collection CAPES / Agrégation

CAPES
AGREG

MATHS

PROBABILITÉS et STATISTIQUES

pour le CAPES et l'Agrégation interne

deuxième édition

Jérôme ESCOFFIER

ellipses

PROBABILITÉS ET STATISTIQUES pour le CAPES externe et l'Agrégation interne de Mathématiques

CAPES/AGREG

MATHÉMATIQUES

PROBABILITÉS ET STATISTIQUES pour le CAPES externe et l'Agrégation interne de Mathématiques

deuxième édition

Jérôme ESCOFFIER
Professeur agrégé en classes préparatoires
au lycée Saint-Joseph d'Avignon

ellipses

ISBN 978-2-7298-5850-6

32, rue Bargue 75740 Paris cedex 15

www.editions-ellipses.fr

À Paul et Mathieu

Avant-propos

Cet ouvrage est le fruit de ma participation aux Jurys des Concours ainsi qu'à la préparation des candidats.
Il propose aux candidats préparant le CAPES de Mathématiques un cours de Probabilités correspondant au programme de l'Écrit, ainsi que des exercices, tous corrigés, sur chaque chapitre.
Il s'adresse également aux candidats à l'Agrégation Interne, dont le programme du concours comprend celui du CAPES et quelques points spécifiques (signalés comme tels) qui sont détaillés : fonctions génératrices, théorème de transfert pour une variable à densité, vecteurs à densité et loi forte des grands nombres.
Le dernier chapitre est spécifiquement destiné aux candidats au CAPES : les leçons d'Oral du CAPES relatives aux dénombrements et aux probabilités sont traitées et commentées, ainsi que celle portant sur les séries statistiques doubles, qui fait l'objet d'un complément indispensable.

Les Probabilités posent souvent des difficultés aux candidats car certains n'ont pas suivi de cours de Probabilités après le Lycée, tandis que beaucoup d'autres ont abordé cette discipline directement avec le point de vue de la théorie de la mesure.
Le cours est traité conformément au programme, c'est-à-dire sans référence à la théorie de la mesure. Les outils utilisés (fonction Γ, intégrale de Gauss, suites et séries, intégrales généralisées, algèbre bilinéaire...) font l'objet à chaque occasion de rappels et de développements classiques, permettant ainsi aux candidats de réviser des points fondamentaux d'analyse et d'algèbre.
Deux points hors programme m'ont semblé indispensables : la tribu des boréliens et les familles sommables ; ils sont traités en annexe.

J'espère que ce livre permettra aux candidats d'aborder les Probabilités avec une base solide.

Ce livre pourra également intéresser tout professeur ayant à enseigner les Probabilités au Lycée ou dans les deux premières années de l'Enseignement Supérieur.

Les lecteurs qui le souhaitent peuvent me contacter à l'adresse suivante : jerome.escoffier@yahoo.fr ou s'adresser aux Editions Ellipses qui transmettront.

Je tiens à remercier vivement mes collègues D. Barcelo, L. Bouttier et G. Laget pour leur relecture attentive et leurs commentaires ainsi que B. Escoffier pour son aide très précieuse en LaTeX. Un très grand merci également à Christine pour son soutien.

Je salue enfin les Éditions Ellipses pour la qualité de leur travail.

Jérôme ESCOFFIER

Sommaire

Chapitre I : Espaces probabilisés

Les probabilités sont une branche des Mathématiques dont l'objet est l'étude des phénomènes aléatoires. Historiquement, il s'agissait essentiellement des jeux de hasard et des problèmes d'espérance de vie pour des calculs de rente.
L'approche mathématique de ces problèmes n'apparaît réellement qu'au 17e siècle avec Pascal et Fermat, puis au siècle suivant avec Bayes et Laplace.
L'élaboration d'un cadre mathématique rigoureux est très récente et elle est due à Kolmogorov, qui a axiomatisé le calcul des probabilités (fondements du calcul des probabilités, 1933) et a permis en particulier l'utilisation de la théorie de la mesure.

I.1 - Expérience aléatoire et univers

Définition 1.1 *On appelle expérience aléatoire une expérience sur un système dont le résultat n'est pas connu d'avance et peut varier si on répète cette expérience.*

Exemple 1.1 *Jeter une pièce de monnaie, lancer des dés, prélever des boules dans une urne, lancer une fléchette en direction d'une cible, observer la position d'une particule dans un liquide...*

Le résultat, par hypothèse unique, de la réalisation de l'expérience aléatoire est noté ω.

Définition 1.2 *On appelle univers l'ensemble des résultats possibles. Il est noté Ω.*

Exemple 1.2 *On effectue deux jets successifs d'un dé :* $\Omega = [\![1;6]\!]^2$.

Exemple 1.3 *On lance indéfiniment une pièce de monnaie : un résultat possible est alors une suite de l'ensemble $\{F;P\}$, et donc $\Omega = \{F;P\}^{\mathbb{N}}$.*

La difficulté vient du fait qu'il est possible, pour une même expérience aléatoire, de définir plusieurs univers, suivant ce que l'on entend par le terme « résultat possible ».

Par exemple, pour une expérience aléatoire consistant à prélever une boule dans une urne contenant 2 boules rouges et 3 boules noires, on peut considérer qu'un résultat possible est une couleur ($\Omega = \{N; R\}$) ou l'une des 5 boules ($\Omega = \{R_1; R_2; N_1; N_2; N_3\}$).
De même, pour le lancer d'une fléchette contre une cible, on peut considérer comme résultat possible le point d'impact ($\Omega = \mathbb{R}^2$, après avoir muni le plan d'un repère), ou la trajectoire suivie par la fléchette ($\Omega = \mathcal{C}([0;1];\mathbb{R}^3)$, ensemble des applications continues de $[0;1]$ dans $\mathbb{R}^3$).

Remarque : si on répète la même expérience aléatoire d'univers Ω, on pourra choisir comme univers Ω^n dans le cas de n répétitions, et $\Omega^{\mathbb{N}}$ si on la répète indéfiniment.

I.2 - Événements

C'est une propriété $\mathcal{E}$ énonçable (c'est-à-dire accessible à l'expérience), vérifiée ou non selon le résultat obtenu. On l'identifie à l'ensemble des résultats de l'expérience pour lesquels elle est vérifiée :

$$\mathcal{E} \text{ est identifiée à } \{\omega \in \Omega / \mathcal{E} \text{ est vraie lorsque } \omega \text{ est réalisé}\}.$$

En pratique, on ne s'intéresse souvent qu'à un sous-ensemble d'événements. C'est le cas par exemple lorsque $\Omega = \mathbb{R}$: on considère les parties boréliennes, et pas toutes les parties de $\mathbb{R}$ (Voir annexe 1).

Notation : on note $\mathcal{A}$ l'ensemble des événements relatifs à l'expérience aléatoire. On a donc par définition $\mathcal{A} \subset \mathcal{P}(\Omega)$.

On souhaite pouvoir définir certaines opérations sur les événements, et on utilise la correspondance suivante entre vocabulaires probabiliste et ensembliste :

Terminologie probabiliste	Terminologie ensembliste	Notation
Événement certain	Ensemble entier	Ω
Événement impossible	Ensemble vide	$\emptyset$
Événement élémentaire	Singleton	$\{\omega\}$
Événement contraire de A	Complémentaire de A	$\overline{A}$
A ou B	Réunion de A et B	$A \cup B$
A et B	Intersection de A et B	$A \cap B$
A implique B	A inclus dans B	$A \subset B$
A et B incompatibles	A et B disjoints	$A \cap B = \emptyset$
ω réalise A	ω appartient à A	$\omega \in A$

On exige que $\mathcal{A}$ contienne l'ensemble vide et l'univers, et qu'il soit stable *pour les opérations de réunion et d'intersection finies ou dénombrables, ainsi que par passage au complémentaire :*

i $\Omega \in \mathcal{A}$ et $\emptyset \in \mathcal{A}$

ii $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$

iii $(\forall i \in I \subset \mathbb{N}, A_i \in \mathcal{A}) \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{A}$

iv $(\forall i \in I \subset \mathbb{N}, A_i \in \mathcal{A}) \Rightarrow \bigcap_{i \in I} A_i \in \mathcal{A}$

c'est-à-dire que $\mathcal{A}$ soit une tribu de parties de Ω.

Remarque : ii et iii impliquent iv (utiliser les lois de De Morgan).
Si l'on s'intéresse à un ensemble d'événements $\mathcal{F}$ qui ne forme pas une tribu, on posera $\mathcal{A} = \mathcal{T}(\mathcal{F})$, tribu engendrée par $\mathcal{F}$ (c'est la plus petite tribu, au sens de l'inclusion, qui contient $\mathcal{F}$).

Exemple 1.4

1. *Si Ω est fini ou dénombrable, on s'intéresse naturellement aux événements élémentaires $\{\omega_i\}_{i \in I}$. La tribu engendrée est alors $\mathcal{P}(\Omega)$.*
2. *Si $\mathcal{F}$ est constitué d'un nombre fini ou dénombrable d'événements $(A_i)_{i \in I}$, qui forment une partition de Ω, la tribu engendrée est exactement l'ensemble des réunions quelconques d'événements A_i. Par exemple, en considérant une partition $\mathcal{F} = \{A, \overline{A}\}$ de Ω, on a $\mathcal{A} = \{\emptyset, \Omega, A, \overline{A}\}$.*
3. *Si $\Omega = \mathbb{R}$ ou $\Omega = \mathbb{R}^n$, on considère souvent comme tribu $\mathcal{A}$ la tribu des boréliens (engendrée par les ouverts, ou bien par les pavés $\prod_{i=1}^{n} [a_i; b_i]$, avec pour tout $i \in [\![1; n]\!], a_i \leq b_i$, ou bien, dans le cas de $\mathbb{R}$, par les ouverts $]-\infty; a[, a \in \mathbb{R}$). On ne peut pas prendre $\mathcal{P}(\mathbb{R})$ (voir annexe 1).*

Définition 1.3 *On appelle événement tout élément de la tribu $\mathcal{A}$.*

Définition 1.4 *On appelle événement élémentaire tout singleton (partie constituée d'un seul élément) de la tribu $\mathcal{A}$.*

Exemple 1.5 *On lance deux fois un dé à six faces. On pose $\Omega = [\![1; 6]\!]^2$ et $\mathcal{A} = \mathcal{P}(\Omega)$.*
On considère A : « la somme obtenue est supérieure ou égale à 11 ».
On a : $A = \{(5; 6); (6; 5); (6; 6)\}$ et A est un événement ($A \in \mathcal{A}$).
On considère B : « la somme obtenue est divisible par 3 et par 4 ».
On a : $B = \{(6; 6)\}$ et B est un événement élémentaire.

Attention : ne pas confondre résultat possible ($\omega \in \Omega$) et événement élémentaire ($\{\omega\} \in \mathcal{P}(\Omega)$). En particulier, on verra, lorsqu'on aura défini une probabilité P, que l'écriture $P(\omega)$ n'a aucun sens, et on prendra bien garde à écrire $P(\{\omega\})$.

Définition 1.5 *Le couple $(\Omega; \mathcal{A})$, où Ω est l'univers et $\mathcal{A}$ une tribu de parties de Ω, est appelé espace probabilisable.*

I.3 - Probabilité

Il s'agit d'affecter à chaque événement A un poids $P(A)$ indiquant sa « chance » d'être réalisé si l'on effectue l'expérience aléatoire.

Des considérations relatives à l'expérience peuvent conduire, dans le cas où Ω est fini, à affecter des poids de probabilité identiques à chaque événement élémentaire : ce sera l'hypothèse d'équiprobabilité, par exemple lorsqu'on jette un dé équilibré, que l'on tire des cartes dans un jeu bien battu, que l'on prélève des boules indiscernables...

Une autre approche est l'approche « fréquentiste » : on se donne un événement fixé A, on répète n fois l'expérience aléatoire et on note n_A le nombre de fois où l'événement A a été réalisé. La fréquence de réalisation de A au cours de ces n répétitions est donc : $\frac{n_A}{n}$. On démontrera que, lorsque n tend vers l'infini, $\frac{n_A}{n}$ fluctue de moins en moins, autour d'une valeur limite f_A (c'est la loi faible des grands nombres, voir chapitre VIII), appelée fréquence de réalisation de A.
La limite principale de cette approche est qu'elle ne concerne que des expériences aléatoires aisément répétables, et dans des conditions identiques.
Cependant, l'application f possède les propriétés élémentaires suivantes :

1. $\forall A \in \mathcal{A}, f_A \geq 0$ (positivité)
2. $f_\Omega = 1$ (totalité)
3. $\forall (A; B) \in \mathcal{A}^2, A \cap B = \emptyset \Rightarrow f_{A \cup B} = f_A + f_B$ (additivité)

Par analogie avec ces trois propriétés de la fréquence, on définit ce qu'est une probabilité de manière purement axiomatique (présentation due à Kolmogorov), sans référence à une quelconque observation de la réalisation de l'expérience aléatoire :

Définition 1.6 *Étant donné un espace probabilisable $(\Omega; \mathcal{A})$, on appelle probabilité sur $(\Omega; \mathcal{A})$ toute application $P : \mathcal{A} \to \mathbb{R}$ satisfaisant aux trois axiomes suivants :*

1. $\forall A \in \mathcal{A}, P(A) \geq 0$ *(positivité)*
2. $P(\Omega) = 1$ *(totalité)*
3. *Pour toute suite* (A_n) *d'éléments deux à deux disjoints,*
$$P\left(\bigcup_n A_n\right) = \sum_{n=0}^{+\infty} P(A_n) \quad \textit{(additivité)}$$

Remarque : avec le vocabulaire de la théorie de la mesure, P est donc une mesure positive finie, de masse totale égale à 1.

Définition 1.7 *Soit* $A \in \mathcal{A}$.
Si $P(A) = 0$, *on dit que* A *est un événement presque impossible.*
Si $P(A) = 1$, *on dit que* A *est un événement presque certain.*

Remarque : on se place dans le cas d'un univers fini, $\Omega = \{\omega_1; \cdots; \omega_n\}$, et on s'intéresse aux événements élémentaires en posant comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$. La donnée de n nombres réels positifs $p_1, \cdots, p_n$ de somme égale à 1 permet de définir une probabilité P sur $\mathcal{A}$ en posant :

i $\forall i \in [\![1; n]\!], P(\{\omega_i\}) = p_i$

ii $\forall A \in \mathcal{P}(\Omega), P(A) = \sum\limits_{i/\omega_i \in A} p_i$ (puisque P doit être additive)

Prenons le cas du jet d'un dé, où l'on pose $\Omega = [\![1; 6]\!]$.
La donnée de
$$p_1 = p_2 = p_3 = p_4 = p_5 = \frac{1}{10} \text{ et } p_6 = \frac{5}{10}$$
permet de définir une probabilité P comme décrit ci-dessus.
Si on note A l'événement « le résultat est pair », alors :
$$P(A) = \sum_{i/\omega_i \in A} p_i = p_2 + p_4 + p_6 = \frac{7}{10}$$

On procède de la même façon lorsque Ω est dénombrable, en se donnant une suite de réels positifs $(p_i)_{i \in \mathbb{N}}$ tels que la série $\sum\limits_{i \geq 0} p_i$ converge vers 1.

Cas particulier important : **le cas d'équiprobabilité (Ω fini, de cardinal n).**

Des considérations relatives à l'expérience (dé équilibré, boules indiscernables, ...) peuvent conduire à définir une probabilité P sur $\mathcal{P}(\Omega)$ telle que :
$$P(\{\omega_1\}) = P(\{\omega_2\}) = \cdots = P(\{\omega_n\})$$

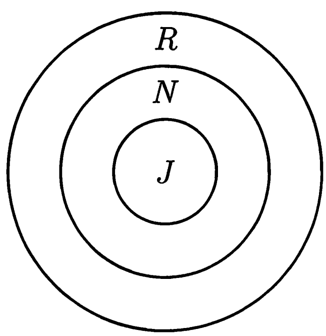
On a alors : $\forall i \in [\![1;n]\!], P(\{\omega_i\}) = \frac{1}{n}$, et P est entièrement déterminée, puisqu'elle est connue sur les n événements élémentaires $\{\omega_i\}$. En effet :

$$\begin{aligned}\forall A \in \mathcal{P}(\Omega), P(A) &= P\Big(\bigcup_{\omega \in A}\{\omega\}\Big)\\ &= \sum_{\omega\in A} P(\{\omega\}) \text{ (additivité de } P)\\ &= \sum_{\omega\in A}\frac{1}{n}\\ &= \frac{\mathrm{Card}(A)}{n}\end{aligned}$$

Définition 1.8 *L'application P définie ci-dessus est appelée la probabilité uniforme.*

Dans ce cas, les calculs de probabilité se ramènent à des problèmes de dénombrement, puisque $P(A) = \dfrac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$.

Exemple de probabilité non uniforme : on lance une fléchette contre la cible suivante :



On note la couleur obtenue : on pose donc $\Omega = \{J; N; R\}$.
Si la probabilité de chacun des trois événements élémentaires est proportionnelle à l'aire de chacune des trois parties colorées, on a :

$$P(\{J\}) = \frac{1}{9}; P(\{N\}) = \frac{3}{9}; P(\{R\}) = \frac{5}{9}$$

L'exemple 1.6 constitue aussi un cas où P n'est pas uniforme.

Propriété 1.1 *Toute probabilité P possède les propriétés suivantes :*

1. $P(\emptyset) = 0$

2. $\forall (A; B) \in \mathcal{A}^2, A \subset B \Rightarrow P(A) \leq P(B)$

3. $\forall A \in \mathcal{A}, P(A) \in [0; 1]$ *et* $P(A) = 1 - P(\overline{A})$

4. *Formule du crible de Poincaré :* $\forall (A_1; \cdots; A_n) \in \mathcal{A}^n$,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\ &+ \cdots \\ &+ (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \\ &+ \cdots \\ &+ (-1)^{n-1} P(A_1 \cap \cdots \cap A_n) \end{aligned}$$

5. *Si* $(A_i)_{i \in \mathbb{N}}$ *est une suite croissante d'événements (au sens de l'inclusion), alors la suite de réels* $(P(A_i))_{i \in \mathbb{N}}$ *est croissante, et :*

$$\lim_{n \to +\infty} P(A_n) = P\left(\bigcup_{i=0}^{+\infty} A_i\right)$$

6. *Si* $(A_i)_{i \in \mathbb{N}}$ *est une suite décroissante d'événements (au sens de l'inclusion), alors la suite de réels* $(P(A_i))_{i \in \mathbb{N}}$ *est décroissante, et :*

$$\lim_{n \to +\infty} P(A_n) = P\left(\bigcap_{i=0}^{+\infty} A_i\right)$$

Remarque : pour $n = 2$, la formule de Poincaré s'écrit :

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Pour $n = 3$, elle s'écrit :

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

Démonstration des propriétés de P :

1. On utilise l'additivité de P :

$$\begin{aligned} P(\Omega \cup \emptyset) &= P(\Omega) + P(\emptyset) \\ \Rightarrow P(\Omega) &= P(\Omega) + P(\emptyset) \\ \Rightarrow P(\emptyset) &= 0 \end{aligned}$$

2. On décompose B : $A \subset B \Rightarrow B = (B \cap \overline{A}) \cup A$ (union disjointe), d'où :

$$\begin{aligned} P(B) &= P(B \cap \overline{A}) + P(A) \\ \Rightarrow P(B) - P(A) &= P(B \cap \overline{A}) \\ \Rightarrow P(B) - P(A) &\geq 0 \text{ (par positivité de } P) \end{aligned}$$

3. D'après la propriété précédente :

$$\begin{aligned} A \subset \Omega &\Rightarrow P(A) \leq P(\Omega) \\ &\Rightarrow P(A) \leq 1 \end{aligned}$$

et d'autre part :

$$A \cup \overline{A} = \Omega \Rightarrow P(A) + P(\overline{A}) = 1$$

4. On démontre la formule par récurrence sur n (voir l'exercice 3.1 pour une autre démonstration, utilisant les propriétés de l'espérance). Soit P_n la propriété : $\forall (A_1; \cdots; A_n) \in \mathcal{A}^n$,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) = &\sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\ &+ \cdots \\ &+ (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \\ &+ \cdots \\ &+ (-1)^{n-1} P(A_1 \cap \cdots \cap A_n) \end{aligned}$$

Pour $n = 1$: il est clair que P_1 est vraie.
Avant de montrer que : $\forall n \in \mathbb{N}, P_n \Rightarrow P_{n+1}$, on va démontrer que P_2 est vraie car la méthode sera identique pour montrer l'hérédité :

$$\begin{aligned} A_1 \cup A_2 &= A_1 \cup (A_2 \cap \overline{A}_1) \text{ (union disjointe)} \\ \Rightarrow P(A_1 \cup A_2) &= P(A_1) + P(A_2 \cap \overline{A}_1) \\ \text{Or : } A_2 &= (A_2 \cap A_1) \cup (A_2 \cap \overline{A}_1) \text{ (union disjointe)} \\ \Rightarrow P(A_2) &= P(A_2 \cap A_1) + P(A_2 \cap \overline{A}_1) \\ \text{d'où : } P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \end{aligned}$$

On démontre à présent que : $\forall n \in \mathbb{N}, P_n \Rightarrow P_{n+1}$.

$$\begin{aligned} \bigcup_{i=1}^{n+1} A_i &= \left(\bigcup_{i=1}^{n} A_i\right) \cup \left(A_{n+1} \cap \overline{\bigcup_{i=1}^{n} A_i}\right) \text{ (union disjointe)} \\ \Rightarrow P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i\right) + P\left(A_{n+1} \cap \overline{\bigcup_{i=1}^{n} A_i}\right) \end{aligned}$$

Or A_{n+1} se décompose sous la forme suivante :

$$A_{n+1} = \left(A_{n+1} \cap \bigcup_{i=1}^{n} A_i\right) \cup \left(A_{n+1} \cap \overline{\bigcup_{i=1}^{n} A_i}\right) \text{ (union disjointe)}$$

On a donc :

$$P\Big(\bigcup_{i=1}^{n+1} A_i\Big) = P\Big(\bigcup_{i=1}^{n} A_i\Big) + P(A_{n+1}) - P\Big(A_{n+1} \cap (\bigcup_{i=1}^{n} A_i)\Big)$$

On développe ensuite en utilisant P_n :

$$\begin{aligned} P\Big(A_{n+1} \cap (\bigcup_{i=1}^{n} A_i)\Big) &= P\Big(\bigcup_{i=1}^{n}(A_i \cap A_{n+1})\Big) \\ &= \sum_{i=1}^{n} P(A_i \cap A_{n+1}) \\ &\quad - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{n+1}) + \cdots \\ &\quad +(-1)^{n-1} P(A_1 \cap \cdots \cap A_{n+1}) \end{aligned}$$

Donc finalement :

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n+1} A_i\Big) &= \sum_{1 \leq i \leq n+1} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n+1} P(A_{i_1} \cap A_{i_2}) \\ &\quad + \cdots \\ &\quad +(-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n+1} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \\ &\quad + \cdots \\ &\quad +(-1)^{n} P(A_1 \cap \cdots \cap A_{n+1}) \end{aligned}$$

5. La suite de réels $(P(A_n))$ est croissante (d'après la propriété 2), et majorée par 1 (d'après la propriété 3). Elle est donc convergente, et sa limite est inférieure ou égale à 1 (c'est un résultat important d'analyse : une suite croissante majorée de réels converge, à savoir démontrer).
L'idée de la démonstration repose sur l'introduction d'une suite d'événements deux à deux disjoints, ce qui va permettre d'utiliser l'additivité de P.
La suite (A_n) étant croissante (pour l'inclusion), on a, pour tout entier n, $\bigcup_{i=0}^{n} A_i = A_n$; on définit une suite d'événements $(B_i)_{i \in \mathbb{N}}$ par :

$$B_0 = A_0 \text{ et } \forall i \geq 1, B_i = A_i \backslash A_{i-1}$$

(la notation $\backslash$ signifie « privé de », autrement dit $A\backslash B = A \cap \overline{B}$).
On a :

$$\forall n \geq 0, \bigcup_{i=0}^{n} A_i = \bigcup_{i=0}^{n} B_i$$

En effet : soit $x \in \bigcup_{i=0}^{n} A_i$ et soit $i_0 = \min\{i \in [\![0;n]\!] / x \in A_i\}$.

