UNIVERSITÉ PARIS DESCARTES
Licence Informatique 2e année

Notes du cours

Probablilités et Statistiques pour l'Informatique.

Version du 5 octobre 2015.

La version la plus récente de ces notes de cours se trouve sur la page
`http://w3.mi.parisdescartes.fr/~glaunes/Proba3/`

Table des matières

Chapitre 1

Probabilités, événements

1.1 Rappels sur les notations mathématiques

1.1.1 Ensembles

Un ensemble est un regroupement d'éléments. Par exemple, on note $\mathbf{N}$ l'ensemble de tous les entiers $0, 1, 2, \ldots$. Le symbole $\in$ permet de noter l'appartenance à un ensemble : par exemple, "$3 \in \mathbf{N}$" signifie "3 est un élément de $\mathbf{N}$", ou encore "3 appartient à $\mathbf{N}$".

Pour décrire un ensemble à partir de ses éléments, on utilise les accolades. Par exemple,
•$A = \{1, 5, 3\}$: A est l'ensemble formé par les nombres 1, 3, et 5. Bien noter que dans cette notation **l'ordre ne compte pas** : $\{1, 5, 3\} = \{1, 3, 5\} = \{3, 1, 5\}$.
•$\mathbf{N} = \{0, 1, 2, 3, \ldots\}$: $\mathbf{N}$ est l'ensemble des nombres 0, 1, 2, etc.
•$B = \{0, 1, 4, 9, \ldots\}$: $\mathbf{N}$ est l'ensemble des carrés des entiers.

Souvent on utilise des expressions logiques entre les accolades pour décrire un ensemble plus précisément. Par exemple, pour le dernier ensemble B, on écrira plutôt :

$$B = \{n^2, n \in \mathbf{N}\},$$

ce qui se lit : "B est l'ensemble des n^2 **tels que** n appartient à $\mathbf{N}$".
Ou encore :

$$B = \{n \in \mathbf{N}, n = k^2, k \in \mathbf{N}\},$$

c'est-à-dire : "B est l'ensemble des entiers n **tels que** $n = k^2$, avec k n'importe quel entier". Dans une telle notation la première virgule fait office de "tel que", tandis que les suivantes se lisent "et" ou "avec". A noter aussi que les lettres utilisées (n, k) dans les expressions sont complètement interchangeables. On aurait pu aussi bien écrire

$$B = \{k \in \mathbf{N}, k = n^2, n \in \mathbf{N}\},$$

ou encore

$$B = \{i \in \mathbf{N}, i = j^2, j \in \mathbf{N}\}.$$

Autre exemple : $C = \{x \in \mathbf{R}, x \leq 3, x \geq 2\}$: ensemble des nombres réels supérieurs à 2 et inférieurs à 3 : c'est l'intervalle $[2, 3]$.

L'**ensemble vide** ne contient aucun élément. Il est noté $\varnothing$. On a par exemple $\{x \in \mathbf{R}, x \leq 2, x \geq 3\} = \varnothing$.

Le **cardinal** d'un ensemble est le nombre de ses éléments. Il est noté "Card" ou "$\sharp$". Par exemple,
- $\mathrm{Card}\{1, 5, 3\} = \sharp\{1, 5, 3\} = 3$,
- $\sharp(\varnothing) = 0$.

Lorsqu'un ensemble n'a qu'un seul élément, comme par exemple l'ensemble $\{4\}$, on l'appelle **singleton**. Attention, il est essentiel de distinguer éléments et ensembles : on écrit $4 \in \{4\}$ (le nombre 4 appartient à l'ensemble $\{4\}$) mais surtout pas $\{4\} = 4$!

1.1.2 Ensembles finis, dénombrables, non dénombrables

Un ensemble est dit **fini** lorsqu'il possède un nombre fini d'éléments. Par exemple $\{1, 5, 3\}$ (3 éléments), ou encore, si n est n'importe quel entier, $\{1, 2, \ldots, n\}$ (n éléments). Sinon c'est un ensemble infini.

Un ensemble est dit **dénombrable** s'il est infini mais que l'on peut énumérer ses éléments, c'est-à-dire attribuer un numéro unique à chacun de ses éléments, ou encore en termes mathématiques : trouver une bijection de $\mathbf{N}$ vers cet ensemble. Par exemple,
- $\mathbf{N}$ est dénombrable (puisque $\mathbf{N} = \{0, 1, 2, 3, 4, \ldots\}$)
- $\mathbf{Z}$ est dénombrable ($\mathbf{Z} = \{0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots\}$.)
- $\mathbf{Q}$ est dénombrable (mais la numérotation n'est pas évidente).

Enfin, quelques exemples d'ensembles **infinis non dénombrables** : $\mathbf{R}$, $[0, 1]$, $\mathbf{R} \setminus \mathbf{Q}$.

1.1.3 Sous-ensembles

On dit qu'un ensemble E est une **partie** d'un autre ensemble F lorsque tous les éléments de E appartiennent aussi à F. On dit encore que E est un **sous-ensemble** de F ou que E est **inclus** dans F, et on note $E \subset F$. Par exemple, $\{3, 5\} \subset \{1, 5, 3\}$, ou bien $\{5\} \subset \{1, 5, 3\}$. (Mais surtout pas $5 \subset \{1, 5, 3\}$: 5 est un élément, pas un ensemble).
Par convention, $\varnothing$ est toujours sous-ensemble de n'importe quel ensemble.

L'**union** de E et F est l'ensemble des éléments appartenant à E **ou** à F, noté $E \cup F$. On a ainsi $E \cup F = \{x, x \in E \text{ ou } x \in F\}$.

L'**intersection** de E et F est l'ensemble des éléments appartenant à E **et** à F, noté $E \cap F$. On a ainsi $E \cap F = \{x, x \in E, x \in F\}$.

Si l'on ne considère que des sous-ensembles d'un même ensemble Ω (ce sera le cas par la suite), on peut parler d'ensemble **complémentaire** d'un sous-ensemble E de Ω : il regroupe tous les éléments de Ω ne faisant pas partie de E, et on le note E^c. Par exemple si $\Omega = \mathbf{R}$, $[2,4]^c =]-\infty, 2[\cup]4, \infty[$.

Quelques propriétés : Commutativité :

$$E \cup F = F \cup E \qquad \text{et} \qquad E \cap F = F \cap E.$$

Associativité :

$$(E \cup F) \cup G = E \cup (F \cup G) \qquad \text{et} \qquad (E \cap F) \cap G = E \cap (F \cap G).$$

Distributivité :

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G) \qquad \text{et} \qquad (E \cap F) \cup G = (E \cup G) \cap (F \cup G).$$

Passage au complémentaire :

$$(E \cup F)^c = E^c \cap F^c \qquad \text{et} \qquad (E \cap F)^c = E^c \cup F^c.$$

1.1.4 Ensemble des parties d'un ensemble

On peut regrouper tous les sous-ensembles possibles d'un ensemble E : c'est l'ensemble des parties de E, noté $\mathcal{P}(E)$. Par exemple,

$$\mathcal{P}(\{1,5,3\}) = \{\varnothing, \{1\}, \{3\}, \{5\}, \{1,3\}, \{5,3\}, \{1,5\}, \{1,3,5\}\}.$$

Attention, les sous-ensembles de E sont des éléments de $\mathcal{P}(E)$: écrire $F \subset E$ est équivalent à écrire $F \in \mathcal{P}(E)$.

1.1.5 Couples, triplets, n-uplets et suites

Pour désigner des suites ordonnées d'éléments, on utilise les parenthèses :

- $(1,5,3)$ est le **triplet** formé des nombres 1, 5 puis 3. Bien noter que dans ce cas l'ordre compte : $(1,5,3) \neq (1,3,5)$.
- $(2.45, 3)$ est le **couple** formé du nombre réel 2.45 et de l'entier 3.
- $(1,3,5,\ldots,19,21)$ est le 10-uplet des 10 premiers entiers impairs.

Pour désigner les ensembles de couples, de triplets, etc, on utilise la notation $\times$ (prononcer "croix") :

- $\{1,3,5\} \times \{1,3,5\}$ (noté aussi $\{1,3,5\}^2$) est l'ensemble des couples d'éléments de $\{1,3,5\}$. On a ainsi

  $$\{1,3,5\}^2 = \{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\}$$

  On peut remarquer que $Card(\{1,3,5\}^2) = 9 = 3^2 = (Card\{1,3,5\})^2$.

- $\mathbf{N} \times \mathbf{N}$ est l'ensemble des couples de nombres entiers, ce qui peut s'écrire $\mathbf{N} \times \mathbf{N} = \{(a, b), a \in \mathbf{N}, b \in \mathbf{N}\}$. On écrit aussi $\mathbf{N} \times \mathbf{N} = \mathbf{N}^2$.
- $\mathbf{N} \times \mathbf{N} \times \mathbf{N} = \mathbf{N}^3$: ensemble des triplets d'entiers.
- $\mathbf{N}^{10}$: ensemble des 10-uplets d'entiers. Par exemple $(1, 3, 5, \ldots, 19, 21) \in \mathbf{N}^{10}$.
- $\mathbf{R} \times \mathbf{N} = \{(x, n), x \in \mathbf{R}, n \in \mathbf{N}\}$. Par exemple, $(2.45, 3) \in \mathbf{R} \times \mathbf{N}$.
- $\mathbf{R}^n = \{(x_1, x_2, \ldots, x_n), x_1 \in \mathbf{R}, \ldots, x_n \in \mathbf{R}\}$: n-uplets de nombres réels (si n est un entier fixé).

Lorsque la suite a une infinité d'éléments, on parle de **suite** tout court, et on note avec des indices de position : par exemple $(1, 1/2, 1/3, 1/4, \ldots)$ est la suite $(u_n)_{n \in \mathbf{N}}$ avec $u_n = 1/n$. On note $\mathbf{N}^{\mathbf{N}}$ l'ensemble des suites de nombres entiers, $\mathbf{R}^{\mathbf{N}}$ l'ensemble des suites de nombres réels, etc.

1.2 Combinatoire

1.2.1 Cardinal de l'ensemble des couples ou des n-uplets

Si $\mathrm{Card}(A) < \infty$ et $\mathrm{Card}(B) < \infty$,

$$\mathrm{Card}(A \times B) = \mathrm{Card}(A)\mathrm{Card}(B).$$

Plus généralement si $\mathrm{Card}(A_1) < \infty, \mathrm{Card}(A_2) < \infty, \ldots, \mathrm{Card}(A_n) < \infty$, alors

$$\mathrm{Card}(A_1 \times A_2 \times \cdots \times A_n) = \mathrm{Card}(A_1)\mathrm{Card}(A_2) \cdots \mathrm{Card}(A_n).$$

En particulier si $\mathrm{Card}(A) < \infty$,

$$\mathrm{Card}(A^n) = (\mathrm{Card}(A))^n.$$

1.2.2 Cardinal de l'ensemble des parties

Si $\mathrm{Card}(A) < \infty$,

$$\mathrm{Card}(\mathcal{P}(A)) = 2^{\mathrm{Card}(A)}.$$

1.2.3 Ensemble des combinaisons de p éléments parmi n

Soit A un ensemble contenant n éléments, et p un entier. On note $\mathcal{P}_p(A)$ l'ensemble des sous-ensembles de A contenant p éléments. Par exemple si $A = \{1, 3, 5\}$,

$$\begin{aligned}
\mathcal{P}_0(A) &= \{\varnothing\}, \\
\mathcal{P}_1(A) &= \{\{1\}, \{3\}, \{5\}\}, \\
\mathcal{P}_2(A) &= \{\{1, 3\}, \{3, 5\}, \{1, 5\}\}, \\
\mathcal{P}_3(A) &= \{\{1, 3, 5\}\}.
\end{aligned}$$

On remarque ici que $\mathrm{Card}(\mathcal{P}_0(A)) = 1$, $\mathrm{Card}(\mathcal{P}_1(A)) = 3$, $\mathrm{Card}(\mathcal{P}_2(A)) = 3$, et $\mathrm{Card}(\mathcal{P}_3(A)) = 1$. De manière générale le cardinal de $\mathcal{P}_p(A)$ lorsque A contient n éléments est noté $\binom{n}{p}$ et vaut

$$\binom{n}{p} = \frac{n!}{(n-p)!p!}.$$

Rappelons quelques formules classiques sur ce nombre :

$$\binom{n}{p} = \binom{n}{n-p}, \qquad \binom{n}{1} = n, \qquad \binom{n}{0} = \binom{n}{n} = 1,$$

$$\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}.$$

1.2.4 Ensemble des permutations d'un ensemble à n éléments

Une permutation d'un ensemble A à n éléments est un n-uplet formé des éléments de A sans répétition. L'ensemble de ces permutations est noté $\mathcal{S}(A)$. Autrement dit on forme $\mathcal{S}(A)$ en considérant tous les ordres possibles pour les éléments de A. Par exemple,

$$\mathcal{S}(\{1,3,5\}) = \{(1,3,5),(1,5,3),(3,1,5),(3,5,1),(5,1,3),(5,3,1)\}.$$

Le cardinal de A, c'est-à-dire le nombre de permutations de n éléments, est égal à

$$\mathrm{Card}(\mathcal{S}(A)) = n!.$$

1.2.5 Ensemble des arrangements de p éléments parmi n

Soit toujours A un ensemble à n éléments, et p un entier. On note $\mathcal{S}_p(A)$ l'ensemble des p-uplets formés d'éléments de A, sans répétition. Par exemple si $A = \{1,3,5\}$,

$$\begin{aligned}
\mathcal{S}_1(A) &= \{1,3,5\},\\
\mathcal{S}_2(A) &= \{(1,3),(3,1),(3,5),(5,3),(1,5),(5,1)\},\\
\mathcal{S}_3(A) &= \{(1,3,5),(1,5,3),(3,1,5),(3,5,1),(5,1,3),(5,3,1)\} = \mathcal{S}(A).
\end{aligned}$$

On voit que pour construire $S_p(A)$, on forme, à partir de chaque élément de $\mathcal{P}_p(A)$, les p-uplets obtenus en permutant ses éléments. Comme il y a $p!$ telles permutations, le nombre total d'éléments de $S_p(A)$ est égal à

$$\mathrm{Card}(\mathcal{S}_p(A)) = p!\,\mathrm{Card}(\mathcal{P}_p(A)) = p!\binom{n}{p} = \frac{n!}{(n-p)!}.$$

1.3 Événements et probabilité

1.3.1 Exemple introductif

On lance un dé. Quelle est la probabilité que le résultat soit un nombre pair supérieur ou égal à trois ?

Ce problème est très simple et peut se résoudre de tête : le dé doit tomber sur 4 ou 6 pour que le résultat soit à la fois pair et supérieur à 3 ; on peut donc répondre directement qu'il y a 2 chance sur 6 pour que ça arrive.

Nous allons voir comment introduire des notations mathématiques rigoureuses générales permettant de résoudre des problèmes tels que celui-ci, ou d'autres plus complexes pour lesquels l'intuition ne suffit pas. Cependant pour simplifier la compréhension, nous allons suivre ce premier exemple.

Il s'agit de trouver la probabilité qu'un certain **événement** se produise, en l'occurrence l'événement

$$A = \text{"le résultat du dé est un nombre pair supérieur ou égal à 3"}.$$

On note cette probabilité $P(A)$. Par convention la **probabilité** d'un événement est un nombre réel compris entre 0 et 1 (c'est-à-dire appartenant à l'intervalle $[0, 1]$), 0 signifiant que l'événement n'a aucune chance de se réaliser, et 1 qu'il est certain de se réaliser.

1.3.2 Modèle fondamental des probabilités

L'idée fondamentale de la modélisation mathématique des probabilités consiste à assimiler les événements à des sous-ensembles d'un grand ensemble Ω appelé l'**univers**. L'événement Ω correspond à tout événement qui se réalise forcément. Par exemple l'événement "Le dé tombe sur une face" est égal à Ω. On aura donc naturellement

$$P(\Omega) = 1.$$

Inversement, l'ensemble vide $\varnothing$ correspond à tout événement qui ne peut pas se réaliser, comme par exemple "le résultat du dé est un nombre à la fois pair et impair". On aura donc

$$P(\varnothing) = 0.$$

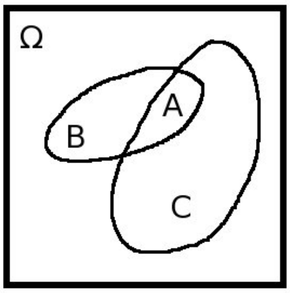
1.3.3 Intersections d'événements

L'événement A se décompose naturellement de la manière suivante :

$$A = \text{"le résultat du dé est un nombre pair" et "le résultat du dé est supérieur ou égal à 3"},$$

ou encore, si l'on a défini $B =$ "le résultat du dé est un nombre pair" puis $C =$ "le résultat du dé est supérieur ou égal à 3", on écrira simplement

$$A = B \text{ et } C.$$



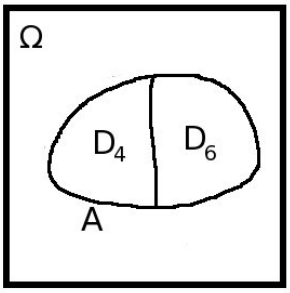


FIGURE 1.1 – L'événement A représenté comme intersection des événements B et C puis comme union des événements D_4 et D_6.

En termes de sous-ensembles de Ω, le "et" de l'égalité précédente se traduit par une intersection. Plutôt que $A = B$ et C, on écrira en fait :

$$A = B \cap C.$$

Malheureusement cette décomposition ne permet pas directement de calculer la probabilité de l'événement A car il n'existe pas de formule générale reliant cette probabilité à celles de B et de C.

1.3.4 Unions d'événements et probabilité d'une union

Un dé a six faces, numérotées de 1 à 6, et par conséquent pour que le résultat soit pair, il faut qu'il tombe sur 2, 4, ou 6. Donc pour qu'il soit pair et supérieur à 3, il faut qu'il tombe sur 4 ou 6. On peut donc formuler différemment l'événement A :

$$A = \text{"le résultat du dé est 4 ou 6"}$$

ou encore :

$$A = \text{"le résultat du dé est 4" ou "le résultat du dé est 6"}$$

Par conséquent A est composé de deux événements élémentaires, que l'on peut noter D_4, D_6, avec $D_i =$ "le résultat du dé est i". On a donc

$$A = D_4 \text{ ou } D_6.$$

Les "ou" dans l'égalité précédente se traduisent alors par des unions de sous-ensembles. On écrira donc en fait :

$$A = D_4 \cup D_6.$$

Pour calculer la probabilité d'une union d'événements on dispose des formules suivantes :

Probabilité de l'union de deux événements - formule générale. *Soient E et F deux événements. Alors*

$$\boxed{P(E \cup F) = P(E) + P(F) - P(E \cap F).}$$

Probabilité d'une union disjointe.

- *Soient E et F deux événements. Si E et F sont* **incompatibles** *(c'est-à-dire disjoints : $E \cap F = \varnothing$) alors*

$$P(E \cup F) = P(E) + P(F).$$

- *Soient $E_1, \ldots, E_n$ plusieurs événements. S'ils sont* **deux à deux disjoints** *($E_i \cap E_j = \varnothing$ pour tous $i \neq j$) alors*

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n).$$

Dans notre exemple les événements D_4 et D_6 sont disjoints. En effet l'événement $D_4 \cap D_6$ correspond à : "Le dé tombe sur 4 et sur 6", ce qui est impossible. Ainsi $D_4 \cap D_6 = \varnothing$. On peut donc écrire

$$P(A) = P(D_4) + P(D_6).$$

Finalement on sait que la probabilité de tomber sur chaque face est de $\frac{1}{6}$. Ainsi $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Interprétation graphique : Les formules précédentes se retrouvent facilement en dessinant un diagramme tel que celui de la figure 1.1, et en imaginant que les probabilités des événements sont les aires des sous-ensembles correspondants (en supposant que Ω est un carré de côté 1 par exemple).

