \dots, A_n)$ un système complet d'évènements. Soit B un évènement. Alors,

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i).$$

Si chaque A_i est de probabilité non nulle, on a de plus :

$$P(B) = \sum_{i=1}^{n} P(A_i) P_{A_i}(B).$$

Remarque 2.7
On convient que, si A_i est de probabilité nulle, alors $P(A_i)P_{A_i}(B) = 0$. Alors, la formule ci-dessus peut être appliquée y compris quand les A_i ne sont peut-être pas tous de probabilité non nulle.

Corollaire 2.8
Soient A et B deux évènements. Alors,

$$P(B) = P(A \cap B) + P(\overline{A} \cap B).$$

Si de plus, $0 < P(A) < 1$, on a :

$$P(B) = P(A)P_A(B) + P(\overline{A})P_{\overline{A}}(B).$$

Exercice 2.9
On considère trois urnes U_1, U_2 et U_3. La première contient 0 boule blanche et 1 boule rouge. La deuxième contient 1 boule blanche et 2 boules rouges. La troisième contient 2 boules blanches et 3 boules rouges.
On fait l'expérience consistant à choisir une urne de façon équiprobable, puis à tirer une boule dans l'urne, de façon équiprobable. Quelle est la probabilité d'obtenir une boule blanche ?

Théorème 2.10 (Formule de Bayes)
Soient A et B deux évènements de probabilité non nulle. Alors

$$P_B(A) = \frac{P_A(B)P(A)}{P(B)}.$$

Remarque 2.11
En dépit de son apparence anecdotique, cette formule est fondamentale et une bonne compréhension de ce qu'elle dit permet d'éliminer beaucoup de contre-intuitions probabilistes courantes. On renvoie aux épisodes 26 et 28 de l'excellente chaîne *Hygiène mentale* pour une présentation de l'inférence bayésienne.

Proposition 2.12 (Bayes et probabilités totales)
Si $(A_1, \ldots, A_n)$ est un système complet d'évènements de probabilité non nulle et si B est un évènement de probabilité non nulle, alors :

$$\forall j \in [\![1, n]\!] : P_B(A_j) = \frac{P_{A_j}(B)P(A_j)}{\sum_{i=1}^{n} P_{A_i}(B)P(A_i)}.$$

Exercice 2.13
D'après l'épisode 26 de la chaîne Hygiène mentale.
Bob, amateur de jeu de rôles, possède 5 dés : un dé à 4 faces, un dé à 6 faces, un dé à 8 faces, un dé à 12 faces et un dé à 20 faces. Il fait l'expérience consistant à choisir aléatoirement un dé (de façon équiprobable) et à le lancer. Sachant qu'il a obtenu un 7, quelle est la probabilité qu'il ait lancé le dé à 8 faces ?

2.2 Évènements indépendants

Définition 2.14 (Évènements indépendants)
Deux évènements A et B de Ω sont indépendants si $P(A \cap B) = P(A)P(B)$.

Remarques 2.15
- Un évènement de probabilité 0 (ou de probabilité 1) est indépendant de tout autre évènement.
- Si B est de probabilité non nulle, A et B sont indépendants ssi $P(A) = P_B(A)$. L'indépendance signifie donc que la connaissance que l'évènement B s'est réalisé n'influe pas sur la probabilité que l'évènement A se réalise (et inversement).

Attention !
On ne confondra pas les notions d'évènements incompatibles et d'évènements indépendants. Si A et B sont incompatibles, on a $A \cap B = \emptyset$ et donc $P(A \cap B) = 0$. Donc, le seul cas où deux évènements incompatibles sont indépendants est le cas où au moins l'un des deux est de probabilité nulle.

