

Probabilités

Maxime Forriez^{1,2, a} and Maxime Forriez^{1,2,a}

¹ Sorbonne université, 2, rue Francis de Croisset, 75 018 Paris

² Institut de géographie, 191, rue Saint-Jacques, Bureau 105,
75 005 Paris,

^amaxime.forriez@sorbonne-universite.fr

24 septembre 2025

1 Les axiomes de Kolmogorov

Le calcul des probabilités est l'étude des phénomènes aléatoires ou non déterministes. La théorie moderne des probabilités repose sur les **quatre axiomes de A. N. Kolmogorov**.

La validité du modèle mathématique pour une **expérience** donnée dépend de la similitude des probabilités attribuées avec les fréquences relatives réelles. Ainsi, probabilités et statistiques se lient, notamment avec des tests et des estimations.

Une **expérience aléatoire** a un résultat impossible à connaître à l'avance.

Exemple. Une population statistique permet de réaliser une infinité d'expériences aléatoires.

L'ensemble de tous les résultats S est appelé un **univers**, ou **ensemble fondamental**, ou **ensemble des possibilités**. Il correspond à l'ensemble des réalisations potentielles. Tout élément appartenant à S est une **éventualité** e . Un sous-ensemble de S est appelé un **événement** E . Tout événement possède un **complément** (ou un **contraire**), c'est-à-dire si e n'est pas réalisé, noté $\neg E$ ou E^C ou S/E . Une éventualité peut aussi être désignée comme un événement élémentaire.

\emptyset est l'**événement impossible**.

S est l'**événement certain**.

Soient A et B deux événements alors :

- $A \cup B$ est l'événement qui se produit si A est réalisé, ou si B est réalisé, si A et B sont réalisés ;

— $A \cap B$ est l'événement qui se produit si A et B sont réalisés.

Deux événements sont **incompatibles** si deux événements A et B sont **dis-joints**. Lorsque deux événements A et B sont disjoints, c'est-à-dire lorsque $A \cap B = \emptyset$, on dit qu'ils **s'excluent mutuellement**. Dit autrement, A et B s'excluent mutuellement lorsqu'ils ne peuvent pas se réaliser simultanément.

Si l'univers S est **fini** ou **infiniment dénombrable**, alors chaque sous-ensemble S est un événement.

Si l'univers S n'est **pas dénombrable**, certains sous-ensembles de S ne peuvent être des événements.

Tous les événements forment une σ -algèbre ε de sous-ensembles de S . L'algèbre des événements renvoie aux opérations ensemblistes (union, intersection, complémentaire, *etc.*).

Soient S un univers, ε la famille des événements, et Pr une fonction à valeurs réelles définies sur ε . On dit que Pr est une **fonction de probabilité**, et que $\text{Pr}(A)$ est la probabilité de l'événement A .

Par exemple, pour une expérience de jets de dé, S possède six éventualités que l'on peut noter : $e_1 = 1, e_2 = 2, e_3 = 3, e_4 = 4, e_5 = 5, e_6 = 6$, c'est-à-dire que $S = \{1, 2, 3, 4, 5, 6\}$, et représente un **ensemble fini**. Si on définit un événement E tel que $E = \{e_1, e_2, e_3\}$, alors son contraire est $E^C = \{e_4, e_5, e_6\}$.

Si on définit l'événement A comme l'ensemble des chiffres pairs du dé, $A = \{e_2, e_4, e_6\}$. Si on définit l'événement B comme l'ensemble des chiffres impairs du dé, $B = \{e_1, e_3, e_5\}$. $A \cap B = \emptyset$, donc A et B sont incompatibles.

Une **probabilité** est un nombre compris entre 0 et 1, 0 étant l'événement impossible, et 1, l'événement certain (axiome 1). La probabilité d'un événement est la proportion de cas pour lesquels l'événement se manifeste si l'expérience est répétée **indépendamment** et **sous les mêmes conditions** à l'infini. La probabilité correspond à la limite de la fréquence relative.

$$\text{Pr}(E) = \lim_{n \rightarrow +\infty} f_n \quad (1)$$

où f_n est la fréquence relative de l'événement E . Cette remarque est fondamentale pour comprendre les liens entre les probabilités et les distributions statistiques.

Plus simplement, une probabilité correspond au rapport entre le nombre de cas réalisés (ou favorables) pour un événement donné et le nombre total d'événements possibles dans l'univers.

N.B. 1 La probabilité de l'événement impossible E , c'est-à-dire l'ensemble vide, est nulle.

$$\text{Pr}(E) = \text{Pr}(\emptyset) = 0 \quad (2)$$

N.B. 2 La probabilité de l'événement certain E vaut 1 (axiome 2).

$$\text{Pr}(E) = \text{Pr}(\S) = 1 \quad (3)$$

Si A et B sont des événements qui s'excluent mutuellement, alors :

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \quad (4)$$

Cela se généralise avec A_n événements s'excluant mutuellement (axiome 3).

$$\Pr(A_1 \cup A_2 \cup \dots \cup A_n) = \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_n) \quad (5)$$

En résumé, on note les probabilité $\Pr(e_i)$ avec $i \in \mathbb{N}$. Les axiomes de Kolmogorov sont : 1. la positivité de $\Pr(e_i)$; l'**échelle** de $\Pr(e_i)$ comprise entre 0 et 1; l'**additivité de deux événements incompatibles**.

1.1 La notion de tribu

Soit X un ensemble et soit A une partie de X , une **tribu** vérifie trois axiomes :

— A n'est pas vide ;

$$A \neq \emptyset \quad (6)$$

— A est stable par complémentaire.

$$\forall B \in A, C_A^B \in A \quad (7)$$

— A est stable par union dénombrable.

$$\text{si } \forall n \in \mathbb{N}, B_n \in A, \text{ alors } \bigcup B_n \in A. \quad (8)$$

Dit autrement, une **tribu** est une algèbre d'ensembles stable par réunion dénombrable.

Le couple (X, A) est appelé **espace mesurable** ou **espace probabilisable**. Dans un contexte probabiliste, les parties de X qui appartiennent à la tribu A sont appelées des événements ; dans le cadre d'une mesure, des ensembles mesurables.

Il faut juste retenir que, dans le cadre probabiliste, la tribu normale est la **tribu de Borel** (ou tribu borélienne). Géométriquement, la tribu de Borel correspond à la droite des réels. La notion ne servira pas dans cet exposé, mais le lecteur peut la rencontrer dans des textes d'approfondissement.

Émile
Borel
(1871-
1956)

1.1.1 Ensembles probabilisés finis

Soit S un univers fini, $S = \{a_1, a_2, \dots, a_n\}$. Si, à chaque point $a_i \in S$, on attribue un réel p_i , appelé **probabilité de** a_i , vérifiant :

1. chaque p_i est positif ou nul ($p_i \geq 0$) ;

2. la somme p_i est égale à 1 :

$$p_1 + p_2 + \dots + p_n = 1 \quad (9)$$

alors on dit que l'on est en présence d'un **ensemble probabilisé fini**.

On définit alors la probabilité $\Pr(A)$ d'un événement quelconque A comme étant la somme des probabilités des points de A . Pour des raisons de commodité, on écrit $\Pr(a_i)$ au lieu de $\Pr(\{a_i\})$.

Exercice type Trois chevaux A , B et C participent à une course. A a deux fois plus de chances que B de gagner. B a deux fois plus de chances que C de gagner.

— Quelles sont les probabilités respectives $\Pr(A)$, $\Pr(B)$ et $\Pr(C)$ de gagner des trois chevaux ?

■ On pose $\Pr(C) = p$.

■ On sait que $\Pr(B) = 2p$.

■ On sait que $\Pr(A) = 2\Pr(B) = 4p$.

■ Or $\Pr(A) + \Pr(B) + \Pr(C) = 1$

$$4p + 2p + p = 1 \Leftrightarrow p = \frac{1}{7} \quad (10)$$

■ On en déduit que :

$$\Pr(A) = \frac{4}{7} \quad (11)$$

et

$$\Pr(B) = \frac{2}{7} \quad (12)$$

et

$$\Pr(C) = \frac{1}{7} \quad (13)$$

— Quelle est la probabilité pour que B **ou** C soit vainqueur ?

$$\Pr(\{B, C\}) = \Pr(B) + \Pr(C) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7} \quad (14)$$

1.1.2 Ensembles finis équiprobables

Si, dans un univers S , chaque élément possède la même probabilité, alors il s'agit d'un **espace équiprobable** (ou uniforme).