Voici une autre relation importante découlant des formules précédentes :

$$\boxed{\textit{Si } E \subset F \textit{ alors } P(E) \leq P(F).}$$

Encore une fois, ceci se comprend très facilement sur un schéma (cf. figure 1.2)

1.3.5 Partition de Ω

Des événements $E_1, \ldots, E_n$ forment une **partition** de l'univers Ω s'ils sont deux à deux incompatibles et que leur réunion est égale à Ω. Autrement dit : un et un seul de ces événements se réalise. Alors on a la relation :

$$P(E_1) + \cdots + P(E_n) = 1.$$

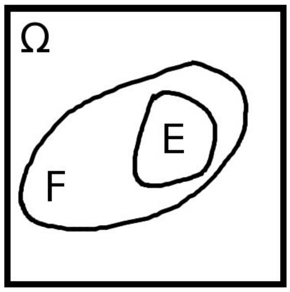


FIGURE 1.2 – Si $E \subset F$ alors $P(E) \leq P(F)$.

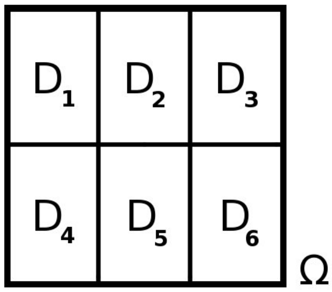


FIGURE 1.3 – Les événements D_1 à D_6 forment une partition de Ω.

Cette relation s'obtient en appliquant simplement la formule de probabilité d'une union disjointe, en se rappelant que $P(\Omega) = 1$:

$$1 = P(\Omega) = P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + \cdots + P(E_n).$$

Dans l'exemple du lancer de dé, les événements $D_1, \ldots, D_6$ forment une partition de Ω. Puisque chacun de ces événements a pour probabilité $\frac{1}{6}$, on vérifie bien la formule : $\frac{1}{6} + \cdots + \frac{1}{6} = 6 \times \frac{1}{6} = 1$ (voir figure 1.3). Dans ce cas on voit que non seulement ces ensembles forment une partition de Ω, mais qu'aussi ils ont tous la même probabilité. C'est ce qu'on appelle l'équiprobabilité :

Définition. *On dit qu'il y a* **équiprobabilité** *entre des événements $E_1, \ldots, E_n$ s'ils forment une partition de Ω et que leurs probabilités sont toutes égales : $P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$.*

Dans l'exemple du lancer de dé, les 6 événements $D_1, D_2, \ldots, D_6$ sont ainsi équiprobables. Par contre si on considère un dé truqué (certaines faces ont plus de chances d'apparaître que d'autres), alors ces événements $D_1, D_2, \ldots, D_6$ formeront toujours une partition mais ne seront plus équiprobables. Supposons par exemple que le dé soit truqué de sorte que la probabilité de chaque face soit proportionnelle au résultat. Ceci signifie que le dé a

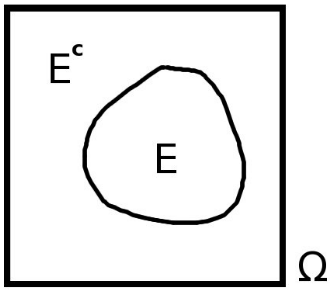


FIGURE 1.4 – E et E^c forment une partition de Ω.

deux fois plus de chances de tomber sur 2 que sur 1 ; trois fois plus de chances de tomber sur 3 que sur 1, etc. On peut alors calculer ces probabilités en écrivant que la somme de ces probabilités doit faire 1 :

$$P(D_1) + P(D_2) + \cdots + P(D_6) = 1.$$

Or d'après ce qu'on vient de voir, $P(D_2) = 2P(D_1), P(D_3) = 3P(D_1)$, etc. Ainsi,

$$P(D_1) + 2P(D_1) + 3P(D_1) + \cdots + 6P(D_1) = 1,$$

$$P(D_1)(1 + 2 + 3 + \cdots + 6) = 21P(D_1) = 1,$$

et ainsi $P(D_1) = \frac{1}{21}, P(D_2) = \frac{2}{21}, \ldots, P(D_6) = \frac{6}{21}$.

1.3.6 Événement complémentaire

L'**événement complémentaire** d'un événement E est l'événement correspondant à l'ensemble complémentaire E^c. Il s'agit de la négation de l'événement E : E^c se réalise si et seulement si E ne se réalise pas. On a la formule

$$\boxed{P(E^c) = 1 - P(E),}$$

qui provient simplement du fait que E et E^c forment une partition de Ω.

La relation suivante est très utile :

Soient E et F deux événements. Alors

$$\boxed{P(E) = P(E \cap F) + P(E \cap F^c).}$$

On comprend facilement cette relation en dessinant des patates (cf. figure 1.5). La preuve mathématique est assez simple aussi ; la voici :

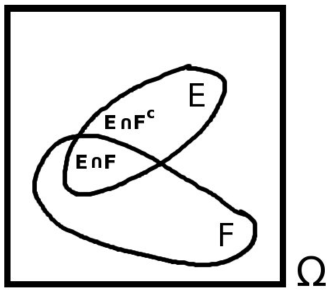


FIGURE 1.5 – $P(E) = P(E \cap F) + P(E \cap F^c)$.

Preuve. On va utiliser la formule de probabilité d'une union disjointe. On écrit d'abord A comme une union :

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

On vérifie que cette union est disjointe :

$$(A \cap B) \cap (A \cap B^c) = A \cap B \cap A \cap B^c = \varnothing$$

On applique donc la formule de probabilité d'une union disjointe :

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

□

Chapitre 2

Probabilités conditionnelles et indépendance d'événements

2.1 Probabilités conditionnelles

On tire deux cartes successivement et sans remise d'un jeu de 32 cartes. Quelle est la probabilité que ce soient deux piques ?

La réponse "intuitive" est la suivante : il y a 8 chances sur 32 de tirer un pique pour la première carte, puis 7 chances sur 31 pour la deuxième (puisqu'il reste 31 cartes dont 7 piques), donc la probabilité demandée est égale à $\frac{8}{32} \times \frac{7}{31}$.

La notion de probabilité conditionnelle permet de formuler rigoureusement une telle réponse.

Définition (Probabilité conditionnelle)**.** *Soient E et F deux événements. On note $P(E|F)$ (ce qui se lit "probabilité de E sachant F") la probabilité que l'événement E se réalise sachant que l'événement F est réalisé.*

Autrement dit, à partir d'une expérience probabiliste où E et F sont aléatoires, on change de situation : on suppose que F n'est plus aléatoire mais réalisé, et on cherche à calculer la probabilité de E dans cette nouvelle situation. Sur un schéma ensembliste (figure 2.1), ce changement de situation probabiliste consiste à considérer l'ensemble F comme le nouvel univers, à la place de Ω, et donc à ne plus considérer ce qui est en dehors de F. Autrement dit, la probabilité d'un événement E sera obtenue en ne retenant que la probabilité de sa partie contenue dans le nouvel univers, c'est-à-dire l'intersection $E \cap F$, et en divisant par $P(F)$, afin que dans la nouvelle situation, la probabilité de l'univers F soit égale à 1.

FIGURE 2.1 – Schéma du changement de situation probabiliste correspondant au conditionnement par rapport à un événement F. A gauche, situation initiale dans l'univers Ω ; à droite, situation conditionnée à F : l'événement F devient le nouvel univers, et seule la partie de E comprise dans F est considérée.

Ceci permet de justifier la formule fondamentale suivante, qui est en fait la définition mathématique de la probabilité conditionnelle :

$$\boxed{P(E|F) = \frac{P(E \cap F)}{P(F)}.}$$

On voit par cette formule que $P(E|F)$ n'est défini que si $P(F) \neq 0$; ce qui est cohérent : on ne peut pas supposer que F est réalisé s'il n'a aucune chance de se réaliser.

remarque : Attention, $E|F$ ne signifie rien en soit, ce n'est pas un événement. Seule l'expression $P(E|F)$ a un sens. En fait la probabilité conditionnelle $P(E|F)$ se note parfois $P_F(E)$, notation moins usuelle mais bien plus juste car elle traduit précisément le changement de situation probabiliste évoqué plus haut : la fonction P_F est la fonction de probabilité de la nouvelle situation, en remplacement de la fonction P initiale.

Voici maintenant comment exprimer rigoureusement la réponse à l'exemple précédent : soient A l'événement "La première carte est un pique", B l'événement "La deuxième carte est un pique", et C l'événement "Les deux cartes tirées sont des piques". On a donc $C = A \cap B$. D'après la formule des probabilités conditionnelles, on a

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

La probabilité de tirer un pique en premier est de $P(A) = \frac{8}{32}$. Maintenant si A est réalisé, alors il reste 31 cartes dont 7 piques, et donc la probabilité de tirer à nouveau un pique est de $\frac{7}{31}$: ainsi $P(B|A) = \frac{7}{31}$. Finalement on obtient

$$P(C) = P(B|A)P(A) = \frac{7}{31} \times \frac{8}{32} = \frac{7}{124}.$$

remarque : Toujours penser à vérifier que la probabilité calculée est bien un nombre compris entre 0 et 1.

2.2 Formule des probabilités totales

On dispose de deux urnes. La première contient deux boules noires et une boule blanche ; et la deuxième contient trois boules noires et deux boules blanches. On tire au hasard une boule de la première urne et on la place dans la deuxième urne. Puis on tire au hasard une boule de la deuxième urne. Quelle est la probabilité que cette deuxième boule soit blanche ?

On note B_1 l'événement "la première boule tirée est blanche", et B_2 l'événement "la deuxième boule tirée est blanche". On cherche donc à calculer $P(B_2)$. La formule suivante, vue au premier chapitre, est le point de départ du calcul :

$$P(B_2) = P(B_2 \cap B_1) + P(B_2 \cap B_1^c).$$

Autrement dit on décompose l'événement "la deuxième boule tirée est blanche" en deux : "la deuxième boule tirée est blanche et la première était blanche", et "la deuxième boule tirée est blanche et la première n'était pas blanche (donc était noire)". On utilise alors la définition mathématique des probabilités conditionnelles pour calculer $P(B_2 \cap B_1)$ et $P(B_2 \cap B_1^c)$, ce qui donne :

$$P(B_2) = P(B_2|B_1)P(B_1) + P(B_2|B_1^c)P(B_1^c).$$

C'est la formule des probabilités totales. A présent toutes les probabilités à droite de l'égalité peuvent se calculer :

- Pour déterminer $P(B_2|B_1)$ on se place dans la situation où B_1 est réalisé, c'est-à-dire que la première boule tirée est blanche. Dans ce cas la deuxième urne contiendra trois boules blanches et trois boules noires, et donc la probabilité de tirer une boule blanche sera de $\frac{3}{6} = \frac{1}{2}$. Ainsi on a montré que $P(B_2|B_1) = \frac{1}{2}$.
- Pour déterminer $P(B_2|B_1^c)$ on se place dans la situation inverse : la première boule tirée est noire, donc la deuxième urne contient deux boules blanches et quatre boules noires. Ainsi $P(B_2|B_1^c) = \frac{2}{6} = \frac{1}{3}$.
- $P(B_1)$ est la probabilité que la première boule tirée soit blanche : $P(B_1) = \frac{1}{3}$.
- Enfin $P(B_1^c) = 1 - P(B_1) = \frac{2}{3}$.

Finalement on trouve donc

$$P(B_2) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{7}{18}.$$

On a démontré et utilisé dans ce calcul la formule des probabilités totales :

Formule des probabilités totales, cas simple. *Soient E et F deux événements, avec $P(F)$ et $P(F^c)$ non nuls. Alors*

$$\boxed{P(E) = P(E|F)P(F) + P(E|F^c)P(F^c).}$$

La version générale de la formule consiste à conditionner non plus par seulement deux événements (F et F^c) pour le calcul de $P(E)$, mais par un nombre n :

Formule des probabilités totales, cas général. *Soient E un événement et $F_1, F_2, \ldots, F_n$ des événements formant une partition de Ω, avec $P(F_i) \neq 0$ pour tout i. Alors*

$$\boxed{P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n).}$$

En voici un exemple d'utilisation (cas $n = 3$) :

Pour traiter une maladie, les médecins disposent de trois nouveaux médicaments MA, MB, MC. Dans un premier cas, les médecins prescrivent indifféremment l'un des trois médicaments pour chaque traitement. Dans un deuxième cas, ils commencent à connaître mieux ces médicaments, et prescrivent MA dans 50% des cas, MB dans 30% des cas, et MC dans 20% des cas. En fait les taux de réussite de ces médicaments sont respectivement de 98%, 96% et 95%. Calculer la probabilité d'échec du traitement dans chaque cas.

On va noter A l'événement "recevoir le médicament MA", et de même B et C. On cherche la probabilité de E="échec du traitement". Il faut bien comprendre les données du problème : les taux de réussite correspondent en fait à des probabilités conditionnelles : "si on reçoit le médicament MA, alors on guérit dans 98% des cas". On a donc $P(E^c|A) = 0.98$, et de même $P(E^c|B) = 0.96$ et $P(E^c|C) = 0.95$. On va donc utiliser la formule des probabilités totales en remarquant que A, B et C forment une partition de Ω (car un et un seul des trois médicaments est administré lors d'un traitement) :

$$P(E^c) = P(E^c|A)P(A) + P(E^c|B)P(B) + P(E^c|C)P(C).$$

<u>1er cas</u> : $P(A) = P(B) = P(C) = \frac{1}{3}$. Donc $P(E^c) = 0.98 \times \frac{1}{3} + 0.96 \times \frac{1}{3} + 0.95 \times \frac{1}{3} = 0.963$. Finalement $P(E) = 1 - P(E^c) = 0.037$.
<u>2e cas</u> : $P(A) = 0.5$, $P(B) = 0.3$ et $P(C) = 0.2$. Donc $P(E^c) = 0.98 \times 0.5 + 0.96 \times 0.3 + 0.95 \times 0.2 = 0.968$, et $P(E) = 0.032$.

2.3 Formule de Bayes

Revenons au problème des deux urnes, et cherchons à répondre à cette deuxième question :

Si la deuxième boule tirée est blanche, quelle est la probabilité que la première boule tirée ait été blanche aussi ?

Ici on demande de calculer $P(B_1|B_2)$. Cette probabilité conditionnelle ne peut pas être trouvée directement comme c'est le cas pour $P(B_2|B_1)$. En effet ici on cherche la probabilité d'un événement en conditionnant par rapport à un événement qui en découle, ce qui va contre l'intuition. Il faut donc faire une manipulation pour "retourner" la probabilité conditionnelle :

$$P(B_1|B_2) = \frac{P(B_1 \cap B_2)}{P(B_2)} = \frac{P(B_2|B_1)P(B_1)}{P(B_2)}.$$

A présent on peut effectuer le calcul :

$$P(B_1|B_2) = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{7}{18}} = \frac{3}{7}.$$

Nous avons donc utilisé la formule suivante, qui est la base de la formule de Bayes :

Soient E et F deux événements avec $P(E)$ et $P(F)$ non nuls. Alors

$$\boxed{P(F|E) = \frac{P(E|F)P(F)}{P(E)}}.$$

La formule de Bayes est simplement la combinaison de cette formule et de la formule des probabilités totales pour le calcul de $P(E)$:

Formule de Bayes, cas simple. *Soient E et F deux événements avec $P(E)$, $P(F)$ et $P(F^c)$ non nuls. Alors*

$$\boxed{P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}}.$$

Dans l'exercice des deux urnes, cette formule aurait permis de répondre à la deuxième question directement. A présent voici la version générale de la formule :

Formule de Bayes, cas général. *Soient E et F deux événements, et $F_1, F_2, \ldots, F_n$ des événements formant une partition de Ω, avec $P(E)$, $P(F)$ et $P(F_i)$ non nuls pour tout i. Alors*

$$\boxed{P(F|E) = \frac{P(E|F)P(F)}{P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n)}}.$$

remarque : Le cas simple correspond au cas $n = 2, F_1 = F$ et $F_2 = F^c$.

2.4 Indépendance

2.4.1 Indépendance de deux événements

La définition intuitive de l'indépendance est la suivante : deux événements sont indépendants lorsque le résultat de l'un n'influence pas le résultat de l'autre. Autrement dit si E et F sont ces deux événements, le fait de supposer que F est réalisé ne change pas la probabilité de réalisation de E et inversement. On a donc :

$$P(E|F) = P(E) \quad \text{et} \quad P(F|E) = P(F).$$

Or $P(E|F) = \frac{P(E \cap F)}{P(F)}$ et $P(F|E) = \frac{P(E \cap F)}{P(F)}$, et par conséquent on voit que les deux conditions ci-dessus se résument en une seule : $P(E \cap F) = P(E)P(F)$. C'est la définition mathématique de l'indépendance :

Définition. *Deux événements E et F sont* **indépendants** *si et seulement si*

$$P(E \cap F) = P(E)P(F).$$

remarque : Cette définition a l'avantage de ne pas supposer $P(E)$ et $P(F)$ non nuls, à la différence de la double condition précédente. En fait si $P(E) = 0$ ou si $P(F) = 0$, la formule est toujours vérifiée. Autrement dit : un événement de probabilité nulle est toujours indépendant de tout autre événement.

En pratique c'est souvent l'intuition qui permet de décider si deux événements sont indépendants ou pas. La formule $P(E \cap F) = P(E)P(F)$ est alors utilisée pour faire des calculs.

exemple 1 : *On lance une pièce de monnaie deux fois de suite sur une table. Quelle est la probabilité d'obtenir deux fois face ?*

On note $A =$"on obtient face au premier lancer" et $B =$"on obtient face au deuxième lancer". Ici ces deux événements sont clairement indépendants : le fait d'obtenir face en premier ne change pas la probabilité d'obtenir face au deuxième. Par conséquent la probabilité demandée ("obtenir deux fois face") sera : $P(A \cap B) = P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

exemple 2 : *A Paris il pleut en moyenne un jour sur deux en octobre. Quelle est la probabilité qu'il pleuve à la fois le* 15 *et le* 16 *octobre ?*

Ici on notera $A =$"il pleut le 15 octobre" et $B =$"il pleut le 16 octobre", et on cherche donc à calculer $P(A \cap B)$. Ici il est sans doute faux de penser que les deux événements A et B sont indépendants car en général beau et mauvais temps évoluent sur des échelles de temps plus grandes que la journée : autrement dit il y a plus de chances que le 16 il fasse un temps similaire à celui du 15 ; et donc s'il a plu le 15 il y aura une plus grande probabilité qu'il pleuve aussi le 16. Par conséquent même si on connait par hypothèse $P(A) = P(B) = \frac{1}{2}$, on ne peut pas en déduire $P(A \cap B)$ du fait de cette non-indépendance.