Remarque 2.16
Dans certains exercices, l'hypothèse d'indépendance est implicite. Par exemple, quand on lance deux dés, il est en général sous-entendu que les résultats des deux dés sont indépendants

Définition 2.17 (Indépendance mutuelle)
Soient $A_1, \ldots, A_n$ des évènements de Ω. On dit qu'ils sont mutuellement indépendants (ou simplement indépendants) si pour tous indices distincts $i_1, \ldots, i_k \in [\![1, n]\!]$,

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k}).$$

Attention !
Cette propriété est plus forte *a priori* que l'indépendance deux à deux des évènements (qu'on emploie rarement), comme le montre l'exemple suivant.

Exercice 2.18
On considère deux lancers successifs (et indépendants) d'un dé classique. On considère A, B et C les évènements :
- A : le premier résultat est pair ;
- B : le deuxième résultat est pair ;
- C : les deux résultats ont même parité.

Montrer que A, B et C sont deux à deux indépendants mais qu'ils ne sont pas mutuellement indépendants.

Attention !
Pour montrer l'indépendance de n évènements $A_1, \ldots, A_n$, il ne suffit pas de montrer que $P(A_1 \cap \cdots \cap A_n) = P(A_1) \ldots P(A_n)$. Un contre-exemple (un peu bête) consiste à prendre trois évènements A, B et C tels que A et B ne sont pas indépendants et C est de probabilité nulle. Alors,

$$P(A \cap B \cap C) = P(A)P(B)P(C) = 0.$$

Proposition 2.19
Soient $A_1, \ldots, A_n$ des évènements mutuellement indépendants. Pour tout $i \in [\![1, n]\!]$, on note B_i un évènement égal à A_i ou à $\overline{A_i}$. Alors, les évènements $B_1, \ldots, B_n$ sont mutuellement indépendants.

Exercice 2.20 (Fonction indicatrice d'Euler)
Soit $n \geq 2$ un entier. On munit l'univers $\Omega = [\![1, n]\!]$ de la probabilité uniforme. On note $p_1, \ldots, p_r$ les facteurs premiers de n. Pour tout entier d divisant n, on note A_d l'évènement

$$A_d = \{1 \leq k \leq n \mid d \text{ divise } k\}$$

et B l'évènement $B = \{1 \leq k \leq n \mid k \text{ et } n \text{ sont premiers entre eux}\}$.

1. Calculer $P(A_d)$.
2. Montrer que les évènements $A_{p_1}, \dots, A_{p_r}$ sont mutellement indépendants.
3. Montrer que $P(B) = \displaystyle\prod_{i=1}^{r} \big(1 - \frac{1}{p_i}\big)$.

3 Couples de variables aléatoires

3.1 Loi conjointe, loi marginale

Définition 3.1 (Loi conjointe de deux variables aléatoires)
Soit $X : (\Omega, P) \to E$ et $Y : (\Omega, P) \to F$ deux variables aléatoires sur (Ω, P). La variable (X, Y) à valeurs dans $E \times F$, donnée par

$$\forall \omega \in \Omega : (X, Y)(\omega) = \big(X(\omega), Y(\omega)\big)$$

est un couple de variables aléatoires. La loi de la variable aléatoire (X, Y) est appelée la loi conjointe des variables aléatoires X et Y.

Proposition 3.2
La loi conjointe de X et Y (à valeurs dans E et F) est entièrement déterminée par la donnée, pour tous x de E et y de F, de $P\big((X = x) \cap (Y = y)\big)$.

Remarque 3.3
On représente souvent ces valeurs dans un tableau à deux dimensions. On peut numéroter les lignes par les éléments x de E, les colonnes par les éléments y de F ; dans la case de coordonnée (x, y), on inscrit $P\big((X = x) \cap (Y = y)\big)$.
Comme les évènements $(X = x) \cap (Y = y)$ forment un système complet d'évènements sur Ω, la somme de toutes valeurs inscrites dans le tableau est égale à 1. De plus la somme des valeurs inscrites dans la ligne x est égale $P(X = x)$ et la somme des valeurs inscrites dans la colonne y est égale à $P(Y = y)$.