Si $i_0 = 0$, $x \in A_0 = B_0$, donc $x \in \bigcup_{i=0}^{n} B_i$.

Si $i_0 > 0$, $x \in A_{i_0}$ et $x \notin A_{i_0-1}$, donc $x \in B_{i_0} \subset \bigcup_{i=0}^{n} B_i$.

D'où :

$$\bigcup_{i=0}^{n} A_i \subset \bigcup_{i=0}^{n} B_i$$

L'autre inclusion est évidente puisque pour tout entier i, $B_i \subset A_i$. On en déduit, grâce à l'additivité de P :

$$P\left(\bigcup_{i=0}^{+\infty} A_i\right) = P\left(\bigcup_{i=0}^{+\infty} B_i\right) = \sum_{i=0}^{+\infty} P(B_i)$$

On revient ensuite aux sommes partielles :

$$\begin{aligned}
\sum_{i=0}^{+\infty} P(B_i) &= \lim_{n\to+\infty} \sum_{i=0}^{n} P(B_i) \\
&= \lim_{n\to+\infty} P\left(\bigcup_{i=0}^{n} B_i\right) \\
&= \lim_{n\to+\infty} P\left(\bigcup_{i=0}^{n} A_i\right) \\
&= \lim_{n\to+\infty} P(A_n)
\end{aligned}$$

6. Passer au complémentaire.

Remarque : les propriétés qui viennent d'être démontrées sont communes à toutes les mesures positives (sauf la troisième pour laquelle il faut supposer en plus que la mesure de Ω vaut 1).

Définition 1.9 *Le triplet $(\Omega; \mathcal{A}; P)$ est appelé espace probabilisé.*

I.4 - Probabilité conditionnelle

On se donne un événement B qui est réalisé et, disposant de cette information, on souhaite modifier la probabilité affectée aux événements A de la tribu $\mathcal{A}$ (ceux qui ont une intersection vide avec B auront alors naturellement une probabilité nulle, et B aura bien sûr une probabilité égale à 1).
On ne change donc pas d'espace probabilisable (c'est toujours le couple $(\Omega; \mathcal{A})$), on modifie seulement l'application P.

Définition 1.10 *Soit B un événement de probabilité non nulle. On appelle probabilité conditionnelle à B, ou probabilité sachant B, associée à P l'application :*

$$P_B : \begin{cases} \mathcal{A} \to \mathbb{R} \\ A \mapsto \dfrac{P(A \cap B)}{P(B)} \end{cases}$$

Notation : on note aussi $P(A/B)$ pour $P_B(A)$. On sera alors attentif au fait que c'est une simple notation, et que A/B n'est pas un événement. On utilise dans la suite indifféremment l'une de ces deux notations.

Propriété 1.2 *L'application P_B est une probabilité sur* $(\Omega; \mathcal{A})$.

Démonstration : laissée au lecteur (vérifier les trois conditions).

Cas particulier de l'équiprobabilité :
on se place dans le cas où l'univers Ω est fini, la tribu $\mathcal{A}$ est égale à $\mathcal{P}(\Omega)$ et P est la probabilité uniforme. Soit B un événement de probabilité non nulle, distinct de Ω. On a alors :

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{\dfrac{\operatorname{Card}(A \cap B)}{\operatorname{Card}(\Omega)}}{\dfrac{\operatorname{Card}(B)}{\operatorname{Card}(\Omega)}} = \frac{\operatorname{Card}(A \cap B)}{\operatorname{Card}(B)}$$

On en déduit la valeur de P_B sur les événements élémentaires :

$$\begin{aligned} \forall \omega \in \overline{B}, P_B(\{\omega\}) &= 0 \\ \forall \omega \in B, P_B(\{\omega\}) &= \frac{1}{\operatorname{Card}(B)} \end{aligned}$$

La probabilité P_B n'est donc pas uniforme, mais sa restriction à $\mathcal{P}(B)$ est uniforme, et elle est bien sûr toujours nulle sur les événements élémentaires qui ne sont pas inclus dans B (voir l'exemple 1.6).

Propriété 1.3 *Soit B un événement de probabilité non nulle. Alors :*

$$\forall A \in \mathcal{A}, P(A \cap B) = P(A/B)P(B)$$

Par récurrence, on obtient, sous réserve que toutes les probabilités conditionnelles qui suivent soient bien définies, la formule des probabilités enchaînées :

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= P(A_1/(A_2 \cap \dots \cap A_n))P(A_2/(A_3 \cap \dots \cap A_n)) \\ &\quad \dots P(A_{n-1}/A_n)P(A_n) \end{aligned}$$

I.5 - Formule des probabilités totales

Définition 1.11 *On appelle système complet d'événements toute famille d'événements $(A_i)_{i \in I}$ (I étant une partie non vide de $\mathbb{N}$) telle que :*

1. $\forall i \in I, A_i \neq \emptyset$
2. $\forall (i;j) \in I^2, i \neq j \Rightarrow A_i \cap A_j = \emptyset$
3. $\bigcup_{i \in I} A_i = \Omega$

Propriété 1.4 *Soit $(A_i)_{i \in I}$ un système complet d'événements. Alors on a la formule des probabilités totales :*

$$\forall A \in \mathcal{A}, P(A) = \sum_{i \in I} P(A/A_i)P(A_i)$$

(sous réserve que pour tout i, $P(A_i)$ soit non nulle, afin que les probabilités conditionnelles soient bien définies).

Démonstration : comme $(A_i)_{i \in I}$ est un système complet, on peut écrire $A = A \cap \left(\bigcup_{i \in I} A_i\right)$ et on a donc la décomposition de A sous forme d'union disjointe :

$$A = \bigcup_{i \in I} \left(A \cap A_i\right)$$

Par additivité de P, $P(A) = \sum_{i \in I} P(A \cap A_i) = \sum_{i \in I} P(A/A_i)P(A_i)$.

Remarque : on écrit souvent la formule des probabilités totales relativement au système complet $\{A_1; \overline{A}_1\}$:

$$P(A) = P(A/A_1)P(A_1) + P(A/\overline{A}_1)P(\overline{A}_1)$$

Exemple 1.6 *On considère une urne U_1 contenant deux boules blanches et une boule noire, et une urne U_2 contenant une boule blanche et une boule noire. On choisit une urne au hasard puis on prélève une boule dans cette urne. Les boules sont indiscernables au toucher.*
Quelle est la probabilité de tirer une boule blanche ?
On commence par poser l'univers :

$$\Omega = \{(U_1; B_1); (U_1; B_2); (U_1; N_1); (U_2; B_3); (U_2; N_2)\}$$

On note B l'événement « la boule tirée est blanche », E_1 « la boule tirée est extraite de l'urne U_1 » et E_2 « la boule tirée est extraite de l'urne U_2 ».
D'après l'énoncé, $P(E_1) = P(E_2)$ et comme le système $\{E_1; E_2\}$ est complet, $P(E_1) + P(E_2) = 1$. Par conséquent, $P(E_1) = P(E_2) = \frac{1}{2}$.
L'énoncé permet ensuite de définir les probabilités conditionnelles P_{E_1} et P_{E_2} : les boules étant indiscernables, P_{E_1} prend la même valeur sur chaque événement élémentaire appartenant à $\mathcal{P}(E_1)$, et elle est nulle sur chaque événement élémentaire appartenant à $\mathcal{P}(\overline{E}_1)$.
On a donc :

$$P_{E_1}(\{(U_1; B_1)\}) = P_{E_1}(\{(U_1; B_2)\}) = P_{E_1}(\{(U_1; N_1)\}) = \frac{1}{3}$$

et :

$$P_{E_1}(\{(U_2; B_3)\}) = P_{E_1}(\{(U_2; N_2)\}) = 0$$

Le même raisonnement permet de définir P_{E_2} :

$$P_{E_2}(\{(U_2; B_3)\}) = P_{E_2}(\{(U_2; N_2)\}) = \frac{1}{2}$$

et :

$$P_{E_2}(\{(U_1; B_1)\}) = P_{E_2}(\{(U_1; B_2)\}) = P_{E_2}(\{(U_1; N_1)\}) = 0$$

La probabilité P est alors entièrement déterminée puisqu'on a, par la formule des probabilités totales :

$$\begin{aligned} \forall A \in \mathcal{P}(\Omega), P(A) &= P_{E_1}(A)P(E_1) + P_{E_2}(A)P(E_2) \\ &= P_{E_1}(A) \times \frac{1}{2} + P_{E_2}(A) \times \frac{1}{2} \end{aligned}$$

On l'applique à l'événement B :

$$P(B) = P_{E_1}(B) \times \frac{1}{2} + P_{E_2}(B) \times \frac{1}{2} = \frac{2}{3} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{7}{12}$$

On obtient de la même façon la valeur de P sur chacun des 5 événements élémentaires :

$$P(\{(U_1; B_1)\}) = P(\{(U_1; B_2)\}) = P(\{(U_1; N_1)\}) = \frac{1}{6}$$

et :

$$P(\{(U_2;B_3)\}) = P(\{(U_2;N_2)\}) = \frac{1}{4}$$

En particulier, on remarquera bien que P n'est pas la probabilité uniforme !

I.6 - Formule de Bayes

Elle était appelée « formule de la probabilité des causes », car elle permet de calculer $P(A/B)$ connaissant $P(B/A)$.

Propriété 1.5 *Soient A et B deux événements de probabilité non nulle. Alors on peut écrire :*

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \quad \textit{(formule de Bayes)}$$

Démonstration : $P(A \cap B) = P(A/B)P(B)$ et $P(B \cap A) = P(B/A)P(A)$, puis $P(A \cap B) = P(B \cap A)$.

Cas particulier : lorsqu'on a un système complet $(A_i)_{i \in I}$, on obtient :

$$\begin{aligned} \forall n \in I, P(A_n/A) &= \frac{P(A/A_n)P(A_n)}{P(A)} \\ &= \frac{P(A/A_n)P(A_n)}{\sum_{i \in I} P(A/A_i)P(A_i)} \end{aligned}$$

Dans le cas d'un système complet $\{A_1; \overline{A}_1\}$:

$$P(A_1/A) = \frac{P(A/A_1)P(A_1)}{P(A/A_1)P(A_1) + P(A/\overline{A}_1)P(\overline{A}_1)}$$

Exemple 1.7 *On considère une urne U_1 contenant deux boules blanches et une boule noire, et une urne U_2 contenant une boule blanche et une boule noire. On choisit une urne au hasard puis on prélève une boule dans cette urne. Les boules sont indiscernables au toucher. On obtient une boule blanche.*
Quelle est la probabilité que la boule soit extraite de l'urne U_1 ?
On reprend l'univers et les notations précédentes :

$$\Omega = \{(U_1;B_1);(U_1;B_2);(U_1;N_1);(U_2;B_3);(U_2;N_2)\}$$

On note B l'événement « la boule tirée est blanche », E_1 « la boule tirée est extraite de l'urne U_1 », et E_2 « la boule tirée est extraite de l'urne U_2 ». Une probabilité P a été définie, via les probabilités conditionnelles P_{E_1} et

P_{E_2} (voir exemple 1.6).
On applique la formule de Bayes en utilisant le système complet $\{E_1; E_2\}$:

$$\begin{aligned} P(E_1/B) &= \frac{P(B/E_1)P(E_1)}{P(B/E_1)P(E_1)+P(B/E_2)P(E_2)} \\ &= \frac{\frac{2}{3}\times\frac{1}{2}}{\frac{2}{3}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}} = \frac{4}{7} \end{aligned}$$

I.7 - Indépendance de deux événements

On traduit le fait que la réalisation de l'un n'a pas d'influence sur la réalisation de l'autre.

Définition 1.12 *A est dit indépendant de B si $P(A/B) = P(A)$ ou si $P(B) = 0$.*

Le fait d'ajouter le cas $P(B) = 0$ permet d'avoir la condition nécessaire et suffisante suivante :

Propriété 1.6 *A est indépendant de B si et seulement si $P(A \cap B) = P(A)P(B)$.*

Démonstration : on suppose que A est indépendant de B.
Si $P(B) = 0$, alors $P(A \cap B) = 0$ car $(A \cap B) \subset B$.
Si $P(B) \neq 0$, alors $P(A/B) = P(A)$ donc $P(A \cap B) = P(A)P(B)$.
Réciproquement, on suppose que $P(A \cap B) = P(A)P(B)$.
Si $P(B) = 0$, alors d'après la définition A est indépendant de B.
Si $P(B) \neq 0$, alors $\dfrac{P(A \cap B)}{P(B)} = P(A)$, donc $P(A/B) = P(A)$ et d'après la définition, on a bien que A est indépendant de B.

Remarque : la propriété précédente montre que la relation est symétrique, et on dit alors que A et B sont indépendants.

Exemple 1.8 *On jette un dé équilibré.*
On pose $\Omega = [\![1; 6]\!]$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P la probabilité uniforme.
Soient les événements suivants : $A = \{2; 4; 6\}$, $B = \{5; 6\}$ et $C = \{5\}$. On a immédiatement :
$P(A) = \dfrac{1}{2}$, $P(B) = \dfrac{1}{3}$, $P(C) = \dfrac{1}{6}$, $P(A \cap B) = \dfrac{1}{6}$, $P(A \cap C) = 0$ *et* $P(B \cap C) = \dfrac{1}{6}$.
On en conclut que A et B sont indépendants, mais que A et C ainsi que B et C ne le sont pas.

Attention : ne pas confondre incompatibilité ($A \cap B = \emptyset$) et indépendance ($P(A \cap B) = P(A)P(B)$).
En particulier, on notera bien que la notion d'indépendance, à la différence de celle d'incompatibilité, dépend de la probabilité P.

Exercice : reprendre l'exemple consistant à jeter un dé, et définir une autre probabilité P' pour laquelle A et B ne sont pas indépendants.

Propriété 1.7

$$\begin{aligned} A \text{ et } B \text{ indépendants} &\Leftrightarrow \overline{A} \text{ et } B \text{ indépendants} \\ &\Leftrightarrow A \text{ et } \overline{B} \text{ indépendants} \\ &\Leftrightarrow \overline{A} \text{ et } \overline{B} \text{ indépendants} \end{aligned}$$

Démonstration : laissée au lecteur (en cas de difficulté, voir la correction de l'exercice 1.6).

I.8 - Indépendance de n événements

Définition 1.13 *On dit que n événements $A_1, \cdots, A_n$ sont mutuellement indépendants si :*

$$\forall I \subset [\![1; n]\!], I \neq \emptyset, P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Ils sont dits indépendants dans leur ensemble si :

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

La définition s'étend au cas d'une suite d'événements (A_n), en considérant l'indépendance mutuelle des événements de toute sous famille finie.

Remarque : l'indépendance mutuelle entraîne l'indépendance des événements 2 à 2, et la réciproque est fausse, comme le montre l'exemple suivant :

Exemple 1.9 *On lance deux fois une pièce de monnaie équilibrée.*
On pose $\Omega = \{P; F\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P la probabilité uniforme.
On considère les événements suivants : A « le premier jet a donné Pile », B « le deuxième jet a donné Pile », et C « les deux jets ont donné le même résultat ».
On a : $P(A) = P(B) = P(C) = \frac{1}{2}, P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$, et $P(A \cap B \cap C) = \frac{1}{4}$. Les événements A, B et C ne sont donc pas mutuellement indépendants, mais ils sont indépendants deux à deux.

Exercice : étant donnés n événements, combien doit-on vérifier d'égalités pour montrer qu'ils sont mutuellement indépendants ?
(réponse : $2^n - n - 1$).

I.9 - Énoncé des exercices

Exercice 1.1 *Soient* $(\Omega_1; \mathcal{P}(\Omega_1); P_1)$ *et* $(\Omega_2; \mathcal{P}(\Omega_2); P_2)$ *deux espaces probabilisés avec* Ω_1 *et* Ω_2 *finis.*

a) Montrer que l'application $P : \{(\omega_1; \omega_2)\} \mapsto P_1(\{\omega_1\})P_2(\{\omega_2\})$ *et telle que* $\forall A \in \mathcal{P}(\Omega_1 \times \Omega_2), P(A) = \sum_{\omega \in A} P(\{\omega\})$ *est une probabilité sur l'espace probabilisable produit :* $(\Omega_1 \times \Omega_2; \mathcal{P}(\Omega_1 \times \Omega_2))$.
P est appelée probabilité produit.

b) Montrer que si P_1 *et* P_2 *sont uniformes, alors* P *est uniforme.*

Exercice 1.2 *6 personnes doivent prendre place dans 3 voitures pouvant chacune recevoir de 0 à 6 passagers. Chaque personne choisit un véhicule au hasard.*

a) Décrire Ω.

b) Quelle est la probabilité que dans chaque voiture montent exactement 2 personnes ?

Exercice 1.3 *Une urne contient n boules (indiscernables au toucher) numérotées par les entiers de 1 à n. On extrait successivement avec remise N boules.*

a) Décrire l'univers.

b) Quelle est la probabilité d'avoir obtenu au moins 2 numéros distincts ?

c) Quelle est la probabilité d'avoir obtenu au moins une fois les boules 1, 2 et 3 ?

Exercice 1.4 *Soient A et B deux événements tels que :*

$$P(A) = \frac{3}{8}, P(B) = \frac{1}{2} \text{ et } P(A \cap B) = \frac{1}{4}$$

Calculer $P(\overline{A}/\overline{B})$.

Exercice 1.5 *Soient A et B deux événements indépendants tels que :*

$$P(A) = \frac{1}{2} \text{ et } P(A \cup B) = \frac{2}{3}$$

Calculer $P(\overline{B}/A)$.

Exercice 1.6

a) Vérifier que :

$$\begin{aligned} A \text{ et } B \text{ indépendants} &\Leftrightarrow \overline{A} \text{ et } B \text{ indépendants} \\ &\Leftrightarrow A \text{ et } \overline{B} \text{ indépendants} \\ &\Leftrightarrow \overline{A} \text{ et } \overline{B} \text{ indépendants} \end{aligned}$$

b) Établir que deux événements A et B sont indépendants si et seulement si :

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

Exercice 1.7 *On considère trois urnes : U_1 composée de deux boules noires et deux boules rouges, U_2 composée d'une noire et deux rouges, U_3 composée d'une noire et trois rouges (les boules sont indiscernables au toucher). On tire une boule dans U_1, une boule dans U_2, et on les met dans U_3, puis on tire une boule dans U_3. On constate que cette boule est noire. Calculer la probabilité que la boule tirée dans U_1 ait été rouge.*

Exercice 1.8 *On suppose qu'un couple a la même probabilité d'avoir à chaque naissance un garçon ou une fille. On considère un couple ayant 2 enfants.*

a) Quelle est la probabilité qu'ils aient une fille sachant qu'ils ont un garçon ?

b) Quelle est la probabilité qu'ils aient une fille sachant que l'on a rencontré par hasard l'un des deux enfants, que c'était un garçon et qu'il était l'aîné ?

Exercice 1.9 *Soient A, B et C trois événements indépendants 2 à 2, avec $P(C) > 0$.*
Donner une condition nécessaire et suffisante sur A, B et C pour que A et B soient indépendants relativement à la probabilité conditionnelle P_C.

Exercice 1.10 *On jette une paire de dés non pipés. Calculer la probabilité que la somme obtenue soit supérieure ou égale à 10, sachant que l'un au moins a donné 5.*

Exercice 1.11 *Problème posé à Pascal par le chevalier de Méré.*
« Qu'est ce qui est le plus probable : sortir au moins un 6 en lançant 4 fois un dé ou au moins un double 6 en lançant 24 fois deux dés ? »
(Préciser l'univers dans chacun des cas au préalable).

Exercice 1.12 *Problème posé à Galilée par le Duc de Toscane.*
« Pourquoi, quand on effectue trois lancers de dé, obtient-on plus souvent la somme 10 que la somme 9, bien que chacune soit obtenue de 6 manières différentes ? »

Exercice 1.13 *Probabilités et arithmétique.*
Soit n un entier supérieur ou égal à 2. On note :

$$\phi(n) = \mathrm{Card}\{1 \leq p \leq n / pgcd(p;n) = 1\}$$

Soit $n = \prod_{i=1}^{k} p_i^{\alpha_i}$ la décomposition de n en produit de facteurs premiers. Le but est de démontrer que :

$$\phi(n) = n \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)$$

On tire au hasard un entier compris entre 1 et n. On note A l'événement « le nombre obtenu est premier avec n », et pour $i \in [\![1;k]\!]$, on note A_i l'événement « le nombre obtenu est divisible par p_i ».

1. *Définir un espace probabilisé modélisant l'expérience aléatoire.*
2. *Exprimer $P(A)$ en fonction de $\phi(n)$ et n.*
3. *Montrer que $\forall i \in [\![1;k]\!], P(A_i) = \dfrac{1}{p_i}$.*
4. *Montrer que les événements $A_1, \cdots, A_k$ sont indépendants (dans leur ensemble).*
5. *Exprimer A en fonction des A_i et en déduire que :*

$$\phi(n) = n \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)$$

Exercice 1.14 *Matrices stochastiques.*
Une particule se déplace sur les 3 sommets d'un triangle (ABC) de la façon suivante, à l'issue de chaque seconde : lorsqu'elle est en A, elle y reste avec une probabilité de 0.25, elle va en B avec une probabilité de 0.5, et en C avec une probabilité de 0.25. Lorsqu'elle est en B, elle va en A avec une probabilité de 0.5, et va en C avec une probabilité de 0.5. Lorsqu'elle est en C, elle va toujours en B. On suppose qu'elle est en A à l'instant initial.
Quelle est la probabilité qu'elle soit en B au bout de 60 secondes ?
On note A_n l'événement : « la particule est en A à l'instant n », $a_n = P(A_n)$ (mêmes notations pour B et C), et $X_n = \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}$.