Il faudrait disposer d'autres informations pour répondre à la question.

exemple 3 : *On lance un dé. Quelle est la probabilité que le résultat soit un nombre pair supérieur ou égal à trois ?*

Il s'agit ici du premier exemple du chapitre 1. On avait noté B="le résultat est un nombre pair", C="le résultat est ≥ 3", et la probabilité cherchée est $P(A)$ avec $A = B \cap C$. En principe B et C n'ont aucune raison d'être indépendants puisqu'ils concernent le même lancer de dé. Pourtant on a $P(B) = \frac{1}{2}$, $P(C) = \frac{2}{3}$, et on avait trouvé $P(A) = P(B \cap C) = \frac{1}{3}$. On a donc $P(B \cap C) = P(B)P(C)$, et donc B et C sont mathématiquement indépendants. C'est une sorte d'indépendance fortuite, qui va contre l'intuition.

Modifions à présent légèrement l'énoncé : *On lance un dé. Quelle est la probabilité que le résultat soit un nombre pair supérieur ou égal à quatre ?*
Ici l'analyse est la même, à savoir que les événements B et D="le résultat est ≥ 4" n'ont pas de raison d'être indépendants ; et le calcul montre en effet que $P(B)P(D) = \frac{1}{2} \times \frac{1}{2} \neq P(B \cap D) = \frac{1}{3}$; c'est-à-dire que B et D ne sont pas indépendants.

Pour résumer, l'intuition peut nous dire si des événements sont indépendants, mais à l'inverse on ne peut jamais être certain que des événements ne sont pas indépendants sans faire le calcul.

2.4.2 Indépendance et événements complémentaires

Il est facile de montrer que

$$\begin{aligned} E \text{ et } F \text{ indépendants} &\Leftrightarrow E \text{ et } F^c \text{ indépendants,} \\ &\Leftrightarrow E^c \text{ et } F \text{ indépendants,} \\ &\Leftrightarrow E^c \text{ et } F^c \text{ indépendant.} \end{aligned}$$

Intuitivement, on est simplement en train de dire que le fait que E se réalise/ne se réalise pas est indépendant du fait que F se réalise/ne se réalise pas.

2.4.3 Indépendance de n événements

On généralise la notion d'indépendance pour plus de 2 événements. Voici d'abord la version à 3 événements :

Définition. *Trois événements E, F, G sont dits* **indépendants** *(on dit aussi* **mutuellement indépendants***) lorsque les quatre relations suivantes sont vérifiées :*

$$P(E \cap F) = P(E)P(F),$$

$$P(E \cap G) = P(E)P(G),$$

$$P(F \cap G) = P(F)P(G),$$
$$P(E \cap F \cap G) = P(E)P(F)P(G).$$

Pourquoi a-t-on besoin de toutes ces relations ? On pourrait penser que les deux premières ou les trois premières suffisent et entraînent les autres. Mais ceci est faux, comme le montre le contre-exemple suivant :

Exemple : *On tire deux fois un dé à six faces. Les événements suivants sont-ils indépendants ?*
$A =$ *"le premier dé tombe sur un nombre pair",*
$B =$ *"le deuxième dé tombe sur un nombre impair",*
$C =$ *"les deux dés ont même parité".*

Il est clair que A et B sont indépendants. A et C sont aussi indépendants : en effet, d'une part la probabilité de C vaut $P(C) = \frac{1}{2}$; d'autre part si l'on suppose que A est réalisé (le premier dé est pair), alors les deux dés auront même parité si le deuxième est aussi pair, donc avec probabilité $\frac{1}{2}$. Ainsi $P(C|A) = P(C) = \frac{1}{2}$ et donc A et C sont indépendants. Par le même raisonnement on peut voir que B et C sont aussi indépendants. Pour résumer on a les trois relations (indépendances deux-à-deux)

$$P(A \cap B) = P(A)P(B),$$

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

Cependant il est facile de voir que $P(A \cap B \cap C) = 0$ (les trois événements ne peuvent se réaliser ensemble) ; et donc que $P(A \cap B \cap C)$ est différent de $P(A)P(B)P(C)$. Ainsi les trois événements ne sont pas indépendants puisqu'il manque la dernière relation, alors qu'ils sont indépendants deux-à-deux.

Passons à présent à l'indépendance d'un nombre quelconque d'événements :

Définition. *Des événements* $E_1, E_2, E_3, \ldots$ *sont dits* **indépendants**, *ou* **mutuellement indépendants**, *si la réalisation d'un certain nombre d'entre-eux n'influence pas la réalisation des autres. Ceci équivaut à ce que toutes les relations suivantes soient vérifiées :*

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_p}) = P(E_{i_1}) \times P(E_{i_2}) \times \cdots \times P(E_{i_p}),$$

pour tous indices tels que $1 \leq i_1 < i_2 < \cdots < i_p$.

2.4.4 Expériences aléatoires indépendantes

Des expériences aléatoires $\mathcal{E}_1 \ldots, \mathcal{E}_n$ sont indépendantes lorsque le résultat de l'une d'entre-elles n'influe pas sur le résultat des autres. Ceci signifie qu'en choisissant pour chacune de ces expériences $\mathcal{E}_i$ n'importe quel événement E_i qui lui est relatif, on obtient des événements $E_1, \ldots, E_n$ indépendants.

En pratique c'est directement l'intuition (ou une hypothèse explicite) qui permet de décider si des expériences aléatoires sont indépendantes. On peut alors en déduire l'indépendance d'événements qui en découlent.

exemple *: On lance trois dés, et on note $A =$ "le 1er dé tombe sur 3 ", $B =$ "le deuxième dé tombe sur un nombre pair", et $C =$ "le 3e dé tombe sur un nombre impair". Calculer $P(A \cap B \cap C)$.*

Les trois lancers de dé sont ici clairement des expériences aléatoires indépendantes. On en déduit que A, B et C sont indépendants, puisque chacun de ces événements se réfère à un lancer différent. Ainsi $P(A \cap B \cap C) = P(A)P(B)P(C) = \frac{1}{6} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{24}$.

2.5 Double conditionnement, indépendance conditionnelle

Comme on l'a vu dès le début de ce chapitre, conditionner par rapport à un événement G signifie se placer dans une nouvelle situation probabiliste dans laquelle G est réalisé. Cette nouvelle situation change les probabilités des événements... Les probabilités calculées conditionnellement à G obéissent aux même règles que les probabilités calculées dans la configuration initiale. Pour cette raison on note parfois $P(E|G) = P_G(E)$ la probabilité d'un événement E sachant G. En termes mathématiques, on dit que P_G est une **probabilité**, au même titre que la probabilité de référence P, et obéit aux même règles.

Voici quelques exemples de notions et formules découlant de ce principe :

Probabilité conditionnelle d'une union *Soient E, F, G trois événements, avec $P(G) \neq 0$. Alors*

$$P(E \cup F|G) = P(E|G) + P(F|G) - P(E \cap F|G).$$

Double conditionnement *Soient E, F, G trois événements, avec $P(F \cap G) \neq 0$. Alors*

$$P(E|F \cap G) = \frac{P(E \cap F|G)}{P(F|G)}.$$

Cette formule devient claire si on la comprend comme la définition, sous l'hypothèse que G est réalisé, de la probabilité conditionnelle de E sachant F. Autrement, elle se démontre très facilement à partir des définitions :

Preuve.

$$P(E|F\cap G)=\frac{P(E\cap F\cap G)}{P(F\cap G)}=\frac{P(E\cap F|G)P(G)}{P(F|G)P(G)}=\frac{P(E\cap F|G)}{P(F|G)}.$$

□

Formule des probabilités totales conditionnelle : *Soient E, F, G trois événements avec $P(F\cap G)\neq 0$. Alors*

$$P(E|G)=P(E|F\cap G)P(F|G)+P(E|F^c\cap G))P(F^c|G).$$

La notion d'indépendance conditionnelle est souvent utile :

Indépendance conditionnelle : *Soient E, F, G trois événements, avec $P(G)\neq 0$. On dit que E et F sont* **indépendants conditionnellement** *à G lorsque, sous l'hypothèse que G est réalisé, E et F sont indépendants. Autrement dit,*

$$P(E\cap F|G)=P(E|G)P(F|G).$$

Des événements conditionnellement indépendants n'ont aucune raison d'être indépendants. Voici un exemple de cette situation.

Exemple : *Un tribunal doit traiter d'affaires de meurtres. A la fin du procès deux jurés doivent se prononcer et l'accusé sera condamné si les deux jurés se prononcent pour la condamnation. On suppose que lorsque l'accusé a réellement commis un meurtre, le procès lui est logiquement défavorable, si bien que chaque juré le déclarera coupable avec une probabilité de 0.7 et leurs deux avis sont alors indépendants. Lorsqu'il n'est pas coupable, leurs deux avis sont aussi indépendants et la probabilité qu'il soit déclaré coupable est alors de 0.2 pour chacun des deux jurés. On suppose que 60% des accusés sont effectivement coupables.*

Un accusé est jugé. On note les événements $A=$"le premier juré le déclare coupable" et $B=$"le deuxième juré le déclare coupable".

1. *Calculer la probabilité qu'il soit condamné.*
2. *Les événements A et B sont-ils indépendants ?*

1. On doit calculer $P(A\cap B)$. Traduisons d'abord les hypothèses. Notons $M=$"l'accusé est un meurtrier". On a :

$$P(A|M)=P(B|M)=0.7,\ \text{et}\ P(A|M^c)=P(B|M^c)=0.2.$$

Les hypothèses d'indépendance de l'énoncé sont ici des indépendances conditionnelles : conditionnellement au fait que l'accusé est un meurtrier (c'est-à-dire que M est réalisé), les deux avis sont indépendants, et donc A et B sont indépendants. On a donc

$$P(A \cap B|M) = P(A|M)P(B|M).$$

De même A et B sont indépendants conditionnellement à M^c :

$$P(A \cap B|M^c) = P(A|M^c)P(B|M^c).$$

Enfin on a $P(M) = 0.6$. A présent on peut calculer :

$$\begin{aligned} P(A \cap B) &= P(A \cap B|M)P(M) + P(A \cap B|M^c)P(M^c) \quad \text{(formule des probabilités totales)}, \\ &= P(A|M)P(B|M)P(M) + P(A|M^c)P(B|M^c)P(M^c), \\ &= 0.7 \times 0.7 \times 0.6 + 0.2 \times 0.2 \times 0.4 = \frac{31}{100}. \end{aligned}$$

2. On doit comparer $P(A \cap B)$ à $P(A)P(B)$. On a :

$$P(A) = P(A|M)P(M) + P(A|M^c)P(M^c) = 0.7 \times 0.6 + 0.2 \times 0.4 = \frac{1}{2},$$

et de même $P(B) = \frac{1}{2}$. Ainsi $P(A)P(B) = \frac{1}{4} \neq P(A \cap B)$. A et B ne sont donc pas indépendants.

Chapitre 3

Variables aléatoires et lois

3.1 Définitions générales

3.1.1 Variables aléatoires

On appelle **variable aléatoire** tout nombre réel aléatoire, c'est-à-dire dont la valeur dépend du résultat d'une expérience probabiliste. Par exemple :

On lance un dé. Soit X le résultat obtenu.

Ici X est une variable aléatoire et les valeurs possibles de X sont $1, 2, 3, 4, 5, 6$. Pour chacune de ces valeurs, X a une certaine probabilité de lui être égal. Ici en fait on peut donner directement les probabilités des événements $"X = 1", "X = 2", \dots, "X = 6"$: on a $P("X = 1") = P("X = 2") = \dots = P("X = 6") = \frac{1}{6}$.

Remarque : Un nombre réel fixé (c'est-à-dire non aléatoire) peut être vu comme une variable aléatoire ayant une probabilité 1 d'être égale à la valeur considérée. Par exemple le nombre $x = 2$ sera identifié à une variable X telle que $P(X = 2) = 1$.

Remarques sur les notations :

- Par convention, les variables aléatoires sont en général notées avec des lettres capitales (X,Y,T,etc.) pour les différencier des nombres réels non aléatoires.
- Pour noter les événements relatifs à une variable aléatoire X, comme par exemple $"X = 1", "X \leq 2", "0 \leq X \leq 4"$, on utilise souvent plutôt les crochets : $[X = 1], [X \leq 2], [0 \leq X \leq 4]$.
- De même, plutôt que $P("X = 1")$ ou $P([X = 1])$ on écrira simplement $P(X = 1)$ ou $P[X = 1]$.

3.1.2 Support d'une variable aléatoire

Le **support** d'une variable aléatoire est l'ensemble des ses valeurs possibles. C'est la première chose à préciser lorsqu'on considère une variable aléatoire. On notera $\mathcal{S}(X)$ le support d'une variable aléatoire X.

Exemple 1 : On reprend l'exemple précédent : X est le résultat d'un lancer de dé. Le support de X est alors $\{1, 2, 3, 4, 5, 6\}$.

Exemple 2 : Soit X le nombre de jours avant la prochaine pluie (en supposant que ça va forcément arriver). Alors $X \in \{0, 1, 2, \ldots\} = \mathbf{N}$.

Exemple 3 : On lance une balle, et on note X la distance parcourue par la balle avant de s'arrêter. Alors on aura $X \in [0, d]$ (en supposant qu'il y a une distance maximale d possible), ou bien simplement $X \in [0, +\infty[$.

On vient en fait de voir trois types de support différents avec ces trois exemples : support fini pour le premier, support infini dénombrable pour le second, support infini non dénombrable pour le troisième. Cette distinction est essentielle en probabilités, car les calculs de probabilités vont s'effectuer de façon complètement différentes suivant les cas.

Définition. *Une* **variable aléatoire discrète** *est une variable aléatoire dont le support est un ensemble fini ou infini dénombrable.*

3.1.3 Fonction de répartition d'une variable aléatoire

Définition. *La* **fonction de répartition** *d'une variable aléatoire X est la fonction définie pour tout $t \in \mathbf{R}$ par*

$$\boxed{F_X(t) = P(X \leq t).}$$

Autrement dit, $F_X(t)$ est la probabilité de l'événement "la valeur de X est inférieure ou égale à t".

Proposition. *On a $F_X(t) \in [0, 1]$ pour tout $t \in \mathbf{R}$; et F_X est une fonction croissante.*

Preuve. $F_X(t) \in [0, 1]$ car c'est une probabilité. De plus, si $t \leq u$, on a $[X \leq t] \subset [X \leq u]$ et donc $P(X \leq t) \leq P(X \leq u)$, c'est-à-dire $F_X(t) \leq F_X(u)$. Donc F_X est croissante. □

3.2 Variables aléatoires discrètes

Loi d'une variable discrète. *Donner la loi d'une variable aléatoire discrète X, c'est calculer les probabilités $P(X = x)$ pour toutes les valeurs x possibles prises par X (autrement dit pour tous les x appartenant au support de X).*

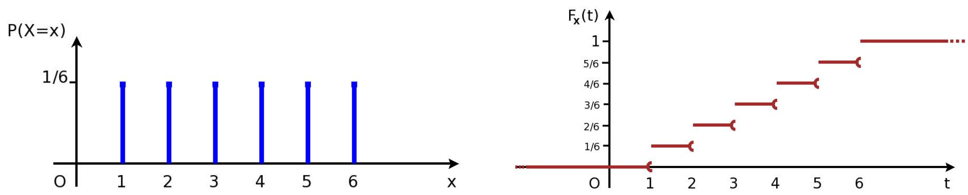


FIGURE 3.1 – Loi et fonction de répartition d'un lancer de dé

3.2.1 Quelques exemples

Exemple 1 : X est le résultat du lancer d'un dé. On a vu que $P(X = 1) = P(X = 2) = \cdots = P(X = 6) = \frac{1}{6}$: c'est la loi de X. On peut la présenter sous forme de tableau :

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

On peut aussi présenter ce résultat sous la forme d'un graphique, ou diagramme en bâtons (figure 3.1 à gauche). Enfin on peut aussi tracer le graphe de la fonction de répartition F_X (figure 3.1 à droite). On voit ici que la fonction de répartition est constante par morceaux : elle présente des sauts pour les valeurs du support de X, mais reste constante entre deux de ces valeurs. Ce sera toujours le cas pour des variables discrètes.

Lorsque toutes les probabilités formant la loi de X sont égales, comme dans l'exemple du dé, on parle de loi uniforme. C'est l'exemple le plus simple de variable aléatoire.

Définition. *On dit que la variable X suit la* **loi uniforme** *sur $\{x_1, x_2, \ldots, x_n\}$ lorsque le support de X est égal à $\{x_1, x_2, \ldots, x_n\}$ et que $P(X = x_i) = \frac{1}{n}$ pour tout $1 \leq i \leq n$.*

Voici un deuxième exemple élémentaire de variable aléatoire discrète :

Exemple 2 *: On lance trois pièces de monnaie. Soit X le nombre de "Face" obtenu. Quelle est la loi de X ?*

Le support de X est ici $\{0, 1, 2, 3\}$. On doit donc calculer $P(X = 0), P(X = 1), P(X = 2)$ et $P(X = 3)$.

Définissons F_1="la première pièce tombe sur Face" ; et de même F_2 et F_3. On peut clairement supposer que les trois lancers de pièce sont indépendants ici ; et donc que F_1, F_2, F_3 sont indépendants.

- $[X = 0] = F_1^c \cap F_2^c \cap F_3^c$ donc $P(X = 0) = P(F_1^c)P(F_2^c)P(F_3^c)$ grâce à l'indépendance. Ainsi $P(X = 0) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$.
- $[X = 1] = (F_1 \cap F_2^c \cap F_3^c) \cup (F_1^c \cap F_2 \cap F_3^c) \cup (F_1^c \cap F_2^c \cap F_3)$. Cette union est disjointe,

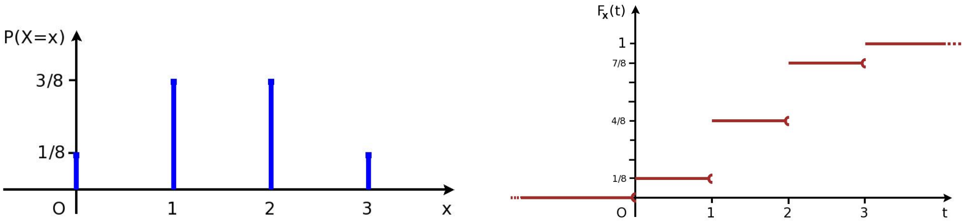


FIGURE 3.2 – Loi et fonction de répartition de X pour l'exemple 2

donc on peut additionner les probabilités :

$$\begin{aligned} P(X=1) &= P(F_1 \cap F_2^c \cap F_3^c) + P(F_1^c \cap F_2 \cap F_3^c) + P(F_1^c \cap F_2^c \cap F_3), \\ &= P(F_1)P(F_2^c)P(F_3^c) + P(F_1^c)P(F_2)P(F_3^c) + P(F_1^c)P(F_2^c)P(F_3), \\ & \qquad \text{(par indépendance)} \\ &= (\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}) \times 3 = \frac{3}{8}. \end{aligned}$$

- $[X=2] = (F_1 \cap F_2 \cap F_3^c) \cup (F_1^c \cap F_2 \cap F_3) \cup (F_1 \cap F_2^c \cap F_3)$.

$$\begin{aligned} P(X=1) &= P(F_1 \cap F_2 \cap F_3^c) + P(F_1^c \cap F_2 \cap F_3) + P(F_1 \cap F_2^c \cap F_3), \text{ (union disjointe)} \\ &= P(F_1)P(F_2)P(F_3^c) + P(F_1^c)P(F_2)P(F_3) + P(F_1)P(F_2^c)P(F_3), \text{ (indépendance)} \\ &= (\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}) \times 3 = \frac{3}{8}. \end{aligned}$$

- $[X=3] = F_1 \cap F_2 \cap F_3$ donc $P(X=0) = P(F_1)P(F_2)P(F_3) = \frac{1}{8}$ par indépendance.