Définition 3.4 (Lois marginales d'un couple de variables aléatoires)
Soit Z une variable aléatoire à valeurs dans un produit $E \times F$. On écrit $Z = (X, Y)$, avec X et Y des variables aléatoires dans E et F. Les lois des variables aléatoires X et Y sont les lois marginales de Z.

Proposition 3.5
La loi d'une variable aléatoire Z à valeurs dans un produit $E \times F$ détermine les lois marginales de Z.

Attention !
La réciproque est fausse. Les lois de X et Y nous apprennent très peu sur la loi conjointe de (X, Y).

Exercice 3.6
On considère deux lancers successifs d'une pièce équilibrée. On note X la variable aléatoire valant 1 (resp. 0) si la pièce tombe sur Pile (resp. Face) au premier lancer ; de même pour Y au deuxième lancer. Écrire sous forme de tableau les lois conjointes des couples (X, X), $(X, 1-X)$ et (X, Y).

3.2 Lois conditionnelles

Définition 3.7 (Loi conditionnelle)
Soient X et Y deux variables aléatoires à valeurs dans E et F, soit $x \in E$. On suppose que $P(X = x) > 0$. La loi conditionnelle de Y sachant $(X = x)$ est la loi de Y pour la probabilité conditionnelle $P_{(X=x)}$.

Remarques 3.8

- La loi conditionnelle de Y sachant $(X = x)$ est donc déterminée par les valeurs $P_{(X=x)}(Y = y)$, où $y \in F$; c'est-à-dire par les valeurs $\dfrac{P\big((Y = y) \cap (X = x)\big)}{P(X = x)}$.
- Si la loi conjointe du couple (X, Y) a été écrite dans un tableau bi-dimensionnel, la loi conditionnelle de Y sachant $(X = x)$ se lit en considérant les valeurs de la ligne x et en normalisant par la somme de ces valeurs.

Proposition 3.9
Soient X et Y deux variables aléatoires sur Ω. La loi conjointe de (X, Y) est déterminée par la loi de X et la donnée, pour tout $x \in X(\Omega)$ tel que $P(X = x) > 0$, des lois conditionnelles de Y sachant $(X = x)$. En particulier, ces données déterminent la loi de Y. On a :

$$\forall y \in Y(\Omega) : P(Y = y) = \sum_{x \in X(\Omega)} P(X = x) P_{(X=x)}(Y = y).$$

3.3 Variables aléatoires indépendantes

Définition 3.10 (Variables aléatoires indépendantes)
Deux variables aléatoires X et Y définies Ω sont indépendantes si pour tout $A \subset X(\Omega)$ et pour tout $B \subset Y(\Omega)$, les évènements $(X \in A)$ et $(Y \in B)$ sont indépendants.

Remarque 3.11
Dans les exercices, si deux expériences sont réalisées successivement (ou simultanément, mais sans influence l'une sur l'autre), on considère que les résultats de ces expériences donnent des variables aléatoires indépendantes.

Théorème 3.12
Deux variables aléatoires X et Y sont indépendantes ssi

$$\forall x \in X(\Omega), \forall y \in Y(\Omega) : P\big((X = x) \cap (Y = y)\big) = P(X = x) P(Y = y).$$

Corollaire 3.13
Si deux variables aléatoires X et Y sont indépendantes, alors la loi conjointe de (X, Y) est déterminée par les lois de X et de Y

Proposition 3.14
Si X et Y sont indépendantes, et si f et g sont des fonctions définies respectivement sur $X(\Omega)$ et $Y(\Omega)$, alors les variables aléatoires $f(X)$ et $g(Y)$ sont elles aussi indépendantes.