1. *A l'aide de la formule des probabilités totales, déterminer* $A \in \mathcal{M}_3(\mathbb{R})$ *telle que :* $\forall n \in [\![0; 59]\!], X_{n+1} = AX_n$.
2. *A est appelée matrice stochastique. Quelle remarque peut-on faire sur ses coefficients ?*
3. *Montrer que 1 est valeur propre de A, et en déduire son spectre.*
4. *Diagonaliser A et déterminer* X_{60}.

I.10 - Correction des exercices

Exercice 1.1

a) Soient $\Omega_1 = \{a_i, 1 \leq i \leq \operatorname{Card}\Omega_1\}$ et $\Omega_2 = \{b_j, 1 \leq j \leq \operatorname{Card}\Omega_2\}$. On montre que P vérifie les trois axiomes.
Positivité : c'est clair puisque P_1 et P_2 sont positives.
Totalité :

$$\begin{aligned} P(\Omega_1 \times \Omega_2) &= \sum_{\omega \in \Omega_1 \times \Omega_2} P(\{\omega\}) \\ &= \sum_{i,j} P(\{(a_i; b_j)\}) \\ &= \sum_{i,j} P_1(\{a_i\}) P_2(\{b_j\}) \\ &= \sum_i P_1(\{a_i\}) \sum_j P_2(\{b_j\}) \\ &= P_1(\Omega_1) P_2(\Omega_2) \\ &= 1 \end{aligned}$$

Additivité : il suffit de vérifier l'additivité de P pour la réunion de deux événements disjoints. Ω étant fini, $\mathcal{P}(\Omega)$ est également fini (quel est son cardinal ?), et on étendra le résultat par une récurrence immédiate à toute union finie d'événements disjoints.
Soient A et B deux événements disjoints :

$$\begin{aligned} P(A \cup B) &= \sum_{\omega \in A \cup B} P(\{\omega\}) \\ &= \sum_{\omega \in A} P(\{\omega\}) + \sum_{\omega \in B} P(\{\omega\}) \\ &\quad \text{(A et B sont disjoints et les deux sommes sont finies)} \\ &= P(A) + P(B) \end{aligned}$$

b) Les événements élémentaires de $\mathcal{P}(\Omega_1 \times \Omega_2)$ sont les singletons, c'est-à-dire les éléments de la forme $\{(a; b)\}$ où $a \in \Omega_1$ et $b \in \Omega_2$.

On calcule la valeur de P sur les singletons :

$$\begin{aligned} P(\{(a;b)\}) &= P_1(\{a\})P_2(\{b\}) \\ &= \frac{1}{\operatorname{Card}\Omega_1} \times \frac{1}{\operatorname{Card}\Omega_2} \\ &= \frac{1}{\operatorname{Card}(\Omega_1 \times \Omega_2)} \end{aligned}$$

Remarque : on considère une expérience aléatoire pour laquelle on choisit comme espace probabilisé $(\Omega; \mathcal{P}(\Omega); P)$, P étant la probabilité uniforme. On répète n fois cette expérience. Pour munir l'espace probabilisable $(\Omega^n; \mathcal{P}(\Omega^n))$ d'une probabilité, on choisit la probabilité produit, qui n'est rien d'autre que la probabilité uniforme d'après la question b). C'est cette démarche que l'on utilise implicitement dans de nombreux exercices, dès que l'on répète une expérience aléatoire pour laquelle on est en situation d'équiprobabilité.

Exercice 1.2

a) Soient V_1, V_2 et V_3 les trois véhicules. On a 6 répétitions de la même expérience aléatoire (choisir l'un des 3 véhicules). On pose donc : $\Omega = \{V_1; V_2; V_3\}^6$.

b) On choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, car Ω est fini et on veut que la tribu contienne chaque événement élémentaire. On prend comme probabilité P la probabilité uniforme (tirages équiprobables).
Soit A l'événement « dans chaque véhicule prennent place exactement 2 personnes ».
Puisque P est uniforme, $P(A) = \dfrac{\operatorname{Card} A}{\operatorname{Card}\Omega}$.
Calcul de $\operatorname{Card} A$: A est l'ensemble des 6-uplets où apparaissent exactement 2 fois V_1, 2 fois V_2 et 2 fois V_3. On a donc $\begin{pmatrix} 6 \\ 2 \end{pmatrix}$ façons de placer les 2 V_1, puis $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ façons de placer ensuite les 2 V_2 et enfin une seule façon de placer les 2 V_3. Par conséquent : $\operatorname{Card} A = \begin{pmatrix} 6 \\ 2 \end{pmatrix}\begin{pmatrix} 4 \\ 2 \end{pmatrix} = 90$.
Comme $\operatorname{Card}(\Omega) = 3^6$, on obtient $P(A) = \dfrac{90}{81}$.

Exercice 1.3

a) On répète N fois la même expérience aléatoire (prélever une boule parmi n), et on pose donc $\Omega = \{1; \cdots; n\}^N$. On choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité P la probabilité uniforme.

b) Soit A l'événement « le résultat obtenu comporte au moins 2 numéros distincts ». On calcule la probabilité de $\overline{A}$ (qui est l'événement « le résultat obtenu comporte N fois le même numéro »).

$$P(\overline{A}) = \frac{\text{Card}(\overline{A})}{\text{Card}(\Omega)} = \frac{n}{n^N}$$

et donc $P(A) = 1 - P(\overline{A}) = 1 - \dfrac{n}{n^N}$.

c) On note, pour $i \in [\![1;3]\!]$, A_i l'événement « le résultat obtenu comporte au moins une fois le numéro i », et on cherche $P(A_1 \cap A_2 \cap A_3)$. On raisonne sur les événements contraires :

$$\begin{aligned} P(\overline{A_1 \cap A_2 \cap A_3}) &= P(\overline{A}_1 \cup \overline{A}_2 \cup \overline{A}_3) \\ &= P(\overline{A}_1) + P(\overline{A}_2) + P(\overline{A}_3) - P(\overline{A}_1 \cap \overline{A}_2) \\ &\quad -P(\overline{A}_1 \cap \overline{A}_3) - P(\overline{A}_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) \end{aligned}$$

Un calcul simple de dénombrement donne :

$$P(\overline{A}_1) = P(\overline{A}_2) = P(\overline{A}_3) = \frac{(n-1)^N}{n^N}$$

$$P(\overline{A}_1 \cap \overline{A}_2) = P(\overline{A}_1 \cap \overline{A}_3) = P(\overline{A}_2 \cap \overline{A}_3) = \frac{(n-2)^N}{n^N}$$

$$P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) = \frac{(n-3)^N}{n^N}$$

On obtient donc :

$$P(A_1 \cap A_2 \cap A_3) = 1 - 3\frac{(n-1)^N}{n^N} + 3\frac{(n-2)^N}{n^N} - \frac{(n-3)^N}{n^N}$$

Exercice 1.4

$P(\overline{B})$ étant non nul, $P(\overline{A}/\overline{B})$ est bien défini. On a :

$$P(\overline{A}/\overline{B}) = \frac{P(\overline{A} \cap \overline{B})}{P(\overline{B})} = \frac{P(\overline{A \cup B})}{P(\overline{B})}$$

Comme $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \dfrac{3}{8} + \dfrac{1}{2} - \dfrac{1}{4} = \dfrac{5}{8}$, on obtient :

$$P(\overline{A}/\overline{B}) = \frac{1 - \frac{5}{8}}{1 - \frac{1}{2}} = \frac{3}{4}$$

Exercice 1.5

$P(A)$ étant non nul, $P(\overline{B}/A)$ est bien défini. On a :

$$P(\overline{B}/A) = \frac{P(\overline{B} \cap A)}{P(A)} = \frac{P(A) - P(B \cap A)}{P(A)}$$

et comme A et B sont indépendants,

$$P(\overline{B}/A) = \frac{P(A) - P(B)P(A)}{P(A)} = 1 - P(B)$$

Calcul de $P(B)$:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)P(B) \end{aligned}$$

d'où $P(B) = \dfrac{P(A \cup B) - P(A)}{1 - P(A)} = \dfrac{1}{3}$, et donc $P(\overline{B}/A) = \dfrac{2}{3}$.

Exercice 1.6

a) Si A et B sont indépendants, alors :

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

d'où l'indépendance de $\overline{A}$ et B. En appliquant le résultat précédent à B et $\overline{A}$, on a directement que si $\overline{A}$ et B sont indépendants alors $\overline{A}$ et $\overline{B}$ sont indépendants. On en déduit de la même façon que si $\overline{A}$ et $\overline{B}$ sont indépendants, alors A et $\overline{B}$ sont indépendants, puis que A et B sont indépendants. On a donc montré les 3 équivalences (par permutation circulaire).

b) Si A et B sont indépendants, alors, d'après la propriété démontrée au a) :

$$\begin{aligned} P(A \cap B)P(\overline{A} \cap \overline{B}) &= P(A)P(B)P(\overline{A})P(\overline{B}) \\ &= P(A \cap \overline{B})P(\overline{A} \cap B) \end{aligned}$$

Réciproquement, si $P(A \cap B)P(\overline{A} \cap \overline{B}) = P(A \cap \overline{B})P(\overline{A} \cap B)$, alors on a :

$$\begin{aligned} &P(A \cap B)(1 - P(A) - P(B) + P(A \cap B)) \\ &= (P(A) - P(A \cap B)) \times (P(B) - P(A \cap B)) \end{aligned}$$

Il reste en développant : $P(A \cap B) = P(A)P(B)$, ce qui entraîne l'indépendance de A et B.

Exercice 1.7

On note, pour $i \in [\![1;3]\!]$, R_i l'événement « la boule tirée dans l'urne i est rouge » et N_i l'événement « la boule tirée dans l'urne i est noire ». On veut calculer $P(R_1/N_3)$.

D'après la formule de Bayes : $P(R_1/N_3) = \dfrac{P(N_3/R_1)P(R_1)}{P(N_3)}$. On calcule $P(N_3/R_1)$:

$$\begin{aligned}
P(N_3/R_1) &= \frac{P(N_3 \cap R_1)}{P(R_1)} \\
&= \frac{P(N_3 \cap R_1 \cap R_2) + P(N_3 \cap R_1 \cap N_2)}{P(R_1)} \\
&\quad \text{(formule des probabilités totales)} \\
&= \frac{P(N_3/(R_2 \cap R_1))P(R_2/R_1)P(R_1)}{P(R_1)} \\
&\quad + \frac{P(N_3/(N_2 \cap R_1))P(N_2/R_1)P(R_1)}{P(R_1)} \\
&= P(N_3/(R_2 \cap R_1))P(R_2/R_1) + P(N_3/(N_2 \cap R_1))P(N_2/R_1) \\
&= \frac{1}{6} \times \frac{2}{3} + \frac{2}{6} \times \frac{1}{3} \\
&= \frac{2}{9}
\end{aligned}$$

On calcule de la même façon $P(N_3)$: par la formule des probabilités totales,

$$P(N_3) = P(N_3 \cap R_1 \cap R_2) + P(N_3 \cap R_1 \cap N_2) + P(N_3 \cap N_1 \cap R_2) + P(N_3 \cap N_1 \cap N_2)$$

puis on développe chacun des 4 éléments par la formule des probabilités enchaînées :

$$\begin{aligned}
P(N_3 \cap R_1 \cap R_2) &= P(N_3/(R_1 \cap R_2))P(R_2/R_1)P(R_1) \\
&= \frac{1}{6} \times \frac{2}{3} \times \frac{2}{4}
\end{aligned}$$

On obtient finalement :

$$P(N_3) = \frac{1}{6} \times \frac{2}{3} \times \frac{2}{4} + \frac{2}{6} \times \frac{1}{3} \times \frac{2}{4} + \frac{2}{6} \times \frac{2}{3} \times \frac{2}{4} + \frac{3}{6} \times \frac{1}{3} \times \frac{2}{4} = \frac{11}{36}$$

D'où $P(R_1/N_3) = \dfrac{\frac{2}{9} \times \frac{2}{4}}{\frac{11}{36}} = \dfrac{4}{11}$.

Exercice 1.8

On pose $\Omega = \{F; G\}^2$, on choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité P la probabilité uniforme.

a) On note A l'événement « le couple a une fille » et B l'événement « le couple a un garçon ». P étant uniforme, $P(A/B) = \dfrac{\text{Card}(A \cap B)}{\text{Card}(B)} = \dfrac{2}{3}$.

b) On note G_1 l'événement « l'aîné est un garçon » :

$$P(A/G_1) = \frac{\text{Card}(A \cap G_1)}{\text{Card}(G_1)} = \frac{1}{2}$$

Exercice 1.9

$$\begin{aligned} P((A \cap B)/C) &= P(A/C)P(B/C) \\ \Leftrightarrow \frac{P(A \cap B \cap C)}{P(C)} &= \frac{P(A \cap C)}{P(C)} \times \frac{P(B \cap C)}{P(C)} \\ \Leftrightarrow \frac{P(A \cap B \cap C)}{P(C)} &= \frac{P(A)P(C)}{P(C)} \times \frac{P(B)P(C)}{P(C)} \\ \Leftrightarrow P(A \cap B \cap C) &= P(A)P(B)P(C) \end{aligned}$$

Comme les trois événements sont supposés indépendants deux à deux, on a la condition nécessaire et suffisante : A, B et C sont mutuellement indépendants pour P.

Exercice 1.10

On pose $\Omega = [\![1;6]\!]^2$, on choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité P la probabilité uniforme. Soit A l'événement « la somme obtenue est supérieure ou égale à 10 », et soit B « l'un au moins des deux dés a donné 5 ». Comme P est uniforme, $P(A/B) = \dfrac{\text{Card}(A \cap B)}{\text{Card}(B)} = \dfrac{3}{11}$.

Exercice 1.11

Premier cas : 4 lancers successifs d'un dé (équilibré).
On répète 4 fois la même expérience aléatoire et on pose donc $\Omega = [\![1;6]\!]^4$. On choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité P la probabilité uniforme. Soit A l'événement « le résultat obtenu comporte au moins un 6 ». On raisonne sur l'événement contraire :

$$P(\overline{A}) = \frac{\text{Card}(\overline{A})}{\text{Card}(\Omega)} = \frac{5^4}{6^4}$$

On a donc $P(A) = 1 - \dfrac{5^4}{6^4} \simeq 0.518$.
Deuxième cas : 24 lancers successifs de 2 dés (équilibrés).
On répète 24 fois la même expérience aléatoire et on pose donc $\Omega = ([\![1;6]\!]^2)^{24}$. On choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité P la probabilité uniforme. Soit A l'événement « le résultat obtenu

comporte au moins un double 6 ». On raisonne encore sur l'événement contraire :

$$P(\overline{A}) = \frac{\text{Card}(\overline{A})}{\text{Card}(\Omega)} = \frac{35^{24}}{36^{24}}$$

On a donc $P(A) = 1 - \frac{35^{24}}{36^{24}} \simeq 0.491$.

Exercice 1.12

On répète 3 fois la même expérience aléatoire et on pose donc $\Omega = [\![1;6]\!]^3$. On choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité P la probabilité uniforme. On note S_{10} « la somme des 3 éléments du triplet vaut 10 » et S_9 « la somme des 3 éléments du triplet vaut 9 ».

S_{10} est composé des éléments $(1;3;6), (1;4;5), (2;3;5), (2;4;4), (2;6;2)$ et $(3;3;4)$, plus tous les triplets obtenus en permutant les éléments de ces 6 triplets. Pour chacun des triplets $(1;3;6), (1;4;5)$ et $(2;3;5)$, on a $3!$ permutations, et pour chacun des triplets $(2,4,4), (2;6;2)$ et $(3;3;4)$, on a 3 permutations. Par conséquent :

$$\text{Card}(S_{10}) = 3 \times (3!) + 3 \times 3 = 27$$

En dénombrant les éléments de S_9 de la même façon, on obtient :

$$\text{Card}(S_9) = 3 \times (3!) + 2 \times 3 + 1 = 25$$

On en conclut que $P(S_9) < P(S_{10})$.

Exercice 1.13

1. On pose $\Omega = \{1; \cdots; n\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et on prend pour P la probabilité uniforme (ce qui traduit le fait que l'on est en situation d'équiprobabilité, chaque singleton $\{p\}$, $1 \leq p \leq n$, étant affecté d'une probabilité égale à $\frac{1}{n}$).
2. Comme P est uniforme, $P(A) = \frac{\text{Card}\, A}{\text{Card}\, \Omega} = \frac{\phi(n)}{n}$.
3. On a $P(A_i) = \frac{\text{Card}\, A_i}{\text{Card}\, \Omega}$. Or $A_i = \{\lambda p_i / 1 \leq \lambda \leq \frac{n}{p_i}\}$, donc $\text{Card}\, A_i = \frac{n}{p_i}$, et par conséquent $P(A_i) = \frac{1}{p_i}$.
4. l'événement $A_1 \cap \cdots \cap A_k$ est réalisé si et seulement si le nombre obtenu est divisible par chacun des p_i. Comme les p_i sont premiers entre eux, ceci équivaut à dire que le nombre obtenu est divisible par le produit des p_i (Théorème de Gauss). D'où :

$$A_1 \cap \cdots \cap A_k \quad = \quad \{\lambda p_1 \cdots p_k / 1 \leq \lambda \leq \frac{n}{p_1 \cdots p_k}\}$$

Par conséquent, $\mathrm{Card}(A_1 \cdots A_k) = \dfrac{n}{p_1 \cdots p_k}$, et donc :

$$P(A_1 \cap \cdots \cap A_k) = \frac{1}{p_1 \cdots p_k} = P(A_1) \cdots P(A_k)$$

Les événements $A_1,\dots,A_k$ sont donc indépendants (dans leur ensemble).

5. Le nombre obtenu est premier avec n si et seulement si il n'est divisible par aucun des p_i, donc $A = \bigcap_{i=1}^{k} \overline{A_i}$. Or l'indépendance des A_i entraîne l'indépendance de leur complémentaire, donc :

$$P(A) = \prod_{i=1}^{k} P(\overline{A_i}) = \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)$$

Comme $P(A) = \dfrac{\phi(n)}{n}$, on en déduit l'égalité souhaitée :

$$\phi(n) = n \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)$$

Exercice 1.14

1. A_n, B_n, C_n formant un système complet d'événements, on a :

$$\begin{aligned} P(A_{n+1}) &= P_{A_n}(A_{n+1})P(A_n) + P_{B_n}(A_{n+1})P(B_n) + P_{C_n}(A_{n+1})P(C_n) \\ &= 0.25P(A_n) + 0.5P(B_n) \end{aligned}$$

En procédant de la même façon pour $P(B_{n+1})$ et $P(C_{n+1})$, on obtient :

$$\begin{pmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \end{pmatrix} = \begin{pmatrix} 0.25 & 0.5 & 0 \\ 0.5 & 0 & 1 \\ 0.25 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}$$

2. Les coefficients de A sont tous compris entre 0 et 1 (puisque ce sont des valeurs de P_{A_n}, P_{B_n} et P_{C_n}), et la somme des coefficients de chaque colonne vaut 1 (pour la première : $P_{A_n}(A_{n+1}) + P_{A_n}(B_{n+1}) + P_{A_n}(C_{n+1}) = 1$).

3. A et ${}^tA$ ont le même spectre et d'après la remarque précédente, $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ est vecteur propre de ${}^tA$, associé à la valeur propre 1. Comme $\det(A) = 0$ et $\mathrm{tr}(A) = 0.25$, on en déduit le spectre de A :

$$Sp(A) = \{-0.75; 0; 1\}$$

4. On recherche les sous-espaces propres. On obtient :

$$E_{-0.75} = Vect\left\{\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}\right\}, E_0 = Vect\left\{\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}\right\},$$

$$\text{et} \quad E_1 = Vect\left\{\begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}\right\}$$

A est donc diagonalisable (on le savait puisqu'elle possède 3 valeurs propres distinctes) et $A = PDP^{-1}$, avec :

$$P = \begin{pmatrix} 1 & 2 & 2 \\ -2 & -1 & 3 \\ 1 & -1 & 2 \end{pmatrix} \text{ et } D = \begin{pmatrix} -0.75 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

On en déduit $X_{60} = PD^{60}P^{-1}X_0$. Comme on a :

$$P^{-1} = \frac{1}{21}\begin{pmatrix} 1 & -6 & 8 \\ 7 & 0 & -7 \\ 3 & 3 & 3 \end{pmatrix} \text{ et } X_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

on obtient finalement :

$$X_{60} = \frac{1}{21}\begin{pmatrix} (-0.75)^{60} + 6 \\ -2 \times (-0.75)^{60} + 9 \\ (-0.75)^{60} + 6 \end{pmatrix}$$

La probabilité que la particule se trouve en B vaut donc environ $\frac{3}{7}$.

Chapitre II : Variables aléatoires : généralités

II.1 - Définition

Il s'agit d'une application X dont la valeur dépend du résultat obtenu lors de l'expérience aléatoire. C'est donc une application

$$X : \begin{cases} \Omega \to \mathbb{R} \text{ ou } \mathbb{R}^n \text{ (variable réelle ou vectorielle)} \\ \omega \mapsto X(\omega) \end{cases}$$

Exemple 2.1 *On jette deux dés, et on pose* $\Omega = [\![1;6]\!]^2$.

$$X : \omega = (a;b) \mapsto a+b \textit{ (somme des deux dés)}$$

$$Y : \omega = (a;b) \mapsto \begin{cases} 0 & \textit{si } a \textit{ et } b \textit{ impairs} \\ 1 & \textit{si } a \textit{ et } b \textit{ pairs} \\ 2 & \textit{sinon} \end{cases}$$

Exemple 2.2 *On lance une fléchette contre une cible, et on pose* $\Omega = \mathbb{R}^2$.

$$D : \omega = (x;y) \mapsto \sqrt{x^2+y^2}$$

(distance euclidienne du point d'impact au centre de la cible)

Exemple 2.3 *Étant donné un événement* $A \in \mathcal{A}$, *on définit son indicatrice* $\phi_A : \omega \mapsto \begin{cases} 1 & \textit{si } \omega \in \mathcal{A} \\ 0 & \textit{sinon} \end{cases}$.

Il est naturel d'exiger que l'on puisse calculer la probabilité que « X tombe dans l'intervalle $[a;b]$ » (ou dans le pavé $\prod_{i=1}^{n}[a_i;b_i]$). Il faut donc que : $X^{-1}([a;b]) \in \mathcal{A}$. On rappelle que pour toute partie B de $\mathbb{R}$, on pose :

$$X^{-1}(B) = \{\omega \in \Omega / X(\omega) \in B\}$$

C'est l'image réciproque de la partie B, il n'est pas question de bijection réciproque de X dans la notation $X^{-1}(B)$. Comme les intervalles $[a;b]$ engendrent $\mathcal{B}_{\mathbb{R}}$, tribu des boréliens de $\mathbb{R}$ (voir annexe 1), cela signifie que X est une application mesurable. On pose donc comme définition :

Définition 2.1 *On appelle variable aléatoire réelle sur $(\Omega; \mathcal{A})$ toute application $X : \Omega \to \mathbb{R}$ telle que :*

$$\forall B \in \mathcal{B}_{\mathbb{R}}, X^{-1}(B) \in \mathcal{A}$$

Notation : on notera v.a.r. pour variable aléatoire réelle.