La loi de X est en fait un exemple de loi binomiale. Le cas général sera vu un peu plus loin. La figure 3.2 montre le graphique de cette loi ainsi que la fonction de répartition.

Exemple 3 *: On lance un dé et on recommence indéfiniment. Soit X le rang d'apparition du premier 6 (par exemple $[X=5]$ ="on obtient 6 pour la première fois au 5^e lancer"). Déterminer la loi de X.*

Le support de X est ici $\{1, 2, 3, \ldots\} = \mathbf{N}^*$. En toute rigueur il faudrait ici remarquer que X n'est pas nécessairement défini en envisageant le cas où le 6 n'apparaît jamais ; ou bien rajouter la valeur $+\infty$ au support en définissant l'événement $[X = +\infty]$ ="le 6 n'apparaît jamais". Cependant on peut montrer que la probabilité de cet événement est nulle, et donc qu'il n'y a pas lieu de le prendre en compte.

Pour tout $n \geq 1$, notons A_n l'événement "le n^e lancer vaut 6". A_n ne dépend que du résultat du n^e lancer, et il est clair que les lancers sont des expériences indépendantes. Par conséquent les A_n sont des événements indépendants.

- $[X=1] = A_1$, donc $P(X=1) = P(A_1) = \frac{1}{6}$,
- $[X=2] = A_1^c \cap A_2$, donc $P(X=2) = P(A_1^c)P(A_2)$ par indépendance. Donc $P(X = 2) = \frac{5}{6} \times \frac{1}{6}$.

- $[X = 3] = A_1^c \cap A_2^c \cap A_3$ donc $P(X = 3) = P(A_1^c)P(A_2^c)P(A_3)$ grâce à l'indépendance. Donc $P(X = 3) = (\frac{5}{6})^2 \times \frac{1}{6}$.
- Plus généralement, on aura $[X = n] = A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c \cap A_n$, donc $P(X = n) = P(A_1^c)P(A_2^c) \dots P(A_{n-1}^c)P(A_n)$, toujours grâce à l'indépendance. Ainsi $P(X = n) = (\frac{5}{6})^{n-1}\frac{1}{6}$.

Finalement cette dernière formule est valable pour tout $n \geq 1$; on a donc ainsi bien déterminé la loi de X. Cette loi s'appelle la loi géométrique de paramètre $\frac{1}{6}$.

Définition. *On dit qu'une variable aléatoire suit la* **loi géométrique** *de paramètre p, où p est un nombre $p \in [0, 1]$ fixé, si le support de X est égal à $\mathbf{N}^*$ et que*

$$P(X = n) = (1 - p)^{n-1}p.$$

3.2.2 Quelques lois usuelles

On a déjà défini les lois **uniformes** et **géométriques**. Voici quelques autres lois classiques :

Loi de Bernoulli. C'est la loi la plus élémentaire : 0 ou 1.

Définition. *Une variable aléatoire X suit la* **loi de Bernoulli** *de paramètre $p \in [0, 1]$ si $X \in \{0, 1\}$ et $P(X = 1) = p$, $P(X = 0) = 1 - p$. On note $X \sim \mathcal{B}(p)$.*

Loi binomiale. C'est la loi du nombre de succès lors de n essais indépendants d'une même expérience probabiliste. Nous allons la calculer sur un exemple simple.

exemple *: On lance n dés et on note X le nombre de fois que l'on obtient* 6*. Quelle est la loi de X ?*

Cet exemple est simplement la généralisation de l'exemple 2 de la section précédente. Tout d'abord le support de X est $\{0, 1, 2, \dots, n\}$. Notons A_i l'événement "On obtient 6 au i-ème lancer". Tous ces événements sont indépendants. On a :

$[X = 0] = A_1^c \cap A_2^c \cap \cdots \cap A_n^c$ donc $P(X = 0) = P(A_1^c)P(A_2^c) \cdots P(A_n^c) = (\frac{5}{6})^n$.

Pour l'événement $[X = 1]$ on peut le décomposer ainsi : $[X = 1] = (A_1 \cap A_2^c \cap \cdots \cap A_n^c) \cup (A_1^c \cap A_2 \cap \cdots \cap A_n^c) \cup \cdots \cup (A_1^c \cap A_2^c \cap \cdots \cap A_n)$, donc

$$\begin{aligned}
P(X = 1) &= P(A_1 \cap A_2^c \cap \cdots \cap A_n^c) + P(A_1^c \cap A_2 \cap \cdots \cap A_n^c) + \cdots + P(A_1^c \cap A_2^c \cap \cdots \cap A_n), \\
&= P(A_1)P(A_2^c) \cdots P(A_n^c) + P(A_1^c)P(A_2) \cdots P(A_n^c) + \cdots + P(A_1^c)P(A_2^c) \cdots P(A_n), \\
&= \frac{1}{6}(\frac{5}{6})^{n-1} + \frac{1}{6}(\frac{5}{6})^{n-1} + \cdots + \frac{1}{6}(\frac{5}{6})^{n-1} = n\frac{1}{6}(\frac{5}{6})^{n-1}.
\end{aligned}$$

Pour $[X = 2]$ et les suivants, on procède de la même manière, mais il devient compliqué de l'écrire in-extenso. On voit qu'à chaque fois on décompose l'événement $[X = k]$ en considérant tous les cas possibles pour les k succès envisagés. Pour chacun de ces cas la probabilité sera la même : $(\frac{1}{6})^k(\frac{5}{6})^{n-k}$; et il reste donc à compter le nombre de ces cas. Il s'agit de compter le nombre de choix possibles pour les positions des k succès parmi les n essais. Il y en a $\binom{n}{k}$. Ainsi $P(X=k)=\binom{n}{k}(\frac{1}{6})^k(\frac{5}{6})^{n-k}$.

Définition. *Une variable aléatoire suit la loi binomiale de paramètres $n \in \mathbf{N}$ et $p \in [0,1]$ lorsque $X \in \{0,1,2,\dots,n\}$ et*

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

La loi binomiale de paramètres n et p se note $\mathcal{B}(n,p)$.

La figure 3.3 montre des exemples de lois binomiales pour différents paramètres.

Loi de Poisson. La loi de Poisson sera vu plus en détail par la suite. Nous donnons simplement la définition ici :

Définition. *Soit $\lambda \geq 0$. Une variable X suit la* **loi de Poisson** *de paramètre λ si pour tout entier $k \geq 0$,*

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}.$$

Le support de X est ici $\mathbf{N}$. La loi de Poisson de paramètre λ est notée $\mathcal{P}(\lambda)$.

3.2.3 Les événements élémentaires $[X=x]$

1) Les événements $[X=x]$, considérés pour tous les x appartenant au support, forment une partition de Ω. En effet, X prend nécessairement une et une seule de ces valeurs, ce qui prouve bien que la réunion des $[X=x]$ est égale à Ω, et que $[X=x]\cap[X=y]=\varnothing$ si $x \neq y$. Par conséquent la somme totale des $P(X=x)$ doit toujours être égale à 1, ce qui s'écrit, en notant $\mathcal{S}$ le support de X,

$$\sum_{x\in\mathcal{S}} P(X=x)=1.$$

Il faut toujours penser à le vérifier lorsqu'on calcule une loi. Pour les exemples précédents, on a :

- pour le dé : $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=1$,
- pour l'exemple 2 : $\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$,
- pour la loi géométrique :

$$\sum_{n=1}^{+\infty}(1-p)^{n-1}p=p\sum_{n=0}^{+\infty}(1-p)^n=p\frac{1}{1-(1-p)}=1.$$

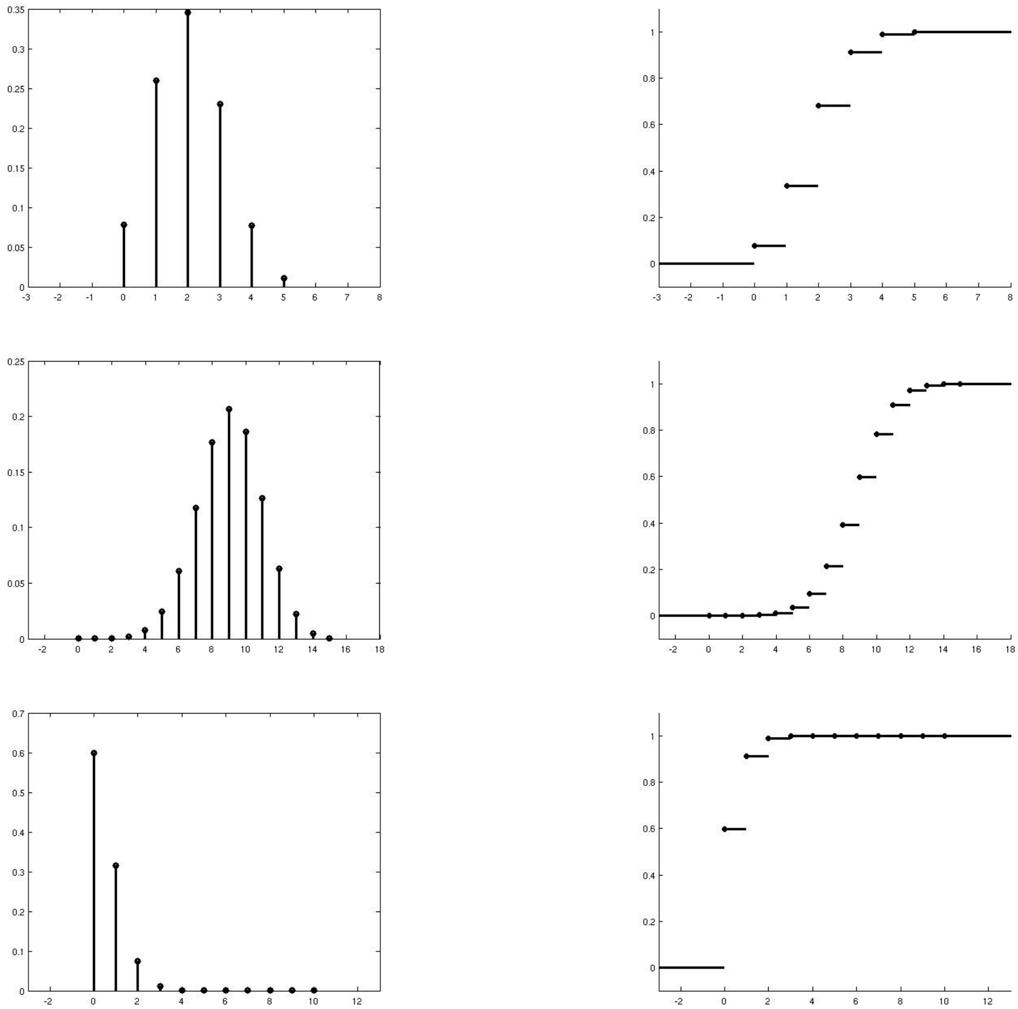


FIGURE 3.3 – lois binomiales (à gauche) et fonctions de répartition F_X correspondantes (à droite) pour différents paramètres : 1e ligne : $n = 5, p = 0.4$; 2e ligne : $n = 15, p = 0.6$; 3e ligne : $n = 10, p = 0.05$.

2) De plus ces événements $[X = x]$ sont les événements élémentaires pour la variable X, au sens où tout événement relatif à X s'exprime comme une union de ces événements, et sa probabilité est la somme des $P(X = x)$ correspondants.

exemple : L'événement "le dé tombe sur un nombre pair supérieur à 3" est égal à $[X = 4] \cup [X = 6]$ si X est le résultat du dé. Sa probabilité est donc de $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

exemple : Si X est une variable discrète quelconque, l'événement $[X \leq t]$ est l'union des $[X = x]$ pour tous les x appartenant au support de X tels que $x \leq t$. On peut ainsi donner une formule pour la fonction de répartition d'une variable discrète :

$$F_X(t) = \sum_{x \in \mathcal{S}, x \leq t} P(X = x).$$

3.3 Variables aléatoires à densité

3.3.1 Exemple introductif

On jette un stylo sur une table, et on note X l'angle entre 0 et π qu'il forme avec le bord de la table. Quelle est la loi de X ?

L'ensemble des valeurs possibles pour cette variable aléatoire est $[0, \pi]$. En suivant l'idée vue auparavant, on voudrait donc chercher à calculer tous les $P(X = \alpha)$ pour $\alpha \in [0, \pi]$. En fait on verra que $P(X = \alpha)$ sera toujours égal à 0, ce qui signifie que quelle que soit la valeur de α, le mikado n'a aucune chance de former exactement l'angle α avec la table. On est obligé pour donner un sens aux probabilités ici, de considérer des intervalles et non des valeurs uniques, et de regarder les probabilités que X soit dans ces intervalles. Par exemple on peut raisonnablement penser ici que $P(0 \leq X \leq \frac{\pi}{2}) = \frac{1}{2}$, ou que $P(\frac{\pi}{4} \leq X \leq \frac{\pi}{2}) = \frac{1}{4}$. Plus généralement, on peut raisonnablement supposer que la probabilité que X appartienne à un intervalle $[\alpha, \beta]$ correspond à la proportion d'angles compris dans cet intervalle, c'est-à-dire que

$$P(\alpha \leq X \leq \beta) = \frac{\beta - \alpha}{\pi}.$$

Ceci permet de caractériser entièrement la variable X car la probabilité de tout événement lié à X peut se calculer à partir de cette formule.

On va néanmoins ici pouvoir donner une valeur de probabilité associée à un angle α donné en considérant un petit intervalle $[\alpha - \varepsilon, \alpha + \varepsilon]$: on a

$$P(\alpha - \varepsilon \leq X \leq \alpha + \varepsilon) = \frac{(\alpha + \varepsilon) - (\alpha + \varepsilon)}{\pi} = \frac{2\varepsilon}{\pi}.$$

On obtient ce qu'on souhaite en prenant la limite pour ε tendant vers 0 de

$$\frac{P(\alpha - \varepsilon \leq X \leq \alpha + \varepsilon)}{2\varepsilon},$$

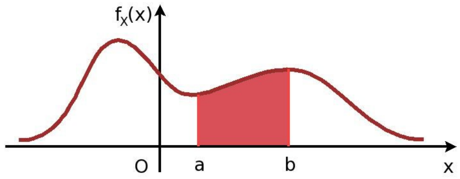


FIGURE 3.4 – Graphe d'une densité de probabilité f_X. La partie colorée correspond à $P(a \leq X \leq b)$.

c'est-à-dire du rapport entre la probabilité de l'intervalle et la longueur de l'intervalle. Ici ce rapport vaut $\frac{1}{\pi}$, donc sa limite aussi. Cette valeur $\frac{1}{\pi}$ est appelée densité de X en α. Inversement, si l'on connaît la densité de X en tout point $x \in [0, \pi]$ on obtiendra une probabilité du type $P(\alpha \leq X \leq \beta)$ en calculant l'intégrale de cette densité sur l'intervalle $[\alpha, \beta]$.

3.3.2 Définition

Une variable aléatoire X est dite **à densité** *lorsqu'il existe une fonction $f_X : \mathbf{R} \to \mathbf{R}^+$ telle que*

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx \qquad \textit{pour tous } a, b \in \mathbf{R}, \quad a \leq b.$$

Cette fonction f_X est appelée **densité** *de X.*

remarque : on peut prendre $a = -\infty$ ou $b = +\infty$ dans cette formule.

Interprétation graphique : La probabilité $P(a \leq X \leq b)$ correspond à l'aire du domaine situé sous le graphe de f_X entre les abscisses a et b (voir figure 3.4).

remarque 1 : on a bien $P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)dx = 0$.

remarque 2 : $\int_{-\infty}^{+\infty} f_X(x)dx = P(X \in \mathbf{R}) = 1$. Il faut toujours penser à le vérifier.

remarque 3 : Calculer la loi d'une variable à densité, c'est calculer sa densité.

3.3.3 Loi uniforme sur un intervalle

C'est l'exemple le plus simple de loi à densité.

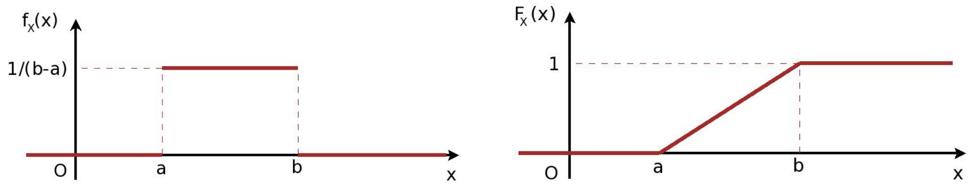


FIGURE 3.5 – Graphes de la densité et de la fonction de répartition de la loi uniforme sur un intervalle $[a, b]$.

Définition. *La loi uniforme sur un intervalle* $[\alpha, \beta]$ *est la loi de densité*

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} \text{ si } x \in [\alpha, \beta], \\ 0 \text{ sinon.} \end{cases}$$

On note $X \sim \mathcal{U}([\alpha, \beta])$ *("X suit la loi uniforme sur* $[\alpha, \beta]$*").*

exemple 1 : Dans le premier exemple, il est raisonnable de supposer que X suit la loi uniforme sur $[0, \pi]$. On aura donc par exemple :

$$P\left(X \leq \frac{\pi}{2}\right) = \int_0^{\frac{\pi}{2}} \frac{1}{\pi} dx = \frac{1}{2},$$

ou encore :

$$P\left(\frac{\pi}{4} \leq X \leq \frac{\pi}{2}\right) = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \frac{1}{\pi} dx = \frac{1}{4}.$$

3.3.4 Fonction de répartition d'une variable aléatoire à densité

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t) dt.$$

Ainsi F_X est une primitive de la fonction densité f_X.

exemple : Si X suit la loi uniforme sur $[\alpha, \beta]$, on a

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} \text{ si } x \in [\alpha, \beta], \\ 0 \text{ sinon,} \end{cases}$$

et donc

$$F_X(x) = \begin{cases} 0 \text{ si } x \leq \alpha \\ \frac{x-\alpha}{\beta-\alpha} \text{ si } x \in [\alpha, \beta], \\ 1 \text{ si } x \geq \beta \end{cases}$$

La densité et la fonction de répartition d'une loi uniforme sont représentées sur la figure 3.5.

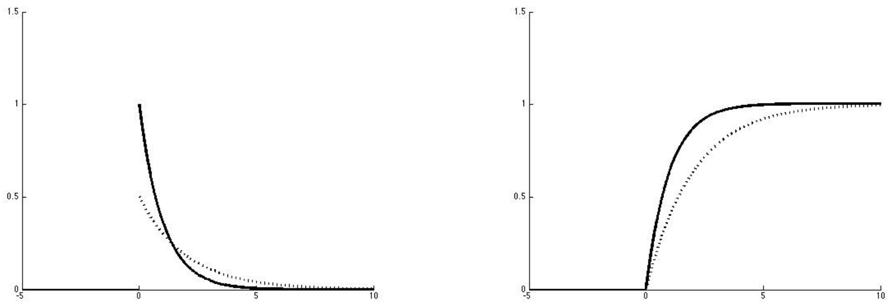


FIGURE 3.6 – Graphes de la densité et de la fonction de répartition de la loi exponentielle pour $a = 1$ (traits pleins) et $a = 0.5$ (pointillés).

3.3.5 Loi exponentielle

Soit $a > 0$ un réel. On dit que X suit la loi exponentielle de paramètre a si elle admet la densité

$$f_X(x) = ae^{-ax}1_{\mathbf{R}^+}(x) = \begin{cases} ae^{-ax} \text{ si } x \geq 0, \\ 0 \text{ sinon.} \end{cases}$$

Sa fonction de répartition est :

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-ax} \text{ si } x \geq 0, \\ 0 \text{ sinon.} \end{cases}$$

La figure 3.6 montre densités et fonctions de répartition de la loi exponentielle pour deux valeurs différentes du paramètre a.