Soit l'univers $S = \{e_1, e_2, \dots, e_n\}$. L'**équiprobabilité** correspond au cas où chaque éventualité dispose de la même probabilité. Ce sont des **probabilités symétriques** qui sont très fréquentes aux niveaux des jeux de hasard comme les dés.

$$\Pr(e_1) = \Pr(e_2) = \dots = \Pr(e_n) = p \quad (15)$$

Dit autrement,

$$\Pr(S) = \Pr(e_1) + \Pr(e_2) + \dots + \Pr(e_n) = np = 1 \quad (16)$$

Les probabilités se réduisent à une seule identique pour chaque éventualité. Elle vaut :

$$p = \frac{1}{n} = \frac{1}{\text{card}E} \quad (17)$$

Dit autrement, si l'univers S contient n éléments, la probabilité de chaque élément est : $\frac{1}{n}$. Si un événement A est formé de r éléments, sa probabilité est :

$$\Pr(A) = r \times \frac{1}{n} = \frac{r}{n} \quad (18)$$

avec r le nombre de cas favorables à la réalisation de l'événement A et n le nombre de cas possibles de l'univers S .

N.B. 1. La moyenne arithmétique pose cette hypothèse.

N.B. 2. « Choisir un élément au hasard dans l'ensemble S » signifie que S est un espace équiprobable, ce qui signifie que chaque élément de S a la même probabilité.

1.1.3 Ensembles fondamentaux infinis

S est un univers infiniment dénombrable.

$$S = \{a_1, a_2, \dots\} \quad (19)$$

Comme dans le cas fini, on obtient un espace probabilisé en attribuant à chaque $a_i \in S$ un nombre réel p_i ; appelé sa probabilité et tel que :

1. $p_i \geq 0$;
2. $p_1 + p_2 + \dots = \sum_{i=1}^{\infty} p_i = 1$ La probabilité $\Pr(A)$ d'un événement quelconque A est alors la somme de tous ses éléments.

1.1.4 Ensembles discrets et continus

On dit qu'un ensemble probabilisé **fini ou infini dénombrable** est **discret**, tandis qu'un ensemble probabilisé **non dénombrable** est dit **non discret** (ou continu).

1.2 L'addition de probabilités

Les probabilités de deux événements peuvent s'additionner sous certaines conditions. Cela arrive lorsque l'on s'intéresse à l'union de deux ou de plusieurs événements.

1. **Théorème.** La probabilité de l'événement contraire correspond à la différence entre 1 et la probabilité de l'événement.

$$\Pr(E^C) = 1 - \Pr(E) \quad (20)$$

2. **Axiome 3.** Si deux événements E et F sont incompatibles, c'est-à-dire que leur intersection est nulle, que ce sont deux ensembles disjoints, la probabilité qu'au moins un des deux survienne est égale à la somme de leurs probabilités respectives.

$$\Pr(E \text{ ou } F) = \Pr(E) + \Pr(F) \quad (21)$$

- **Théorème.** Si $E \subset F$, alors $\Pr(E) \leq \Pr(F)$.
- **Théorème.** Soient deux événements E et F , alors :

$$\Pr(\bar{F}) = \Pr(E) - \Pr(E \cap F) \quad (22)$$

Théorème. À l'opposé, si les événements sont compatibles, c'est-à-dire que leur intersection est non nulle, l'addition des probabilités donne une fausse solution. Il faut nécessairement lui ôter les valeurs de l'intersection.

$$\Pr(E \text{ ou } F) = \Pr(E) + \Pr(F) - \Pr(E \text{ et } F) \quad (23)$$

Cela correspond à la probabilité que l'un ou l'autre des événements se produisent. La formule reprend celle des quatre cardinaux.

Corollaire. E , F et G sont des événements quelconques :

$$\begin{aligned} \Pr(E \cup F \cup G) = & \Pr(E) + \Pr(F) + \Pr(G) - \Pr(E \cap F) - \Pr(E \cap G) \\ & - \Pr(F \cap G) + \Pr(E \cap F \cap G) \end{aligned} \quad (24)$$

N.B. Lors d'un cas pratique, il faut tenter de décomposer le problème en éléments incompatibles.

1.3 La multiplication de probabilités

La multiplication de deux probabilités intervient lorsque l'on souhaite calculer la probabilité d'une intersection de deux événements, c'est-à-dire la chance que les deux arrivent simultanément lors d'une expérience.

1.3.1 Événement dépendant et événement indépendant

Deux événements sont **indépendants** si la connaissance préalable que le premier arrivera ne change pas la probabilité de l'autre. Les résultats d'un tirage avec remise sont indépendants – la remise ne change pas le nombre total d'événements possibles –, tandis que les résultats d'un tirage sans remise sont en général dépendants – la non remise change le nombre total d'événements possibles d'un tirage à l'autre.

Si deux événements E et F sont indépendants, la probabilité que tous les deux se produisent simultanément est le produit des deux probabilités de E et de F .

$$\Pr(E \text{ et } F) = \Pr(E) \times \Pr(F) \quad (25)$$

Cela implique que la probabilité de l'un ne modifie pas la probabilité de l'autre lorsque l'autre est réalisé.

N.B. Ne pas confondre indépendant et incompatible.

1.3.2 Probabilité conditionnelle (ou épreuves successives)

Toute probabilité n'est bien définie qu'en fonction des connaissances disponibles. Cela implique en particulier que la valeur d'une probabilité peut changer et doit s'adapter si davantage de connaissances sont acquises au cours d'une expérience. Ce concept de connaissances partielles est une idée pivot du calcul de probabilité. C'est dans ce cadre qu'interviennent les probabilités conditionnelles.

La **probabilité conditionnelle** de l'événement E en connaissant (ou en sachant) F , notée $\Pr(E/F)$, est donnée par :

$$\Pr(E/F) = \frac{\Pr(E \cap F)}{\Pr(F)} \quad (26)$$

c'est-à-dire que la probabilité que deux événements se produisent vaut :

$$\Pr(E \cap F) = \Pr(E/F) \times \Pr(F) \quad (27)$$

C'est le **théorème des probabilités composées**.

Théorème. Sur un univers S , soient deux événements E et F , alors :

$$\Pr(E/F) = \frac{\text{nombre de réalisations possibles de } E \text{ et } F}{\text{nombre de réalisations possibles de } F} \quad (28)$$

Comme $E \cap F = F \cap E$, on peut en déduire que :

$$\Pr(F \cap E) = \Pr(F) \times \Pr(E/F) \quad (29)$$

Corollaire. Pour des événements quelconques E_1, E_2, \dots, E_n

$$\Pr(E_1 \cap E_2 \cap \dots \cap E_n) = \Pr(E_1) \times \Pr(E_2/E_1) \times \Pr(E_3/E_1 \cap E_2) \dots \Pr(E_n/E_1 \cap E_2 \cap \dots \cap E_{n-1}) \quad (30)$$

Il est alors possible de reformuler la définition de l'indépendance de deux événements. Deux événements E et F sont indépendants, si la probabilité conditionnelle de E sachant F vaut la probabilité de E , et si la probabilité conditionnelle de F sachant E vaut la probabilité de F .

$$\begin{cases} \Pr(E/F) = \Pr(E) \\ \Pr(F/E) = \Pr(F) \end{cases} \quad (31)$$

On dit qu'un événement F est **indépendant** d'un événement E si la probabilité pour que F se produise n'est pas influencée par le fait que E se soit ou ne se soit pas produit. Dit autrement, F est indépendant de E si la probabilité F est égale à la probabilité conditionnelle de F sachant que E s'est produit.

$$\Pr(F) = \Pr(F/E) \quad (32)$$

Si on remplace $\Pr(F/E)$ par $\Pr(F)$ dans le théorème de la multiplication :

$$\Pr(E \cap F) = \Pr(E) \Pr(F/E) \quad (33)$$

on obtient l'équation définissant formellement l'indépendance entre A et B :

$$\Pr(E \cap F) = \Pr(E) \Pr(F) \quad (34)$$

Dans le cas contraire, les événements sont dépendants.

Attention ! Très fréquemment, on postulera que deux événements sont indépendants d'après la nature de l'expérience.