Exemple 2.4 *Les variables X et Y de l'exemple 2.1 sont des v.a.r. sur $(\mathbb{R}^2; \mathcal{B}_{\mathbb{R}^2})$; la variable D de l'exemple 2.2 est continue, donc l'image réciproque de tout borélien de $\mathbb{R}$ est un borélien de $\mathbb{R}^2$, et D est donc une v.a.r. ; enfin, la variable ϕ_A de l'exemple 2.3 est aussi une v.a.r. sur $(\Omega; \mathcal{A})$, puisque $A \in \mathcal{A}$.*

Remarque : lorsque $\mathcal{A} = \mathcal{P}(\Omega)$ (par exemple lorsque Ω est fini et que l'on veut que la tribu $\mathcal{A}$ contienne tous les événements élémentaires), la condition $\forall B \in \mathcal{B}_{\mathbb{R}}, X^{-1}(B) \in \mathcal{P}(\Omega)$ est automatiquement vérifiée.

Propriété 2.1 *L'ensemble des variables aléatoires réelles définies sur un même espace $(\Omega; \mathcal{A})$, muni des lois usuelles, a une structure d'algèbre.*
En particulier, la somme et le produit de deux variables aléatoires réelles sont des variables aléatoires réelles.

Définition 2.2 *Soit X une variable aléatoire réelle. Les événements du type $X^{-1}([a;b[)$, $X^{-1}(]a;b])$, $X^{-1}([a;b])$ et $X^{-1}(\{a\})$ sont appelés événements liés à X.*
La tribu engendrée par ces événements s'appelle tribu engendrée par X ; c'est par définition une tribu incluse dans $\mathcal{A}$, notée $\mathcal{A}_X$.

Comme les intervalles $[a;b[$ engendrent la tribu des boréliens $\mathcal{B}_{\mathbb{R}}$ (voir annexe 1), on a $\mathcal{A}_X = X^{-1}(\mathcal{B}_{\mathbb{R}})$.

II.2 - Loi de probabilité d'une variable aléatoire

Définition 2.3 *On appelle loi de probabilité (ou distribution de probabilité) d'une v.a.r. X définie sur un espace $(\Omega; \mathcal{A}; P)$ l'application*

$$\mu : \begin{cases} \mathcal{B}_{\mathbb{R}} \to [0;1] \\ B \mapsto P(X^{-1}(B)) \end{cases}$$

Remarque : par définition, $X^{-1}(B) = \{\omega \in \Omega / X(\omega) \in B\}$, et cet événement est souvent noté « $X \in B$ ».

Propriété 2.2 *L'application μ ainsi définie est une probabilité sur l'espace probabilisable $(\mathbb{R}; \mathcal{B}_{\mathbb{R}})$. Elle est appelée probabilité image de P par X.*

Démonstration : vérifier les trois axiomes de la définition d'une probabilité (en cas de difficulté, voir l'exercice 2.2).

Notation : l'application μ est parfois notée P_X.

Exemple 2.5 *On lance un dé équilibré, on pose $\Omega = [\![1;6]\!]$, $\mathcal{A} = \mathcal{P}(\Omega)$ et on prend pour P la probabilité uniforme. On réalise un gain nul si l'on obtient 1, un gain de 1 € si l'on obtient 2, 3 ou 4 et de 2 € si le résultat est 5 ou 6. On note X la v.a.r. égale au gain obtenu. On a donc $X(\Omega) = \{0;1;2\}$. La loi de X, c'est-à-dire l'application μ définie plus haut, est entièrement déterminée par $P(X = 0)$, $P(X = 1)$ et $P(X = 2)$, c'est-à-dire par les poids de probabilité affectés aux éléments de son univers image $X(\Omega)$. Par exemple pour $B = [0,7;2]$:*

$$\begin{aligned}
P(X \in B) &= P(X^{-1}(B)) \\
&= P(X^{-1}(\{1\}) \cup X^{-1}(\{2\})) \\
&= P(X^{-1}(\{1\}) + P(X^{-1}(\{2\})) \\
&= P(\{2;3;4\}) + P(\{5;6\}) \\
&= P(\{2\}) + \cdots + P(\{6\}) \\
&= \frac{5}{6}
\end{aligned}$$

Propriété 2.3 *Si X est une v.a.r. sur $(\Omega;\mathcal{A})$ et si g est bien définie et continue par morceaux sur $X(\Omega)$, alors $g \circ X$ est une v.a.r. sur $(\Omega;\mathcal{A})$.*

Démonstration : soit $B \in \mathcal{B}_{\mathbb{R}}$. Comme g est continue par morceaux, $g^{-1}(B) \in \mathcal{B}_{\mathbb{R}}$, et comme X est une v.a.r. sur $(\Omega;\mathcal{A})$, $X^{-1}(g^{-1}(B)) \in \mathcal{A}$, c'est-à-dire $(g \circ X)^{-1}(B) \in \mathcal{A}$.

II.3 - Fonction de répartition

Définition 2.4 *Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega;\mathcal{A};P)$. On appelle fonction de répartition de X l'application*

$$F : \begin{cases} \mathbb{R} \to [0;1] \\ x \mapsto P(X \leq x) \end{cases}$$

Dans l'exemple précédent 2.5, on a facilement :

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ \dfrac{1}{6} & si \quad 0 \leq x < 1 \\ \dfrac{2}{3} & si \quad 1 \leq x < 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

Propriété 2.4 *Toute fonction de répartition F possède les propriétés suivantes :*

1. *F est croissante.*
2. $\lim\limits_{x \to -\infty} F(x) = 0$ *et* $\lim\limits_{x \to +\infty} F(x) = 1$.
3. *F est continue à droite en tout point x de $\mathbb{R}$.*
4. *F a une limite à gauche en tout point x de $\mathbb{R}$, et :*

$$\forall x \in \mathbb{R}, F(x) - \lim_{t \to x^-} F(t) = P(X = x)$$

Démonstration des propriétés de F :

1. Soit $y \geq x$: $F(y) - F(x) = P(x < X \leq y) \geq 0$, et F est donc croissante.
2. Comme F est monotone (en l'occurrence croissante), on a l'équivalence suivante :

$$\lim_{x \to -\infty} F(x) = 0 \Longleftrightarrow \lim_{n \to +\infty} F(-n) = 0$$

(bon exercice de manipulation de la définition de la limite, voir exercice 2.1 en cas de difficulté).
Or $F(-n) = P(X \in]-\infty; -n])$. Notons, pour tout $n \in \mathbb{N}$, $A_n = X^{-1}(]-\infty; -n])$. La suite d'événements (A_n) est décroissante au sens de l'inclusion et on a donc, grâce aux propriétés de P démontrées au chapitre I, que la suite de réels $(P(A_n))$ est convergente et :

$$\begin{aligned} \lim_{n \to +\infty} P(A_n) &= P\left(\bigcap_{n=0}^{+\infty} A_n\right) \\ \Rightarrow \lim_{n \to +\infty} F(-n) &= P(\emptyset) = 0 \end{aligned}$$

Même démarche pour la limite en $+\infty$.

3. Soit $x \in \mathbb{R}$ fixé. Comme F est croissante, il suffit d'étudier la convergence de la suite de réels $\left(F\left(x + \frac{1}{n}\right)\right)$ pour étudier l'existence et la valeur de la limite à droite de F en x. Par définition :

$$F\left(x + \frac{1}{n}\right) = P\left(X \in \left]-\infty; x + \frac{1}{n}\right]\right)$$

En notant $A_n = X^{-1}\left(\left]-\infty; x + \frac{1}{n}\right]\right)$, on obtient une suite d'événements décroissante pour l'inclusion, donc la suite de réels $(P(A_n))$ converge et :

$$\begin{aligned} \lim_{n \to +\infty} P(A_n) &= P\left(\bigcap_{n=0}^{+\infty} A_n\right) \\ \Rightarrow \lim_{n \to +\infty} F\left(x + \frac{1}{n}\right) &= P\left(X^{-1}(]-\infty; x])\right) = F(x) \end{aligned}$$

4. Là encore, il suffit d'étudier la suite $\left(F\left(x-\frac{1}{n}\right)\right)$:

$$F\left(x-\frac{1}{n}\right)=P\left(X\in\left]-\infty;x-\frac{1}{n}\right]\right)$$

En notant $A_n = X^{-1}(\left]-\infty; x-\frac{1}{n}\right])$, on obtient une suite d'événements croissante pour l'inclusion, donc la suite de réels $(P(A_n))$ converge et :

$$\begin{aligned}\lim_{n\to+\infty} P(A_n) &= P\left(\bigcup_{n=0}^{+\infty} A_n\right)\\ \Rightarrow \lim_{n\to+\infty} F\left(x-\frac{1}{n}\right) &= P(X^{-1}(]-\infty;x[)) = P(X<x)\end{aligned}$$

F a donc une limite à gauche en x, et :

$$F(x)-\lim_{t\to x^-} F(t) = P(X\leq x) - P(X<x) = P(X=x)$$

Théorème 2.1 *Toute fonction* $F:\mathbb{R}\to[0;1]$ *telle que :*

1. F est croissante

2. $\lim_{x\to-\infty} F(x)=0$ et $\lim_{x\to+\infty} F(x)=1$

3. F est continue à droite

est la fonction de répartition d'une v.a.r. X.

Ce théorème est admis.

Propriété 2.5 *La donnée de F caractérise la loi de probabilité μ.*

Démonstration : si on connaît F, alors :

$$\forall x<y, \mu(]x;y]) = P(X\in]x;y]) = F(y)-F(x)$$

On connaît donc μ sur les parties du type $]x;y]$, et comme ces parties engendrent la tribu des boréliens $\mathcal{B}_{\mathbb{R}}$ (voir annexe 1), μ est connue sur $\mathcal{B}_{\mathbb{R}}$.

Les v.a.r. au programme sont de deux types :
les variables discrètes, lorsque $X(\Omega)$ est fini ou dénombrable. Dans les cas usuels, $X(\Omega)$ pourra s'écrire sous la forme $\{x_i, i\in I\}$, avec $I=\{1;...;n\}$ ou $I=\mathbb{N}$, et la suite (x_i) strictement croissante. F sera alors constante sur chaque intervalle $]-\infty;x_1[$, $]x_1;x_2[$, ...
Les variables continues, lorsque F est continue et peut s'écrire sous la

forme $F(x) = \int_{-\infty}^{x} f(t)dt$, où f est positive, d'intégrale sur $\mathbb{R}$ égale à 1 et discontinue sur un ensemble fini de points.
Cependant, il existe des v.a.r. qui ne sont ni continues, ni discrètes. Par exemple, si X est égale au temps d'attente pour passer à un feu qui est vert 30 secondes, puis orange et rouge 30 secondes, on a $X(\Omega) = [0; 30]$ (donc X n'est pas discrète) et sa fonction de répartition vaut 0 sur $]-\infty; 0[$, puis $0.5 + \frac{x}{60}$ sur $[0; 30]$ et 1 sur $[30; +\infty[$. X n'est donc pas non plus continue.

II.4 - Vecteurs aléatoires

Définition 2.5 *On appelle vecteur aléatoire toute application :*

$$X : \left\{ \begin{array}{ccc} \Omega & \to & \mathbb{R}^n \\ \omega & \mapsto & \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} \end{array} \right.$$

où les X_i sont des variables aléatoires réelles.

Exemple 2.6 *On jette deux dés, et on pose $\Omega = [\![1; 6]\!]^2$.*

$$X : \omega = (a; b) \mapsto \begin{pmatrix} \max(a; b) \\ a + b \end{pmatrix}$$

est un vecteur aléatoire.

On donne ensuite la définition générale de l'indépendance de variables aléatoires, qui sera caractérisée plus simplement dans les deux cas au programme : les variables discrètes et les variables à densité.

Définition 2.6 *Deux v.a.r. X et Y sont indépendantes si :*

$$\forall (B_1; B_2) \in (\mathcal{B}_{\mathbb{R}})^2, P(X \in B_1 \cap Y \in B_2) = P(X \in B_1)P(Y \in B_2)$$

En particulier, pour tout couple $(x; y) \in \mathbb{R}^2$,

$$P(X \in]-\infty; x] \cap Y \in]-\infty; y]) = P(X \in]-\infty; x])P(Y \in]-\infty; y])$$

On a la même définition pour n v.a.r. :

Définition 2.7 *Les v.a.r. $X_1, \cdots, X_n$ sont indépendantes si pour tout ensemble d'indices I inclus dans $[\![1; n]\!]$, on a, pour toute famille de boréliens $(B_i)_{i \in I}$:*

$$P\left(\bigcap_{i \in I} X_i \in B_i\right) = \prod_{i \in I} P(X_i \in B_i)$$

On étend la définition à une suite de variables aléatoires en considérant l'indépendance de toute sous-famille finie.

II.5 - Énoncé des exercices

Exercice 2.1 *Retour sur les propriétés de F. Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction croissante.*

1. *Montrer que si $\lim_{x \to +\infty} F(x)$ existe, alors $\lim_{n \to +\infty} F(n)$ existe (et que les deux limites sont égales !). Donner un contre-exemple dans le cas où F n'est pas monotone.*
2. *Montrer que si F est la fonction de répartition d'une v.a.r. X, alors :*

$$\lim_{x \to +\infty} F(x) = 1$$

3. *Montrer que $\lim_{x \to x_0^+} F(x)$ existe si et seulement si $\lim_{n \to +\infty} F\left(x_0 + \frac{1}{n}\right)$ existe (et que les deux limites sont alors égales). C'est ce résultat qui a été utilisé pour démontrer la propriété sur la limite à droite d'une fonction de répartition. Donner un contre-exemple dans le cas où F n'est pas monotone.*
4. *Montrer que si F est la fonction de répartition d'une v.a.r. X, alors l'ensemble de ses points de discontinuité est fini ou dénombrable.*

Exercice 2.2 *Montrer que la loi de probabilité d'une v.a.r. X (l'application μ) est une probabilité sur l'espace probabilisable $(\mathbb{R}; \mathcal{B}_{\mathbb{R}})$.*

Exercice 2.3 *Soient X et Y deux v.a.r. de fonctions de répartition respectives F et G définies par :*

$$F(x) = \begin{cases} 0 & si \quad x < -1 \\ \dfrac{1}{3} & si \quad -1 \leq x < 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

$$G(x) = \begin{cases} 0 & si \quad x \leq 0 \\ x^2 & si \quad 0 \leq x \leq \dfrac{1}{2} \\ 1 - \dfrac{3}{4} e^{\frac{1}{2} - x} & si \quad x \geq \dfrac{1}{2} \end{cases}$$

1. *Déterminer les réels x pour lesquels $P(X = x) > 0$ et les réels y pour lesquels $P(Y = y) > 0$.*

2. *Soit $T = aX + b$ (a et b réels fixés, $a > 0$). Déterminer la fonction de répartition de T.*
3. *On suppose X et Y indépendantes, et on pose $Z = \max(X;Y)$, $U = \min(X;Y)$. Déterminer la fonction de répartition de Z ainsi que celle de U.*

Exercice 2.4 *Soient F une fonction de répartition continue et h un réel strictement positif fixé. Montrer que G définie par :*

$$\forall x \in \mathbb{R}, G(x) = \frac{1}{h}\int_x^{x+h} F(t)dt$$

est une fonction de répartition.

Exercice 2.5 *Démontrer que $\mathbb{Z}$ est dénombrable et proposer une variable aléatoire dont la fonction de répartition F a pour ensemble de points de discontinuité $\mathbb{Z}$.*

II.6 - Correction des exercices

Exercice 2.1

1. On vérifie tout d'abord que si la fonction F tend vers l en $+\infty$, alors la suite $(F(n))$ converge vers l :

$$F(x) \xrightarrow[x\to+\infty]{} l \Rightarrow (\forall\epsilon > 0, \exists A > 0, x \geq A \Rightarrow |F(x) - l| < \epsilon)$$

Soit $N = E(A) + 1$: si $n \geq N$, alors $n \geq A$ et donc $|F(n) - l| < \epsilon$.

Remarque : la croissance de F ne joue aucun rôle.

Réciproquement :

$$F(n) \xrightarrow[n\to+\infty]{} l \Rightarrow (\forall\epsilon > 0, \exists N \in \mathbb{N}, n \geq N \Rightarrow |F(n) - l| < \epsilon)$$

Soit $x \geq N$:
F étant croissante, on a $F(x) \geq F(N)$, et donc : $F(x) > l - \epsilon$.
D'autre part, on a : $x \leq E(x) + 1$. Comme $x \geq N$, $E(x) + 1$ est un entier supérieur ou égal à N, d'où :

$$F(x) \leq F(E(x) + 1) < l + \epsilon$$

Conclusion : $x \geq N \Rightarrow |F(x) - l| < \epsilon$, et donc : $F(x) \xrightarrow[x\to+\infty]{} l$.
On a le même résultat si F décroît.

Dans le cas où F n'est pas monotone, l'équivalence est fausse. On peut prendre par exemple F définie sur $\mathbb{R}$ par :

$$F(x) = \sin(2\pi x)$$

La suite $(F(n))$ tend vers 0 et F n'a pas de limite (considérer la suite $\left(F\left(n+\frac{1}{4}\right)\right)$ par exemple).

2. Si F est la fonction de répartition d'une v.a.r. X, elle est croissante, donc il suffit de montrer que :

$$F(n) \xrightarrow[n\to+\infty]{} 1$$

Par définition : $F(n) = P(X \in]-\infty; n])$.
La suite d'événements $A_n = X^{-1}(]-\infty; n])$ est croissante pour l'inclusion donc la suite de réels $(P(A_n))$ converge et :

$$\lim_{n\to+\infty} P(A_n) = P\left(\bigcup_{n=0}^{+\infty} A_n\right) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$$

3. Il s'agit, comme à la question précédente, de manipuler la définition de la limite. Si la fonction F tend vers l à droite en x_0 :

$$\forall \epsilon > 0, \exists \alpha > 0, x \in]x_0; x_0+\alpha[\Rightarrow |F(x) - l| < \epsilon$$

Soit N tel que $\frac{1}{N} < \alpha$. Si $n \geq N$, alors $\frac{1}{n} < \alpha$, donc $x_0 + \frac{1}{n} \in]x_0; x_0+\alpha[$. Par conséquent, $\left|F\left(x_0+\frac{1}{n}\right) - l\right| < \epsilon$, et donc la suite $\left(F\left(x_0+\frac{1}{n}\right)\right)$ converge vers l.

Remarque : la croissance de F n'intervient pas.

Réciproquement, si la suite $\left(F\left(x_0+\frac{1}{n}\right)\right)$ converge vers l :

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, n \geq N \Rightarrow \left|F\left(x_0+\frac{1}{n}\right) - l\right| < \epsilon$$

Soit $x \in \left]x_0; x_0+\frac{1}{N}\right[$:
comme F est croissante, $F(x) \leq F\left(x_0+\frac{1}{N}\right)$, et donc $F(x) < l + \epsilon$.
Soit n tel que $x_0+\frac{1}{n} \leq x$: comme F est croissante, $F\left(x_0+\frac{1}{n}\right) \leq F(x)$.
De plus, $n \geq N$ (puisque $x_0 + \frac{1}{n} \leq x < x_0 + \frac{1}{N}$), ce qui implique que

$F\left(x_0 + \dfrac{1}{n}\right) > l - \epsilon$. On a donc $F(x) > l - \epsilon$.
Conclusion :

$$F(x) \xrightarrow[x \to x_0^+]{} l$$

L'équivalence est fausse si F n'est pas monotone.
Par exemple, F définie par :

$$F(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \sin\left(\dfrac{2\pi}{x}\right) & si \quad x > 0 \end{cases}$$

est telle que la suite $\left(F\left(\dfrac{1}{n}\right)\right)$ tend vers 0 et la fonction F n'a pas de limite en 0^+ (poser $x = \dfrac{1}{n + \frac{1}{4}}$).

4. D'après les propriétés d'une fonction de répartition, F est discontinue en x_0 si et seulement si :

$$\underbrace{\lim_{x \to x_0^+} F(x) - F(x_0)}_{\textbf{hauteur du saut}} > 0$$

Soit D_n l'ensemble des points où la hauteur du saut est supérieure ou égale à $\dfrac{1}{n}$: $D_n = \left\{x_0 \in \mathbb{R} / \lim_{x \to x_0^+} F(x) - F(x_0) \geq \dfrac{1}{n}\right\}$. Comme F est croissante et comprise entre 0 et 1, la somme des hauteurs des sauts est inférieure ou égale à 1, donc D_n comporte au plus n éléments.
Soit D l'ensemble des points de discontinuité de F :

$$D = \bigcup_{n=1}^{+\infty} D_n$$

D est donc une union dénombrable d'ensembles finis, et il est donc fini ou dénombrable.

Exercice 2.2

Pour montrer que μ est une probabilité sur $(\mathbb{R}; \mathcal{B}_{\mathbb{R}})$, on vérifie les trois axiomes.
Positivité : $\forall B \in \mathcal{B}_{\mathbb{R}}, \mu(B) = P(X^{-1}(B)) \geq 0$.
Totalité : $\mu(\mathbb{R}) = P(X \in \mathbb{R}) = P(\Omega) = 1$.