3.4 Lois quelconques

Une variable aléatoire non discrète n'est pas nécessairement à densité. Voici un exemple d'une telle situation :

Une machine à remplir les bouteilles est défectueuse : elle verse dans chaque bouteille (de 75cL) une quantité aléatoire de boisson comprise entre 0 *et* 1 *litre.*

On peut définir deux variables aléatoires liées à cette expérience : X la quantité de boisson versée par la machine, et Y la quantité de boisson contenue dans la bouteille. On a en fait les relations

$$\begin{cases} Y = X \text{ si } X \leq 0.75 \\ Y = 0.75 \text{ si } X > 0.75 \end{cases}$$

En l'absence d'autre précision, on peut considérer que X suit une loi uniforme sur $[0, 1]$. Il s'agit donc bien d'une loi à densité. En revanche Y n'est pas une variable discrète (ses

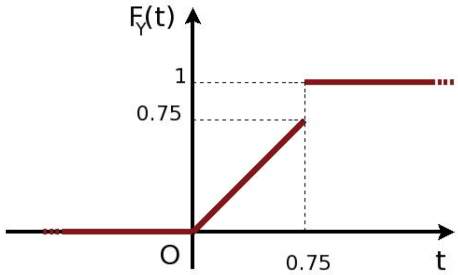


FIGURE 3.7 – exemple de la machine à remplir les bouteilles : graphe de la fonction de répartition de la loi de Y.

valeurs possibles correspondent à l'intervalle $[0, 0.75]$), et n'est pas non plus une variable à densité car $P(Y = 0.75) = P(X > 0.75) = 0.25 \neq 0$. On dit que la loi de Y possède un **atome** en $x = 0.75$.

On peut calculer facilement la fonction de répartition de Y :

$$F_Y(t) = \begin{cases} 0 \text{ si } t \leq 0, \\ t \text{ si } 0 < t < 0.75, \\ 1 \text{ si } t \geq 0.75. \end{cases}$$

Cette fonction est représentée sur la figure 3.7

3.5 Espérance et variance d'une variable aléatoire

3.5.1 Quelques rappels sur les sommes, séries et intégrales

Convergence absolue.

Une **série** $\displaystyle\sum_{n=0}^{+\infty} a_n$ est dite **convergente** si les sommes $\displaystyle\sum_{n=0}^{N} a_n$ convergent lorsque N tend vers $+\infty$, et on note alors $\displaystyle\sum_{n=0}^{+\infty} a_n = \lim_{N \to +\infty} \sum_{n=0}^{N} a_n$.

Une **série** $\displaystyle\sum_{n=0}^{+\infty} a_n$ est dite **absolument convergente** lorsque $\displaystyle\sum_{n=0}^{+\infty} |a_n|$ est une série convergente. Dans ce cas $\displaystyle\sum_{n=0}^{+\infty} a_n$ est convergente, et de plus il est permis de sommer les termes a_n dans n'importe quel ordre : $\displaystyle\sum_{n=0}^{+\infty} a_n = a_0 + a_1 + a_2 + a_4 + \cdots = a_3 + a_{10} + a_8 + a_5 + a_4 + a_7 + \cdots$ (on somme tous les a_n mais dans un ordre différent). Pour bien mettre en valeur cette propriété, on pourra noter cette somme $\sum_{n\in\mathbf{N}} a_n$: somme des a_n pour tous les $n \in \mathbf{N}$, quel que soit l'ordre d'énumération.

De la même manière une intégrale $\int_a^b f(x)dx$ est dite **absolument convergente** lorsque $\int_a^b |f(x)|dx$ est une intégrale convergente.

En probabilités on ne s'intéressera qu'aux séries et intégrales absolument convergentes.

Quelques techniques de calcul.

- L'indice utilisé dans une notation $\sum$ est "muet", c'est-à-dire qu'il n'a pas d'existence en dehors de cette somme. On peut donc le remplacer par n'importe quelle autre lettre :

$$\sum_{k=1}^{n} a_k = \sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j.$$

- Décalage d'indices :

$$\begin{aligned} \sum_{i=0}^{n-1} a_{i+1} &= \sum_{k=1}^{n} a_k \qquad \text{(en posant k=i+1)} \\ &= \sum_{i=1}^{n} a_i \qquad \text{(puisque l'indice est "muet")} \end{aligned}$$

  Pour une série on aura de même

$$\sum_{i=0}^{+\infty} a_{i+1} = \sum_{i=1}^{+\infty} a_i.$$

- Factorisation : $\sum_{i=0}^{n} x a_i = x \sum_{i=0}^{n} a_i$
- Série géométrique :

$$\sum_{i=0}^{n} x^i = \frac{1-x^{n+1}}{1-x} \qquad \text{si } x \neq 1.$$

$$\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x} \qquad \text{si } |x| < 1.$$

  Par exemple,

$$\sum_{i=3}^{n} x^i = \sum_{i=0}^{n-3} x^{i+3} = x^3 \sum_{i=0}^{n-3} x^i = x^3 \frac{1-x^{n-2}}{1-x}.$$

- Séries entières : Dans certains cas on peut obtenir un développement en série entière d'une fonction $f(x)$:

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n.$$

Par exemple

$$\frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n \qquad \text{pour } x \in]-1,1[,$$

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \qquad \text{pour tout } x \in \mathbf{R}.$$

On a alors le droit de dériver la série entière :

$$f'(x) = \sum_{n=1}^{+\infty} n a_n x^{n-1} = \sum_{n=0}^{+\infty} (n+1) a_{n+1} x^n.$$

Cette propriété est très utile pour les calculs d'espérance et de variance.

3.5.2 Définitions

Espérance

L'espérance d'une variable aléatoire X est la moyenne des valeurs prises par la variable, pondérées par leurs probabilités. On la note $E(X)$. Avant de voir des définitions plus précises, voyons un exemple :

exemple : On lance un dé. Si on obtient 6 on reçoit 8 euros ; sinon on perd 2 euros. On note G le gain obtenu. Quelle est l'espérance de ce gain ?

La loi de G est très simple : $G \in \{-2, 8\}$, et on a $P(G=-2) = \frac{5}{6}$ et $P(G=8) = \frac{1}{6}$. La moyenne pondérée des gains est donc

$$E(G) = (-2) \times \frac{5}{6} + 8 \times \frac{1}{6} = -\frac{1}{3} = -0.33.$$

Cette espérance est négative, ce qui signifie que le jeu est plutôt défavorable.

On verra plus tard que l'espérance correspond aussi à l'idée de valeur moyenne obtenue lorsqu'on répète un grand nombre de fois la même expérience : intuitivement, dans l'exemple précédent, si on joue à ce jeu un grand nombre n de fois, on perdra environ $n \times \frac{1}{3}$ euros.

Définition. *L'***espérance** *d'une variable discrète X de support $\mathcal{S}(X)$ est :*

$$\boxed{E(X) = \sum_{x \in \mathcal{S}(X)} x P(X=x),}$$

lorsque cette somme est bien définie.

Lorsque le support est infini, cette somme est une série infinie ; il peut donc arriver qu'elle ne soit pas absolument convergente : l'espérance n'est alors pas définie.

Définition. *L'***espérance** *d'une variable aléatoire X de densité f sur* $\mathbf{R}$ *est :*

$$\boxed{E(X) = \int_{\mathbf{R}} x f(x) dx,}$$

lorsque cette intégrale est bien définie.

Ici encore cette intégrale peut ne pas être absolument convergente ; dans ce cas l'espérance n'est pas définie.

Quelques propriétés de l'espérance.

- L'espérance est une fonction linéaire : si a, b sont des réels, et X, Y des variables aléatoires,
$$\boxed{\text{E(aX+bY)=aE(X)+bE(Y)}}.$$
En particulier $E(X+Y) = E(X) + E(Y)$, et $E(X-Y) = E(X) - E(Y)$.
<u>**Attention** : La loi de $X+Y$ ne se calcule pas en faisant la somme des lois de X et de Y (ce qui n'aurait aucun sens).</u>
- Si $a \in \mathbf{R}$, $E(a) = a$ puisque a est non aléatoire : autrement dit, a s'interprète comme une variable aléatoire prenant la valeur a avec probabilité 1.
- Soient X une variable aléatoire, et $g : \mathbf{R} \to \mathbf{R}$ une fonction quelconque. L'espérance de $g(X)$ se calcule ainsi :
$$E(g(X)) = \sum_{x \in \mathcal{S}(X)} g(x) P(X = x) \qquad \text{si } X \text{ est une variable discrète,}$$
$$E(g(X)) = \int_{\mathbf{R}} g(x) f_X(x) dx \qquad \text{si } X \text{ est une variable à densité.}$$
Ici encore ces sommes ou ces intégrales peuvent ne pas être absolument convergentes. Par exemple il se peut très bien que $E(g(X))$ n'existe pas alors que $E(X)$ existe.

Variance

La variance permet de mesurer l'écart des valeurs de la variable par rapport à l'espérance :

Définition. *La* **variance** *d'une variable aléatoire X est*

$$\boxed{V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2.}$$

Montrons que cette dernière égalité est vraie : on a

$$E((X-E(X))^2) = E(X^2 + E(X)^2 - 2XE(X)) = E(X^2) + E(E(X)) - 2E(XE(X)),$$

par la propriété de linéarité. A présent $E(X)$ est un nombre réel ; on peut donc écrire $E(E(X)) = E(X)$, et $E(XE(X)) = E(X)E(X) = E(X)^2$. Ainsi

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

La variance est un nombre positif qui peut être infini, même lorsque l'espérance est définie.

Définition. *L'* **écart-type** *d'une variable aléatoire X est la racine carrée de sa variance :*

$$\boxed{\sigma(X) = \sqrt{V(X)}.}$$

3.5.3 Exemples de calculs

Espérance et variance d'une variable géométrique

Une variable X de loi géométrique de paramètre $p \in [0,1]$ a pour loi :

$$P(X=n) = (1-p)^{n-1}p \qquad \text{pour } n \geq 1.$$

Calculons $E(X)$:

$$\begin{aligned} E(X) &= \sum_{n\geq 1} nP(X=n) \\ &= \sum_{n\geq 1} n(1-p)^{n-1}p \\ &= p\sum_{n\geq 1} n(1-p)^{n-1}. \end{aligned}$$

Notons $f(x) = \sum_{n\geq 1} nx^{n-1}$. $f(x)$ est la dérivée de $\sum_{n\geq 1} x^n = \sum_{n\geq 0} x^n - x^0 = \frac{1}{1-x} - 1$, donc $f(x) = \left(\frac{1}{1-x} - 1\right)' = \frac{1}{(1-x)^2}$. Ainsi

$$E(X) = pf(1-p) = p\frac{1}{p^2} = \frac{1}{p}.$$

$$\boxed{E(X) = \frac{1}{p}}$$

Calculons $V(X)$:

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= \sum_{n\geq 1} n^2 P(X=n) - \frac{1}{p^2} \\ &= p\sum_{n\geq 1} n^2(1-p)^{n-1} - \frac{1}{p^2}. \end{aligned}$$

Notons $f(x) = \sum_{n\geq 1} n^2 x^{n-1}$. On a

$$\begin{aligned} f(x) &= \sum_{n\geq 1} n(n+1)x^{n-1} - \sum_{n\geq 1} nx^{n-1} \\ &= \left(\sum_{n\geq 1} x^{n+1}\right)'' - \left(\sum_{n\geq 1} x^n\right)' \\ &= \left(\sum_{n\geq 2} x^{n}\right)'' - \left(\sum_{n\geq 1} x^n\right)' \\ &= \left(\frac{1}{1-x} - 1 - x\right)'' - \left(\frac{1}{1-x} - 1\right)' \\ &= \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2}, \end{aligned}$$

et donc

$$V(X) = pf(1-p) - \frac{1}{p^2} = p\left(\frac{2}{p^3} - \frac{1}{p^2}\right) - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p}.$$

$$\boxed{V(X) = \frac{1-p}{p^2}}$$

Espérance et variance d'une variable binomiale

Une variable X de loi binomiale de paramètres $n \in \mathbf{N}$ et $p \in [0,1]$ a pour loi :

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{pour } 0 \leq k \leq n.$$

Calculons $E(X)$:

$$E(X) = \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k}.$$

Tout d'abord le terme $k=0$ de cette somme est nul, donc

$$E(X)=\sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k}.$$

De plus pour $k\geq 1$,

$$\begin{aligned} k\binom{n}{k} &= k\frac{n!}{k!(n-k)!}=\frac{n!}{(k-1)!(n-k)!}=n\frac{(n-1)!}{(k-1)!(n-k)!} \\ &= n\frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \\ &= n\binom{n-1}{k-1}. \end{aligned}$$

Ainsi

$$\begin{aligned} E(X) &= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k} &=& \quad n\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n-k} \\ &= n\sum_{k=0}^{n-1}\binom{n-1}{k}p^{k+1}(1-p)^{n-(k+1)} &=& \quad np\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-(k+1)} \\ &= np\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{(n-1)-k} &=& \quad np(p+(1-p))^{n-1}. \end{aligned}$$

$$\boxed{E(X)=np}$$

Calculons $V(X)$:

$$V(X) \;=\; E(X^2)-E(X)^2=E(X^2)-n^2p^2.$$

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{n}k^2P(X=k) \;=\; \sum_{k=0}^{n}k^2\binom{n}{k}p^k(1-p)^{n-k} \;=\; \sum_{k=1}^{n}k^2\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n}k(k-1)\binom{n}{k}p^k(1-p)^{n-k}+\sum_{k=1}^{n}k\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=2}^{n}k(k-1)\binom{n}{k}p^k(1-p)^{n-k}+\sum_{k=1}^{n}k\binom{n}{k}p^k(1-p)^{n-k}. \end{aligned}$$

Le deuxième terme est égal à $E(X)$, donc à np. De plus pour $k\geq 2$,

$$\begin{aligned} k(k-1)\frac{n!}{(k!(n-k)!} &= \frac{n!}{(k-2)!(n-k)!} \\ &= n(n-1)\frac{(n-2)!}{(k-2)!((n-2)-(k-2)!} \\ &= n(n-1)\binom{n-2}{k-2}. \end{aligned}$$

Ainsi

$$\begin{aligned}
E(X^2) &= n(n-1)\sum_{k=2}^{n}\binom{n-2}{k-2}p^k(1-p)^{n-k}+np \\
&= n(n-1)\sum_{k=0}^{n-2}\binom{n-2}{k}p^{k+2}(1-p)^{n-k-2}+np \\
&= n(n-1)p^2\sum_{k=0}^{n-2}\binom{n-2}{k}p^{k}(1-p)^{n-2-k}+np \\
&= n(n-1)p^2(p+(1-p))^{n-2} = n(n-1)p^2+np.
\end{aligned}$$

Finalement,

$$V(X) = E(X^2)-np = n(n-1)p^2+np-n^2p^2.$$

$$\boxed{V(X) = np(1-p)}$$

Espérance et variance d'une variable de loi exponentielle

Soit X une variable aléatoire de loi exponentielle de paramètre $a > 0$. X admet la densité

$$f_X(x) = ae^{-ax}1_{\mathbf{R}^+} = \begin{cases} ae^{-ax} \text{ si } x \geq 0, \\ 0 \text{ sinon.} \end{cases}.$$

Calculons $E(X)$:

$$E(X) = \int_{\mathbf{R}} xf_X(x)dx = \int_0^{+\infty} xae^{-ax}dx.$$

On fait une intégration par parties avec $u = x$, $u' = 1$ et $v' = ae^{-ax}$, $v = -e^{-ax}$:

$$\begin{aligned}
E(X) &= \left[x(-e^{-ax})\right]_0^{+\infty} - \int_0^{+\infty}(-e^{-ax})dx \\
&= 0+\int_0^{+\infty} e^{-ax}dx = \left[-\frac{1}{a}e^{-ax}\right]_0^{+\infty} = \frac{1}{a}.
\end{aligned}$$

$$\boxed{E(X) = \frac{1}{a}}$$

Calculons $V(X)$:

$$V(X) = E(X^2)-E(X)^2 = \int_{\mathbf{R}} x^2 f_X(x)dx - \frac{1}{a^2} = \int_0^{+\infty} x^2ae^{-ax}dx - \frac{1}{a^2}.$$

On pose $u = x^2$, $u' = 2x$ et $v' = ae^{-ax}$, $v = -e^{-ax}$:

$$\begin{aligned}
V(X) &= \left[x^2(-e^{-ax})\right]_0^{+\infty} - \int_0^{+\infty} 2x(-e^{-ax})dx - \frac{1}{a^2} \\
&= 0 + 2\int_0^{+\infty} xe^{-ax}dx - \frac{1}{a^2} \\
&= 2\left[x(-\frac{1}{a}e^{-ax})\right]_0^{+\infty} - 2\int_0^{+\infty}(-\frac{1}{a}e^{-ax})dx - \frac{1}{a^2} \\
&= 0 + \frac{2}{a}\int_0^{+\infty} e^{-ax}dx - \frac{1}{a^2} \\
&= \frac{2}{a}\left[-\frac{1}{a}e^{-ax}\right]_0^{+\infty} - \frac{1}{a^2} \\
&= \frac{2}{a}\frac{1}{a} - \frac{1}{a^2} = \frac{1}{a^2}.
\end{aligned}$$

$$\boxed{V(X) = \frac{1}{a^2}}$$

3.6 Variables aléatoires indépendantes

C'est sans doute la notion la plus utilisée en probabilités et statistiques.

Définition. *Deux variables aléatoires X et Y sont dites* **indépendantes** *lorsque tout événement $[X \in I]$ avec $I \subset \mathbf{R}$ est indépendant de tout événement $[Y \in J]$ avec $J \subset \mathbf{R}$. Autrement dit , $\forall I, J \subset \mathbf{R}$,*

$$P(X \in I, Y \in J) = P(X \in I)P(Y \in J).$$

On note parfois $X \perp\!\!\!\perp Y$. ("X indépendant de Y").

En pratique, c'est souvent l'intuition qui permet de décider si deux variables sont indépendantes. On peut alors utiliser la définition comme une propriété : tout événement relatif à une des deux variables sera indépendant de tout événement relatif à l'autre.

exemple : On lance deux dés. Soit X le résultat du premier et Y le résultat du second. Alors X et Y sont indépendantes.

On généralise facilement cette définition à un nombre quelconque, voire une infinité, de variables aléatoires :

Définition. *Des variables aléatoires $X_1, X_2, X_3, \ldots$ sont* **indépendantes** *(ou mutuellement indépendantes) si pour tous sous-ensembles $I_1, I_2, I_3, \ldots$ de $\mathbf{R}$, les événements $[X_i \in I_i]$ sont indépendants. Soit encore : pour tous $I_1, I_2, I_3, \ldots \subset \mathbf{R}$,*

$$P(X_1 \in I_1, X_2 \in I_2, \ldots, X_n \in I_n) = P(X_1 \in I_1)P(X_2 \in I_2)\cdots P(X_n \in I_n).$$

3.7 Exemples de calculs de lois utilisant l'indépendance

3.7.1 Somme de variables indépendantes

exemple 1 : *On lance deux dés. Calculer la loi de la somme S des résultats obtenus.*

$S \in \{2, 3, 4, \ldots, 12\}$. Si on note X et Y les résultats des deux dés, X et Y sont des variables aléatoires indépendantes, de même loi uniforme sur $\{1, 2, \ldots, 6\}$: $P(X = i) = P(Y = i) = \frac{1}{6}$.