On dit que **trois** événements E, F et G sont indépendants si :

1. les événements sont indépendants deux à deux :

$$\begin{aligned} \Pr(E \cap F) &= \Pr(E) \Pr(F) \\ \Pr(E \cap G) &= \Pr(E) \Pr(G) \\ \Pr(F \cap G) &= \Pr(F) \Pr(G) \end{aligned} \quad (35)$$

2. $\Pr(E \cap F \cap G) = \Pr(E) \Pr(F) \Pr(G)$

1.3.3 Principe de probabilité totale

Il est possible de décomposer la probabilité de tout autre événement en somme. Soit E un événement, il vérifie les propriétés suivantes par définition :

$$E \subset S \quad (36)$$

$$E \cup E^C = S \quad (37)$$

$$E \cap E^C = \emptyset \quad (38)$$

Soit F un autre événement, il est défini par :

$$F \subset S \quad (39)$$

$$F = F \cap E = F \cap (E \cup E^C) = (F \cap E) \cup (F \cap E^C) \quad (40)$$

Dit autrement, la probabilité de F vaut alors :

$$\Pr(F) = \Pr((F \cap E) \cup (F \cap E^C)) \quad (41)$$

$$\Pr(F) = \Pr((F \cap E)) + \Pr((F \cap E^C)) \quad (42)$$

$$\Pr(F) = \Pr(E) \times \Pr(F/E) + \Pr(E^C) \times \Pr(F/E^C) \quad (43)$$

$$\Pr(F) = \Pr(E) \times (\Pr(F/E) + \Pr(F/E^C)) \quad (44)$$

Cette équation se généralise assez facilement en :

$$\Pr(F) = \sum_{k=1}^n \Pr(E_k) \times \Pr(F/E_k) \quad (45)$$

avec une partition de l'univers S telle que $E_i \cap E_j = \emptyset$ avec $i \neq j, i \in [1, n]$ et $j \in [1, n]$.

1.3.4 Théorème de Bayes

Le diagramme en arbre est utile pour représenter des expériences avec une partition naturelle de l'ensemble fondamental, ou bien des expériences qui se font en étapes. La loi de la probabilité totale permet de calculer la probabilité d'un événement quelconque. Le théorème de Bayes renverse ce problème. Au lieu de demander avec quelle probabilité un résultat se produit lors de l'expérience, on s'interroge sur la manière dont un résultat particulier a pu se produire.

Soient des événements A_1, A_2, \dots, A_n formant une partition de l'univers S . Les événements A_i s'excluent mutuellement et leur réunion est S .

Soit B un autre événement quelconque dans S (Fig. 1), alors $B = S \cap B$, or $S = A_1 \cup A_2 \cup \dots \cup A_n$

$$\begin{aligned} B &= (A_1 \cup A_2 \cup \dots \cup A_n) \cap B \\ B &= (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B) \end{aligned} \quad (46)$$

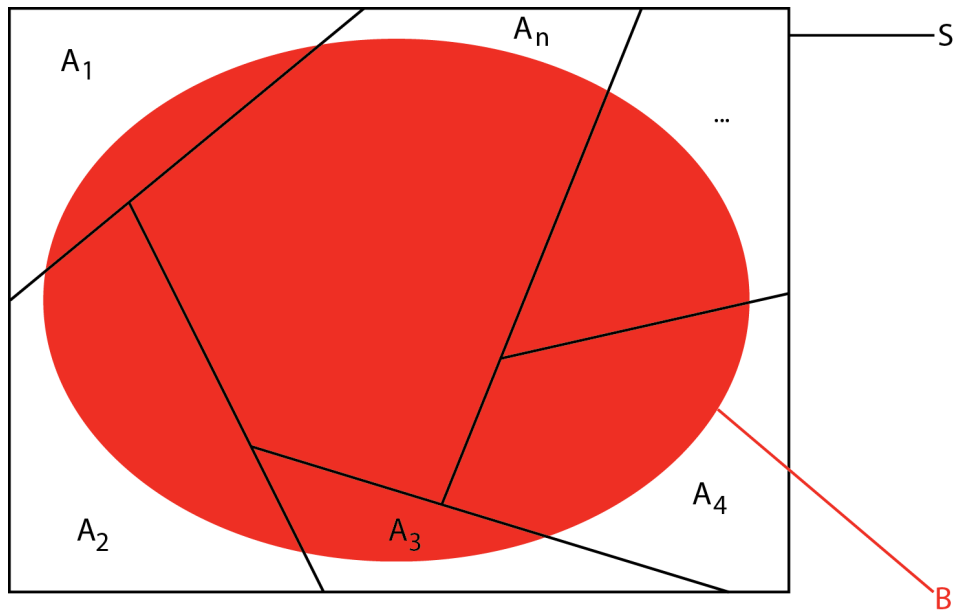


FIGURE 1 – Partition des ensembles A_i dans S et B dans B

$A_i \cap B$ s'excluent mutuellement. On peut dire que :

$$\Pr(B) = \Pr(A_1 \cap B) + \Pr(A_2 \cap B) + \dots + \Pr(A_n \cap B) \quad (47)$$

D'après le théorème de la multiplication :

$$\Pr(B) = \Pr(A_1) \Pr(B/A_1) + \Pr(A_2) \Pr(B/A_2) + \dots + \Pr(A_n) \Pr(B/A_n) \quad (48)$$

Pour un i donné, la probabilité conditionnelle de A_i en sachant B réalisé, est définie par :

$$\Pr(A_i/B) = \frac{\Pr(A_i \cap B)}{\Pr(B)} \quad (49)$$

Or la partition représentée par $\Pr(B)$ est connue, et permet d'énoncer le **théorème de Bayes**.

Théorème de Bayes. Soient A_1, A_2, \dots, A_n , une partition de S , et B un événement quelconque pour tout i , on a alors :

$$\Pr(A_i/B) = \frac{\Pr(A_i) \Pr(B/A_i)}{\sum_{j=1}^n (\Pr(A_j) \Pr(B/A_j))} \quad (50)$$

Par comprendre de quoi il s'agit, on peut reprendre l'expérience de Bayes. Soient trois urnes contenant des boules rouges. L'urne n° 1, notée A_1 , contient de 2 boules rouges sur 5. L'urne n° 2, notée A_2 , contient 3 boules rouges sur 5. L'urne n° 3, notée A_3 , contient 1 boule rouge sur 5. Avec un dé, on décide que si le lancer est 1, 2 ou 3, on pioche dans A_1 , si le lancer est 4 ou 5, on pioche dans A_2 , et si le lancer est 6, on pioche dans A_3 . Sans connaître le résultat du dé, quelle est la probabilité de tirer une boule rouge ?

Dans ce cas, les événements sont incompatibles. Pour répondre à cette question, on calcule une **probabilité a posteriori**, définie comme le rapport entre une probabilité conditionnelle et l'application du théorème des probabilités totales, c'est-à-dire :

$$\Pr(E_k/F) = \frac{\Pr(E_k) \times \Pr(F/E_k)}{\sum_{i=1}^n \Pr(E_i) \times \Pr(F/E_i)} \quad (51)$$

où $\Pr(E_k)$ est une **probabilité a priori**. Le souci est qu'elle est souvent inconnue. Le théorème est de fait impossible à mettre en œuvre. La probabilité *a priori* relève souvent d'une **probabilité subjective** (ou personnelle).

Toutefois, il est possible d'obtenir des probabilités *a priori* par des chroniques historiques (crues) ou un historique Internet (traitement des spams). Avec la massification des données, la statistique bayésienne et les réseaux bayésiens ont fait leur retour en force dans les applications.

La technique de la capture-recapture Soit N l'ensemble des poissons d'un lac, pour lequel n est inconnu. Comment estimer n ?

Soit n_1 une capture de poissons (c'est-à-dire un échantillon de n). On met une bague à ses poissons et on les rejette dans le lac.

Soit n_2 une deuxième capture de poissons (c'est-à-dire un nouvel échantillon de N). Dans ce cas, il existe une portion de poissons déjà bagués représentée par le nombre k .

Intuitivement, la portion $\frac{n_1}{n}$ est identique à $\frac{k}{n_2}$.

$$\frac{n_1}{n} = \frac{k}{n_2} \Rightarrow n = \frac{n_1 n_2}{k} \quad (52)$$

Si l'on formule en termes de probabilités la deuxième pêche, $\Pr(k/n)$ est le rapport entre les cas favorables et le nombre de cas total.