Additivité : soit (B_n) une suite d'éléments de $\mathcal{B}_{\mathbb{R}}$ deux à deux disjoints.

$$\begin{aligned}
\mu\Big(\bigcup_n B_n\Big) &= P\Big(X^{-1}\Big(\bigcup_n B_n\Big)\Big) \\
&= P\Big(\bigcup_n X^{-1}(B_n)\Big) \\
&= \sum_{n=0}^{+\infty} P(X^{-1}(B_n))(\text{additivité de } P) \\
&= \sum_{n=0}^{+\infty} \mu(B_n)
\end{aligned}$$

Exercice 2.3

1. On sait que pour tout x dans $\mathbb{R}$: $F(x) - \lim_{t\to x^-} F(t) = P(X = x)$. On en déduit immédiatement que $P(X = -1) = \frac{1}{3}$, $P(X = 2) = \frac{2}{3}$, et : $\forall x \in \mathbb{R}\backslash\{-1; 2\}, P(X = x) = 0$.
D'autre part, G étant continue sur $\mathbb{R}$: $\forall y \in \mathbb{R}, P(Y = y) = 0$.
2. La fonction de répartition s'obtient directement à partir de celle de X :

$$\begin{aligned}
P(T \leq t) &= P(aX + b \leq t) \\
&= P\Big(X \leq \frac{t-b}{a}\Big) \\
&= F\Big(\frac{t-b}{a}\Big) \\
&= \begin{cases} 0 & si \quad t < -a + b \\ \frac{1}{3} & si \quad -a + b \leq t < 2a + b \\ 1 & si \quad t \geq 2a + b \end{cases}
\end{aligned}$$

3. On utilise l'indépendance de X et Y :

$$\begin{aligned}
P(Z \leq z) &= P(X \leq z \cap Y \leq z) \\
&= P(X \leq z)P(Y \leq z) \\
&= F(z)G(z) \\
&= \begin{cases} 0 & si \quad z \leq 0 \\ \frac{1}{3}z^2 & si \quad 0 \leq z \leq \frac{1}{2} \\ \frac{1}{3}\Big(1 - \frac{3}{4}e^{\frac{1}{2}-z}\Big) & si \quad \frac{1}{2} \leq z < 2 \\ 1 - \frac{3}{4}e^{\frac{1}{2}-z} & si \quad z \geq 2 \end{cases}
\end{aligned}$$

Pour la fonction de répartition de U, on passe à l'événement contraire :

$$\begin{aligned}
P(U \leq u) &= 1 - P(U > u) \\
&= 1 - P(X > u \cap Y > u) \\
&= 1 - P(X > u)P(Y > u) \\
&= 1 - (1 - F(u))(1 - G(u)) \\
&= \begin{cases} 0 & si \quad u < -1 \\ \dfrac{1}{3} & si \quad -1 \leq u \leq 0 \\ 1 - \dfrac{2}{3}(1 - u^2) & si \quad 0 \leq u \leq \dfrac{1}{2} \\ 1 - \dfrac{1}{2}e^{\frac{1}{2}-u} & si \quad \dfrac{1}{2} \leq u < 2 \\ 1 & si \quad 2 \leq u \end{cases}
\end{aligned}$$

Exercice 2.4

Comme F est continue, G est dérivable, et :

$$\forall x \in \mathbb{R}, G'(x) = \frac{1}{h}(F(x+h) - F(x))$$

Comme F est croissante, G' est positive, et donc G est croissante.
De plus, G étant dérivable, elle est continue, et en particulier continue à droite.
Enfin, on a pour tout réel x : $\forall t \in [x; x+h], F(x) \leq F(t) \leq F(x+h)$, ce qui implique (intégrer sur $[x; x+h]$, puis multiplier par $\frac{1}{h}$) :

$$F(x) \leq G(x) \leq F(x+h)$$

En passant à la limite, il vient : $\lim\limits_{x \to -\infty} G(x) = 0$ et $\lim\limits_{x \to +\infty} G(x) = 1$.
G est donc bien une fonction de répartition.

Exercice 2.5

Il faut montrer que $\mathbb{Z}$ est en bijection avec $\mathbb{N}$: l'application

$$f : n \mapsto \begin{cases} 2n - 1 & si \quad n \geq 1 \\ -2n & si \quad n \leq 0 \end{cases}$$

est une bijection de $\mathbb{Z}$ vers $\mathbb{N}$.
Il suffit ensuite de définir une variable aléatoire telle que $X(\Omega) = \mathbb{Z}$, dont les poids de probabilité forment une famille sommable de somme égale à 1.
On peut poser par exemple :

$$\forall n \in \mathbb{Z}^*, P(X = n) = \frac{2}{\pi^2 n^2} \text{ et } P(X = 0) = \frac{1}{3}$$

On a bien :

$$\sum_{n\in\mathbb{Z}} P(X=n) = 2\sum_{n=1}^{+\infty} \frac{2}{\pi^2 n^2} + \frac{1}{3} = \frac{4}{\pi^2} \times \frac{\pi^2}{6} + \frac{1}{3} = 1$$

et l'ensemble des points de discontinuité de la fonction de répartition de X est bien $\mathbb{Z}$ puisque :

$$\forall x \in \mathbb{R}, F(x) = P(X \leq x) = \sum_{n\leq[x]} P(X=n)$$

(fonction en escalier).

Chapitre III : Variables aléatoires discrètes

III.1 - Définition

Définition 3.1 *Une v.a.r. X est dite discrète si $X(\Omega)$ est fini ou dénombrable.*

Un ensemble dénombrable est un ensemble en bijection avec $\mathbb{N}$. X est donc discrète si $X(\Omega)$ est de la forme $\{x_i, 1 \leq i \leq n\}$ (cas fini) ou de la forme $\{x_i, i \in \mathbb{N}\}$ (cas dénombrable).

Remarque : $\mathbb{Q}$ est dénombrable, donc une v.a.r. X qui prend toutes les valeurs de $\mathbb{Q}$ est discrète : ses valeurs ne sont donc pas nécessairement « isolées ».

Exemple 3.1 *Une urne contient 3 boules rouges et 4 boules noires. On extrait successivement avec remise 2 boules de l'urne (on choisit comme univers $\Omega = \{R_1; R_2; R_3; N_1; N_2; N_3; N_4\}^2$ et comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$). On mise au départ 10 €, et on gagne 8 € par boule rouge obtenue. Soit X la v.a.r. prenant pour valeur le gain final. On a : $X(\Omega) = \{-10; -2; +6\}$.*

Exemple 3.2 *On considère le tournoi suivant : la première épreuve oppose les joueurs J_0 et J_1, puis le vainqueur dispute la deuxième épreuve contre J_2, puis le vainqueur dispute la troisième épreuve contre J_3, et c. (Cf Capes 2006 épreuve 1...). Le tournoi s'arrête dès qu'un joueur a remporté deux duels consécutifs. Si X est la variable aléatoire égale au nombre de duels alors disputés, on a $X(\Omega) = [\![2; +\infty[$.*

III.2 - Loi de probabilité

Soit X une v.a.r. discrète, d'univers image $X(\Omega) = \{x_i, i \in I\}$ (où I est une partie de $\mathbb{N}$). D'après la définition vue au chapitre précédent, il s'agit de déterminer l'application $\mu : \left\{ \begin{array}{l} \mathcal{B}_{\mathbb{R}} \to [0;1] \\ B \mapsto P(X^{-1}(B)) \end{array} \right.$. Or cette application est entièrement déterminée dès que l'on connaît $(P(X = x_i))_{i \in I}$.
En effet, on a, par l'additivité de P :

$$\forall B \in \mathcal{B}_{\mathbb{R}}, \mu(B) = P(X \in B) = P\left(\bigcup_{i / x_i \in B} (X = x_i) \right) = \sum_{i / x_i \in B} P(X = x_i)$$

D'où la définition :

Définition 3.2 *Donner la loi d'une v.a.r. discrète X, c'est donner $X(\Omega)$, et pour tout $x_i \in X(\Omega)$, la valeur de $P(X = x_i)$.*

Exemple 3.3 *On reprend le jeu avec mise initiale de 10 €. On suppose les boules indiscernables au toucher et on choisit pour P la probabilité uniforme. X a pour loi :*

$X(\Omega)$	-10	-2	6
$P(X = x_i)$	$\frac{16}{49}$	$\frac{24}{49}$	$\frac{9}{49}$

Propriété 3.1 *Puisque P vérifie l'axiome de totalité, on a évidemment :*

$$\sum_{i \in I} P(X = x_i) = 1$$

Démonstration : il suffit d'écrire que les événements $X^{-1}(\{x_i\}), i \in I$, sont disjoints et que leur réunion est égale à Ω.

Remarque : ne pas confondre la probabilité P, qui a été installée sur l'espace probabilisable $(\Omega; \mathcal{A})$ et la probabilité μ, qui est la loi de X.
Par exemple, pour le jeu avec mise initiale de 10 €, P est uniforme et μ ne l'est pas.

Propriété 3.2 *Soit I une partie non vide de $\mathbb{N}$.*
La donnée de deux familles de réels $(x_i)_{i \in I}$ et $(p_i)_{i \in I}$ tels que :
1. $\forall i \in I, p_i \geq 0$
2. $\sum_{i \in I} p_i = 1$

permet de définir la loi de probabilité d'une v.a.r. discrète X en posant :

$$X(\Omega) = \{x_i; i \in I\} \text{ et } \forall i \in I, P(X = x_i) = p_i$$

Démonstration : X est bien discrète puisque I est une partie non vide de $\mathbb{N}$, et on vérifie que :

$$\mu : \begin{cases} \mathcal{B}_{\mathbb{R}} \to [0;1] \\ B \mapsto P(X^{-1}(B)) \end{cases}$$

est une probabilité.

Dans le cas où I n'est pas finie, la condition 2 est que la famille (p_i) soit sommable, de somme 1, ce qui équivaut à la convergence de la série $\sum p_i$ vers 1, puisque les p_i sont positifs (voir annexe 2).
Rappel : dans le cas d'une suite de réels quelconques (u_n), la sommabilité de la famille équivaut à la convergence absolue de la série.

Exemple 3.4 *Soit X d'univers image $X(\Omega) = \mathbb{N}$ telle que*

$$\forall n \in \mathbb{N}, P(X = n) = e^{-2}\frac{2^n}{n!}$$

On a bien la positivité des $P(X = n)$ et $\sum_{n=0}^{+\infty} P(X = n) = 1$.
(Vérifiez que vous savez montrer la convergence de la série, et le fait que sa somme vaut 1...).

III.3 - Fonction de répartition

Comme on l'a vu au chapitre précédent, c'est la fonction :

$$F : \left\{ \begin{array}{l} \mathbb{R} \to [0;1] \\ x \mapsto P(X \leq x) \end{array} \right.$$

Dans le cas où X prend des valeurs ordonnées $x_1 < x_2 < \cdots$, la fonction F est alors constante sur chaque intervalle $]x_i; x_{i+1}[$: on a une fonction en escalier.
En effet, si on considère deux réels x et y tels que $x_i < x \leq y < x_{i+1}$, on a :

$$F(y) - F(x) = P(X \in]x; y]) = P(\emptyset) = 0$$

Exemple 3.5 *Si on reprend le jeu avec mise initiale de 10 €, on obtient :*

$$F(x) = \left\{ \begin{array}{lll} 0 & si & x < 10 \\ \dfrac{16}{49} & si & -10 \leq x < -2 \\ \dfrac{40}{49} & si & -2 \leq x < 6 \\ 1 & si & 6 \leq x \end{array} \right.$$

III.4 - Espérance

Il s'agit de la moyenne des valeurs de X pondérées par leur probabilité d'apparition :

Définition 3.3 *Soit X une v.a.r. discrète.*
On appelle espérance de X le réel, noté $E(X)$:

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

lorsque $X(\Omega) = \{x_1; \cdots; x_n\}$, *ou :*

$$E(X) = \sum_{i=1}^{+\infty} x_i P(X = x_i)$$

lorsque $X(\Omega) = \{x_i; i \geq 1\}$ *et que cette série est absolument convergente.*

Dans ce dernier cas, on ne peut pas exiger seulement la convergence de la série : on veut définir l'espérance de X indépendamment de la façon dont on a numéroté $X(\Omega)$ (c'est-à-dire indépendamment de la bijection : $\mathbb{N} \to X(\Omega)$). On exige donc la convergence commutative de la série, ce qui équivaut à son absolue convergence, ou encore à la sommabilité de la famille $(x_i P(X = x_i))_{i \in \mathbb{N}}$ (voir annexe 2).

Remarque : dans le premier cas, lorsque $X(\Omega)$ est fini, $E(X)$ est donc le barycentre de la famille de points pondérés $(x_i; P(X = x_i))_{i=1,\cdots,n}$.

Exemple 3.6 *En reprenant le jeu des 10 €,*

$$E(X) = (-10) \times \frac{16}{49} + (-2) \times \frac{24}{49} + 6 \times \frac{9}{49} = -\frac{22}{7}$$

Exemple 3.7 *Soit* X *de loi :* $X(\Omega) = \mathbb{N}$ *et* $\forall n \in \mathbb{N}, P(X = n) = e^{-2}\frac{2^n}{n!}$.
La série $\sum_{n \geq 0} nP(X = n)$ *est convergente (utiliser par exemple la règle de D'Alembert), donc* $E(X)$ *existe, et :*

$$\sum_{n=0}^{+\infty} nP(X = n) = \sum_{n=0}^{+\infty} ne^{-2}\frac{2^n}{n!} = e^{-2}\sum_{n=1}^{+\infty} \frac{2^n}{(n-1)!} = e^{-2} \times 2e^2 = 2$$

Exemple 3.8 *Soit* X *de loi* $X(\Omega) = \mathbb{N}^*$ *et* $\forall n \in \mathbb{N}^*, P(X = n) = \frac{6}{\pi^2 n^2}$.
La définition de la loi est cohérente, puisque la série $\sum_{n \geq 1} \frac{1}{n^2}$ *converge vers* $\frac{\pi^2}{6}$. *(ce résultat peut s'obtenir par exemple grâce aux séries de Fourier, en développant la fonction* $f : \mathbb{R} \to \mathbb{R}$, 2π *périodique et telle que :*

$$\forall x \in [-\pi; \pi], f(x) = 1 - \frac{x^2}{\pi^2}$$

puis en appliquant le théorème de Dirichlet en $x = \pi$*).*
Comme la série $\sum_{n \geq 1} nP(X = n)$ *a son terme général égal à* $\frac{6}{\pi^2 n}$, *elle diverge et* X *n'a donc pas d'espérance.*

Propriété 3.3 *Si $X(\Omega)$ possède un maximum et un minimum, alors $E(X)$ existe et :*

$$x_{min} \leq E(X) \leq x_{max}$$

Démonstration : dans le cas dénombrable : soit $M = \max(|x_{min}|; |x_{max}|)$. On a pour tout i, $|x_i| \leq M$, donc la série à termes positifs $\sum_{i \geq 0} p_i |x_i|$ a ses sommes partielles majorées par la constante $\sum_{i=0}^{+\infty} p_i M = M$. Elle est donc convergente, et $E(X)$ existe. De plus :

$$\begin{aligned} \forall i \in \mathbb{N}; x_{min} &\leq x_i \leq x_{max} \\ \Rightarrow \sum_{i=0}^{+\infty} p_i x_{min} &\leq \sum_{i=0}^{+\infty} p_i x_i \leq \sum_{i=0}^{+\infty} p_i x_{max} \end{aligned}$$

D'où l'encadrement souhaité (puisque $\sum_{i=0}^{+\infty} p_i = 1$).
Même démarche dans le cas fini.

Théorème 3.1 *Si Ω est fini ou dénombrable, alors :*

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

sous réserve de convergence absolue.

Ce théorème est essentiel car il permet de montrer facilement la linéarité de l'opérateur E, ce qui n'a rien d'évident avec les formules de la définition. De plus, il permet de faire le lien avec la théorie de la mesure, où l'on écrirait :

$$E(X) = \int_{\Omega} X dP$$

Démonstration : on note : $A_i = \{\omega \in \Omega / X(\omega) = x_i\}$. Alors :

$$P(X = x_i) = P(X^{-1}(\{x_i\})) = P(A_i) = \sum_{\omega \in A_i} P(\{\omega\})$$

On obtient donc :

$$\begin{aligned}
\sum_i x_i P(X = x_i) &= \sum_i x_i \sum_{\omega \in A_i} P(\{\omega\}) \\
&= \sum_i \sum_{\omega \in A_i} x_i P(\{\omega\}) \\
&= \sum_i \sum_{\omega \in A_i} X(\omega) P(\{\omega\}) \text{ car } \omega \in A_i \Rightarrow X(\omega) = x_i \\
&= \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \text{ car } \bigcup_i A_i = \Omega
\end{aligned}$$

Ce résultat a peu d'intérêt pour le calcul pratique de l'espérance : dans le cas du jeu des 10 €, on a une somme de 49 termes, au lieu d'une somme de 3 termes si l'on utilise directement la définition. Par contre, il permet de démontrer :

Propriété 3.4 *Soient X et Y deux v.a.r. discrètes sur Ω qui possèdent une espérance, et soit $\lambda \in \mathbb{R}$.*
Alors $X + \lambda Y$ est une v.a.r. discrète sur Ω qui possède une espérance, et :

$$E(X + \lambda Y) = E(X) + \lambda E(Y)$$

Autrement dit, l'opérateur E est une forme linéaire sur l'espace vectoriel des v.a.r. discrètes sur Ω d'espérance finie.

Démonstration : la série $\sum_{\omega \in \Omega} (X(\omega) + \lambda Y(\omega)) P(\{\omega\})$ est absolument convergente comme somme de deux séries absolument convergentes, et on peut séparer la somme en deux (puisque chacune a bien un sens) :

$$\begin{aligned}
\sum_{\omega \in \Omega} (X(\omega) + \lambda Y(\omega)) P(\{\omega\}) &= \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) + \lambda \sum_{\omega \in \Omega} Y(\omega) P(\{\omega\}) \\
&= E(X) + \lambda E(Y)
\end{aligned}$$

La démonstration qui précède nécessite que Ω soit fini ou dénombrable, sinon l'indexation du symbole $\sum$ n'a plus de sens.
Le résultat (linéarité de l'espérance) reste vrai même si Ω n'est ni fini, ni dénombrable.
L'idée est de décomposer canoniquement les v.a.r. discrètes comme combinaison de fonctions caractéristiques. On note :

$$\begin{aligned}
X(\Omega) &= \{x_i; i \in I\}, \text{ où } I \text{ est une partie finie ou dénombrable de } \mathbb{N}, \\
\text{et } A_i &= \{\omega / X(\omega) = x_i\}
\end{aligned}$$

On a alors la décomposition de X : $X = \sum_{i \in I} x_i \mathbf{1}_{A_i}$.
Le membre de droite est une série de fonction qui converge simplement vers la fonction X, puisque si l'on fixe $\omega \in \Omega$, alors ω appartient à un et un seul des A_i, et la somme ne comporte donc qu'un terme, qui vaut x_i, c'est-à-dire $X(\omega)$.
On a par définition de l'espérance :

$$E(X) = \sum_{i \in I} x_i P(X = x_i) = \sum_{i \in I} x_i P(A_i)$$

On décompose également Y : $Y = \sum_{j \in J} y_j \mathbf{1}_{B_j}$. On obtient alors une décomposition de $X + Y$:

$$X + Y = \sum_{i \in I; j \in J} (x_i + y_j) \mathbf{1}_{A_i \cap B_j}$$

En effet : $\forall \omega \in \Omega, \exists!(i; j) / \omega \in A_i$ et $\omega \in B_j$, et on a alors bien :

$$(X + Y)(\omega) = x_i + y_j = \sum_{i \in I; j \in J} (x_i + y_j) \mathbf{1}_{A_i \cap B_j}(\omega)$$

D'après la décomposition que l'on vient d'obtenir :

$$\begin{aligned}
E(X+Y) &= \sum_{i,j} (x_i + y_j) P(A_i \cap B_j) \\
&= \sum_{i,j} x_i P(A_i \cap B_j) + \sum_{i,j} y_j P(A_i \cap B_j) \\
&= \sum_i x_i \sum_j P(A_i \cap B_j) + \sum_j y_j \sum_i P(A_i \cap B_j) \\
&= \sum_i x_i P(A_i) + \sum_j y_j P(B_j) \\
&= E(X) + E(Y)
\end{aligned}$$

III.5 - Composition par une fonction

On considère une v.a.r. X discrète, d'univers image $X(\Omega) = \{x_i; i \in I\}$ où I est une partie non vide de $\mathbb{N}$, et g une fonction continue par morceaux définie sur un intervalle J contenant $X(\Omega)$. Alors $g \circ X(\Omega) = \{g(x_i); i \in I\}$ est fini ou dénombrable, donc $g \circ X$ est une v.a.r. discrète (on savait que c'était une variable aléatoire réelle par la propriété 2.3).

Le résultat le plus important est le théorème de transfert :

Théorème 3.2 *Sous les hypothèses précédentes, on a :*

$$E(g(X)) = \sum_i g(x_i)P(X = x_i)$$

sous réserve de convergence absolue.

Démonstration : on note $A_i = \{\omega \in \Omega / X(\omega) = x_i\}$. Dans le cas où Ω est fini ou dénombrable, on a directement, sous réserve de convergence absolue :

$$\begin{aligned} E(g(X)) &= \sum_{\omega \in \Omega} g(X(\omega))P(\{\omega\}) \\ &= \sum_i \sum_{\omega \in A_i} g(x_i)P(\{\omega\}) \\ &= \sum_i g(x_i) \sum_{\omega \in A_i} P\{\omega\}) \\ &= \sum_i g(x_i)P(A_i) \end{aligned}$$

Dans le cas général, on utilise, comme pour la linéarité de l'espérance, la décomposition de X : $X = \sum_i x_i \mathbf{1}_{A_i}$. On a alors une décomposition de $g(X)$ (qui n'est pas nécessairement « canonique », car les $g(x_i)$ ne sont pas nécessairement distincts) :

$$g(X) = \sum_i g(x_i)\mathbf{1}_{A_i}$$

et donc : $E(g(X)) = \sum_i g(x_i)P(A_i)$.

Exemple 3.9 *En reprenant le jeu des 10 € :*

$$E(X^2) = (-10)^2 \times \frac{16}{49} + (-2)^2 \frac{24}{49} + 6^2 \times \frac{9}{49} = \frac{2020}{49}$$

III.6 - Variance et écart-type

Définition 3.4 *On appelle moment d'ordre 2 de X l'espérance, si elle existe, de la variable X^2.*
D'après le théorème de transfert, c'est donc le réel, s'il existe :

$$E(X^2) = \sum_i x_i^2 P(X = x_i)$$

Définition 3.5 *On appelle variance de X l'espérance, si elle existe, de la variable $(X - E(X))^2$:*

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

C'est donc le réel, sous réserve d'existence :

$$E((X - E(X))^2) = \sum_i (x_i - E(X))^2 P(X = x_i)$$

La variance est aussi appelée moment centré d'ordre 2.

Remarque : son existence suppose en particulier celle de $E(X)$.
On mesure la somme des carrés des écarts entre les valeurs de X et la moyenne, pondérés par les poids de probabilité. C'est donc un indicateur de « dispersion » autour de $E(X)$.

Propriété 3.5 *Si X admet une variance, alors :*

1. $V(X) \geq 0$
2. $V(X) = 0 \Leftrightarrow X$ *est constante*

En effet : $V(X) = 0 \Rightarrow \forall i, (x_i - E(X))^2 P(X = x_i) = 0$. On a donc que tous les x_i sont égaux (à $E(X)$). (En toute rigueur, c'est le cas sauf lorsque $P(X = x_i) = 0$, c'est-à-dire hors d'un ensemble de mesure nulle : X est donc constante « presque partout »).

Propriété 3.6 *X admet une variance si et seulement si X admet un moment d'ordre 2.*

Démonstration : il n'y a pas de problème d'existence si $X(\Omega)$ est fini.
Si $X(\Omega)$ est dénombrable : l'existence d'une variance suppose par définition celle de $E(X)$, et l'existence de $E(X^2)$ implique celle de $E(X)$ (essayez de le démontrer, voir exercice 3.5 si difficulté).
Soit $N \geq 1$. On a :

$$\sum_{i=1}^{N}(x_i - E(X))^2 p_i = \sum_{i=1}^{N} x_i^2 p_i - 2E(X)\sum_{i=1}^{N} x_i p_i + (E(X))^2$$

On en déduit l'équivalence entre la convergence de la série $\sum_{i\geq 1} x_i^2 p_i$ et celle de $\sum_{i\geq 1}(x_i - E(X))^2 p_i$.