$[S = 2] = [X = 1] \cap [Y = 1]$ donc $P(S = 2) = P(X = 1)P(Y = 1) = \frac{1}{36}$.

$[S = 3] = ([X = 1] \cap [Y = 2]) \cup ([X = 2] \cap [Y = 1])$ donc $P(S = 3) = P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1) = \frac{2}{36}$.

$[S = 4] = ([X = 1] \cap [Y = 3]) \cup ([X = 2] \cap [Y = 2]) \cup ([X = 3] \cap [Y = 1])$ donc $P(S = 4) = P(X = 1)P(Y = 3) + P(X = 2)P(Y = 2) + P(X = 3)P(Y = 1) = \frac{3}{36}$.

Ainsi de suite. On voit donc que pour chaque valeur possible de la somme S, la probabilité correspondante s'obtient en comptant le nombre de façons différentes d'obtenir cette somme avec les deux résultats. On aboutit ainsi à la loi suivante :

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

exemple 2 : *Soient X une variable de loi de Poisson $\mathcal{P}(\lambda)$ et Y de loi de Poisson $\mathcal{P}(\mu)$. On suppose X et Y indépendantes. Calculer la loi de $Z = X + Y$.*

Les trois variables X, Y, Z sont à valeurs dans $\mathbf{N}$. Calculons $P(Z = n)$ pour tout $n \in \mathbf{N}$: on va décomposer l'événement $[Z = n]$ ainsi :

$$[Z = n] = [Z = n, X = 0] \cup [Z = n, X = 1] \cup [Z = n, X = 2] \cup \cdots .$$

Cette union est disjointe, donc on peut additionner les probabilités :

$$\begin{aligned} P(Z = n) &= P(Z = n, X = 0) + P(Z = n, X = 1) + P(Z = n, X = 2) + \cdots \\ &= \sum_{k \geq 0} P(Z = n, X = k). \end{aligned}$$

Or

$$\begin{aligned} P(Z = n, X = k) &= P(X + Y = n, X = k) \\ &= P(k + Y = n, X = k) \\ &= P(Y = n - k, X = k) \\ &= P(Y = n - k)P(X = k), \end{aligned}$$

grâce à l'indépendance de X et Y.

Or $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ donc $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ si $k \geq 0$ (et $P(X=k)=0$ sinon), et $P(Y=n-k) = e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}$ si $n-k \geq 0$, soit $k \geq n$ (et $P(Y=n-k)=0$ sinon). Ainsi on voit que $P(X=k)P(Y=n-k)$ est non nul seulement si $0 \leq k \leq n$, et donc la somme sur k s'arrête en fait à $k=n$:

$$\begin{aligned} P(Z=n) &= \sum_{k \geq 0} P(Z=n, X=k) = \sum_{k=0}^{n} e^{-\lambda}\frac{\lambda^k}{k!} e^{-\mu}\frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-\lambda-\mu}\sum_{k=0}^{n}\frac{1}{k!(n-k)!}\lambda^k\mu^{n-k} \\ &= e^{-(\lambda+\mu)}\frac{1}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda^k\mu^{n-k} \\ &= e^{-(\lambda+\mu)}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k} \\ &= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}. \end{aligned}$$

Ceci correspond à la formule de la loi de Poisson de paramètre $\lambda+\mu$. Ainsi la loi de $X+Y$ est la loi de Poisson de paramètre $\lambda+\mu$.

Le principe de calcul qui vient d'être vu dans ces deux exemples se généralise : il s'agit de calculer la loi de la somme de deux variables discrètes indépendantes :

Proposition. *Soient X et Y deux variables discrètes indépendantes, et $Z = X+Y$. Alors la loi de Z se calcule à partir de celles de X et Y via la formule : pour tout $z \in \mathcal{S}(Z)$,*

$$P(Z=z) = \sum_{x \in \mathcal{S}(X)} P(X=x)P(Y=z-x).$$

Dans le cas de variables à densité, il existe une formule similaire, en remplaçant la somme par une intégrale :

Proposition. *Soient X et Y deux variables indépendantes admettant des densités f_X et f_Y, et $Z = X+Y$. Alors Z est une variable à densité et f_Z se calcule à partir de f_X et f_Y via la formule : pour tout $z \in \mathbf{R}$,*

$$f_Z(z) = \int_{\mathbf{R}} f_X(x)f_Y(z-x)dx.$$

On dit que f_Z est le **produit de convolution** de f_X et f_Y.

Voyons deux exemples d'utilisation de cette formule :

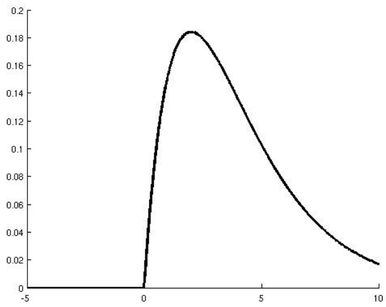


FIGURE 3.8 – Densité de la somme de 2 variables indépendantes exponentielles de même paramètre $a = 0.5$

exemple 3 :*Soient X et Y deux variables indépendantes de lois exponentielles de paramètre a. Calculer la loi de $Z = X + Y$.*

On a $f_X(x) = f_Y(x) = ae^{-ax}$ si $x \geq 0$, 0 sinon. La formule donne :

$$f_Z(z) = \int_{\mathbf{R}} f_X(x) f_Y(z-x) dx.$$

Or $f_X(x)f_Y(z-x)$ vaut $ae^{-ax}ae^{-a(z-x)}$ lorsque $x \geq 0$ et $z - x \geq 0$, soit $x \geq 0$ et $x \leq z$, soit encore $x \in [0, +\infty] \cap [-\infty, z]$. Sinon $f_X(x)f_Y(z-x)$ vaut 0. On a deux cas :

- Si $z < 0$ alors $[0, +\infty] \cap [-\infty, z] = \varnothing$ donc $f_Z(z) = 0$
- Si $z \geq 0$ alors $[0, +\infty] \cap [-\infty, z] = [0, z]$, donc

$$\begin{aligned} f_Z(z) &= \int_0^z ae^{-ax}ae^{-a(z-x)}dx = a^2 \int_0^z e^{-ax-az+ax}dx \\ &= a^2 \int_0^z e^{-az}dx = a^2 e^{-az} \int_0^z 1dx = a^2 z e^{-az}. \end{aligned}$$

Ainsi la densité de Z est $f_Z(z) = a^2 z e^{-az}$ si $z \geq 0$, 0 sinon. Cette fonction est représentée sur la figure 3.8 dans le cas $a = 0.5$.

exemple 4 :*Soient X et Y deux variables indépendantes de lois uniformes sur* $[0, 1]$ *Calculer la loi de $Z = X + Y$.*

X et Y sont des variables à densité et $f_X(x) = f_Y(x) = 1$ si $0 \leq x \leq 1$, 0 sinon. Z est donc une variable à densité d'après la proposition et sa densité est donnée par

$$f_Z(z) = \int_{\mathbf{R}} f_X(x) f_Y(z-x) dx.$$

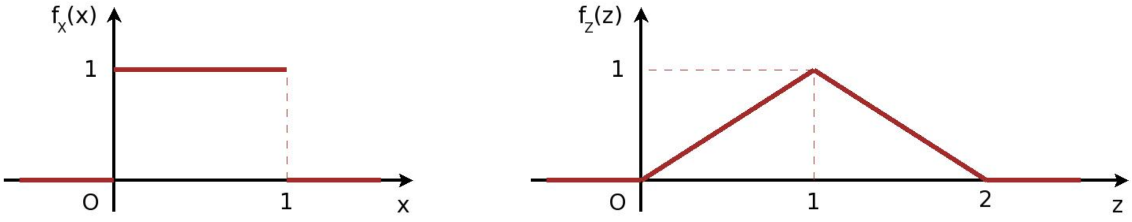


FIGURE 3.9 – à gauche : densité de la loi uniforme sur $[0,1]$. A droite : densité de la somme de deux variables indépendantes de lois uniformes sur $[0,1]$.

Or $f_X(x)f_Y(z-x)$ vaut 1 lorsque $0 \leq x \leq 1$ et $0 \leq z-x \leq 1$, soit $0 \leq x \leq 1$ et $z-1 \leq x \leq z$, soit encore $x \in [0,1] \cap [z-1,z]$. $f_X(x) = 0$ sinon. L'intersection $[0,1] \cap [z-1,z]$ est parfois vide. On a en fait plusieurs cas :

- Si $z < 0$ alors l'intersection est vide donc $f_Z(z) = 0$.
- Si $0 \leq z < 1$ alors $[0,1] \cap [z-1,z] = [0,z]$ donc $f_Z(z) = \int_0^z 1dx = z$.
- Si $1 \leq z \leq 2$ alors $[0,1] \cap [z-1,z] = [z-1,1]$ donc $f_Z(z) = \int_{z-1}^1 1dx = 1-(z-1) = 2-z$.
- Si $z \geq 2$ alors l'intersection est vide donc $f_Z(z) = 0$.

Le graphe de f_Z est représenté sur la figure 3.9.

3.7.2 Maximum ou minimum de variables indépendantes

exemple 3 : *Un planeur tombe dans un grand lac, que l'on suppose de taille carrée, de côté 10km. Le planeur dispose de feux de détresse pour signaler sa position, mais leur portée est de 4 km maximum. On voudrait évaluer la probabilité qu'il soit vu de la côte. On repère la position du planeur par des coordonnées X et Y (avec $0 \leq X \leq 10$ et $0 \leq Y \leq 10$). On suppose que les variables X et Y sont indépendantes et de même loi uniforme sur $[0,10]$, ce qui signifie intuitivement que la planeur a la même probabilité de se retrouver en tel ou tel endroit du lac. On note D la distance du planeur au bord du lac.*

1. *Calculer la loi de D.*
2. *Calculer $P(D \leq 4)$.*

1) La distance au bord du lac est le minimum des distances aux quatre côtés du lac, qui sont égales à $X, Y, 10-X, 10-Y$: ainsi,

$$D = \min\{X, 10-X, Y, 10-Y\}.$$

Le support de D est $[0,5]$. Pour calculer la loi de D nous allons déterminer sa fonction de

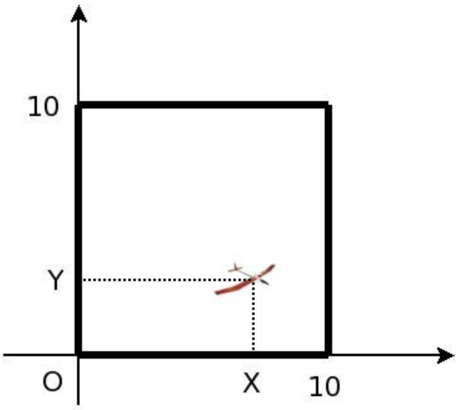


FIGURE 3.10 – Exemple 3 : position du planeur dans le lac

répartition : soit $t \in [0, 5]$:

$$\begin{array}{rcl} F_D(t) & = & P(D \leq t) \\ & = & P(X \leq t \text{ ou } 10 - X \leq t \text{ ou } Y \leq t \text{ ou } 10 - Y \leq t) \\ & = & P([X \leq t] \cup [10 - X \leq t] \cup [Y \leq t] \cup [10 - Y \leq t]) \\ & = & P([X \leq t] \cup [X \geq 10 - t] \cup [Y \leq t] \cup [Y \geq 10 - t]) \\ & = & 1 - P([X > t] \cap [X < 10 - t] \cap [Y > t] \cap [Y < 10 - t]) \\ & = & 1 - P([t < X < 10 - t] \cap [t < Y < 10 - t]). \end{array}$$

L'indépendance de X et Y permet alors d'écrire :

$$F_D(t) = 1 - P(t < X < 10 - t)P(t < Y < 10 - t).$$

Ces probabilités se calculent à partir de la densité de la loi uniforme sur $[0, 10]$:

$$\begin{array}{rcl} F_D(t) & = & 1 - \left(\int_t^{10-t} \frac{1}{10} du \right) \left(\int_t^{10-t} \frac{1}{10} du \right) \\ & = & 1 - \frac{(10-t)-t}{10} \frac{(10-t)-t}{10} = 1 - \frac{(10-2t)^2}{100} = 1 - \frac{100 + 4t^2 - 40t}{100} = \frac{4t(10-t)}{100}. \end{array}$$

On a ainsi calculé $F_D(t)$ pour $0 \leq t \leq 5$. A présent si $t < 0$, $F_D(t) = P(D \leq t) = 0$, et si $t > 5$, $F_D(t) = P(D \leq t) = 1$. Pour résumer, la fonction $F_D(t)$ est donnée par :

$$F_D(t) = \begin{cases} 0 \text{ si } t < 0 \\ \frac{4t(10-t)}{100} \text{ si } 0 \leq t \leq 5 \\ 1 \text{ si } t > 5 \end{cases}$$

Cette fonction est représentée sur la figure 3.11 à gauche.

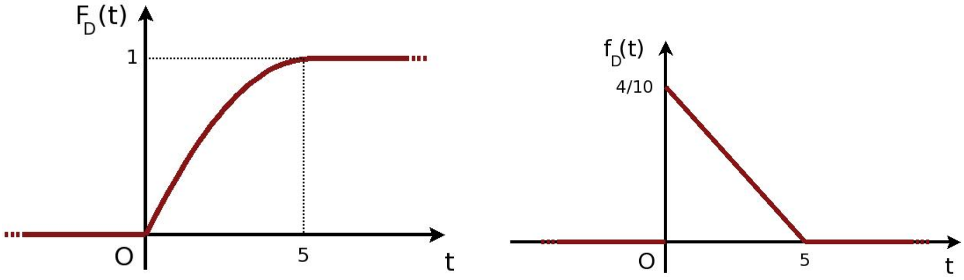


FIGURE 3.11 – exemple du planeur : fonction de répartition et densité de D

A présent pour calculer la densité de D il suffit de dériver cette fonction :

$$f_D(t) = F'_D(t) = \begin{cases} 0 \text{ si } t < 0 \\ \frac{4(10-2t)}{100} \text{ si } 0 \leq t \leq 5 \\ 1 \text{ si } t > 5. \end{cases}$$

Cette fonction est représentée sur la figure 3.11 à droite.

2) $P(D \leq 4) = F_D(4) = \frac{4*4*(10-4)}{100} = 96\%$

Chapitre 4

Lois jointes, lois conditionnelles.

On s'intéresse dans ce chapitre au calculs de probabilités d'événements faisant intervenir plusieurs variables aléatoires qui ne sont pas nécessairement indépendantes. L'idée fondamentale est qu'il ne suffit pas de connaître les lois de toutes les variables pour calculer de telles probabilités.

Exemple : A titre d'exemple, présentons trois expériences similaires : Dans les trois expériences on dispose d'un grand nombre de billes, de deux sacs notés A et B, vides au départ, et d'une pièce de monnaie.

- première expérience : On lance 10 fois successivement la pièce de monnaie ; à chaque fois, si la pièce tombe sur face, on met une bille dans le sac A, sinon on met une bille dans le sac B.
- deuxième expérience : On lance 10 fois successivement la pièce de monnaie ; à chaque fois, si la pièce tombe sur face, on met une bille dans le sac A et une bille dans le sac B, sinon on ne fait rien.
- troisième expérience : On lance 10 fois successivement la pièce de monnaie ; à chaque fois, si la pièce tombe sur face, on met une bille dans le sac A, sinon on ne fait rien. Puis on relance 10 fois la pièce et à chaque fois, si la pièce tombe sur face, on met une bille dans le sac B, sinon on ne fait rien.

Notons X le nombre de billes contenues dans le sac A à la fin de l'expérience, Y le nombre de billes dans le sac B, et Z le nombre de billes total dans les deux sacs. On a donc toujours $Z = X + Y$. Il est facile de voir que quelle que soit l'expérience, X et Y suivent toutes les deux la loi binomiale de paramètres $n = 10$ et $p = 0.5$. En effet chaque sac contient au plus 10 billes et le nombre de billes correspond au nombre de succès lors de 10 lancers successifs indépendants, avec probabilité de succès 0.5. Par contre la loi de Z n'est pas la même dans les trois expériences. En effet dans la premières expérience on a forcément $Z = 10$: on a placé 10 billes en tout. Dans la deuxième expérience, le nombre de billes dans B est égal au nombre de billes dans A, autrement dit $Y = X$, et donc $Z = 2X$. Les valeurs possibles de Z sont donc $0, 2, 4, \ldots, 20$, et sa loi est donnée par $P(Z = 2k) = P(X = k) = \binom{20}{k}0.5^k(1 - 0.5)^{20-k} = 0.5^{20}\binom{20}{k}$. Enfin dans la troisème expérience on peut voir que X et Y sont indépendantes et que la loi de Z est la loi binomiale de paramètres $n = 20$ et $p = 0.5$ (on fait 20 lancers de dé en tout, et on place une bille dans un des deux sacs à chaque fois qu'on tombe sur face).

Cet exemple montre que la loi d'une somme $Z = X + Y$ ne dépend pas uniquement des lois des deux variables X et Y. En fait lorsqu'une variable ou un événement dépend du résultat de deux variables X et Y discrètes, il faut pour faire des calculs connaître les probabilités $P([X = x] \cap [Y = y])$ pour toutes les valeurs possibles de X et Y, et ces probabilités ne se déduisent pas des probabilités $P(X = x)$ et $P(Y = y)$ à moins que les variables soient indépendantes. Ces probabilités $P([X = x] \cap [Y = y])$ forment ce qu'on appelle la loi jointe de X et Y, ou encore la loi du couple (X, Y).

4.1 Couples, triplets, vecteurs aléatoires

Définition. *Un* **couple aléatoire** *est un couple (X, Y) où X et Y sont des variables aléatoires. De même on parle de triplet aléatoire (X, Y, Z) si on considère trois variables X, Y, Z, et plus généralement de vecteur aléatoire, ou n-uplet aléatoire $(X_1, X_2, \ldots, X_n)$, en considérant n variables $X_1, X_2, \ldots, X_n$.*

Définition. *Le* **support d'un couple** *aléatoire (X, Y) est l'ensemble des valeurs prises par (X, Y), c'est-à-dire l'ensemble des couples de valeurs prises par X et Y. On le note $\mathcal{S}(X, Y)$, et il est donc égal à $\mathcal{S}(X) \times \mathcal{S}(Y)$.*

exemples :

- Si X et Y correspondent à des lancers de dés, on a $\mathcal{S}(X) = \mathcal{S}(Y) = \{1, 2, 3, 4, 5, 6\}$ et $\mathcal{S}(X, Y) = \{(x, y), x, y \in \{1, 2, 3, 4, 5, 6\}\} = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, ce qui se note aussi $\{1, 2, 3, 4, 5, 6\}^2$. Le cardinal de $\mathcal{S}(X, Y)$ (nombre de valeurs prises par (X, Y)) est donc $6^2 = 36$.
- Avec 10 lancers de dé, on obtient 10 variables $X_1, \ldots, X_{10}$, et donc un vecteur aléatoire $(X_1, \ldots, X_{10})$. Le support de ce vecteur est l'ensemble $\mathcal{S}((X_1, \ldots, X_{10})) = \{1, 2, 3, 4, 5, 6\}^{10}$, et donc $Card(\mathcal{S}((X_1, \ldots, X_{10}))) = 6^{10}$, soit environ 60 millions de valeurs possibles.