Exemple 3.15 (Fonction d'un couple de variables aléatoires)
Soient deux variables aléatoires X et Y définies sur Ω. On considère une fonction f définie sur $X(\Omega) \times Y(\Omega)$ et on cherche à déterminer la loi de $f(X, Y)$. La loi de $f(X, Y)$ est déterminée par les valeurs

$$P\big(f(X,Y) = z\big) = \sum_{\substack{(x,y)\in(X,Y)(\Omega) \\ f(x,y)=z}} P\big((X = x) \cap (Y = x)\big),$$

où z est dans $f \circ (X, Y)(\Omega)$. En particulier, il est nécessaire de connaître la loi conjointe de (X, Y) pour déterminer la loi de $f(X, Y)$.

Si on suppose maintenant X et Y indépendantes, la formule se simplifie en :

$$P\big(f(X,Y) = z\big) = \sum_{\substack{(x,y)\in(X,Y)(\Omega) \\ f(x,y)=z}} P(X = x)P(Y = y)$$

Exemple 3.16 (Somme de variables aléatoires indépendantes - Produit de convolution)
On suppose désormais X et Y indépendantes et à valeurs réelles. On cherche la loi de $Z = X + Y$. La formule ci-dessus donne

$$\forall z \in \mathbb{R}, P(Z = z) = \sum_{\substack{x,y\in\mathbb{R} \\ x+y=z}} P(X = x)P(Y = y).$$

On dit que la loi de Z est obtenue comme produit de convolution des lois de X et de Y.

Exercice 3.17
On suppose que X et Y sont deux variables aléatoires indépendantes sur Ω, suivant une loi uniforme sur $[\![1, n]\!]$. Calculer la loi de $X + Y$.

3.4 Extension à un n-uplet de variables aléatoires

Définition 3.18 (Loi conjointe)
Soient $X_1, \dots, X_n$ des variables aléatoires définies sur Ω et à valeurs dans des ensembles (finis) $E_1, \dots, E_n$. La loi conjointe des variables aléatoires $X_1, \dots, X_n$ est la loi de la variable aléatoire $(X_1, \dots, X_n)$, à valeurs dans $E_1 \times \cdots \times E_n$.

Remarque 3.19
Elle est déterminée par toutes les valeurs

$$P\big((X_1 = x_1) \cap \cdots \cap (X_n = x_n)\big),$$

où $x_i \in E_i$. On peut imaginer que ces valeurs sont les éléments d'un tableau n-dimensionnel.

Définition 3.20 (Lois marginales)
Soit X une variable aléatoire à valeurs dans un produit $E_1 \times \cdots \times E_n$. On écrit $X = (X_1, \ldots, X_n)$. Les lois des variables aléatoires X_i sont les lois marginales de X.

Remarque 3.21
Elles sont déterminées par la loi de X, selon la formule :

$$\forall x \in E_i : P(X_i = x) = \sum_{\substack{(x_1,\ldots,x_n)\in E_1\times\cdots\times E_n \\ x_i = x}} P\big((X_1 = x_1) \cap \cdots \cap (X_n = x_n)\big).$$

Définition 3.22 (Variables aléatoires indépendantes)
Des variables aléatoires $X_1, \ldots, X_n$ sur Ω, à valeurs dans des ensembles $E_1, \ldots, E_n$ sont *(mutuellement) indépendantes* si, pour tous évènements $A_1, \ldots, A_n$ de $E_1, \ldots, E_n$, les évènements $(X_1 \in A_1), \ldots, (X_n \in A_n)$ sont mutuellement indépendants.

Proposition 3.23
Les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes si, et seulement si :

$$\forall x_1 \in E_1, \ldots, \forall x_n \in E_n : P\big((X_1 = x_1) \cap \cdots \cap (X_n = x_n)\big) = P(X_1 = x_1) \ldots P(X_n = x_n).$$

Théorème 3.24 (Somme de n variables indépendantes de Bernoulli)
Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes à valeurs sans $\{0, 1\}$, suivant une loi de Bernoulli de même paramètre p.
Alors, la variable aléatoire $X_1 + \cdots + X_n$ suit une loi binomiale de paramètres n et p.