On capture k parmi les n_1 poissons marqués, c'est-à-dire $\binom{n_1}{k}$.

Il existe $(n_2 - k)$ poissons non marqués parmi les $(n - n_1)$ poissons non marqués, c'est-à-dire $\binom{n - n_1}{n_2 - k}$.

Le cas total est n_2 parmi n , c'est-à-dire $\binom{n}{n_2}$.

Au final,

$$\Pr(k/n) = \frac{\binom{n_1}{k} \binom{n - n_1}{n_2 - k}}{\binom{n}{n_2}} \quad (53)$$

Le souci est que ce résultat ne correspond pas à celui qui est recherché. Il faut par conséquent appliquer le théorème de Bayes. Une nouvelle fois, la probabilité *a priori* est inconnue.

$$\Pr(n/k) = \frac{\Pr(n) \times \Pr(k/n)}{\sum_{i=1}^n \Pr(n) \times \Pr(k/n)} \quad (54)$$

La notion de **vraisemblance** permet de résoudre le problème.

Une première étape consiste à regarder la probabilité pour $n + 1$.

$$\frac{\Pr(k/n + 1)}{\Pr(k/N)} = \frac{(n + 1 - n_1)(n + 1 - n_2)}{(n + 1 - n_1 - n_2)(n + 1)} = \frac{n_1 n_2}{k} - n \quad (55)$$

avec $\frac{\Pr(k/n+1)}{\Pr(k/n)} \leq 1$. De fait,

$$n \geq \frac{n_1 n_2}{k} - 1 \quad (56)$$

Une seconde étape consiste à regarder la probabilité pour $n - 1$.

$$\frac{\Pr(k/n)}{\Pr(k/n - 1)} = \frac{(n - n_1)(n - n_2)}{(n - n_1 - n_2 + k)n} = \frac{n_1 n_2}{nk} \quad (57)$$

avec $\frac{\Pr(k/n)}{\Pr(k/n-1)} \geq 1$. De fait,

$$n \leq \frac{n_1 n_2}{k} \quad (58)$$

En combinant les deux inéquations, il est **vraisemblable** que :

$$\frac{n_1 n_2}{k} - 1 \leq n^* \leq \frac{n_1 n_2}{k} \quad (59)$$

n^* est une estimation de n . La base de l'échantillonnage repose sur la notion de vraisemblance.

1.3.5 Épreuves indépendantes ou répétées

Soit S un univers fini. Par n **épreuves indépendantes** (ou répétées), on entend l'ensemble probabilisé T formé de n -uples ordonnés d'éléments de S , la probabilité d'un n -uple étant définie comme le produit des probabilités de ses composantes :

$$\Pr(\{s_1, s_2, \dots, s_n\}) = \Pr(s_1) \Pr(s_2) \dots \Pr(s_n) \quad (60)$$

Les épreuves répétées peuvent se représenter par un arbre.

Exercice type Chaque fois que trois chevaux A , B et C courent ensemble, leurs probabilités respectives de gagner sont $\frac{1}{2}$, $\frac{1}{3}$ et $\frac{1}{6}$.

— Cela se traduit par l'univers $S = \{A, B, C\}$ et

$$\begin{array}{l} \Pr(A) \frac{1}{2} \\ \Pr(B) \frac{1}{3} \\ \Pr(C) \frac{1}{6} \end{array} \quad (61)$$

$$\mathbf{N.B.} \quad \Pr(A) + \Pr(B) + \Pr(C) = 1$$

— Si les chevaux courent deux fois, l'univers des deux épreuves répétées est :

$$T = \{\{a, a\}, \{a, b\}, \{a, c\}, \{b, a\}, \{b, b\}, \{b, c\}, \{c, a\}, \{c, b\}, \{c, c\}\} \quad (62)$$

On obtient un arbre pondéré (Fig. 2).

- La probabilité que A gagne toutes les courses est de $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.
- La probabilité que A , puis B , gagnent successivement est de $\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$.
- La probabilité que A , puis C , gagnent successivement est de $\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$.
- La probabilité que B , puis A , gagnent successivement est de $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$.

- La probabilité que B gagne toutes les courses est de $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$.
- La probabilité que B , puis C , gagnent successivement est de $\frac{1}{3} \times \frac{1}{6} = \frac{1}{18}$.
- La probabilité que C , puis A , gagnent successivement est de $\frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$.
- La probabilité que C , puis B , gagnent successivement est de $\frac{1}{6} \times \frac{1}{3} = \frac{1}{18}$.
- La probabilité que C gagne toutes les courses est de $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$.

N.B. $\frac{1}{4} + \frac{1}{6} + \frac{1}{12} + \frac{1}{6} + \frac{1}{9} + \frac{1}{18} + \frac{1}{12} + \frac{1}{18} + \frac{1}{36} = 1$

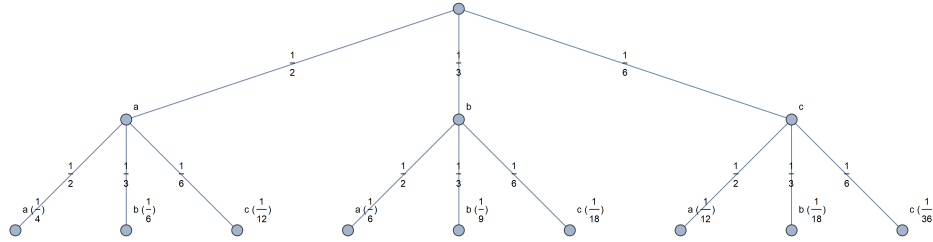


FIGURE 2 – Arbre pondéré présentant les neuf résultats possibles de la course de chevaux

1.3.6 Épreuve de Bernoulli

On **répète** la **même expérience** dans les **mêmes conditions**.

On s'intéresse à chaque fois à la réalisation d'un événement. Chacun de ces expériences s'appelle une **épreuve de J. Bernoulli**. Elle se conclut par un **succès** ou un **échec**.

1. Chaque épreuve a deux issues : succès (S) ou échec (E).
2. Pour chaque épreuve, la probabilité d'un succès est la même.

$$\begin{aligned} \Pr(S) &= p \\ \Pr(E) &= q = 1 - p \end{aligned} \tag{63}$$

3. Les épreuves sont **indépendantes**. Dit autrement, la probabilité d'un succès ne varie pas. Elle ne dépend pas des informations sur les résultats des autres épreuves.

Théorème. Soit un schéma de J. Bernoulli avec p la probabilité d'un succès à une épreuve de J. Bernoulli et n le nombre d'épreuves de J. Bernoulli effectuées. L'univers Ω est :

$$\Omega = \{S, E\} \times \dots \times \{S, E\} = \{S, E\}^n \quad (64)$$

L'épreuve de J. Bernoulli servira dans un prochain chapitre pour introduire la loi de J. Bernoulli, la loi binomiale, la loi multinomiale, *etc.*

2 L'approche des probabilités par les côtes

N'importe quelle probabilité p peut être exprimée comme un pari. On définit la **côte** d à « un contre un » par :

$$d = \frac{p}{p - 1} \quad (65)$$

c'est-à-dire une probabilité valant :

$$p = \frac{d}{d - 1} \quad (66)$$

3 Les quatre espaces de probabilité les plus fréquents

Certains ont déjà été évoqués, mais, avant d'attaquer la notion de loi de probabilité, il est bon de rappeler quatre principaux espaces de travail pour les probabilités :

- l'univers fini ou dénombrable ;
- le cas équiprobable ;
- le schéma binaire « succès » / « échec » ;
- le cas où l'univers correspond aux nombres réels.

4 La notion de variable aléatoire

Une **variable aléatoire** (v.a.) X sur un univers S est une **fonction** de S dans l'ensemble \mathbb{R} des nombres réels, telle que l'inverse de chaque intervalle de \mathbb{R} soit un événement de S . À chaque événement élémentaire s , on associe une quantité $X(s)$.

N.B. 1. Si S est un ensemble discret dans lequel chaque sous-ensemble est un événement, toute fonction à valeurs réelles, définie sur S est une variable aléatoire.

N.B. 2. Si S est un ensemble non dénombrable, certaines fonctions à valeurs réelles définies sur S ne sont pas des variables aléatoires.