4.2 Loi jointe d'un couple de variables discrètes

Définition. *La* **loi d'un couple** *(X, Y) de variables discrètes est l'ensemble des probabilités $P((X, Y) = (x, y) = P(X = x$ et $Y = y)$ (souvent noté $P(X = x, Y = y)$) pour tous les couples $(x, y) \in \mathcal{S}(X, Y)$, c'est-à-dire pour tous les $x \in \mathcal{S}(X)$ et $y \in \mathcal{S}(Y)$.*

On peut présenter cette loi sous forme d'un tableau :

$x \backslash y$	y_1	$\cdots$	y_k	$\cdots$	Total (loi de X)
x_1	$P(X = x_1, Y = y_1)$	$\cdots$	$P(X = x_1, Y = y_k)$	$\cdots$	$P(X = x_1)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
x_n	$P(X = x_n, Y = y_1)$	$\cdots$	$P(X = x_n, Y = y_k)$	$\cdots$	$P(X = x_n)$
$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$
Total (loi de Y)	$P(Y = y_1)$	$\cdots$	$P(Y = y_k)$	$\cdots$	

En faisant les additions de chaque ligne et de chaque colonne, on obtient les lois de X et de Y. On a en fait les formules : pour tout $x \in \mathcal{S}(X)$,

$$P(X = x) = \sum_{y \in \mathcal{S}(Y)} P(X = x, Y = y),$$

qui vient du fait que l'événement $[X = x]$ est l'union disjointe des événements $[X = x] \cap [Y = y]$ pour tous les $y \in \mathcal{S}(Y)$. De même pour tout $y \in \mathcal{S}(Y)$,

$$P(Y = y) = \sum_{x \in \mathcal{S}(X)} P(X = x, Y = y).$$

On appelle les lois de X et de Y les **lois marginales du couple** (X, Y) (puisqu'on les lit à la "marge" du tableau). Souvent on les appelle aussi lois marginales de X et de Y, ce qui n'est pas logique (mais usuel).

En général on ne peut pas retrouver la loi du couple (X, Y) à partir des lois marginales. En revanche, si X et Y sont des variables indépendantes, alors on déduit la loi du couple des lois marginales : $P(X = x, Y = y) = P(X = x)P(Y = y)$.

exemple : On reprend l'exemple des deux sacs A et B. Les trois tableaux suivants correspondent aux trois cas proposés plus haut pour la loi du couple (X_A, X_B) (mais il y a une infinité d'autres possibilités) :

$x_A \backslash x_B$	0	1	
0	0.25	0.25	0.5
1	0.25	0.25	0.5
	0.5	0.5	

$x_A \backslash x_B$	0	1	
0	0	0.5	0.5
1	0.5	0	0.5
	0.5	0.5	

$x_A \backslash x_B$	0	1	
0	0.5	0	0.5
1	0	0.5	0.5
	0.5	0.5	

4.2.1 Formule de calcul d'une espérance

Lorsque l'on connait la loi d'un couple de variables il est possible de calculer l'espérance d'une fonction réelle f de ces deux variables. La formule pour des variables discrètes est :

$$E(f(X, Y)) = \sum_{x \in \mathcal{S}(X)} \sum_{y \in \mathcal{S}(Y)} f(x, y) P(X = x, Y = y).$$

4.3 Loi d'un vecteur de variables discrètes

La loi d'un vecteur $(X_1, X_2, \ldots, X_n)$ de variables discrètes est donnée par les probabilités $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ pour tous les $x_1 \in \mathcal{S}(X_1)$, $x_2 \in \mathcal{S}(X_2)$,... $x_n \in \mathcal{S}(X_n)$. Ces valeurs pourraient se placer dans un tableau à n dimensions (si on pouvait le dessiner !) de même type que celui obtenu pour un couple (X, Y).

4.4 Loi conditionnelle d'une variable

4.4.1 Conditionnement par rapport à un événement

Définition. *Soit X une variable et A un événement tel que $P(A) \neq 0$. La* **loi conditionnelle de X sachant** *A est la loi de X dans la situation où A est réalisé. Si X est une variable discrète, la loi de X sachant A est donnée par les probabilités $P(X = x|A)$ pour tous les $x \in \mathcal{S}(X)$.*

Définition. *L'***espérance conditionnelle de X sachant** *A est l'espérance de X dans la situation où A est réalisé. On la note $E(X|A)$ Dans le cas où X est une variable discrète, on a*

$$E(X|A) = \sum_{x \in \mathcal{S}(X)} xP(X = x|A).$$

4.4.2 Conditionnement par rapport à une autre variable

Soient X et Y deux variables discrètes. Pour tout $y \in \mathcal{S}(Y)$ tel que $P(Y = y) \neq 0$, on peut considérer la loi de X sachant $[Y = y]$, donnée par les valeurs $P(X = x|Y = y)$ pour tous les $x \in \mathcal{S}(X)$. L'espérance conditionnelle de X sachant $[Y = y]$ est donc

$$E(X|Y = y) = \sum_{x \in \mathcal{S}(X)} xP(X = x|Y = y).$$

Lien avec les lois jointes : On a $P(X = x|Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)}$. D plus, la formule des probabilités totales donne :

$$P(X = x) = \sum_{y \in \mathcal{S}(Y)} P(X = x|Y = y)P(Y = y).$$

Espérance conditionnelle de X sachant Y . On a défini $E(X|Y = y)$ pour tout $y \in \mathcal{S}(Y)$. Si maintenant on note $\psi(y) = E(X|Y = y)$, alors on peut définir $E(X|Y) = \psi(Y)$: c'est l'espérance conditionnelle de X sachant Y. Autrement dit, $E(X|Y)$ est la variable aléatoire telle que pour tout $y \in \mathcal{S}(Y)$, $E(X|Y) = E(X|Y = y)$ lorsque $Y = y$. Il faut faire très attention ici car $E(X|Y)$ est une variable aléatoire et non un nombre, à la différence de $E(X)$.

Formule de l'espérance totale :

$$E(E(X|Y)) = E(X).$$

L'intérêt de cette formule est que dans certaines situations, l'espérance de $E(X|Y)$ est plus simple à calculer que celle de X.

4.5 Covariance et corrélation

Définition. *Soient X et Y deux variables aléatoires dont les variances sont définies. La* **covariance** *de X et Y est*

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

La covariance est une mesure incomplète de l'indépendance de deux variables. En effet on a la résultat suivant :

Proposition. *Soient X et Y deux variables indépendantes. Alors* $\mathrm{Cov}(X, Y) = 0$.

Preuve. On fait la preuve pour des variables discrètes. Pour calculer $E(XY)$, on utilise la loi du couple (X, Y) : on fait la somme, pour tous les couples (x, y) de valeurs possibles, de xy multiplié par la probabilité que le couple prenne cette valeur :

$$\begin{aligned}
E(XY) &= \sum_{(x,y)\in\mathcal{S}(X,Y)} xyP(X = x, Y = y), \\
&= \sum_{x\in\mathcal{S}(X),y\in\mathcal{S}(Y)} xyP(X = x, Y = y), \\
&= \sum_{x\in\mathcal{S}(X)} \sum_{y\in\mathcal{S}(Y)} xyP(X = x, Y = y), \\
&= \sum_{x\in\mathcal{S}(X)} \sum_{y\in\mathcal{S}(Y)} xyP(X = x)P(Y = y) \text{ car } X \text{ et } Y \text{ sont indépendantes}, \\
&= (\sum_{x\in\mathcal{S}(X)} xP(X = x))(\sum_{y\in\mathcal{S}(Y)} yP(Y = y)) \\
&= E(X)E(Y).
\end{aligned}$$

□

Attention : La réciproque de la proposition est fausse en général : $cov(X, Y) = 0$ n'implique pas que X et Y sont indépendantes.

Proposition. *Pour toutes variables aléatoires X et Y,*

$$|Cov(X, Y)| \leq \sigma(X)\sigma(Y).$$

Ce résultat est en fait une conséquence immédiate de l'inégalité de Cauchy-Schwartz :

Inégalité de Cauchy-Schwarz Soient X, Y deux variables aléatoires dont les variances sont bien définies. Alors

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

Preuve. Soit $t \in \mathbf{R}$. On pose $f(t) = E((X + tY)^2)$. On a

$$\begin{aligned} f(t) &= E(X^2 + t^2Y^2 + 2tXY) = E(X^2) + t^2E(Y^2) + 2tE(XY), \\ &= E(Y^2)t^2 + 2E(XY)t + E(X^2). \end{aligned}$$

Ainsi $f(t) = at^2 + bt + c$, avec $a = E(Y^2)$, $b = 2E(XY)$, et $c = E(X^2)$. Or $f(t)$ est toujours positif, puisque $f(t) = E((X + tY)^2)$ (espérance d'une variable positive). Donc le polynôme $at^2 + bt + c$ ne peut pas avoir deux racines distinctes réelles (sinon son signe change lorsque t parcourt $\mathbf{R}$), et donc son discriminant est négatif ou nul : $\Delta = b^2 - 4ac \leq 0$. En remplaçant a, b, c par leurs valeurs, on obtient $\Delta = 4E(XY)^2 - 4E(Y^2)E(X^2) \leq 0$, et donc $E(XY)^2 \leq E(Y^2)E(X^2)$, soit encore $E(XY) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$. $\square$

Pour prouver la proposition il suffit d'appliquer cette inégalité à $X - E(X)$ et $Y - E(Y)$.

Définition. *La* **corrélation** *entre deux variables X et Y est définie par*

$$\rho(X, Y) = \frac{\mathrm{C}ov(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\mathrm{C}ov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

On a donc toujours $-1 \leq \rho(X, Y) \leq 1$ d'après la proposition précédente. Lorsque $\rho(X, Y) = 0$, c'est-à-dire lorsque $\mathrm{C}ov(X, Y) = 0$, on dit que X et Y sont décorrélées.

Proposition. *Soient X, Y deux variables de variances finies, et a, b deux réels fixés.*

$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\mathrm{C}ov(X, Y).$$

En particulier on a les formules : $V(X+Y) = V(X)+V(Y)+2\mathrm{C}ov(X, Y)$, $V(X-Y) = V(X) + V(Y) - 2\mathrm{C}ov(X, Y)$. Si X et Y sont décorrélées (donc en particulier si elles sont indépendantes), alors $V(X + Y) = V(X - Y) = V(X) + V(Y)$.

Matrice de variance-covariance. Pour un vecteur aléatoire $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, on peut définir toutes les covariances $\mathrm{C}ov(X_i, X_j)$ pour $1 \leq i \leq n$ et $1 \leq j \leq n$ et les rassembler dans une matrice :

$$\Gamma(\mathbf{X}) = \begin{pmatrix} V(X_1) & \mathrm{C}ov(X_1, X_2) & \mathrm{C}ov(X_1, X_3) & \cdots & \mathrm{C}ov(X_1, X_n) \\ \mathrm{C}ov(X_1, X_2) & V(X_2) & \mathrm{C}ov(X_2, X_3) & \cdots & \mathrm{C}ov(X_2, X_n) \\ \mathrm{C}ov(X_1, X_3) & \mathrm{C}ov(X_2, X_3) & V(X_3) & \cdots & \mathrm{C}ov(X_3, X_n) \\ \vdots & & & \ddots & \vdots \\ \mathrm{C}ov(X_1, X_n) & \mathrm{C}ov(X_2, X_n) & \mathrm{C}ov(X_3, X_n) & \cdots & V(X_n) \end{pmatrix}.$$

Cette matrice est symétrique, donc diagonalisable et à valeurs propres positives.

Notations matricielles pour l'espérance et la variance de vecteurs aléatoires :

Soit $\mathbf{X}$ un vecteur aléatoire : $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. On peut le noter sous forme d'un vecteur colonne : $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$. Son espérance s'écrit alors $E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix}$, et sa matrice de variance covariance est égale à $\Gamma(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^t) - E(\mathbf{X})E(\mathbf{X})^t$. En effet,

$$\mathbf{X}\mathbf{X}^t = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix} = \begin{pmatrix} X_1^2 & X_1X_2 & X_1X_3 & \cdots & X_1X_n \\ X_1X_2 & X_2^2 & X_2X_3 & \cdots & X_2X_n \\ X_1X_3 & X_2X_3 & X_3^2 & \cdots & X_3X_n \\ \vdots & & & \ddots & \vdots \\ X_1X_n & X_2X_n & X_3X_n & \cdots & X_n^2 \end{pmatrix},$$

donc

$$E(\mathbf{X}\mathbf{X}^t) = \begin{pmatrix} E(X_1^2) & E(X_1X_2) & E(X_1X_3) & \cdots & E(X_1X_n) \\ E(X_1X_2) & E(X_2^2) & E(X_2X_3) & \cdots & E(X_2X_n) \\ E(X_1X_3) & E(X_2X_3) & E(X_3^2) & \cdots & E(X_3X_n) \\ \vdots & & & \ddots & \vdots \\ E(X_1X_n) & E(X_2X_n) & E(X_3X_n) & \cdots & E(X_n^2) \end{pmatrix}.$$

et

$$\begin{aligned} E(\mathbf{X})E(\mathbf{X})^t &= \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} \begin{pmatrix} E(X_1) & E(X_2) & \cdots & E(X_n) \end{pmatrix} \\ &= \begin{pmatrix} E(X_1)^2 & E(X_1)E(X_2) & E(X_1)E(X_3) & \cdots & E(X_1)E(X_n) \\ E(X_1)E(X_2) & E(X_2)^2 & E(X_2)E(X_3) & \cdots & E(X_2)E(X_n) \\ E(X_1)E(X_3) & E(X_2)E(X_3) & E(X_3)^2 & \cdots & E(X_3)E(X_n) \\ \vdots & & & \ddots & \vdots \\ E(X_1)E(X_n) & E(X_2)E(X_n) & E(X_3)E(X_n) & \cdots & E(X_n)^2 \end{pmatrix}. \end{aligned}$$

En soustrayant ces deux matrices, on retrouve l'expression de la matrice de variance-covariance.

Chapitre 5

Théorèmes limites et estimation

5.1 Exemples introductifs

Voici deux exemples de situations.

a) *On lance n fois un dé à six faces et on note X_i le résultat du i^e lancer, puis on calcule la moyenne des n lancers $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Qu'obtient-on lorsque n devient grand ?*

L'intuition laisse penser que lorsque n devient grand, cette moyenne devient proche de l'espérance d'un des lancers X_i, c'est-à-dire de $E(X_i) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3.5$. On voudrait donc pouvoir dire que la limite de $\bar{X}_n$ lorsque $n \to +\infty$ vaut 3.5. C'est ce qu'on appelle la loi des grands nombres. Pour cela, il faut d'abord définir ce que signifie converger pour une suite de variables aléatoires, puis prouver qu'il y a bien convergence.

b) *Cette fois le dé dont on dispose est truqué : il a une probabilité p inconnue de tomber sur 6, et $\frac{1-p}{5}$ sur chacune des autres faces. Comment connaître la valeur de p simplement à partir des résultats obtenus en lançant n fois le dé ?*

Si la loi des grands nombres est correcte, on peut l'utiliser ici et dire que la moyenne $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ des résultats est proche de l'espérance $E(X_i) = (1+2+3+4+5)\times\frac{1-p}{5}+6p = 3(1-p)+6p = 3(p+1)$. Or si $\bar{X}_n \simeq 3(p+1)$, c'est que $\frac{1}{3}\bar{X}_n \simeq p+1$, donc que $\frac{1}{3}\bar{X}_n - 1 \simeq p$. Ainsi une valeur approchée de p est donnée par $\frac{1}{3}\bar{X}_n - 1$. On a donc donné une estimation de p, et on dit que $\frac{1}{3}\bar{X}_n - 1$ est un **estimateur** de p car c'est une quantité que l'on peut calculer à partir des résultats des lancers, et qui donne une valeur approchée de la valeur inconnue p.

5.2 Convergence de variables aléatoires

remarque : Il existe plusieurs notions de convergence de variables aléatoires. Dans ce cours nous ne donnerons qu'une seule définition, qui correspond à la convergence presque sûre en théorie des probabilités.

De la même manière que l'on peut s'intéresser à la convergence d'une suite $(x_n)_{n\geq 0}$ de nombres réels déterminés, on peut aussi regarder la convergence d'une suite $(X_n)_{n\geq 0}$ de variables aléatoires, mais cette convergence aura forcément un caractère aléatoire puisque les X_n sont aléatoires. Autrement dit, si X est une autre variable, on peut définir l'événement $[X_n \underset{n\to\infty}{\longrightarrow} X]$ ="la suite $(X_n)_{n\geq 0}$ converge vers X".

Définition. *On dit que $(X_n)_{n\geq 0}$ converge vers X si $P\left(X_n \underset{n\to\infty}{\longrightarrow} X\right) = 1$.*

5.3 Moyenne empirique et loi des grands nombres

5.3.1 Moyenne empirique d'une suite de variables

Définition. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires. On appelle* **moyenne empirique** *(des n premiers termes de la suite) la variable aléatoire*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Proposition. *Soit $(X_n)_{n\geq 0}$ une suite de variables indépendantes et toutes de même loi. On suppose que l'espérance et la variance communes à tous les X_n : $\mu = E(X_n)$, $\sigma^2 = V(X_n)$, sont bien définies. Alors pour tout $n \geq 0$,*

$$E(\bar{X}_n) = \mu \qquad \text{et} \qquad V(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Preuve.

- $E(\bar{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \times n\mu$ puisque tous les X_i ont même espérance. Donc $E(\bar{X}_n) = \mu$.
- $V(\bar{X}_n) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}V(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$ car les X_i sont des variables indépendantes. Donc $V(\bar{X}_n) = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$ puisque tous les X_i ont même variance.

$\square$

Ce résultat est très important car il montre que lorsque n tend vers $+\infty$, la variance de $\bar{X}_n$ tend vers 0, ce qui laisse penser que $\bar{X}_n$ converge vers un nombre non aléatoire, égal à son espérance $E(\bar{X}_n) = \mu$. C'est précisément la loi des grands nombres.

5.3.2 La loi des grands nombres

Théorème 1. *Soit $(X_n)_{n\geq 0}$ une suite de variables indépendantes et toutes de même loi. On suppose que l'espérance $\mu = E(X_n)$ est bien définie. Alors $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge vers μ.*

La preuve de ce théorème est très difficile ; il n'est pas question de la faire dans ce cours.

Voici quelques exemples d'application de la loi des grands nombres :

exemple 1 : Pour une série de lancers de dés (non truqués), si X_i représente le résultat du i^e lancer, alors les X_i sont des variables indépendantes et toutes de même loi (loi uniforme sur $\{1,2,3,4,5,6\}$). L'espérance des X_i est bien définie, donc on peut appliquer la loi des grands nombres : la moyenne $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge vers 3.5 lorque n tend vers $+\infty$: $\bar{X}_n \underset{n\to\infty}{\longrightarrow} 3.5$.

exemple 2 : On suppose qu'à chaque minute il y a une probabilité p qu'un client entre dans une banque (un seul client peut entrer par minute, et toutes les arrivées de clients sont indépendantes) et on note X_n le temps d'arrivée du n^e client. En fait X_n est la somme de n variables Y_i, où Y_i représente le temps d'attente entre les clients $i-1$ et i. Les Y_i sont des variables indépendantes et de même loi géométrique de paramètre p. La loi géométrique admet une espérance, donc on peut appliquer la loi des grands nombres : la moyenne empirique $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}X_n$ converge vers $E(Y_1) = \frac{1}{p}$. Autrement dit, lorsque n est grand, $\frac{1}{n}X_n$ est proche de $\frac{1}{p}$, et donc le temps d'arrivée X_n du n^e client est proche de $\frac{n}{p}$.