Lemme 3.25 (Lemme des coalitions)
Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes à valeurs dans des ensembles $E_1, \ldots, E_n$. Soient f et g des applications de $E_1 \times \cdots \times E_p$ dans F et de $E_{p+1} \times \cdots \times E_n$ dans G. Alors, les variables aléatoires $f(X_1, \ldots, X_p)$ et $g(X_{p+1}, \ldots, X_n)$ sont indépendantes.

Exemple 3.26
Si X, Y, Z, T sont des variables aléatoires indépendantes, alors les variables aléatoires $X+Y$ et $Z \times T$ sont indépendantes.

4 Espérance, variance, covariance

4.1 Espérance

DÉFINITION 4.1 (Espérance d'une variable aléatoire réelle)
Soit X une variable aléatoire réelle. L'espérance de X, notée $E(X)$, est le réel

$$E(X) = \sum_{x \in \mathbb{R}} xP(X = x).$$

Si $E(X) = 0$, on dit que X est centrée.

REMARQUE 4.2
On pourrait étendre la définition au cas où X est à valeurs dans un $\mathbb{R}$-espace vectoriel. Le cas le plus important en pratique est celui où X est à valeurs dans un espace $\mathbb{R}^d$.

EXERCICE 4.3
Calculer l'espérance de X, si X suit une des lois suivantes :

1. Loi uniforme sur dans $[\![a, b]\!]$, avec $a < b$ deux entiers relatifs
2. Loi de Bernoulli de paramètre p
3. Loi binomiale de paramètres n et p

LEMME 4.4
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Alors,

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}).$$

REMARQUE 4.5
Cette formule exprime l'espérance d'une variable aléatoire réelle X comme une somme portant sur Ω plutôt que sur $\mathbb{R}$. Son intérêt est d'abord théorique.

PROPOSITION 4.6
Soient X et Y deux variables aléatoires réelles. Soient λ et μ deux réels.

- ***Linéarité :*** $E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$.
- ***Positivité :*** *si X est à valeurs dans $\mathbb{R}_+$, alors $E(X) \geq 0$.*
- ***Croissance :*** *si $X \leq Y$ (c'est-à-dire si pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$), alors $E(X) \leq E(Y)$.*

REMARQUE 4.7
En utilisant la linéarité de l'espérance, redémontrer que si $X \sim \mathcal{B}(n, p)$, alors $E(X) = np$.

THÉORÈME 4.8 (Formule de transfert)
Soit X une variable aléatoire réelle. Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Alors

$$E\big(f(X)\big) = \sum_{x \in \mathbb{R}} f(x)P(X = x).$$

Attention !
Cette formule montre que l'espérance de $f(X)$ est déterminée par la loi de X. Cependant, la seule connaissance de l'espérance de X ne permet pas le calcul de l'espérance de $f(X)$.

Exemple 4.9
Si X est une variable aléatoire réelle et si r est un entier naturel, on définit $m_r(X)$ par $m_r(X) = E(X^r)$. C'est le moment d'ordre r de X. On a donc :

$$m_r(X) = \sum_{x \in \mathbb{R}} x^r P(X = x).$$

Remarque 4.10
Si $Z = (X, Y)$ est une variable aléatoire à valeurs dans $\mathbb{R}^2$ et si $f : \mathbb{R}^2 \to \mathbb{R}$, on a aussi une formule de transfert pour calculer l'espérance de $f(Z)$:

$$E\big(f(Z)\big) = \sum_{z \in \mathbb{R}^2} f(z) P(Z = z) = \sum_{x, y \in \mathbb{R}} f(x, y) P\big((X = x) \cap (Y = y)\big).$$

On peut généraliser à une variable aléatoire à valeurs dans $\mathbb{R}^d$, où $d \geq 2$.