N.B. 3. On utilise les lettres capitales de la fin de l'alphabet latin pour désigner les variables aléatoires (X, Y, Z) .

Soit $a \in \mathbb{R}$ et $b \in \mathbb{R}$, on note $(X = a)$ l'ensemble des antécédents de a et $(X = b)$ l'ensemble des antécédents de b . On utilise les notations :

- $\Pr(X = a)$ pour $\Pr(\{s \in S : X(s) = a\})$ ce qui se lit « si la variable aléatoire $X = a$, que vaut sa probabilité? » ;
- $\Pr(a \leq X \leq b)$ pour $\Pr(\{s \in S : a \leq X(s) \leq b\})$

En résumé, une variable aléatoire est un nombre réel qui dépend du résultat d'une expérience aléatoire. Une variable aléatoire réelle traduit l'idée de résultat numérique associé au phénomène aléatoire étudié. Néanmoins, il existe des variables aléatoires complexes et vectorielles (ou vecteur aléatoire, plongé dans un espace à n dimensions). Chaque exécution de l'expérience génère une réalisation de variable aléatoire. Les variables aléatoires constituent un espace fondamental d'éléments aléatoires, un tel élément étant défini par référence à une expérience aléatoire.

5 La notion de loi de probabilité

Soit X une variable aléatoire définie sur un univers S à valeurs finies, c'est-à-dire :

$$X(S) = \{x_1, x_2, \dots, x_n\} \quad (67)$$

$X(S)$ devient un espace probabilité si l'on définit la probabilité $\Pr(X = x_i)$ des x_i , que l'on écrira $f(x_i)$. Cette fonction f sur $X(S)$ qui est définie par $f(x_i) = \Pr(X = x_i)$ est ce que l'on appelle la **distribution** ou la **loi de probabilité** de X . On la donne sous la forme d'un tableau (Tab. 1).

x_1	x_2	\dots	x_n
$f(x_1)$	$f(x_2)$	\dots	$f(x_n)$

TABLE 1 – Loi de probabilité de X

Si X est une variable aléatoire ayant la loi de probabilité observée, la moyenne ou l'**espérance mathématique** de X que l'on note $\mathbb{E}(X)$, ou μ_X , ou E , ou μ , est définie par :

$$\mathbb{E}(X) = x_1 f(x_1) + x_2 f(x_2) + \dots + x_n f(x_n) = \sum_{i=1}^n x_i f(x_i) \quad (68)$$

$\mathbb{E}(X)$ est la moyenne pondérée des valeurs que X peut prendre, chaque valeur étant pondérée par sa probabilité.

Proposition sur la linéarité. Soient X une variable aléatoire et a, b , deux réels.

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \quad (69)$$

Théorème du transfert. Soit X une variable aléatoire et f une fonction définie sur $X(\Omega)$, alors :

$$\mathbb{E}(f(X)) = \sum_{k \in X(\Omega)} f(k) \Pr(X = k) \quad (70)$$

La variance de X , notée $\mathbb{V}(X)$, vaut :

$$\mathbb{V}(X) = \sum_{i=1}^n (x_i - \mathbb{E}(X))^2 f(x_i) = E[(X - \mathbb{E}(X))^2] \quad (71)$$

L'écart type de X , noté σ_X , vaut :

$$\sigma_X = \sqrt{\mathbb{V}(X)} \quad (72)$$

Théorème. $\mathbb{V}(\lambda) = \sum_{i=1}^n (x_i^2 f(x_i) - \mathbb{E}(X)^2) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

Théorème. Soit X une variable aléatoire et k un nombre réel, on a alors :

1. $\mathbb{V}(X + k) = \mathbb{V}(X)$;
2. $\mathbb{V}(kX) = k^2 \mathbb{V}(X)$;
3. $\sigma_{X+k} = \sigma_X$;
4. $\sigma_{kX} = |k| \sigma_X$;
5. $\mathbb{V}(X) = 0 \Leftrightarrow X$ est constante égale à $\mathbb{E}(X)$.

N.B. 1. La moyenne est le **centre de gravité du système**.

N.B. 2. La variance est le **moment d'inertie du système**. Variance et écart type mesurent la dispersion de la variable aléatoire autour de sa moyenne $\mathbb{E}(X)$.

N.B. 3. La variable aléatoire centrée-réduite X^* vaut :

- $X^* = \frac{X - \mathbb{E}(X)}{\sigma_X}$;
- $\mathbb{E}(X^*) = 0$ et $\mathbb{V}(X^*) = 1$.

x_1	x_2	\dots
$f(x_1)$	$f(x_2)$	\dots

TABLE 2 – Loi de probabilité de la variable aléatoire discrète X

5.1 Variables aléatoires discrètes

Si X est une variable aléatoire sur S avec un espace infini dénombrable, c'est-à-dire $X(S) = \{x_1, x_2, \dots\}$ (Tab. 2). Ce genre de variables aléatoires est appelée **variable aléatoire discrète**. Comme dans le cas fini, on transforme $X(S)$ en un espace probabilisé, en définissant la probabilité de x_i par $f(x_i) = \Pr(X = x_i)$, et on dit que f est la loi de probabilité de X .

La loi de probabilité f pour une **variable aléatoire discrète** satisfait les conditions :

1. $f(x_i) \geq 0$;
2. $\Pr(X) = \sum_{i=1}^n f(x_i) = 1$.

On définit l'espérance mathématique $\mathbb{E}(X)$ et la variance $\mathbb{V}(X)$ par :

- $\mathbb{E}(X) = x_1 f(x_1) + x_2 f(x_2) + \dots = \sum_{i=1}^n x_i f(x_i)$;
- $\mathbb{V}(X) = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \dots = \sum_{i=1}^n (x_i - \mu)^2 f(x_i)$.

lorsque les séries correspondantes convergent absolument.

La variance $\mathbb{V}(X)$ existe si et seulement si $\mu = \mathbb{E}(X)$ et $\mathbb{E}(X^2)$ existent en même temps, et que, dans ce cas, la formule

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mu^2 \quad (73)$$

est valable, comme dans le cas fini. Lorsque $\mathbb{V}(X)$ existe, on définit l'écart type σ_X , comme dans le cas fini, par :

$$\sigma_X = \sqrt{\mathbb{V}(X)} \quad (74)$$

Les notions de loi de probabilité produit, de variables aléatoires indépendantes et de fonctions de variables aléatoires se transposent directement au cas général. Si X et Y sont définies sur le même univers S , et si $\mathbb{V}(X)$ et $\mathbb{V}(Y)$ existent en même temps, alors la série $\text{cov}(X, Y)$ converge absolument.

5.2 Variables aléatoires continues

Si X est une variable aléatoire dont l'ensemble de valeur $X(S)$ est l'ensemble des points d'un intervalle.

Par définition, des variables aléatoires l'ensemble $\{a \leq X \leq b\}$ est un événement de S , et la probabilité $\Pr(a \leq X \leq b)$ est par conséquent bien définie.

On suppose qu'il existe une fonction continue par morceaux $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que $\Pr(a \leq X \leq b)$ soit égale à l'aire située en dessous de la courbe f , entre $x = a$ et $x = b$.

En termes de calcul intégral,

$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx \quad (75)$$

Dans ce cas, on dit que X est une **variable aléatoire continue**. On appelle la fonction f , la **distribution** ou la **loi de probabilité continue** de X . Cette fonction satisfait les conditions :

1. $f(x) \geq 0$;
2. $\int_{\mathbb{R}} f(x) dx = 1$.

Cela signifie que f est positive ou nulle et que l'aire totale en dessous de la courbe est égale à 1.

On définit l'**espérance mathématique** $\mathbb{E}(X)$ par :

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x) dx \quad (76)$$

lorsqu'elle existe.

On définit les fonctions de variables aléatoires de la même manière que dans le cas discret. On peut montrer que si $Y = \Phi(X)$, alors :

$$\mathbb{E}(Y) = \int_{\mathbb{R}} \Phi(x) f(x) dx \quad (77)$$

lorsque le membre de droite existe.

On définit la variance $\mathbb{V}(X)$ par :

$$\mathbb{V}(X) = \mathbb{E}[(X - \mu)^2] = \int_{\mathbb{R}} (x - \mu)^2 f(x) dx \quad (78)$$

lorsqu'elle existe.