Le calcul pratique se fait souvent avec la formule de Koenig-Huygens :

Théorème 3.3 *Si X possède une variance, alors :*

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= \sum_i x_i^2 P(X = x_i) - \Big(\sum_i x_i P(X = x_i)\Big)^2 \end{aligned}$$

On obtient donc la variance en calculant « la moyenne des carrés moins le carré de la moyenne ».

Démonstration : on utilise la linéarité de l'espérance :

$$\begin{aligned} V(X) &= E(X - E(X))^2 \\ &= E\big(X^2 - 2XE(X) + (E(X))^2\big) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

Exemple 3.10 *Pour le jeu des 10 €, on obtient :*

$$V(X) = (-10)^2 \times \frac{16}{49} + (-2)^2 \times \frac{24}{49} + 6^2 \times \frac{9}{49} - \left(\frac{-22}{7}\right)^2 \simeq 31.35$$

Propriété 3.7 *Soit $(a; b) \in \mathbb{R}^2$. Si X possède une variance, alors la v.a.r. $aX + b$ possède une variance et :*

$$V(aX + b) = a^2 V(X)$$

Démonstration : il faut vérifier que la variable $(aX + b - E(aX + b))^2$ possède une espérance.
Or, par linéarité de E :

$$(aX + b - E(aX + b))^2 = (aX + b - aE(X) - b)^2 = a^2(X - E(X))^2$$

Puisque X possède une variance, on a par définition l'existence de l'espérance de la variable $(X - E(X))^2$, et donc celle de la variable $a^2(X - E(X))^2$.
On a immédiatement :

$$\begin{aligned} V(aX + b) &= E(a^2(X - E(X))^2) \\ &= a^2 E(X - E(X))^2 \\ &= a^2 V(X) \end{aligned}$$

Définition 3.6 *On appelle, lorsque X admet une variance, écart-type de X le réel :*

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

Définition 3.7 *On appelle variable centrée toute variable X telle que : $E(X) = 0$, et variable réduite toute variable X telle que : $V(X) = 1$.*

Exemple 3.11 *La variable $X - E(X)$ est centrée et la variable $\dfrac{1}{\sigma(X)}X$ est réduite.*

Définition 3.8 *On appelle moment d'ordre k l'espérance, si elle existe, de la v.a.r. X^k, c'est-à-dire, d'après le théorème de transfert :*

$$E(X^k) = \sum_i x_i^k P(X = x_i)$$

Remarque : si $X(\Omega)$ est fini, alors X possède des moments de tout ordre. Dans le cas où $X(\Omega)$ est dénombrable, l'existence du moment d'ordre k suppose par définition la convergence absolue de la série $\sum\limits_{i \in \mathbb{N}} x_i^k P(X = x_i)$, et on a la propriété :

Propriété 3.8 *Si X possède un moment d'ordre k, alors tous ses moments d'ordre $k' \leq k$ existent.*

Démonstration : voir exercice 3.5.

III.7 - Énoncé des exercices

Exercice 3.1 *Soit $(\Omega; \mathcal{A}; P)$ un espace probabilisé.*

1. *Soit un événement A : montrer que $P(A) = E(\mathbf{1}_A)$, où $\mathbf{1}_A$ est l'indicatrice de A.*
   Cette égalité sera notamment à la base de la démonstration de l'inégalité de Bienaymé-Tchebychev.
2. *Écrire $\mathbf{1}_{A \cup B}$, $\mathbf{1}_{A \cap B}$ et $\mathbf{1}_{\overline{A}}$ en fonction de $\mathbf{1}_A$ et $\mathbf{1}_B$.*
3. *Démonstration de la formule de Poincaré : on considère n événements $A_1, \cdots, A_n$.*
   i Écrire $\mathbf{1}_{A_1 \cup \cdots \cup A_n}$ en fonction des indicatrices des A_i.
   ii Soit Q un polynôme unitaire à coefficients réels de degré n :

$$Q(X) = X^n + a_{n-1}X^{n-1} + a_{n-2}X^{n-2} + \cdots + a_0$$

On note $\alpha_1, \cdots, \alpha_n$ ses racines (éventuellement complexes, non nécessairement distinctes) et $\sigma_1, \cdots, \sigma_n$ les fonctions symétriques élémentaires des racines, c'est-à-dire :

$$\forall k \in [\![1; n]\!], \sigma_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_k}$$

Développer $(X - \alpha_1)\cdots(X - \alpha_n)$ en faisant apparaître les σ_k.

iii Démontrer alors la formule de Poincaré.

Exercice 3.2 *Soit f définie sur $\mathbb{N}^*$ par :*

$$\forall n > 0, f(n) = \frac{4}{n(n+1)(n+2)}$$

Vérifier que f est la distribution de probabilité d'une v.a.r. X à valeurs dans $\mathbb{N}^*$ (telle que : $\forall n \in \mathbb{N}^*, P(X = n) = f(n)$), et calculer lorsque c'est possible l'espérance et la variance de X.

Exercice 3.3 *Soit X la v.a.r. prenant pour valeurs les entiers naturels congrus à 0 ou 1 modulo 3, et dont la loi de probabilité est donnée par :*

$$\forall n \in X(\Omega), P(X = n) = \alpha 3^{-n}$$

1. *Calculer α.*
2. *Calculer, lorsque c'est possible, l'espérance et la variance de X.*

Exercice 3.4 *Soit X une variable aléatoire dans $\mathbb{N}$. On note F sa fonction de répartition et on suppose que $E(X)$ existe.*

1. *Montrer que : $n(1 - F(n)) \xrightarrow[n \to +\infty]{} 0$.*
2. *Cela suffit-il à assurer la convergence de la série $\sum_{n \geq 0} (1 - F(n))$?*

   c'est-à-dire :

$$u_n = \circ\left(\frac{1}{n}\right) \Rightarrow \sum_{n \geq 0} u_n \text{ converge ?}$$

3. *Démontrer que $1 - F(n)$ est le terme général d'une série convergente, dont la somme est égale à $E(X)$.*

Exercice 3.5 *1. Montrer que si une v.a.r. discrète possède un moment d'ordre 2, alors elle possède un moment d'ordre 1.*

2. De manière générale, montrer que si une v.a.r. discrète admet un moment d'ordre k, alors elle admet des moments d'ordre k' pour tout $k' \leq k$.

Exercice 3.6 *1. Soit f définie sur $\mathbb{R}$, 2π-périodique, paire et telle que :*

$$\forall x \in [o; \pi], f(x) = \cosh x$$

Déterminer la série de Fourier de f, étudier sa convergence et donner sa somme, puis en déduire la valeur de $\displaystyle\sum_{n=0}^{+\infty} \frac{1}{n^2 + 1}$.

2. *Soit X une v.a.r. dont la loi est donnée par :*

$$X(\Omega) = \mathbb{N} \text{ et } \forall n \in \mathbb{N}, P(X = n) = \frac{\alpha}{n^2 + 1}$$

Déterminer α. X admet-elle une espérance ?

Exercice 3.7 *k urnes contiennent chacune n boules indiscernables au toucher numérotées de 1 à n. On prélève une boule dans chaque urne, et on appelle X_n la v.a.r. égale au plus grand des numéros obtenus.*

1. *Déterminer Ω.*
   On choisit comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$ et comme probabilité P la probabilité uniforme.
2. *Préciser $X(\Omega)$ et calculer $P(X = 1)$.*
3. *Calculer, pour $h \in X(\Omega)$, $P(X_n \leq h)$.*
4. *En déduire, pour $h \in X(\Omega)$, $P(X_n = h)$.*
5. *Déterminer un équivalent de $E(X_n)$ lorsque n tend vers $+\infty$ (k fixé) (on ne cherchera pas à calculer explicitement $E(X_n)$).*
   Indication : faire apparaître une somme de Riemann.

Exercice 3.8 *Soit X une v.a.r. discrète admettant un moment d'ordre 2. Déterminer le minimum de la fonction :*

$$\phi : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ a & \mapsto & E((X - a)^2) \end{array} \right.$$

Interpréter le résultat.

Exercice 3.9 *Une urne est composée de 6 boules indiscernables au toucher portant les numéros suivants : 0, 1, 1, 2, 2 et 4. On tire une poignée de n boules et on appelle X_n le produit des numéros obtenus.*

1. *Déterminer, pour $n \in [\![1; 6]\!]$, un espace probabilisé $(\Omega_n; \mathcal{A}_n; P_n)$.*
2. *Déterminer les lois de $X_1, \ldots, X_6$.*
3. *Si l'on veut maximiser le produit obtenu, combien vaut-il mieux tirer de boules ?*

Exercice 3.10 *Une urne contient $n + 1$ boules indiscernables au toucher numérotées comme suit : 1, 1, 2, 3, 4, ..., n. On tire une poignée de N boules et on appelle S la somme des N numéros obtenus.*

1. *Définir $(\Omega; \mathcal{A}; P)$.*

2. Soit X_k la variable définie par :

$$X_k = \begin{cases} k & \textit{si le numéro } k \textit{ est dans la poignée} \\ 0 & \textit{sinon} \end{cases}$$

Déterminer la loi de X_k.

3. Exprimer S en fonction des X_k et en déduire $E(S)$.

III.8 - Correction des exercices

Exercice 3.1

1. On a $\mathbf{1}_A(\Omega) = \{0; 1\}$, et :

$$P(\mathbf{1}_A = 1) = P(\{\omega/\mathbf{1}_A(\omega) = 1\}) = P(\{\omega/\omega \in A\}) = P(A)$$

D'où : $E(\mathbf{1}_A) = 0 \times (1 - P(A)) + 1 \times P(A) = P(A)$.

2. On vérifie facilement que $\mathbf{1}_{A\cap B} = \mathbf{1}_A\mathbf{1}_B$, $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A\mathbf{1}_B$, et que $\mathbf{1}_{\overline{A}} = 1 - \mathbf{1}_A$.

3. i d'après la question précédente :

$$\mathbf{1}_{A_1\cup\cdots\cup A_n} = 1 - \mathbf{1}_{\overline{A_1}\cap\cdots\cap\overline{A_n}} = 1 - (1 - \mathbf{1}_{A_1})\cdots(1 - \mathbf{1}_{A_n})$$

ii On développe :

$$\begin{aligned}(X - \alpha_1)\cdots(X - \alpha_n) &= X^n + X^{n-1}(-\alpha_1 - \cdots - \alpha_n) \\ &\quad + X^{n-2}(\alpha_1\alpha_2 + \cdots \alpha_{n-1}\alpha_n) + \cdots \\ &\quad + (-1)^n\alpha_1\cdots\alpha_n \\ &= X^n + X^{n-1}(-\sigma_1) + X^{n-2}(\sigma_2) + \cdots \\ &\quad + (-1)^n\sigma_n\end{aligned}$$

Remarque : on retrouve en particulier les relations entre coefficients du polynôme et fonctions symétriques élémentaires des racines :

$$\forall k \in [\![0; n]\!], a_k = (-1)^{n-k}\sigma_{n-k}$$

iii On a donc l'identité algébrique suivante :

$$(1 - \alpha_1)\cdots(1 - \alpha_n) = 1 - \sigma_1 + \sigma_2 - \sigma_3 \cdots + (-1)^n\sigma_n$$

En posant : $\forall k \in [\![0; n]\!], \tilde{\sigma}_k = \sum_{1\leq i_1<\cdots<i_k\leq n} \mathbf{1}_{A_{i_1}}\cdots\mathbf{1}_{A_{i_n}}$, on obtient :

$$1 - (1 - \mathbf{1}_{A_1})\cdots(1 - \mathbf{1}_{A_n}) = \tilde{\sigma}_1 - \tilde{\sigma}_2 + \tilde{\sigma}_3 + \cdots (-1)^{n-1}\tilde{\sigma}_n$$

Par linéarité de l'espérance :

$$\begin{aligned} E(\mathbf{1}_{A_1\cup\cdots\cup A_n}) &= E(\tilde{\sigma}_1) - E(\tilde{\sigma}_2) + E(\tilde{\sigma}_3) + \cdots (-1)^{n-1}E(\tilde{\sigma}_n) \\ &= P(A_1) + \cdots + P(A_n) \\ &\quad -P(A_1 \cap A_2) - \cdots - P(A_{n-1} \cap A_n) \\ &\quad \ldots \\ &\quad +(-1)^{n-1}P(A_1 \cap \cdots \cap A_n) \end{aligned}$$

ce qui est la formule de Poincaré, puisque :

$$E(\mathbf{1}_{A_1\cup\cdots\cup A_n}) = P(A_1 \cup \cdots \cup A_n)$$

Suggestion : pour bien voir l'intérêt des fonctions caractéristiques, vous pouvez regarder l'exercice d'algèbre suivant : soit E un ensemble non vide. On munit $\mathcal{P}(E)$ de la loi Δ, appelée différence symétrique, définie par :

$$\forall(A; B) \in (\mathcal{P}(E))^2, A\Delta B = (A \cap \overline{B}) \cup (\overline{A} \cap B)$$

Question : $(\mathcal{P}(E); \Delta)$ est-il un groupe ?
Pour l'associativité, vous verrez vite l'intérêt des fonctions caractéristiques (commencer par exprimer $\mathbf{1}_{A\Delta B}$ en fonction de $\mathbf{1}_A$ et $\mathbf{1}_B$) !

Exercice 3.2
La série $\sum_{n\geq 1} f(n)$ est à termes positifs. Il reste à vérifier qu'elle converge vers

1. Elle est convergente puisque $f(n) \underset{n\to+\infty}{\sim} \frac{4}{n^3}$. Pour calculer sa somme, on décompose $f(n)$ en éléments simples :

$$\begin{aligned} \sum_{n=1}^{N} \frac{4}{n(n+1)(n+2)} &= \sum_{n=1}^{N}\Big(\frac{2}{n} - \frac{4}{n+1} + \frac{2}{n+2}\Big) \\ &= 2\sum_{n=1}^{N}\frac{1}{n} - 4\sum_{n=1}^{N}\frac{1}{n+1} + 2\sum_{n=1}^{N}\frac{1}{n+2} \\ &= 2\sum_{n=1}^{N}\frac{1}{n} - 4\sum_{n=2}^{N+1}\frac{1}{n} + 2\sum_{n=3}^{N+2}\frac{1}{n} \\ &= 2\Big(1 + \frac{1}{2}\Big) - 4\Big(\frac{1}{2} + \frac{1}{N+1}\Big) \\ &\quad +2\Big(\frac{1}{N+1} + \frac{1}{N+2}\Big) \\ &\xrightarrow[N\to+\infty]{} 1 \end{aligned}$$

Attention à bien prendre des sommes partielles pour le calcul, sinon vous écrivez une série convergente comme somme de trois séries divergentes.

Remarque : dans la décomposition en éléments simples :

$$\frac{4}{n(n+1)(n+2)} = \frac{a}{n} + \frac{b}{n+1} + \frac{c}{n+2}$$

la somme des poids $a+b+c$ est nécessairement nulle (multiplier par n et faire tendre n vers $+\infty$), ce qui permet de simplifier le calcul de la somme partielle. C'est le cas pour toutes les séries dont le terme général est une fraction avec des pôles simples au dénominateur et une constante au numérateur.

Conclusion : la suite $(f(n))$ est la distribution de probabilité d'une v.a.r. X ($X(\Omega) = \mathbb{N}^*$ et $P(X=n) = f(n)$).

X possède une espérance puisque $nf(n) \underset{n\to+\infty}{\sim} \frac{4}{n^2}$, ce qui assure la convergence de la série $\sum_{n\geq 1} nf(n)$. On calcule $E(X)$ encore en décomposant en éléments simples :

$$\begin{aligned}
\sum_{n=1}^{N} nP(X=n) &= 4\sum_{n=1}^{N} \frac{1}{(n+1)(n+2)} \\
&= 4\sum_{n=1}^{N}\left(\frac{1}{n+1} - \frac{1}{n+2}\right) \\
&= 4\left(\frac{1}{2} - \frac{1}{N+2}\right) \\
&\xrightarrow[N\to+\infty]{} 2
\end{aligned}$$

Par contre, X ne possède pas de variance puisque $n^2 f(n) \underset{n\to+\infty}{\sim} \frac{4}{n}$, ce qui entraîne la divergence de la série $\sum_{n\geq 1} n^2 f(n)$. ($E(X^2)$ n'existe pas).

Exercice 3.3

Les réels $P(X=n)$ doivent être positifs, donc $\alpha \geq 0$.

La série $\sum_{n\in X(\Omega)} P(X=n)$ est à termes positifs, et toutes ses sommes partielles sont majorées par la constante $\sum_{n=0}^{+\infty} \frac{\alpha}{3^n} = \alpha \frac{1}{1-\frac{1}{3}}$. La série est donc convergente. Sa somme doit être égale à 1 :

$$\sum_{n\in X(\Omega)} P(X=n) = \alpha\Big(\sum_{k=0}^{+\infty}\frac{1}{3^{3k}} + \sum_{k=0}^{+\infty}\frac{1}{3^{3k+1}}\Big) = \alpha\frac{4}{3}\Big(\frac{1}{1-\frac{1}{27}}\Big) = \alpha\frac{18}{13}$$

On pose donc $\alpha = \dfrac{13}{18}$.
L'espérance de X existe si et seulement si la série $\sum\limits_{n\in X(\Omega)} nP(X=n)$ converge. C'est le cas puisque c'est une série à termes positifs dont les sommes partielles sont toutes majorées par la constante : $\dfrac{13}{18}\sum\limits_{n=0}^{+\infty} n\Big(\dfrac{1}{3}\Big)^n$ (cette somme est bien finie, par exemple par la règle de D'Alembert, ou comme dérivée de la série entière de rayon 1 $\sum\limits_{n\geq 0} x^n$, en $x=\dfrac{1}{3}$).
Le calcul de l'espérance utilise la série géométrique dérivée une fois :

$$\forall q \in]-1;1[, \sum_{n=1}^{+\infty} nq^{n-1} = \frac{1}{(1-q)^2}$$

On décompose :

$$\begin{aligned}
\sum_{n\in X(\Omega)} nP(X=n) &= \frac{13}{18}\Big(\sum_{k=0}^{+\infty} 3k\frac{1}{3^{3k}} + \sum_{k=0}^{+\infty}(3k+1)\frac{1}{3^{3k+1}}\Big)\\
&= \frac{13}{18}\Big(\frac{1}{9}\sum_{k=0}^{+\infty} k\Big(\frac{1}{27}\Big)^{k-1} + \frac{1}{27}\sum_{k=0}^{+\infty} k\Big(\frac{1}{27}\Big)^{k-1}\\
&\quad +\frac{1}{3}\sum_{k=0}^{+\infty}\Big(\frac{1}{27}\Big)^{k}\Big)\\
&= \frac{13}{18}\Big(\frac{4}{27}\times\frac{1}{(1-\frac{1}{27})^2} + \frac{1}{3}\times\frac{1}{1-\frac{1}{27}}\Big) = \frac{19}{52}
\end{aligned}$$

Pour la variance, $E(X^2)$ existe car les sommes partielles de la série à termes positifs $\sum\limits_{n\in X(\Omega)} n^2P(X=n)$ sont majorées par la constante $\sum\limits_{n=0}^{+\infty} n^2\Big(\dfrac{1}{3}\Big)^n$, et le calcul se fait en écrivant $n^2 = n(n-1)+n$ et utilisant les séries géométriques dérivées.

Exercice 3.4

1. On majore $n(1-F(n))$:

$$\begin{aligned} n(1-F(n)) &= nP(X>n) \\ &= n\sum_{k=n+1}^{+\infty} P(X=k) \\ &\leq \sum_{k=n+1}^{+\infty} kP(X=k) \xrightarrow[n\to+\infty]{} 0 \end{aligned}$$

(reste d'une série convergente).

2. La condition (nu_n) converge vers 0 n'est pas suffisante. On considère la série $\sum\limits_{n\geq 2} \dfrac{1}{n\ln n}$ (elle fait partie d'un ensemble de séries appelées séries de Bertrand). On a bien : $\dfrac{1}{n\ln n} = \mathrm{o}\left(\dfrac{1}{n}\right)$, mais la série diverge ; on le montre avec le théorème de comparaison séries-intégrales : la fonction

$$g : \left\{ \begin{array}{ccc} [2;+\infty[& \to & \mathbb{R} \\ t & \mapsto & \dfrac{1}{t\ln t} \end{array} \right.$$

est positive, continue et décroissante.
La série $\sum\limits_{n\geq 2} g(n)$ et l'intégrale $\displaystyle\int_2^{+\infty} g(t)dt$ sont donc de même nature.
Or l'intégrale est divergente : soit $x \geq 2$. On a, en intégrant directement :

$$\begin{aligned} \int_2^x \frac{1}{t\ln t}dt &= \int_2^x \frac{\frac{1}{t}}{\ln t}dt \\ &= [\ln(\ln t)]_2^x \\ &= \ln(\ln x) - \ln(\ln 2) \\ &\xrightarrow[x\to+\infty]{} +\infty \end{aligned}$$

La série est donc également divergente.

Remarque : pour une question beaucoup plus simple sur les suites, vous pouvez vérifier que $nu_n \xrightarrow[n\to+\infty]{} 0 \Rightarrow u_n \xrightarrow[n\to+\infty]{} 0$, et que la réciproque est fausse.

3. Soit $N \geq 0$:

$$\begin{aligned}
&\sum_{n=0}^{N}(1-F(n)) \\
= \quad &\sum_{n=0}^{N} P(X>n) \\
= \quad &P(X=1)+\cdots P(X=N+1)+P(X>N+1) \\
&+P(X=2)+\cdots P(X=N+1)+P(X>N+1) \\
&+P(X=3)+\cdots P(X=N+1)+P(X>N+1) \\
&+\cdots+P(X=N+1)+P(X>N+1) \\
= \quad &\sum_{n=1}^{N+1} nP(X=n)+(N+1)P(X>N+1) \\
\xrightarrow[N\to+\infty]{} \quad &\sum_{n=1}^{+\infty} nP(X=n)\,(=E(X))
\end{aligned}$$

car on a montré à la première question que $NP(X>N) \xrightarrow[N\to+\infty]{} 0$.

Exercice 3.5

1. On suppose que $\sum_{i\geq 1} x_i^2 p_i$ converge. On a l'inégalité :

$$\forall x \in \mathbb{R}, |x| \leq x^2+1$$

Comme les séries $\sum_{i\geq 1} x_i^2 p_i$ et $\sum_{i\geq 1} p_i$ sont convergentes, la série $\sum_{i\geq 1} |x_i| p_i$ converge, en vertu du théorème de comparaison pour les séries à termes positifs.