Théorème 4.11 (Inégalité de Markov)
Soit X une variable aléatoire réelle, positive ou nulle. Soit $\alpha > 0$. Alors

$$P(X \geq \alpha) \leq \frac{E(X)}{\alpha}.$$

Remarque 4.12
Si X prend des valeurs réelles quelconques, on peut utiliser cette inégalité avec $|X|$, plutôt que X :

$$P\big(|X| \geq \alpha\big) \leq \frac{E\big(|X|\big)}{\alpha}.$$

Exemple 4.13
Si la moyenne de la classe à un devoir est de 5, au plus un tiers de la classe a dépassé 15.

Corollaire 4.14
Soit X une variable aléatoire réelle, soit $f : \mathbb{R} \to \mathbb{R}_+$ une fonction croissante.
Alors, pour tout $\alpha \in \mathbb{R}$:

$$f(\alpha) P(X \geq \alpha) \leq E[f(X)].$$

Théorème 4.15 (Espérance d'un produit de variables indépendantes)
Soient X et Y deux variables aléatoires réelles indépendantes. Alors, $E(XY) = E(X)E(Y)$.

Exemple 4.16
On considère X une variable aléatoire suivant une loi uniforme sur $\{-1, 0, 1\}$ et Y la

variable aléatoire valant 1 quand $X = 0$ et 0 sinon. On a $E(X) = 0$; $E(Y) = 1/3$ et, comme $XY = 0$, $E(XY) = 0$. Donc, on a bien $E(XY) = E(X)E(Y)$. Mais X et Y ne sont pas indépendantes. On a $P_{(X=0)}(Y = 1) = 1$, mais $P(Y = 1) = 1/3$.
Ceci constitue un contre-exemple à la réciproque du théorème.

4.2 Variance

Définition 4.17 (Variance d'une variable aléatoire réelle)
Soit X une variable aléatoire réelle. On définit sa *variance* $V(X)$ par

$$V(X) = E\Big(\big(X - E(X)\big)^2\Big).$$

Si $V(X) = 1$, on dit que X est réduite.

Remarques 4.18
- La variance de X est donc le moment d'ordre 2 de la variable aléatoire centrée $X - E(X)$.
- Comme espérance d'une variable aléatoire prenant des valeurs positives ou nulles, la variance est positive.
- La variance mesure l'écart d'une variable aléatoire à sa moyenne ; on dit que c'est un indicateur de dispersion.

Définition 4.19 (Écart-type)
L'écart-type $\sigma(X)$ d'une variable aléatoire réelle X est la racine carrée de sa variance $V(X)$.

Remarques 4.20
- L'écart-type de X s'exprime *dans la même unité* que X.
- Étant donnée une variable aléatoire X, la variable aléatoire $X^* = \dfrac{X - E(X)}{\sigma(X)}$ (si $\sigma(X) > 0$) est centrée et réduite. Il est souvent pratique de raisonner avec cette variable aléatoire X^* *doublement normalisée*, plutôt qu'avec X elle-même.

Proposition 4.21
Si X est une variable aléatoire, $V(X) = 0 \iff \exists x_0 \in \mathbb{R} : P(X = x_0) = 1$.
On dit alors que X est presque sûrement constante.

Proposition 4.22
Soient X une variable aléatoire réelle ; soient a, b deux réels. Alors, $V(aX + b) = a^2 V(X)$.

Proposition 4.23 (Formule de König-Huygens)

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

Exercice 4.24
Calculer la variance de X, si X suit une des lois suivantes :

1. Loi uniforme sur $[\![a, b]\!]$, avec $a < b$ deux entiers relatifs
2. Loi de Bernoulli de paramètre p
3. Loi binomiale de paramètres n et p

Théorème 4.25 (Variance d'une somme de variables aléatoires indépendantes)
Si X_1 et X_2 sont deux variables aléatoires réelles indépendantes, alors

$$V(X_1 + X_2) = V(X_1) + V(X_2).$$

Théorème 4.26 (Inégalité de Bienaymé-Tchebychev)
Soit X une variable aléatoire réelle, soit α un réel strictement positif. On note $\mu = E(X)$. Alors, $P\big(|X - \mu| \geq \alpha\big) \leq \dfrac{V(X)}{\alpha^2}$.