De la même manière que dans le cas discret, on peut montrer que $\mathbb{V}(X)$ existe si et seulement si $\mu = \mathbb{E}(X)$ et $\mathbb{E}(X^2)$ existe en même temps, et qu'alors :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mu^2 = \left[\int_{\mathbb{R}} x^2 f(x) dx \right] - \mu^2 \quad (79)$$

Lorsque $\mathbb{V}(X)$ existe, on définit l'écart type σ_X par :

$$\sigma_X = \sqrt{\mathbb{V}(X)} \quad (80)$$

N.B. Les résultats établis pour les variables aléatoires finies seront supposés également valables dans le cas plus général des variables aléatoires discrètes et dans le cas continu.

On dit que des variables aléatoires continues en nombre fini, X, Y, \dots, Z sont **indépendantes** si pour tout intervalle $[a, a'], [b, b'], \dots, [c, c']$

$$\Pr(a \leq X \leq a', b \leq X \leq b', \dots, c \leq X \leq c') = \Pr(a \leq X \leq a') \Pr(b \leq X \leq b') \dots \Pr(c \leq X \leq c') \quad (81)$$

Ainsi, dans le cas continu, les intervalles jouent le même rôle que les points dans le cas discret.

6 La fonction de répartition d'une variable aléatoire

Si X est une variable aléatoire numérique (discrète ou continue), on peut considérer comme événement favorable e , le fait que X soit inférieure ou égale à une valeur x de E . La probabilité de voir se réaliser cet événement est une fonction numérique de x , comprise entre 0 et 1, et que l'on appelle **fonction de répartition** F .

6.1 Variable aléatoire discrète

Soit X une variable aléatoire définie sur E . On appelle **fonction de répartition de X** (ou fonction en non-dépassement), la fonction F définie sur \mathbb{R} par :

$$\forall a \in \mathbb{R}, F(a) = \Pr(X \leq a) = \Pr(e/E) \quad (82)$$

Cette loi de probabilité doit être **monotone** et **non décroissante** (c'est-à-dire strictement croissante) et **varier de 0 à 1** lorsque la valeur x décrit l'intervalle de variation de X : $[x_i, x_{i+1}]$.

Aux limites, lorsque $x \rightarrow -\infty$, $F(x)$ vaut 0. Lorsque $x \rightarrow +\infty$, $F(x)$ vaut 1.

Si la variable aléatoire est réelle et discrète, la fonction F est continue à gauche et non dérivable.

Pour une variable aléatoire discrète, sa représentation graphique est un « **escalier** » (Fig. 3). On dit que la fonction de répartition varie **par plage**.

$$X(E) = \{x_1, x_2, \dots, x_n\} \text{ avec } x_1 < x_2 < \dots < x_n \quad (83)$$

La fonction est définie par :

$$F(x) = \sum_{x_i \leq x} f(x_i) \quad (84)$$

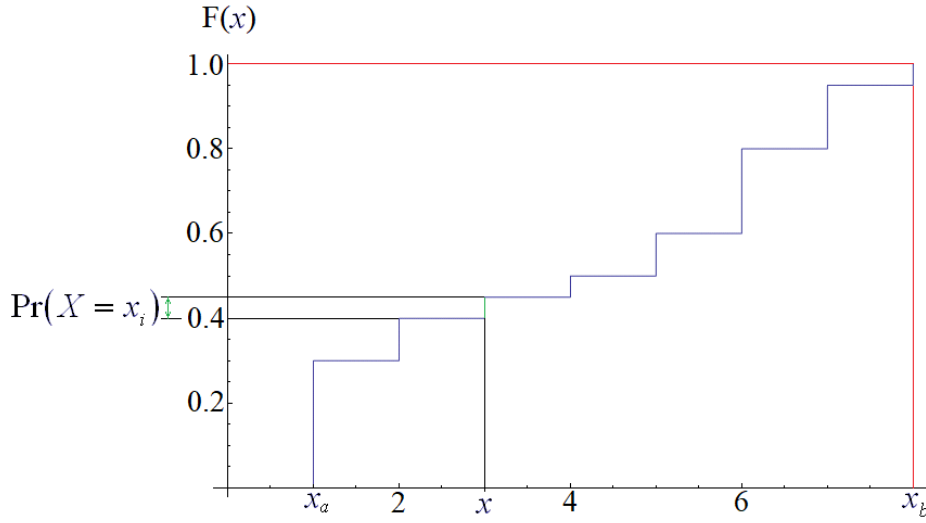


FIGURE 3 – Exemple de fonction de répartition d’une variable aléatoire discrète

$$\begin{aligned}
 x &\in]-\infty, x_1[, (X \leq x) = \emptyset, F(x) = 0 \\
 x &\in [x_1, x_2[, (X \leq x) = (X = x_1), F(x) = \Pr(X = x_1) \\
 x &\in [x_2, x_3[, (X \leq x) = (X = x_1) \cup (X = x_2), \\
 F(x) &= \Pr(X = x_1) + \Pr(X = x_2) \\
 &\dots
 \end{aligned} \tag{85}$$

$$\begin{aligned}
 x &\in [x_n, +\infty[, (X \leq x) = (X = x_1) \cup (X = x_2) \cup \dots \cup (X = x_n), \\
 F(x) &= \Pr(X = x_1) + \Pr(X = x_2) + \dots + \Pr(X = x_n) = 1
 \end{aligned}$$

En pratique, pour l’obtenir, il faut utiliser les **fréquences cumulées** de la variable aléatoire étudiée.

N.B. Deux variables aléatoires ayant la même loi de probabilité ont la même fonction de répartition.

6.2 Variable aléatoire continue

Il existe également des fonctions de répartition de variable continue (Fig. 4). Dans le cas où F est dérivable, la fonction dérivée $f(x)$ de F par rapport à x est appelée la **densité de probabilité**.

$$F(x_i) = \int_{x_a}^{x_i} f(x) \, dx \text{ et } F(x_{i+1}) = \int_{x_a}^{x_{i+1}} f(x) \, dx \tag{86}$$

La probabilité d'être en $F(x_i)$ et $F(x_{i+1})$ vaut alors :

$$\Pr(x_i < X < x_{i+1}) = F(x_{i+1}) - F(x_i) = \int_{x_i}^{x_{i+1}} f(x) dx \quad (87)$$

La probabilité d'atteindre la valeur x_0 est :

$$\Pr(X < x_0) = \int_{-\infty}^{x_0} f(x) dx = F(x_0) \quad (88)$$

Si la variable aléatoire est réelle et continue, la fonction F est continue à droite et dérivable. Elle est monotone croissante, c'est-à-dire $F(a) \leq F(b)$ lorsque $a \leq b$. La limite de F à gauche est 0, et sa limite à droite est 1 :

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad (89)$$

et

$$\lim_{x \rightarrow +\infty} F(x) = 1 \quad (90)$$

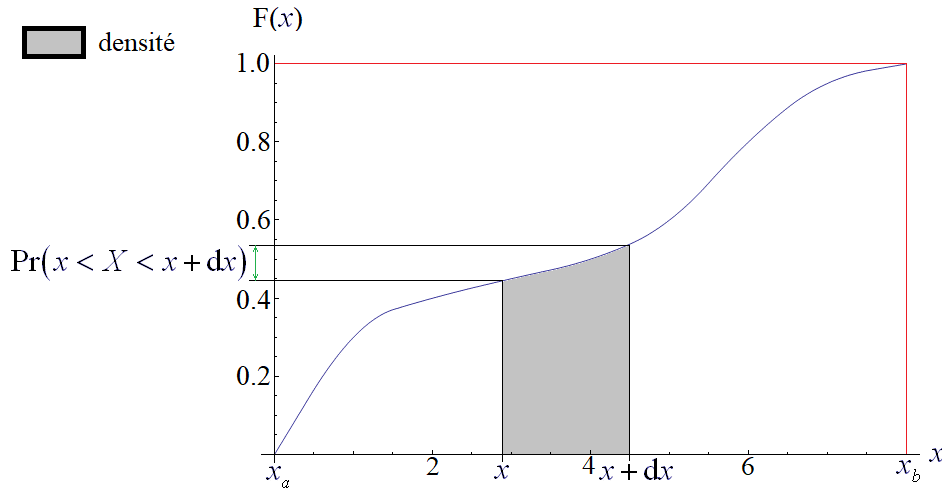


FIGURE 4 – Exemple de fonction de répartition d'une variable aléatoire continue

Remarque 1. Dans le cas d'une variable aléatoire continue, il est impossible d'associer une probabilité non nulle à l'apparition d'une valeur particulière de X .