5.4 Estimation ponctuelle de l'espérance

Supposons à présent que X est le résultat d'une expérience aléatoire, et que l'espérance $E(X) = \mu$ est inconnue, comme dans l'exemple du dé truqué. La loi des grands nombres montre que l'on peut obtenir une valeur approchée de μ en réalisant un grand nombre n de fois la même expérience, puis en calculant la moyenne $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, où X_i correspond au résultat de la i^e expérience. Il s'agit ici d'un problème d'estimation, et nous allons introduire un peu de vocabulaire relatif à ce contexte :

Définition. *Un* **paramètre inconnu** λ *est un nombre réel, fixé mais inconnu, dont on voudrait connaître la valeur.*

Par exemple si l'espérance $\mu = E(X)$ d'une variable est inconnue, μ est un paramètre inconnu. Dans l'exemple du dé truqué, la probabilité p que le dé tombe sur 6 est un paramètre inconnu.

Définition. *Un* **échantillon** $(X_1, X_2, \ldots, X_n)$ *de même loi que X est un vecteur aléatoire tel que toutes les variables X_i sont indépendantes et ont toutes même loi que X.*

Définition. *Un* **estimateur** *du paramètre λ est variable aléatoire L_n fonction de l'échantillon, c'est-à-dire fonction des X_i pour $1 \leq i \leq n$.*

Cette notion seule est donc vide de sens, puisqu'on ne demande pas a priori que l'estimateur L_n soit proche de λ.

Définition. *On dit qu'un estimateur L_n du paramètre λ est* **convergeant** *ou* **consistant** *si L_n converge vers λ.*

Définition. *On dit qu'un estimateur L_n de λ est* **sans biais** *si $E(L_n) = \lambda$.*

En utilisant ce vocabulaire de l'estimation, la loi des grands nombres se traduit donc ainsi : la moyenne $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur du paramètre $\mu = E(X_1)$ convergeant. C'est aussi un estimateur sans biais car $E(\bar{X}_n) = \mu$.

Dans l'exemple du dé truqué, la moyenne $\bar{X}_n$ est donc un estimateur de $\mu = E(X_1)$ convergeant. Cependant dans cet exemple on cherchait un estimateur du paramètre p et non pas de μ. On avait vu que $\mu = 3(p+1)$, donc l'estimateur de p est en fait $P_n = \frac{\bar{X}_n}{3} - 1$. On vérifie facilement qu'il est convergeant et sans biais :

$$P\left(P_n \underset{n\to\infty}{\longrightarrow} p\right) = P\left(\frac{\bar{X}_n}{3} - 1 \underset{n\to\infty}{\longrightarrow} p\right) = P\left(\bar{X}_n \underset{n\to\infty}{\longrightarrow} 3(p+1)\right) = P\left(\bar{X}_n \underset{n\to\infty}{\longrightarrow} \mu\right) = 1,$$

$$E(P_n) = E\left(\frac{\bar{X}_n}{3} - 1\right) = \frac{E(\bar{X}_n)}{3} - 1 = \frac{\mu}{3} - 1 = p.$$

5.5 Estimation ponctuelle de la variance

Cette fois on suppose que la variance $\sigma^2 = V(X)$ d'une variable X est le paramètre inconnu que l'on cherche à estimer. Deux cas peuvent se présenter, suivant que l'espérance est inconnue elle aussi, ou pas.

5.5.1 Estimation avec espérance connue

Si la valeur de μ est connue, on peut l'utiliser pour calculer l'estimateur.

Proposition. $S_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2$ *est un estimateur du paramètre σ^2 convergeant et sans biais.*

Preuve. Si on pose $Y_i = (X_i - \mu)^2$. Les variables Y_i sont indépendantes et identiquement distribuées, et leur espérance est bien définie puisque :

$$E(Y_i) = E((X_i - \mu)^2) = V(X_i) = \sigma^2.$$

On peut donc appliquer aux Y_i la loi des grands nombres : $S_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$ tend vers $E(Y_1) = \sigma^2$ □

5.5.2 Estimation avec espérance inconnue

Si la valeur de μ est inconnue, on ne peut plus l'utiliser dans l'estimation. On va donc remplacer sa valeur par son estimateur :

Proposition. $S'_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ *est un estimateur de* σ^2 *convergeant et biaisé.*

Preuve. On va d'abord écrire S'_n différemment :

$$\begin{aligned}
S'_n &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 + \bar{X}_n^2 - 2X_i\bar{X}_n) \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 + \frac{1}{n}\sum_{i=1}^{n}\bar{X}_n^2 - 2\bar{X}_n\frac{1}{n}\sum_{i=1}^{n}X_i \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 + \frac{1}{n}n\bar{X}_n^2 - 2\bar{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}_n^2.
\end{aligned}$$

On va regarder la convergence de $\frac{1}{n}\sum_{i=1}^{n}X_i^2$, puis celle de $\bar{X}_n^2$:

- D'après la loi des grands nombres $\frac{1}{n}\sum_{i=1}^{n}X_i^2$ converge vers $E(X_1^2)$, qui est égal à $V(X_1) + E(X_1)^2 = \sigma^2 + \mu^2$.
- D'après la loi des grands nombres, $\bar{X}_n$ converge vers μ, et donc $\bar{X}_n^2$ converge vers μ^2 : en effet,
$$P(\bar{X}_n^2 \underset{n\to\infty}{\longrightarrow} \mu^2) = P(\bar{X}_n \underset{n\to\infty}{\longrightarrow} \mu) = 1.$$

Par conséquent S'_n converge vers $\sigma^2 + \mu^2 - \mu^2 = \sigma^2$. Ensuite,

$$\begin{aligned}
E(S'_n) &= E(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}_n^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}E(X_i^2) - E(\bar{X}_n^2) \\
&= \frac{1}{n}n(\sigma^2 + \mu^2) - (\frac{1}{n}\sigma^2 + \mu^2) \\
&= \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Donc l'espérance de $E(S'_n)$ ne vaut pas σ^2 mais $\frac{n-1}{n}\sigma^2$: c'est un estimateur biaisé. $\square$

Puisque $E(S'_n) = \frac{n-1}{n}\sigma^2$, il suffit, pour obtenir un estimateur non biaisé, de prendre $\frac{n}{n-1}S'_n$:

Proposition. $S_n{}'' = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ *est un estimateur de* σ^2 *convergeant et sans biais.*

5.6 Loi normale et théorème central limite

5.6.1 La loi normale

Définition. *Une variable aléatoire* X *suit la loi normale* $\mathcal{N}(\mu, \sigma^2)$ *si elle admet la densité*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

La fonction de répartition $F_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$ de la loi normale n'est pas calculable de façon explicite. Cependant comme cette fonction et sont inverse sont utilisées constamment dans les calculs, des valeurs approchées de F_X et F_X^{-1} pour $X \sim \mathcal{N}(0,1)$ et pour toutes les valeurs de x ont été tabulées depuis longtemps. Dans une application numérique, lors d'un calcul avec la loi normale, on se ramènera toujours à exprimer le résultat en fonction de F_X ou de F_X^{-1} où $X \sim \mathcal{N}(0,1)$. Le résultat suivant montre que l'on peut toujours se ramener à cette loi :

Proposition. *Si* X *suit une loi* $\mathcal{N}(\mu, \sigma^2)$ *et* a, b *sont des réels, alors* $aX + b$ *suit une loi* $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

Par conséquent si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Proposition. *Soient* X *et* Y *deux variables indépendantes telles que* $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ *et* $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. *Alors* $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

5.6.2 Théorème central limite

Théorème 2. *Soit* $(X_n)_{n\geq 0}$ *une suite de variables aléatoires indépendantes et identiquement distribuées, et* X *une variable de loi* $\mathcal{N}(0,1)$. *On suppose que l'espérance* $\mu = E(X_1)$ *et la variance* $\sigma^2 = V(X_1)$ *sont bien définies. Alors pour tous* a, b *réels tels que* $a < b$,

$$P\left(a \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq b\right) \underset{n\to\infty}{\longrightarrow} P(a \leq X \leq b) = \int_a^b f_X(x)dx = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.$$

Autrement dit, ce théorème dit que la loi de

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$$

converge vers celle de X, c'est-à-dire que tout calcul de probabilité effectué sur cette variable peut être approché par un calcul de probabilité sur X.

remarque : l'expression $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$ peut s'écrire de différentes façons : souvent on note $S_n = \sum_{i=1}^{n} X_i$, de sorte que $\bar{X}_n = \frac{1}{n}S_n$, et donc $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}S_n - \mu\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}$.

5.7 Approximations de la loi binomiale

Soit X une variable de loi binomiale $B(n,p)$: on a, pour tout $k \in \{1,\ldots,n\}, P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$. Lorsque n est grand, les coefficients $\binom{n}{k}$ deviennent difficile à calculer numériquement. On cherche donc à approcher la loi binomiale par des lois plus simples à calculer.

5.7.1 Approximation par la loi de Poisson

Proposition. *a) Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires telle que pour tout n la loi de X_n est $B(n,\frac{\lambda}{n})$ où λ est un nombre positif fixé. Soit X une variable de loi de Poisson de paramètre λ. Alors la loi de X_n converge vers la loi de X.*

b) Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires telle que pour tout n la loi de X_n est $B(n,1-\frac{\lambda}{n})$ où λ est un nombre positif fixé. Soit X une variable de loi de Poisson de paramètre λ. Alors la loi de $n-X_n$ converge vers la loi de X.

Preuve.

a)

$$\begin{aligned}
P(X_n=k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\
&= \frac{n!}{k!\,(n-k)!}\frac{\lambda^k}{n^k}\frac{(n-\lambda)^{-k}}{n^{-k}}\left(1-\frac{\lambda}{n}\right)^n\\
&= \frac{\lambda^k}{k!}\frac{n!}{(n-k)!}\frac{1}{(n-\lambda)^k}\left(1-\frac{\lambda}{n}\right)^n\\
&= \frac{\lambda^k}{k!}\frac{n\,(n-1)\cdots(n-k+1)}{(n-\lambda)\,(n-\lambda)\cdots(n-\lambda)}\left(1-\frac{\lambda}{n}\right)^n\\
&= \frac{\lambda^k}{k!}\frac{n}{(n-\lambda)}\frac{(n-1)}{(n-\lambda)}\cdots\frac{(n-k+1)}{(n-\lambda)}\left(1-\frac{\lambda}{n}\right)^n.
\end{aligned}$$

Or $\frac{n}{(n-\lambda)}$ tend vers 1 lorsque n tend vers l'infini, de même que $\frac{(n-1)}{(n-\lambda)},\ldots,\frac{(n-k+1)}{(n-\lambda)}$. De plus $\left(1-\frac{\lambda}{n}\right)^n = e^{n\ln\left(1-\frac{\lambda}{n}\right)} = e^{n\left(-\frac{\lambda}{n}+o\left(\frac{1}{n}\right)\right)} = e^{-\lambda+o(1)} \underset{n\to\infty}{\longrightarrow} e^{-\lambda}$. Ainsi $P\left(X_n=k\right) \underset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}e^{-\lambda} = P\left(X=k\right)$ pour tout $k\geq 0$, et donc la loi de X_n converge vers celle de X.

b) $n-X_n$ suit une loi $B\left(n,\frac{\lambda}{n}\right)$: en effet,
$P\left(n-X_n=k\right) = P\left(X_n=n-k\right) = \binom{n}{n-k}\left(1-\frac{\lambda}{n}\right)^{n-k}\left(\frac{\lambda}{n}\right)^k = \binom{n}{k}\left(1-\frac{\lambda}{n}\right)^{n-k}\left(\frac{\lambda}{n}\right)^k$.
Donc d'après a), la loi de $n-X_n$ converge vers la loi de X.

□

En pratique :

- Si $X\sim B(n,p)$ avec $n\geq 30$ et $p\leq 0.2$, alors on approche la loi de X par une loi de Poisson(np).

- Si $X \sim B(n,p)$ avec $n \geq 30$ et $p \geq 0.8$, alors on approche la loi de $n - X$ par une loi de Poisson($n(1-p)$).

exemple : *A chaque minute un client peut entrer dans un magasin avec une probabilité* $p = 0.05$. *On suppose qu'un seul client peut entrer à chaque minute et que toutes les entrées de clients sont indépendantes. Quelle est la probabilité qu'exactement* 5 *clients entrent en une heure ?*

Soit X le nombre de clients entrés en une heure. X suit une loi binomiale $B(60, 0.05)$. Puisque $60 \geq 30$ et $60 * 0.05 = 3 \leq 5$, on peut approcher la loi de X par une loi de Poisson(3). Par conséquent, $P(X = 5) \simeq e^{-3}\frac{3^5}{5!} \simeq 0.1$.

5.7.2 Approximation par la loi normale

Si X_n suit une loi binomiale $B(n,p)$, alors $X_n = \sum_{i=1}^n Y_i$ où les Y_i sont des variables indépendantes de loi de Bernoulli(p) : $P(Y_i = 1) = p$ et $P(Y_i = 0) = 1 - p$. D'après le théorème central limite, $\frac{\sqrt{n}}{\sigma}(\frac{1}{n}\sum_{i=1}^n Y_i - \mu)$ converge en loi vers $\mathcal{N}(0,1)$, avec $\mu = E(Y_1) = p$ et $\sigma^2 = V(Y_1) = p(1-p)$. Autrement dit, on a le résultat suivant :

Proposition. *Soit* $(X_n)_{n\geq 0}$ *une suite de variables aléatoires telle que pour tout* n *la loi de* X_n *est* $B(n,p)$ *où* p *est fixé. Soit* X *une variable de loi* $\mathcal{N}(0,1)$. *Alors la loi de* $\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\frac{1}{n}X_n - p)$ *converge vers celle de* X.

En pratique : Si $X \sim B(n,p)$ avec $n \geq 30$ et $0.2 < p < 0.8$, alors on approche la loi de $\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\frac{1}{n}X - p)$ par une loi normale $\mathcal{N}(0,1)$, ou encore la loi de X par une loi normale $\mathcal{N}(np, np(1-p))$. ceci vient du fait que si $\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\frac{1}{n}X - p)$ suit une loi $\mathcal{N}(0,1)$, alors X suit une loi $\mathcal{N}(np, np(1-p))$.

exemple : Reprenons l'exemple du magasin, mais supposons cette fois que $p = 0.5$. Alors $60 * 0.5 = 30$ donc on approche la loi de $\frac{\sqrt{60}}{\sqrt{0.5(1-0.5)}}(\frac{1}{60}X - 0.5) = 2\sqrt{60}(\frac{X}{60}X - 0.5)$ par une loi normale $\mathcal{N}(0,1)$. Selon cette approximation, $P(X = 5) \simeq 0$ puisque la loi normale est une loi à densité. Pour obtenir des probabilités non négligeables, on se pose plutôt une question telle que "Quelle est la probabilité qu'en une heure soient entrés plus de 20 clients ?". On a alors $P(X \geq 20) = P(2\sqrt{60}(\frac{X}{60} - 0.5) \geq 2\sqrt{60}(\frac{20}{60} - 0.5) = P(2\sqrt{60}(\frac{X}{60} - 0.5) \geq -\frac{\sqrt{60}}{3}) \simeq 1 - F(-\frac{\sqrt{60}}{3})$, où F est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$.

5.8 Estimation par intervalles de confiance

Définition. *Soit* X *une variable aléatoire et* $(X_1, \ldots, X_n)$ *un échantillon de même loi que* X. *Estimer un paramètre inconnu* λ *par* **intervalle de confiance** *au niveau* $1 - \alpha$ *consiste à trouver deux variables aléatoires* A_n *et* B_n *fonctions de l'échantillon telles que* $P(A_n \leq \lambda \leq B_n) \geq 1 - \alpha$.

5.8.1 Intervalle de confiance grâce à l'inégalité de Bienaymé-Tchebychev

Proposition (Inégalité de Bienaymé-Tchebychev)**.** *Soit une variable aléatoire X de variance finie, et a un réel. Alors*

$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}.$$

Preuve. $V(X) = E((X - E(X))^2)$ et $P(|X - E(X)| \geq a) = P((X - E(X))^2 \geq a^2)$. Donc si on pose $Y = (X - E(X))^2$, l'inégalité à démontrer est équivalente à $P(Y \geq a^2) \leq \frac{E(Y)}{\alpha^2}$. Posons alors $Z = \begin{cases} a^2 \text{ si } Y \geq a^2 \\ 0 \text{ sinon.} \end{cases}$ On a $Z \leq Y$ donc $E(Z) \leq E(Y)$. Or $E(Z) = a^2 P(Y \geq a^2) + 0 \times P(Y < a^2) = a^2 P(Y \geq a^2)$. Donc $a^2 P(Y \geq a^2) \leq E(Y)$. $\square$

Cette inégalité permet d'obtenir un intervalle de confiance pour l'estimation de l'espérance d'une variable X. L'estimateur ponctuel de l'espérance est la moyenne $\bar{X}_n$. On va donc construire un intervalle de confiance autour de cet estimateur. D'après l'inégalité de Bienaymé-Tchebychev, on a $P(|\bar{X}_n - \mu| \geq a) \leq \frac{V(\bar{X}_n)}{a^2}$. Or $V(\bar{X}_n) = \frac{\sigma^2}{n}$. Donc $P(-a \leq \bar{X}_n - \mu \leq a) \leq 1 - \frac{\sigma^2}{na^2}$, soit encore $P(\bar{X}_n - a \leq \mu \leq \bar{X}_n + a) \leq 1 - \frac{\sigma^2}{na^2}$. Supposons que l'on cherche un intervalle de confiance au niveau $1 - \alpha = 0.95$, avec $n = 1000$ et $\sigma = 1$ alors on aura $\alpha = \frac{\sigma^2}{na^2}$, donc $a = \frac{\sigma}{\sqrt{n\alpha}} = 0.1$. L'intervalle de confiance obtenu par l'inégalité de Bienaymé-Tchebychev n'est pas très précis. On peut avoir une meilleure estimation grâce au théorème central limite.

5.8.2 Intervalle de confiance grâce au théorème central limite

D'après le théorème central limite,

$$P\left(-a \leq \frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) \leq a\right) \simeq P\left(-a \leq X \leq a\right) = F_X\left(a\right) - F_X\left(-a\right),$$

où X suit une loi $\mathcal{N}(0,1)$. Or,

$$\begin{aligned} P\left(-a \leq \frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) \leq a\right) &= P\left(-\frac{\sigma}{\sqrt{n}}a \leq \bar{X}_n - \mu \leq \frac{\sigma}{\sqrt{n}}a\right) \\ &= P\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}a \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}a\right). \end{aligned}$$

Pour obtenir un intervalle de confiance, on fixe alors un niveau $1 - \alpha$, et on cherche a tel que $F_X\left(a\right) - F_X\left(-a\right) = 1 - \alpha$. En fait la loi normale étant symétrique, on a $F_X\left(-a\right) = 1 - F_X\left(a\right)$ et donc $F_X\left(a\right) - F_X\left(-a\right) = 2F_X\left(a\right) - 1$, et donc $F_X\left(a\right) - F_X\left(-a\right) = 1 - \alpha$ implique $2F_X\left(a\right) - 1 = 1 - \alpha$ donc $F_X\left(a\right) = 1 - \frac{\alpha}{2}$, donc $a = F_X^{-1}\left(1 - \frac{\alpha}{2}\right)$.