Remarque 4.27
L'inégalité de Bienaymé-Tchebychev quantifie la probabilité qu'une variable aléatoire prenne des valeurs éloignées de son espérance. Autrement dit, elle indique à quel point les valeurs de la variable aléatoire sont concentrées autour de son espérance.
On parle d'inégalité de concentration.

4.3 Covariance

Définition 4.28 (Covariance de deux variables aléatoires réelles)
La covariance de deux variables aléatoires réelles X et Y, notée $\mathrm{Cov}(X, Y)$, est définie par

$$\mathrm{Cov}(X, Y) = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big).$$

Remarques 4.29

- La covariance de X et Y est l'espérance du produit des variables aléatoires centrées $X - E(X)$ et $Y - E(Y)$.
- La covariance de X et X est la variance de X.

Proposition 4.30 (Formule de König-Huygens)

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Définition 4.31 (Variables aléatoires non corrélées)
Si la covariance de X et Y est nulle, on dit que X et Y sont non corrélées.

Remarques 4.32

- Les notions d'indépendance et de non-corrélation étant à peu près synonymes dans le langage courant, il serait sans doute plus approprié de parler de non corrélation linéaire.
- L'indépendance implique la non corrélation mais la réciproque est fausse.

Proposition 4.33
La covariance est une forme bilinéaire symétrique positive sur l'espace vectoriel des variables aléatoires réelles sur un univers probabilisé fini (Ω, P) :

Pour toutes variables aléatoires réelles X, X', Y, Y' et pour tous réels λ, μ :

- $\mathrm{Cov}(X, X) \geq 0$;
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$;
- $\mathrm{Cov}(\lambda X + \mu X', Y) = \lambda\, \mathrm{Cov}(X, Y) + \mu\, \mathrm{Cov}(X', Y)$;
- $\mathrm{Cov}(X, \lambda Y + \mu Y') = \lambda\, \mathrm{Cov}(X, Y) + \mu\, \mathrm{Cov}(X, Y')$.

Théorème 4.34 (Variance d'une somme - cas de deux variables)
Soient X et Y deux variables aléatoires réelles. On a

$$V(X + Y) = V(X) + V(Y) + 2\, \mathrm{Cov}(X, Y).$$

Théorème 4.35 (Variance d'une somme - cas général)
Soient $X_1, \ldots, X_n$ des variables aléatoires réelles. On a

$$V\big(\sum_{i=1}^{n} X_i\big) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

Remarque 4.36
En particulier, si $X_1, \ldots, X_n$ sont des variables aléatoires mutuellement indépendantes (ou seulement deux à deux indépendantes), on a $V\big(\sum_{i=1}^{n} X_i\big) = \sum_{i=1}^{n} V(X_i)$. On retrouve ainsi la formule de la variance d'une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$.

Théorème 4.37 (Inégalité de Cauchy-Schwarz (2ème année))
Soient X et Y deux variables aléatoires réelles. On a $|\,\mathrm{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$, avec égalité ssi :

- *ou bien, $\sigma(X) = 0$;*
- *ou bien, il existe $a, b \in \mathbb{R}$ tels que $Y = a + bX$ presque sûrement (c'est-à-dire que $P(Y = a + bX) = 1$).*

Remarque 4.38
La covariance est donc maximale (en valeur absolue) quand il existe une relation affine entre les deux variable aléatoires. Il y a dans ce cas une forte dépendance entre les deux variables aléatoires.

Remarque 4.39
On peut en fait aussi écrire une inégalité de Cauchy-Schwarz portant sur $E(XY)$.
Si X et Y sont des variables aléatoires réelles, alors

$$E(XY)^2 \leq E(X^2)E(Y^2).$$

La démonstration repose sur les mêmes outils, en travaillant cette fois avec la forme bilinéaire symétrique positive $(X, Y) \mapsto E(XY)$.