Remarque 2. Si l'on considère plusieurs variables aléatoires réelles, la seule connaissance des fonctions de répartition de chacune des variables est insuffisante pour déterminer la dépendance entre ces variables. Dans ce cadre, les probabilités associées deviennent potentiellement conditionnelles.

L'espérance mathématique vaut $\mathbb{E}(X) = \sum_{i=1}^n x \Pr(X = x_i)$ pour une variable discrète, et $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$ pour une variable continue. L'espérance peut être considérée comme le centre de gravité de la distribution de masse $\Pr(x_i)$ ou $\Pr(x)$. Certaines variables aléatoires n'ont pas d'espérance ; la série statistique est divergente dans ce cas.

La variance vaut $\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$ pour une variable discrète, et $\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x) dx$ pour une variable continue. Dans le cadre de deux variables aléatoires, la covariance cov correspond à $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$. Dit autrement, pour un couple de deux variables aléatoires, il existe trois moments différents d'ordre 2 : $\mathbb{V}(X)$, $\mathbb{V}(Y)$ et $\text{cov}(X, Y)$. Il faut remarquer que si X et Y sont indépendantes alors leur covariance est nulle. Tout cela se généralise pour n variables aléatoires

7 La fonction en dépassement d'une variable aléatoire

Soit X une variable aléatoire définie sur E . On appelle **fonction de répartition en dépassement de X** , la fonction G définie sur \mathbb{R} par :

$$\forall x \in \mathbb{R}, G(X) = \Pr(X \geq x) \quad (91)$$

Cette fonction est décroissante au sens large. Elle a une fonction dérivée négative qui ne peut pas définir une probabilité. De fait, elle est rarement utilisée comme fonction de répartition.

8 Inégalité de Bienaymé-Tchebychev

Soient X une variable aléatoire, $\mathbb{E}(X)$, son espérance, et $\mathbb{V}(X)$, sa variance, alors :

$$\forall \varepsilon \in \mathbb{R}_*^+, \Pr(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2} \quad (92)$$

Schématiquement, la formule permet d'estimer la probabilité d'être dans l'intervalle hachuré (Fig. 5).

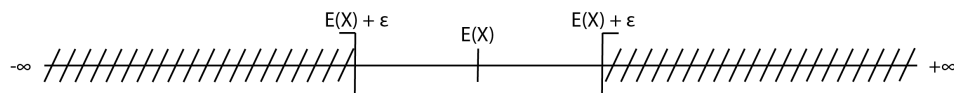


FIGURE 5 – Représentation de l'inégalité de Bienaymé-Tchebychev

Irénée-
Juls
Bie-
naymé
(1796-
1878)
Pafnouti
Tche-
bychev
(1821-
1894)

On utilise plus souvent sa réciproque avec l'événement contraire.

$$- \Pr(|X - \mathbb{E}(X)| \geq \varepsilon) \geq -\frac{\mathbb{V}(X)}{\varepsilon^2} \quad (93)$$

$$1 - \Pr(|X - \mathbb{E}(X)| \geq \varepsilon) \geq 1 - \frac{\mathbb{V}(X)}{\varepsilon^2} \quad (94)$$

$$\Pr(|X - \mathbb{E}(X)| < \varepsilon) \geq 1 - \frac{\mathbb{V}(X)}{\varepsilon^2} \quad (95)$$

La probabilité d'être autour de $\mathbb{E}(X)$ est $\Pr(|X - \mathbb{E}(X)| < \varepsilon)$.

N.B. 1. On utilise parfois l'écart type tel que : $\sigma_X^2 = \mathbb{V}(X)$.

N.B. 2. L'inégalité de Bienaymé-Tchebychev est valable aussi bien dans les cas discrets que continus.

Exercice type Soit $M(\Omega) = [0, 20]$ un ensemble de notes.

— $\mathbb{E}(M) = 12,3$

— $\mathbb{V}(M) = 0,47355$

Quel est le plus grand pourcentage entier p tel que $\Pr(10,3 < M < 14,3) \geq p$?

— On remarque que l'on recherche une valeur centrée telle que $\mathbb{E}(X) + 2$ soit compris entre 10,3 et 14,3.

— On applique l'inégalité de Bienaymé-Tchebychev.

$$\Pr(|M - 12,3| \geq 2) \leq \frac{0,47355}{2^2} \quad (96)$$

Toutefois, on recherche l'événement contraire $\Pr(|M - 12,3| \geq 2)$.

$$- \Pr(|M - 12,3| \geq 2) \geq -\frac{0,47355}{4} \quad (97)$$

$$1 - \Pr(|M - 12,3| \geq 2) \geq 1 - \frac{0,47355}{4} \quad (98)$$

$$\Pr(|M - 12,3| < 2) \geq 1 - \frac{0,47355}{4} \quad (99)$$

$$\Pr(|M - 12,3| < 2) \gtrsim 0,8816 \quad (100)$$

— Le plus grand pourcentage entier p vérifiant la probabilité demandée est :

$$p = 88 \% \quad (101)$$

9 Loi des grands nombres

Soient X_1, X_2, \dots , une suite de variables aléatoires indépendantes, ayant la même loi de probabilité, de moyenne μ et de variance σ^2 .

Soit la **moyenne de l'échantillonnage**.

$$\bar{S}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (102)$$

alors pour tout $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \Pr(|\bar{S}_n - \mu| \geq \varepsilon) = 0 \quad (103)$$

ou

$$\lim_{n \rightarrow +\infty} \Pr(|\bar{S}_n - \mu| < \varepsilon) = 1 \quad (104)$$

N.B. Cela se démontre avec l'inégalité de Bienaymé-Tchebychev.

Attention! Dans le cas de cette loi des grands nombres, la variance des X_i existe, c'est-à-dire qu'elle ne diverge pas. Le théorème est vrai si $\mathbb{E}(X)$ existe.

10 Les probabilités et la certitude : un exemple pour comprendre, les aiguilles de Buffon

Dans un phénomène aléatoire, il n'est pas possible de prévoir le détail des réalisations, mais il est possible de dégager une certitude globale. L'exemple des aiguilles de Buffon permet d'en faire le constat simplement.

Sur un parquet, on jette des aiguilles. C'est un processus purement aléatoire. Il est impossible de prévoir la manière dont les aiguilles vont tomber sur les lattes de parquet (Fig. 6). Toutefois, il est possible d'envisager deux cas : 1. soit l'aiguille touche en tombant une ligne de parquet (cas favorables) ; 2. soit l'aiguille ne touche pas en tombant une ligne de parquet (cas défavorables). L'analyse probabiliste de Buffon démontre que la probabilité $\frac{\text{nombre de cas favorables}}{\text{nombre de lancers}}$ tend à l'infini vers $\frac{2}{\pi}$.

Pour parvenir à ce résultat, il faut fixer les variables utiles (Fig. 7). La longueur de l'aiguille a mesure soit la largeur de la latte l , soit est d'une longueur inférieure $a \leq l$. La distance entre le centre de l'aiguille et la rainure du parquet la plus proche est notée r . L'angle entre l'aiguille et la rainure est noté θ .

Les lancers sont évidemment **indépendants**. Pour illustrer la convergence, le plus simple est d'étude le cas particulier dans lequel $l = a$. Le cas général $a < l$ est un peu plus complexe à étudier. L'objectif de cet exemple étant d'illustrer le raisonnement probabiliste appliqué en statistique, il n'est pas nécessaire d'aller au-delà du cas le plus simple.