2. On a l'inégalité suivante : $\forall x \in \mathbb{R}, |x|^{k-1} \leq |x|^k+1$ (distinguer les cas $|x| \geq 1$ et $|x| \leq 1$). Comme $\sum_{i\geq 1} |x_i|^k p_i$ et $\sum_{i\geq 1} p_i$ sont convergentes, le théorème de comparaison pour les séries à termes positifs assure la convergence de $\sum_{i\geq 1} |x_i|^{k-1} p_i$. Par récurrence finie, on obtient bien que pour tout k' inférieur ou égal à k, X possède un moment d'ordre k'.

Exercice 3.6

1. f étant paire, les coefficients b_n sont tous nuls.
Calcul des coefficients a_n :

$$\forall n \geq 1, a_n \quad = \quad \frac{1}{\pi}\int_{-\pi}^{\pi} \cosh t \cos(nt)dt = \frac{2}{\pi}\underbrace{\int_0^{\pi} \cosh t \cos(nt)dt}_{I}$$

On calcule I par exemple en effectuant une double intégration par parties :

$$\begin{aligned} I &= -\frac{1}{n}\int_0^\pi \sinh t \sin(nt)dt \\ &= -\frac{1}{n}[-\sinh t\frac{\cos(nt)}{n}]_0^\pi - \frac{1}{n}\int_0^\pi \cosh t\frac{\cos(nt)}{n}dt \\ &= \frac{(-1)^n \sinh\pi}{n^2} - \frac{1}{n^2}I \end{aligned}$$

D'où : $I = \dfrac{(-1)^n \sinh\pi}{n^2+1}$ et $a_n = \dfrac{2}{\pi}\dfrac{(-1)^n \sinh\pi}{n^2+1}$.

D'autre part : $\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \cosh t dt = \dfrac{\sinh\pi}{\pi}$. D'où la série de Fourier de f :

$$\frac{\sinh\pi}{\pi} + \sum_{n\geq 1}\frac{2}{\pi}\frac{(-1)^n\sinh\pi}{n^2+1}\cos(nx)$$

Comme f est C^1 par morceaux, le théorème de Dirichlet s'applique (la série de Fourier de f converge pour tout x vers $\dfrac{f(x^-)+f(x^+)}{2}$), et comme f est continue, sa série de Fourier converge en tout point x vers $f(x)$:

$$\forall x \in \mathbb{R}, f(x) = \frac{\sinh\pi}{\pi} + \sum_{n=1}^{+\infty}\frac{2}{\pi}\frac{(-1)^n\sinh\pi}{n^2+1}\cos(nx)$$

Pour $x = \pi$:

$$\sum_{n=1}^{+\infty}\frac{1}{1+n^2} = \left(f(\pi) - \frac{\sinh\pi}{\pi}\right)\frac{\pi}{2\sinh\pi} = \frac{\pi}{2}\coth\pi - \frac{1}{2}$$

Conclusion :

$$\sum_{n=0}^{+\infty}\frac{1}{1+n^2} = \frac{\pi}{2}\coth\pi + \frac{1}{2}$$

2. La série $\displaystyle\sum_{n\geq 0} P(X=n)$ doit être convergente et de somme 1, donc on pose :

$$\alpha = \frac{2}{\pi\coth\pi + 1}$$

La variable X ne possède pas d'espérance car $nP(X=n) \sim \dfrac{\alpha}{n}$, ce qui entraîne la divergence de la série $\displaystyle\sum_{n\geq 0} nP(X=n)$.

Exercice 3.7

1. On pose comme univers : $\Omega = [\![1;n]\!]^k$.
2. On a $X_n(\Omega) = [\![1;n]\!]$, et : $P(X_n = 1) = P(\{(1;\cdots;n)\}) = \dfrac{1}{n^k}$.
3. Soit $h \in [\![1;n]\!]$:

$$\begin{aligned}
P(X_n \leq h) &= P\Bigg(\bigcup_{1\leq\alpha_1\leq h,\cdots,1\leq\alpha_k\leq h} \{(\alpha_1;\cdots;\alpha_k)\}\Bigg)\\
&= P([\![1;h]\!]^k)\\
&= \frac{\mathrm{Card}([\![1;h]\!]^k)}{\mathrm{Card}\,\Omega}\\
&= \frac{h^k}{n^k}
\end{aligned}$$

4. On en déduit :

$$\begin{aligned}
P(X_n = h) &= P(X_n \leq h) - P(X_n \leq h-1)\\
&= \frac{h^k}{n^k} - \frac{(h-1)^k}{n^k}
\end{aligned}$$

5. On obtient alors un équivalent, lorsque n tend vers $+\infty$, de l'espérance grâce aux sommes de Riemann :

$$\begin{aligned}
E(X_n) &= \sum_{h=1}^{n} hP(X = h)\\
&= \sum_{h=1}^{n} \frac{h^{k+1}}{n^k} - \sum_{h=1}^{n-1}(h+1)\frac{h^k}{n^k}\\
&= \frac{n^{k+1}}{n^k} - \sum_{h=1}^{n-1}\frac{h^k}{n^k}\\
&= n\Big(1 - \frac{1}{n}\sum_{h=1}^{n-1}\Big(\frac{h}{n}\Big)^k\Big)\\
&\underset{n\to+\infty}{\sim} n\Big(1 - \int_0^1 t^k dt\Big)\\
&\underset{n\to+\infty}{\sim} n\Big(1 - \frac{1}{k+1}\Big)\\
&\underset{n\to+\infty}{\sim} n\frac{k}{k+1}
\end{aligned}$$

Exercice 3.8

$$\phi(a) = E(X^2 - 2aX + a^2) = E(X^2) - 2aE(X) + a^2$$

ϕ est donc minimale en $a = E(X)$.
L'espérance est donc la meilleure approximation de X par une constante au sens des moindres carrés. En termes d'algèbre bilinéaire, $E(X)$ est le projeté orthogonal de X sur la droite $Vect\{1\}$, au sens du produit scalaire $(X;Y) \mapsto E(XY)$.

Exercice 3.9
Soit $E = \{0;1;1;2;2;4\}$.

1. On pose, pour $n \in [\![1;6]\!]$: $\Omega_n = \{A \subset \mathcal{P}(E) / \operatorname{Card} A = n\}$. On choisit $\mathcal{A}_n = \mathcal{P}(\Omega_n)$, et comme les poignées sont équiprobables (n est fixé), on prend comme probabilité P_n la probabilité uniforme.
2. Le calcul des poids de probabilité se réduit à du dénombrement puisque P_n est uniforme. On a $\operatorname{Card}\Omega_n = \binom{6}{n}$, et on obtient :

$X_1(\Omega_1)$	0	1	2	4
	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

$X_2(\Omega_2)$	0	1	2	4	8
	$\frac{5}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

$X_3(\Omega_3)$	0	2	4	8	16
	$\frac{10}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{1}{20}$

$X_4(\Omega_4)$	0	4	8	16
	$\frac{10}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$

$X_5(\Omega_5)$	0	16
	$\frac{5}{6}$	$\frac{1}{6}$

$X_6(\Omega_6)$	0
	1

3. On compare les espérances :

$$E(X_1) = \frac{5}{3}, E(X_2) = \frac{37}{15}, E(X_3) = \frac{16}{5}$$
$$E(X_4) = \frac{52}{15}, E(X_5) = \frac{8}{3}, E(X_6) = 0$$

Il vaut donc mieux tirer une poignée de 4 boules (espérance la plus élevée).

Exercice 3.10
Soit $E = \{1;1;2;3;\cdots;n\}$.

1. On pose $\Omega = \{A \subset \mathcal{P}(E) / \operatorname{Card} A = N\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et on prend pour P la probabilité uniforme.

2. Si $k=1$:

$$\begin{aligned} P(X_1=0) &= \frac{\mathrm{Card}(X_1^{-1}(\{0\}))}{\mathrm{Card}\,\Omega} \\ &= \frac{\binom{n-1}{N}}{\binom{n+1}{N}} = \frac{(n+1-N)(n-N)}{n(n+1)} \end{aligned}$$

D'où la loi de X_1 :

$X_1(\Omega)$	0	1
	$\dfrac{(n+1-N)(n-N)}{n(n+1)}$	$1-\dfrac{(n+1-N)(n-N)}{n(n+1)}$

Même raisonnement si $k>1$:

$$P(X_k=0) = \frac{\mathrm{Card}(X_k^{-1}(\{0\}))}{\mathrm{Card}\,\Omega} = \frac{\begin{pmatrix} n+1 \\ N-1 \end{pmatrix}}{\begin{pmatrix} n+1 \\ N \end{pmatrix}} = \frac{n+1-N}{n+1}$$

D'où la loi de X_k :

$X_k(\Omega)$	0	k
	$\dfrac{n+1-N}{n+1}$	$\dfrac{N}{n+1}$

3. On a $S=\sum_{k=1}^{n} X_k$, et par linéarité de l'espérance :

$$\begin{aligned} E(S) &= \sum_{k=1}^{n} E(X_k) \\ &= 1-\frac{(n+1-N)(n-N)}{n(n+1)} + \sum_{k=1}^{n} k\left(\frac{N}{n+1}\right) \\ &= 1-\frac{(n+1-N)(n-N)}{n(n+1)} + \frac{Nn}{2} \end{aligned}$$

Chapitre IV : Lois discrètes usuelles

Il s'agit de passer en revue les lois discrètes usuelles du programme, en donnant pour chacune d'elles l'univers image, les poids de probabilité, l'espérance et la variance, ainsi que le contexte classique où on les rencontre.

IV.1 - Loi uniforme sur $[\![1;n]\!]$

Définition 4.1 *X suit la loi uniforme sur $[\![1;n]\!]$ si :*

1. $X(\Omega) = [\![1;n]\!]$

2. $\forall i \in [\![1;n]\!], P(X = i) = \dfrac{1}{n}$

On note alors : $X \sim \mathcal{U}([\![1;n]\!])$.

Il s'agit simplement, comme son nom l'indique, d'une loi dont tous les poids de probabilité sont identiques.

Propriété 4.1 $E(X) = \dfrac{n+1}{2}$ *et* $V(X) = \dfrac{n^2-1}{12}$.

Démonstration :

$$E(X) = \sum_{i=1}^{n} i \times \frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2}$$

Commentaire : sachez redémontrer la valeur de la somme des n premiers entiers (écrire la somme une fois dans un sens et une fois dans l'autre...).

Remarque : comme les poids sont tous égaux, l'espérance est l'isobarycentre des entiers de 1 à n, et vaut donc bien $\dfrac{n+1}{2}$.

Pour la variance :

$$E(X^2) = \sum_{i=1}^{n} i^2 \times \frac{1}{n} = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

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

Commentaire : la valeur de la somme des carrés des n premiers entiers est une bonne application de la formule du binôme (leçon sur les combinaisons)...

IV.2 - Loi de Bernoulli

Définition 4.2 *Soit $p \in [0;1]$. X suit la loi de Bernoulli de paramètre p si :*

1. $X(\Omega) = \{0;1\}$

2. $P(X = 0) = 1 - p$ *et* $P(X = 1) = p$

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

Contexte usuel : on considère un espace probabilisé $(\Omega; \mathcal{A}; P)$ et on s'intéresse à un événement S, appelé « succès ».
On définit alors une v.a.r.

$$X : \omega \mapsto \begin{cases} 1 \text{ si } \omega \in S \\ 0 \text{ sinon} \end{cases}$$

X est donc simplement l'indicatrice du succès S, et si on note $p = P(S)$, X suit une loi de Bernoulli de paramètre p.

Propriété 4.2 $E(X) = p$ *et* $V(X) = p(1-p)$

Remarque : les calculs des moments de X sont facilités par le fait que :

$$\forall n \geq 1, X^n = X$$

IV.3 - Loi binomiale

Définition 4.3 *Soient $n \in \mathbb{N}^*$ et $p \in [0;1]$. X suit la loi binomiale de paramètres n et p si :*

1. $X(\Omega) = [\![0;n]\!]$

2. $\forall k \in [\![0;n]\!], P(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$

On note alors : $X \sim \mathcal{B}(n;p)$.

Contexte usuel : comme pour la loi de Bernoulli, on considère une expérience aléatoire modélisée par un espace probabilisé $(\Omega; \mathcal{A}; P)$ et on s'intéresse à un événement S.
On note $p = P(S)$.
On répète n fois cette expérience aléatoire de manière « indépendante »,

c'est-à-dire que l'on munit l'espace probabilisable produit $(\Omega^n; \mathcal{A}^{\otimes n})$ de la probabilité produit (voir exercice 1.1) notée $P^{\otimes n}$.
On appelle alors X la v.a.r. qui prend pour valeur le nombre de succès obtenus au cours des n réalisations de l'expérience :

$$X : (\omega_1; ...; \omega_n) \mapsto \mathrm{Card}(\{\omega_i / \omega_i \in S\})$$

Déterminons la loi de X. On a évidemment $X(\Omega) = [\![0; n]\!]$.
On calcule ensuite les poids de probabilité, en utilisant la définition de la probabilité produit :

$$\begin{aligned}
P^{\otimes n}(X = 0) &= P^{\otimes n}(\overline{S}; ...; \overline{S}) \\
&= P(\overline{S}) \times ... \times P(\overline{S}) \\
&= (1-p)^n \\
P^{\otimes n}(X = 1) &= P^{\otimes n}((S; \overline{S}; ...; \overline{S}) \cup ... \cup (\overline{S}; ...; \overline{S}; S)) \\
&= P^{\otimes n}(S; \overline{S}; ...; \overline{S}) + ... + P^{\otimes n}(\overline{S}; ...; \overline{S}; S) \\
&= P(S)P(\overline{S})...P(\overline{S}) + ... + P(\overline{S})...P(\overline{S})P(S) \\
&= np(1-p)^{n-1}
\end{aligned}$$

Pour $k \in [\![0; n]\!]$: l'événement $(X = k)$ est la réunion n-uplets où S apparaît k fois et $\overline{S}$ apparaît $n - k$ fois.
Ces événements sont 2 à 2 disjoints, donc par additivité de $P^{\otimes n}$, la probabilité de $(X = k)$ est la somme des probabilités de ces événements.
Ils ont tous pour probabilité $p^k(1-p)^{n-k}$.
On a $\begin{pmatrix} n \\ k \end{pmatrix}$ n-uplets comportant k fois S et $n - k$ fois $\overline{S}$, d'où :

$$P^{\otimes n}(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$$

Propriété 4.3 *$E(X) = np$ et $V(X) = np(1-p)$.*

Démonstration : on donne ici une preuve purement calculatoire. Ces deux résultats peuvent s'obtenir plus élégamment avec les fonctions génératrices (voir fin du chapitre) ou en écrivant X comme une somme de n variables de Bernoulli indépendantes (voir chapitre suivant, « vecteurs aléatoires discrets »).

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

Même type de calcul pour la variance, en écrivant $k^2 = k(k-1) + k$.

Exemple 4.1 *Soit une urne contenant 6 boules noires et 5 boules rouges indiscernables au toucher. On procède à 10 extractions successives d'une boule, avec remise. Quelle est la loi du nombre de boules rouges obtenues ?* *On répète 10 fois la même expérience (puisque les tirages se font avec remise) modélisée par* $\Omega = \{N_1; ...N_6; R_1; ...R_5\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ *et* P *la probabilité uniforme. Le « succès » est ici : « obtenir une boule rouge » soit* $S = \{R_1; ...; R_6\}$, *et donc* $P(S) = \dfrac{5}{11}$. *Par conséquent, en notant* X *la v.a.r. prenant pour valeur le nombre de boules rouges obtenues, on a :* $X \sim \mathcal{B}\left(10; \dfrac{5}{11}\right)$.

IV.4 - Loi hypergéométrique

Définition 4.4 *Soit* $(N; n) \in \mathbb{N}^2$ *tel que* $n \leq N$ *et soit* $p \in]0; 1[$ *tel que* $Np \in \mathbb{N}$.
X suit la loi hypergéométrique de paramètres $(N; n; p)$ *si :*

1. $X(\Omega) = [\![\max(0; n - N(1-p)); \min(n; Np)]\!]$
2. $\forall k \in X(\Omega), P(X = k) = \dfrac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$

On note alors : $X \sim \mathcal{H}(N; n; p)$.

Contexte usuel : on considère une urne contenant N boules (indiscernables) : des boules rouges en proportion p (donc $Np \in \mathbb{N}$) et des boules noires en proportion $1-p$.
On tire simultanément n boules (donc $n \leq N$), et on appelle X la v.a.r. égale au nombre de boules rouges obtenues.

On note U l'ensemble des boules de l'urne.
L'expérience est modélisée par $\Omega = \{A \in \mathcal{P}(U) / \operatorname{Card}(A) = n\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P la probabilité uniforme.

Valeur minimale de X : s'il y a au moins n boules noires, c'est-à-dire si $N(1-p) \geq n$, X peut prendre la valeur 0.
Si le nombre de boules noires est strictement inférieur à n (i.e. $N(1-p) < n$), le nombre de boules rouges minimum que l'on puisse obtenir est $n - N(1-p)$ (on a obtenu toutes les boules noires possibles).
La valeur minimale de X est donc bien : $\max(0; n - N(1-p))$.

Valeur maximale de X : si le nombre de boules rouges est supérieur ou égal à n ($Np \geq n$), on peut obtenir au maximum n boules rouges.
Dans le cas contraire ($Np < n$), on ne peut obtenir que Np boules rouges au maximum.
La valeur maximale de X est donc bien : $\min(n; Np)$.

Remarque : l'ensemble $[\![\max(0; n-N(1-p)); \min(n; Np)]\!]$ est exactement l'ensemble des valeurs de k pour lesquelles $\binom{Np}{k}$ et $\binom{N(1-p)}{n-k}$ sont strictement positifs.

Pour $k \in X(\Omega)$, on a :

$$P(X = k) = \frac{\operatorname{Card}(X = k)}{\operatorname{Card}(\Omega)} \text{ (puisque } P \text{ est uniforme)}$$

Or $\operatorname{Card}(\Omega) = \binom{N}{n}$ (nombre de parties à n éléments d'un ensemble de cardinal N) et $\operatorname{Card}(X = k) = \binom{Np}{k}\binom{N(1-p)}{n-k}$ (nombre de parties contenant k boules rouges prises parmi Np boules rouges multiplié par le nombre de parties contenant $n-k$ boules noires prises parmi $N(1-p)$ boules noires).

Remarque : l'égalité $\sum\limits_{k \in X(\Omega)} P(X = k) = 1$ peut se vérifier à l'aide de la formule de Vandermonde.

Propriété 4.4 *$E(X) = np$ et $V(X) = \dfrac{N-n}{N-1} np(1-p)$.*

IV.5 - Loi géométrique

Définition 4.5 *Soit $p \in]0;1]$. X suit la loi géométrique de paramètre p si :*

1. $X(\Omega) = \mathbb{N}^*$

2. $\forall k \in \mathbb{N}^*, P(X = k) = p(1-p)^{k-1}$

On note alors : $X \sim \mathcal{G}(p)$.

Remarque : on a bien la convergence de la série de terme général $(1-p)^{k-1}$ (série géométrique de raison $| 1-p |< 1$), et

$$\sum_{k=1}^{+\infty} p(1-p)^{k-1} \quad = \quad 1$$

Contexte usuel : on considère une expérience aléatoire modélisée par un espace probabilisé $(\Omega; \mathcal{A}; P)$ et on s'intéresse à un événement S, de probabilité $P(S) = p$.
On répète une infinité de fois cette expérience. L'univers est donc $\Omega^{\mathbb{N}}$, c'est-à-dire l'ensemble des suites d'éléments de Ω. Il faudrait ensuite définir une tribu et une probabilité (notons la $\tilde{P}$), ce qui sort du cadre de ce livre. On appelle alors X la v.a.r. égale au rang d'apparition du premier succès (loi du « temps d'attente du premier succès »).
On a donc bien $X(\Omega) = \mathbb{N}^*$ et :

$$\begin{aligned} \forall k \in \mathbb{N}^*, \tilde{P}(X = k) &= \tilde{P}(\underbrace{\overline{S}; ...; \overline{S}}_{k-1 fois}; S; \Omega; ...) \\ &= P(\overline{S}) \times ... \times P(\overline{S}) \times P(S) \times 1... \\ &= (1-p)^{k-1} p \end{aligned}$$

Propriété 4.5 $E(X) = \dfrac{1}{p}$ *et* $V(X) = \dfrac{1}{p^2}(1-p)$

Démonstration : on donne ici une démonstration directe (bon entraînement sur les séries géométriques dérivées). On retrouvera ces résultats sans calculs fastidieux avec les fonctions génératrices.
Rappel sur les séries géométriques : la série entière $\sum_k x^k$ a pour rayon de convergence 1, et on a :

$$\forall x \in]-1;1[, \sum_{k=0}^{+\infty} x^k \quad = \quad \frac{1}{1-x}$$

Or la somme d'une série entière de rayon R est de classe $\mathcal{C}^\infty$ sur $]-R;R[$, et on peut dériver terme à terme (revoir au besoin ce résultat fondamental d'analyse), ce qui donne ici, pour tout $x \in]-1;1[$:

$$\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}, \text{ et } \sum_{k=2}^{+\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

On a donc :

$$E(X) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p \times \frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

Même type de calcul pour $E(X^2)$, en écrivant $k^2 = k(k-1) + k$ et en utilisant la somme d'une série géométrique dérivée deux fois.

IV.6 - Loi de Poisson

Définition 4.6 *Soit $\lambda > 0$. X suit la loi de Poisson de paramètre λ si :*

1. $X(\Omega) = \mathbb{N}$

2. $\forall k \in \mathbb{N}, P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$

On note alors : $X \sim \mathcal{P}(\lambda)$.

Remarque : on a facilement $\sum_{k=0}^{+\infty} P(X = k) = 1$ (série exponentielle).

Contexte usuel : c'est la « loi limite » d'une loi binomiale lorsque np reste constant et n tend vers $+\infty$.

Soit $\lambda = np$. On considère une v.a.r. X qui suit une loi binomiale $\mathcal{B}\left(n; \dfrac{\lambda}{n}\right)$:

$$\begin{aligned}
\forall k \in [\![0;n]\!], P(X = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \times \frac{n(n-1)\cdots(n-k+1)}{n^k} e^{(n-k)\ln(1-\frac{\lambda}{n})} \\
&\xrightarrow[n\to+\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

La loi de Poisson apparaît donc lorsque l'on compte le nombre de succès au cours d'un nombre d'épreuves très élevé, avec une probabilité p du succès très faible : c'est la loi des « événements rares ».

On peut aussi remarquer que $\dfrac{P(X=n)}{P(X=n-1)} = \dfrac{\lambda}{n}$. La suite $(P(X = n))$

des poids de probabilité est donc croissante tant que $n \leq \lambda$, puis elle est décroissante. Les poids « élevés » sont ceux pour lesquels n est proche de λ, et ils sont ensuite rapidement très faibles, ce qui justifie aussi l'appellation « loi des événements rares ».