Georges
Louis
Le-
clerc,
comte
de
Buffon
(1707-
1788)

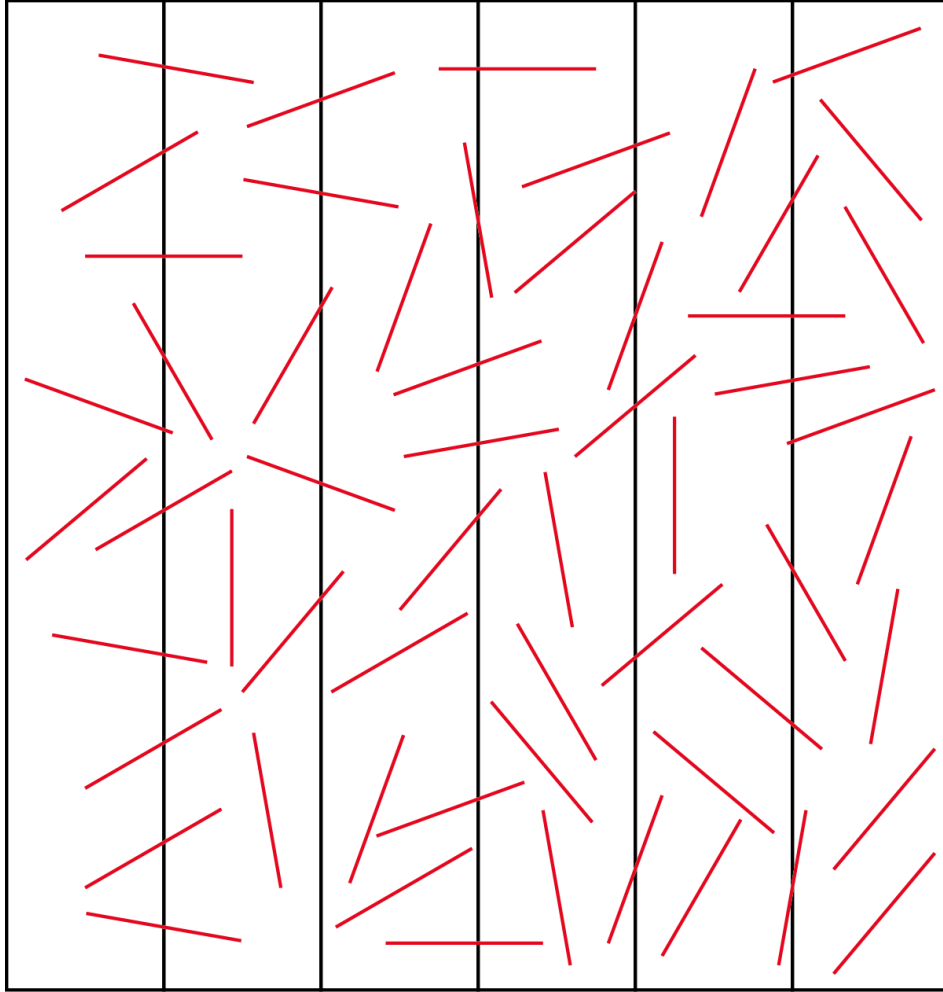


FIGURE 6 – Modèle des aiguilles de Buffon

L'angle θ suit une loi aléatoire uniforme sur $[0, \frac{\pi}{2}]$.

La distance r suit une loi aléatoire uniforme sur $[0, \frac{l}{2}]$.

Soient n lancers, toutes les positions différentes de l'aiguille mises bout à bout forment un polygone à n côtés. Plus n est grand, plus ce polygone se rapprochera d'un cercle. Son périmètre vaut alors $P = n \times a$, et son diamètre $D = \frac{P}{\pi} = \frac{n \times a}{\pi}$. Le problème revient alors à « Combien de rainures parallèles sont coupées par le polygone ? » ou « Combien y a-t-il de rainures à l'intérieur du cercle ? »

Le nombre d'intersections du cercle avec les rainures R vaut :

$$R = \frac{2D}{l} \quad (105)$$

De fait, la probabilité p que l'aiguille coupe une rainure est donnée par :

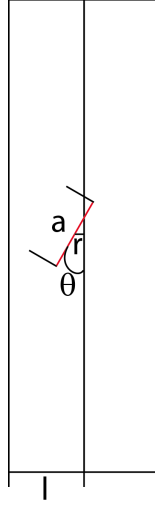


FIGURE 7 – Modèle des aiguilles de Buffon – Définition des variables aléatoires

$$p = \frac{R}{n} = \frac{2D}{l} \times \frac{1}{n} = \frac{2na}{\pi l} \times \frac{1}{n} = \frac{2}{\pi} \times \frac{a}{l} \quad (106)$$

Au cas limite,

$$\sin \theta = \frac{r}{\frac{a}{2}} \Leftrightarrow \frac{a}{2} \sin \theta = r \quad (107)$$

Cela correspond au fait que si l'aiguille touche un point de rainure avec sa pointe sans la chevaucher, alors le triangle rectangle dont l'hypothénuse est la moitié de l'aiguille apparaît, muni d'un côté de longueur r , et d'une portion de la rainure.

Pour les cas défavorables, si l'aiguille est pleinement sur la latte, on a :

$$\frac{a}{2} \sin \theta < r \quad (108)$$

Pour les cas favorables, si l'aiguille chevauche une rainure,

$$\frac{a}{2} \sin \theta > r \quad (109)$$

Comme $l = a$,

$$p = \frac{2}{\pi} \times \frac{a}{l} = \frac{2}{\pi} \quad (110)$$

Il est impossible de prédire dans quelle position tombera l'aiguille, mais il est certain que la probabilité l'aiguille coupe une rainure converge vers la valeur $\frac{2}{\pi}$. Cette expérience a été utilisée pour estimer les décimales de π . Cette méthode reste compliquée à mettre en œuvre ; il faut plus de 10 000 aiguilles pour arriver

à une approximation des deux premières décimales de π . Puisqu'une telle convergence existe, il n'y a pas de raison qu'il n'en existe pas d'autres, et c'est ce que recherche la statistique. Les méthodes statistiques établissent ces convergences (ou divergences), notamment au niveau de la moyenne, de l'écart type, *etc.* Elles permettent de proposer des faits certains dans les processus aléatoires. L'objectif est bien sûr de limiter le risque de se tromper. C'est pour cela que tout paramètre statistique dispose de barres d'erreur, d'intervalle de confiance, *etc.*

11 La construction des lois de probabilité

Une **loi de probabilité** se définit par sa formulation mathématique qui se caractérise par les valeurs des paramètres, qui sont estimées d'après l'échantillon observé, entrant dans la formulation mathématique. Les paramètres concernent la position, la dispersion et la forme de la distribution statistique.

Pour construire les lois de probabilité, la **variable centrée réduite** est souvent utilisée. Néanmoins, cet usage n'est possible que si la moyenne μ et la variance (donc l'écart type σ) existent, ou, à défaut, un paramètre de position et un paramètre de dispersion. Pour obtenir la variable centrée réduite u , il faut en premier lieu centrer la variable sur la moyenne (ou le paramètre de position) :

$$x - \mu \quad (111)$$

puis réduire par l'écart type (ou le paramètre de dispersion) la variable centrée :

$$u = \frac{x - \mu}{\sigma} \quad (112)$$

L'opération permet de placer la moyenne à l'origine de la distribution. Cela permet de repérer facilement la tendance centrale sur un graphique. La moyenne de la variable centrée réduite est par conséquent nulle.

$$\langle u \rangle = 0 \quad (113)$$

L'écart type de la variable centrée réduite, c'est-à-dire l'unité du nouveau système de coordonnées vaut un.

$$\sigma_u = 1 \quad (114)$$

Pour illustrer le cas de distributions sans moyenne et sans écart type, il est possible de citer les lois de Pareto dont l'exposant α est supérieur à deux.

12 Des probabilités aux statistiques...

Certains paramètres de position correspondent clairement à des probabilités. La **médiane** est la valeur dont la probabilité d'être dépassée est $\frac{1}{2}$. Les **quartiles** sont les valeurs dont la probabilité d'être dépassées sont $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{3}{4}$.

La statistique est une application des probabilités. Elle effectue des tests statistiques pour vérifier si les données sont **indépendantes**, **équiprobables**, *etc.* La notion d'indépendance est particulièrement importante si l'on souhaite lier plusieurs variables entre elles. Si les variables sont dépendantes, un test utilisera les probabilités conditionnelles, et le principe des probabilités totales. Elle propose également des **tests de distribution** afin de vérifier la pertinence des paramètres statistiques. Bref, la statistique utilise le vocabulaire probabiliste. Elle établit l'opérationnalité d'un modèle probabiliste par rapport au jeu de données. Encore une fois, ce chapitre ne fait que présenter les notions clés qui seront présentes dans les autres chapitres. Il faut par conséquent bien avoir assimilé ces notions abstraites afin de poursuivre sereinement la lecture de ce texte.