Propriété 4.6 *$E(X) = \lambda$ et $V(X) = \lambda$.*

Démonstration : on a bien la convergence des séries qui suivent (par exemple par la règle de D'Alembert) et :

$$E(X) = \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

Pour la variance :

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{+\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=0}^{+\infty} (k(k-1)+k) e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=0}^{+\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=2}^{+\infty} \frac{\lambda^k}{(k-2)!} + \lambda \\
&= e^{-\lambda}\lambda^2 e^{\lambda} + \lambda \\
&= \lambda^2 + \lambda
\end{aligned}$$

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

IV.7 - Fonction génératrice

Cette partie ne figure qu'au programme officiel de l'Agrégation Interne et pas à celui du CAPES.
Cependant, elle peut intéresser quand même les candidats au CAPES : on manipule des séries entières et on retrouve assez facilement les espérances et variances usuelles.

Définition 4.7 *Soit X une v.a.r. telle que $X(\Omega) \subset \mathbb{N}$.*
On note, pour $k \in X(\Omega)$, $p_k = P(X = k)$ et on pose, pour $k \notin X(\Omega)$, $p_k = 0$.
On appelle fonction génératrice de X l'application :

$$G : t \mapsto \sum_{k=0}^{+\infty} p_k t^k$$

Propriété 4.7 *G est définie au moins sur* $[-1;1]$*, et elle est de classe* C^∞ *au moins sur* $]-1;1[$.

Démonstration : comme la série $\sum_{k\geq 0} p_k$ converge (vers 1), la fonction G est la somme d'une série entière qui converge pour $t=1$, et qui a donc un rayon de convergence $R \geq 1$.

Propriété 4.8 *G est continue au moins sur* $[-1;1]$.

Démonstration :

$$|t| \leq 1 \Rightarrow |p_k t^k| \leq p_k$$

Comme la série $\sum_{k\geq 0} p_k$ converge, la série $\sum_{k\geq 0} p_k t^k$ converge normalement sur $[-1;1]$, et donc sa somme G est continue au moins sur $[-1;1]$ (la convergence normale de la série entraîne la convergence uniforme de la suite des sommes partielles).
La fonction génératrice d'une v.a.r. X permet d'obtenir ses moments :

Théorème 4.1 *X admet une espérance si et seulement si* G' *a une limite à gauche en 1, et dans ce cas :*

$$E(X) = \lim_{t\to 1^-} G'(t)$$

X admet un moment d'ordre 2 (ou une variance) si et seulement si $G^{(2)}$ *a une limite à gauche en 1, et dans ce cas :*

$$E(X^2) = \lim_{t\to 1^-} G^{(2)}(t) + E(X)$$

Démonstration : si la série entière qui définit G a un rayon de convergence $R > 1$, alors G est de classe C^∞ sur $]-R;R[$, donc en particulier en $t=1$, où l'on a :

$$G'(1) = \sum_{k=1}^{+\infty} kp_k = E(X)$$

et :

$$G^{(2)}(1) = \sum_{k=2}^{+\infty} k(k-1)p_k = E(X(X-1))$$

d'où le résultat souhaité.
Le cas qui pose problème est le cas $R=1$. On rappelle le résultat suivant sur les séries entières :

Théorème 4.2 *Soit $\sum a_n x^n$ une série entière de rayon $R > 0$, telle que : $\forall n \in \mathbb{N}, a_n \geq 0$. On note S sa somme. Alors :*
S possède une limite à gauche en R si et seulement si $\sum a_n R^n$ converge, et dans ce cas :

$$\lim_{x \to R^-} S(x) = \sum_{n=0}^{+\infty} a_n R^n$$

Démonstration : si $\sum a_n R^n$ converge, alors $\sum a_n x^n$ converge normalement sur $[-R; R]$. S est donc continue sur $[-R; R]$, donc elle possède une limite à gauche en R qui vaut $S(R)$.
Réciproquement, si S possède une limite à gauche en R :
Comme les a_n sont positifs, on a :

$$\forall x \in [0; R[, \sum_{n=0}^{N} a_n x^n \leq \sum_{n=0}^{+\infty} a_n x^n = S(x)$$

et comme S est croissante (puisque les a_n sont positifs), on a la majoration $S(x) \leq \lim\limits_{t \to R^-} S(t)$, d'où :

$$\forall x \in [0; R[, \sum_{n=0}^{N} a_n x^n \leq \lim_{x \to R^-} S(x)$$

En faisant tendre x vers R dans l'inégalité précédente, il vient :

$$\sum_{n=0}^{N} a_n R^n \leq \lim_{x \to R^-} S(x)$$

La série $\sum a_n R^n$ est à termes positifs et a ses sommes partielles majorées. Elle est donc convergente et :

$$\sum_{n=0}^{+\infty} a_n R^n \leq \lim_{x \to R^-} S(x)$$

De plus, $\forall x \in [0; R[, \sum\limits_{n=0}^{N} a_n x^n \leq \sum\limits_{n=0}^{N} a_n R^n$, ce qui permet d'obtenir, en faisant tendre N vers $+\infty$: $\forall x \in [0; R[, S(x) \leq \sum\limits_{n=0}^{+\infty} a_n R^n$, et donc :

$$\lim_{x \to R^-} S(x) \leq \sum_{n=0}^{+\infty} a_n R^n$$

On applique ce résultat sur les séries entières à la série $\sum kp_k t^{k-1}$ avec $R = 1$. On a $S(t) = G'(t)$ et donc :
G' a une limite à gauche en 1 si et seulement si $\sum kp_k$ converge (c'est-à-dire $E(X)$ existe), et dans ce cas :

$$\lim_{t \to 1^-} G'(t) = \sum_{k=0}^{+\infty} kp_k \ (= E(X))$$

On a de la même façon le résultat pour le moment d'ordre 2 avec la série $\sum k(k-1)t^{k-2}$.

Les exemples qui suivent sont les calculs de fonctions génératrices que l'on obtient pour les lois usuelles, ce qui permet de retrouver aisément espérance et variance :

Exemple 4.2 *Soit $X \sim \mathcal{B}(p)$. On a immédiatement :*

$$\forall t \in \mathbb{R}, G(t) = 1 - p + pt$$

D'où les moments : $E(X) = G'(1) = p$ et $E(X^2) = G^{(2)}(1) + E(X) = p$.

Exemple 4.3 *Soit $X \sim \mathcal{B}(n; p)$.*
On a $\forall k \in [\![0; n]\!], p_k = \binom{n}{k} p^k (1-p)^{n-k}$ et $\forall k > n, p_k = 0$. La fonction G est donc définie sur $\mathbb{R}$ par :

$$\begin{aligned} G(t) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k \\ &= (1 - p + pt)^n \ \textit{(formule du binôme)} \end{aligned}$$

On a évidemment $G(1) = 1$. Comme $G'(t) = np(1-p+pt)^{n-1}$, on obtient $G'(1) = np$, et on retrouve la valeur de l'espérance calculée précédemment. En dérivant deux fois, on va pouvoir obtenir la variance :

$$G^{(2)}(t) = n(n-1)p^2(1-p+pt)^{n-2}$$

donc $G^{(2)}(1) = n(n-1)p^2$, d'où la valeur de $E(X^2)$:

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

On obtient enfin $V(X)$:

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= np(1-p) \end{aligned}$$

Exemple 4.4 *Soit* $X \sim \mathcal{G}(p)$.
La série $\sum_{k\geq 1} p(1-p)^{k-1}t^k$ *converge si et seulement si* $|(1-p)t| < 1$, *et :*

$$\begin{aligned} \forall t \in \left]-\frac{1}{1-p};\frac{1}{1-p}\right[, G(t) &= \sum_{k=1}^{+\infty} p(1-p)^{k-1}t^k \\ &= pt\sum_{k=1}^{+\infty}((1-p)t)^{k-1} \\ &= \frac{pt}{1-(1-p)t} \end{aligned}$$

Le rayon $R = \dfrac{1}{1-p}$ *étant strictement supérieur à 1, on en déduit l'existence de moment de tout ordre, et en particulier :*

$$G'(t) = \frac{p}{(1-(1-p)t)^2} \Rightarrow G'(1) = \frac{1}{p} = E(X)$$

et :

$$G^{(2)}(t) = \frac{2(1-p)p}{(1-(1-p)t)^3} \Rightarrow G^{(2)}(1) = \frac{2(1-p)}{p^2}$$

donc $E(X^2) = \dfrac{2(1-p)}{p^2} + \dfrac{1}{p}$, *et* $V(X) = \dfrac{2(1-p)}{p^2} + \dfrac{1}{p} - \dfrac{1}{p^2} = \dfrac{1-p}{p^2}$.

Exemple 4.5 *Soit* $X \sim \mathcal{P}(\lambda)$.
La série $\sum_{k\geq 0} e^{-\lambda}\frac{\lambda^k}{k!}t^k$ *a un rayon de convergence infini, d'où l'existence de moment de tout ordre, et* $\forall t \in \mathbb{R}, G(t) = \sum_{k=0}^{+\infty} e^{-\lambda}\frac{(\lambda t)^k}{k!} = e^{\lambda(t-1)}$. *On en déduit l'espérance :* $G'(t) = \lambda e^{\lambda(t-1)} \Rightarrow G'(1) = \lambda = E(X)$, *et la variance :*

$$\begin{aligned} G^{(2)}(t) = \lambda^2 e^{\lambda(t-1)} &\Rightarrow G^{(2)}(1) = \lambda^2 \\ &\Rightarrow E(X^2) = \lambda^2 + \lambda \\ &\Rightarrow V(X) = \lambda \end{aligned}$$

Théorème 4.3 *La fonction* G *caractérise la loi suivie par* X.

Démonstration : comme une série entière est la somme de sa série de Taylor, on a :

$$\forall k \in \mathbb{N}, p_k = \frac{G^{(k)}(0)}{k!}$$

La donnée de G permet donc d'obtenir $(p_k)_{k\in\mathbb{N}}$ et on connaît ainsi la loi de X.

IV.8 - Énoncé des exercices

Exercice 4.1 *La variable X suit la loi de Poisson de paramètre λ. On pose :* $Y = \dfrac{1}{(X+1)(X+2)}$. *Calculer l'espérance de Y.*

Exercice 4.2 *Un compteur devrait afficher les valeurs d'une variable aléatoire X suivant une loi $\mathcal{B}(n;p)$. Mais lorsque X est nulle, il affiche un nombre au hasard entre 1 et n. Lorsque X est non nulle, il affiche bien X. Soit Y la variable aléatoire prenant pour valeur le nombre affiché par le compteur.*
Déterminer la loi de Y, ainsi que son espérance et sa variance.

Exercice 4.3 *X est une variable aléatoire qui suit une loi de Poisson de paramètre λ.*

1. *On désigne par Y la variable aléatoire prenant pour valeur :*

$$\begin{cases} 0 & \textit{lorsque} \quad X = 2p+1 \ (p \in \mathbb{N}) \\ p & \textit{lorsque} \quad X = 2p \ (p \in \mathbb{N}) \end{cases}$$

   Calculer la probabilité de l'événement « X prend une valeur paire ».
   Déterminer la loi de Y ainsi que son espérance.

2. *On note Z la variable :* $Z = 4\left[\dfrac{X}{2}\right] - 2X + 1$ *([ ] désigne la partie entière).*
   Déterminer la loi de Z, ainsi que son espérance et sa variance.

Exercice 4.4 *Soit X telle que $X(\Omega) = \mathbb{N}$ et :*

$$\forall n \in \mathbb{N}^*, P(X = n) = \frac{2}{n} P(X = n-1)$$

Déterminer la loi de X.

Exercice 4.5 *Loi de Pascal.*
Une urne contient N boules indiscernables au toucher : des blanches en proportion p (donc Np boules blanches) et des noires en proportion $q = 1 - p$. On extrait avec remise n boules. A chaque tirage, on appelle succès l'obtention d'une boule blanche et on s'intéresse au temps d'attente du $r^{\text{ième}}$ succès, c'est-à-dire au rang du tirage pour lequel apparaît la $r^{\text{ième}}$ boule blanche.

1. *Soit $k \in \mathbb{N}^*$. On appelle A l'événement : « on a obtenu $r-1$ succès au cours de $k-1$ tirages ».*
   Déterminer $P(A)$.

2. *En déduire la loi de X_r, v.a.r. égale au rang du tirage où apparaît le $r^{\text{ième}}$ succès.*
   On dit que X_r suit la loi de Pascal.

3. *A l'aide des dérivées de la fonction :*

$$\begin{array}{rcl}]-1;1[& \to & \mathbb{R} \\ x & \mapsto & \displaystyle\sum_{k=0}^{+\infty} x^k \end{array}$$

   vérifier que $\displaystyle\sum_{k \in X(\Omega)} P(X_r = k) = 1$, et calculer, si elle existe, l'espérance de X_r.

Exercice 4.6 *Allumettes de Banach.*
Un fumeur a dans sa poche gauche et dans sa poche droite une boîte contenant N allumettes. Pour allumer sa cigarette, il choisit à chaque fois au hasard de se servir dans l'une de ses poches.

1. *On considère le moment où, pour la première fois, il s'aperçoit que l'une des boîtes est vide. Soit X le nombre d'allumettes restant dans l'autre. Déterminer la loi de X.*

2. *On s'intéresse à l'instant où la première boîte est vide, mais non encore reconnue vide par le fumeur. Soit Y le nombre d'allumettes restant dans l'autre. Déterminer la loi de Y et en déduire que :*

$$\sum_{k=1}^{N} 2^{k+1} \binom{2N-k-1}{N-1} = 2^{2N}$$

Exercice 4.7 *Une personne souhaitant rentrer chez elle a un trousseau de n clefs. Déterminer le nombre moyen d'essais nécessaires dans chacun des deux cas suivants :*

1. *La personne élimine après chaque essai la clef qui n'a pas convenu.*

2. *La personne remet dans le trousseau après chaque essai la clef qui n'a pas convenu.*

IV.9 - Correction des exercices

Exercice 4.1

D'après le théorème de transfert, l'espérance de Y existe si la série

$$\sum_{k\geq 0} \frac{1}{(k+1)(k+2)} P(X=k)$$

converge. C'est le cas puisque $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$, et on a :

$$E(Y) = \sum_{k=0}^{+\infty} \frac{1}{(k+1)(k+2)} e^{-\lambda}\frac{\lambda^k}{k!} = \frac{e^{-\lambda}}{\lambda^2}\sum_{k=0}^{+\infty}\frac{\lambda^{k+2}}{(k+2)!} = \frac{e^{-\lambda}}{\lambda^2}(e^{\lambda}-1-\lambda)$$

Exercice 4.2

Comme $X(\Omega) = [\![0;n]\!]$, on a $Y(\Omega) = [\![1;n]\!]$. Soit $k \in [\![1;n]\!]$. On utilise la formule des probabilités totales :

$$\begin{aligned} P(Y=k) &= \sum_{i=0}^{n} P(Y=k/X=i)P(X=i) \\ &= P(Y=k/X=0)P(X=0) + P(Y=k/X=k)P(X=k) \\ &= \frac{1}{n}(1-p)^n + \binom{n}{k} p^k(1-p)^{n-k} \end{aligned}$$

L'existence de l'espérance et de la variance de Y est claire puisque $Y(\Omega)$ est fini.

$$E(Y) = \sum_{k=1}^{n} kP(Y=k) = \frac{1}{n}(1-p)^n\sum_{k=1}^{n} k + \sum_{k=1}^{n} k\binom{n}{k} p^k(1-p)^{n-k}$$

On reconnaît dans le deuxième terme l'espérance de X, d'où :

$$E(Y) = (1-p)^n \times \frac{n+1}{2} + np$$

Le calcul de $E(Y^2)$ se fait de la même manière (on reconnaît dans le deuxième terme $E(X^2)$), et on obtient :

$$E(Y^2) = (1-p)^n\frac{(n+1)(2n+1)}{6} + np(1-p) + n^2p^2$$

puis on en déduit $V(Y)$.

Exercice 4.3

1. Comme $X \sim \mathcal{P}(\lambda)$:

$$\begin{aligned} P(\text{« X prend une valeur paire »}) &= \sum_{k=0}^{+\infty} P(X = 2k) \\ &= \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^{2k}}{(2k)!} \\ &= e^{-\lambda} \cosh \lambda \end{aligned}$$

On a $Y(\Omega) = \mathbb{N}$, et : $\forall k > 0, P(Y = k) = P(X = 2k) = e^{-\lambda} \dfrac{\lambda^{2k}}{(2k)!}$.
Pour $k = 0$:

$$\begin{aligned} P(Y = 0) &= P(\text{ « X prend une valeur impaire »}) + P(X = 0) \\ &= 1 - e^{-\lambda}(\cosh \lambda) + e^{-\lambda} \end{aligned}$$

L'espérance et la variance existent (vérifier la convergence des séries), et :

$$\begin{aligned} E(Y) &= \sum_{k=1}^{+\infty} kP(Y = k) \\ &= \sum_{k=0}^{+\infty} ke^{-\lambda} \frac{\lambda^{2k}}{(2k)!} \\ &= \frac{\lambda e^{-\lambda}}{2} \sum_{k=1}^{+\infty} \frac{\lambda^{2k-1}}{(2k-1)!} \\ &= \frac{\lambda e^{-\lambda}}{2} \sinh \lambda \end{aligned}$$

puis en faisant apparaître des séries usuelles :

$$V(Y) = \frac{1}{8} \lambda^2 e^{-2\lambda} (\sinh(2\lambda) + 2)$$

2. Si X prend une valeur paire, alors $Z = 1$ et si X prend une valeur impaire, alors $Z = 4\dfrac{X-1}{2} - 2X + 1 = -1$, d'où $Z(\Omega) = \{-1; 1\}$.

$$P(Z = 1) = P(\text{« X prend une valeur paire »}) = e^{-\lambda} \cosh \lambda$$

et $P(Z = -1) = 1 - e^{-\lambda} \cosh \lambda$. L'espérance et la variance sont évidemment bien définies, et :

$$E(Z) = e^{-\lambda} \cosh \lambda - (1 - e^{-\lambda} \cosh \lambda) = e^{-\lambda}(2 \cosh(\lambda) - e^{\lambda}) = e^{-2\lambda}$$

Z^2 est constante égale à 1, donc $E(Z^2) = 1$ et on en déduit la variance $V(Z) = 1 - e^{-4\lambda}$.

Exercice 4.4

Par une récurrence immédiate, on a : $\forall n \geq 1, P(X = n) = \dfrac{2^n}{n!}P(X = 0)$. On calcule $P(X = 0)$:

$$\sum_{n=0}^{+\infty} P(X = n) = P(X = 0) \sum_{n=0}^{+\infty} \frac{2^n}{n!} = P(X = 0)e^2$$

On a donc $P(X = 0) = e^{-2}$.

Conclusion : $\forall n \in \mathbb{N}, P(X = n) = \dfrac{2^n}{n!}e^{-2}$, et on reconnaît une loi de Poisson : $X \sim \mathcal{P}(2)$.

Exercice 4.5

1. On reconnaît le contexte usuel de la loi binomiale. La probabilité d'avoir $r - 1$ succès au cours de $k - 1$ répétitions est donc :

$$P(A) = \binom{k-1}{r-1} p^{r-1}q^{k-r}$$

2. On a bien sûr $X_r(\Omega) = [\![r; +\infty[$. Soit $k \geq r$. On note S_k l'événement « on a obtenu un succès au $k^{ième}$ essai » :

$$P(X_r = k) = P(S_k \cap A) = P(S_k)P(A) = \binom{k-1}{r-1} p^r q^{k-r}$$

3. On vérifie que la somme vaut 1 :

$$\begin{aligned}\sum_{k=r}^{+\infty} \binom{k-1}{r-1} p^r q^{k-r} &= \frac{p^r}{(r-1)!} \underbrace{\sum_{k=r}^{+\infty} \frac{(k-1)!}{(k-r)!} q^{k-r}}_{\text{dérivée d'ordre } (r-1) \text{ de } \sum q^k} \\ &= \frac{p^r}{(r-1)!} \times \frac{(r-1)!}{(1-q)^r} \\ &= 1\end{aligned}$$

L'espérance se calcule de manière analogue, et on obtient :

$$E(X) = \frac{p^r}{(r-1)!} \times \frac{r!}{(1-q)^{r+1}} = \frac{r}{p}$$

Exercice 4.6

On est dans le cas de la répétition de la même expérience aléatoire : mettre la main dans la poche gauche ou dans la poche droite. On pose $\Omega = \{D; G\}^{2N}$, $\mathcal{A} = \mathcal{P}(\Omega)$, et on prend comme probabilité la probabilité produit (pour rendre compte que les répétitions se font de manière « indépendantes »).

1. On a évidemment $X(\Omega) = [\![0; N]\!]$. Soit $k \in [\![0; N]\!]$. S'il a vidé la poche droite :

$$\begin{aligned}&P((\underbrace{D \text{ ou } G; \cdots; D \text{ ou } G}_{N \text{ fois } D \text{ et } N-k \text{ fois } G}; D; D \text{ ou } G; \cdots; D \text{ ou } G))\\ &= \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k} \times \frac{1}{2}\end{aligned}$$

On a bien sûr le même résultat s'il a vidé la poche gauche, d'où :

$$P(X = k) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}$$

2. On a $Y(\Omega) = [\![0; N]\!]$. Soit $k \in [\![0; N]\!]$.
Si c'est la poche droite :

$$\begin{aligned}&P((\underbrace{D \text{ ou } G; \cdots; D \text{ ou } G}_{N-1 \text{ fois } D \text{ et } N-k \text{ fois } G}; D; D \text{ ou } G; \cdots; D \text{ ou } G))\\ &= \binom{2N-k-1}{N-1}\left(\frac{1}{2}\right)^{2N-k-1} \times \frac{1}{2}\end{aligned}$$

Même résultat si c'est la poche gauche, d'où :

$$P(Y = k) = \binom{2N-k-1}{N-1}\left(\frac{1}{2}\right)^{2N-k-1}$$

Pour le calcul de la somme : on a nécessairement $\sum_{k=0}^{N} P(Y = k) = 1$, soit :

$$\sum_{k=0}^{N} \binom{2N-k-1}{N-1} 2^{k+1} \quad = \quad 2^{2N}$$

Exercice 4.7

On note B la bonne clef et $M_1, \cdots, M_{n-1}$ les mauvaises clefs.

1. Ω est l'ensemble des permutations des éléments $B, M_1, \cdots, M_{n-1}$. On prend pour tribu $\mathcal{A} = \mathcal{P}(\Omega)$ et pour probabilité P la probabilité uniforme (on suppose le choix des clefs fait au hasard). Soit X la variable aléatoire qui à une permutation associe le rang d'apparition de B dans cette permutation.
On a évidemment $X(\Omega) = [\![1; n]\!]$. On calcule les poids de probabilité :
$$\forall k \in [\![1; n]\!], P(X = k) = \frac{\operatorname{Card}(X = k)}{\operatorname{Card}(\Omega)} = \frac{(n-1)!}{n!} = \frac{1}{n}$$
Par conséquent, $X \sim \mathcal{U}_{[\![1;n]\!]}$ et donc $E(X) = \dfrac{n+1}{2}$.
2. On pose comme univers $\Omega = \{B; M_1; \cdots; M_{n-1}\}^{\mathbb{N}}$. Soit X la variable égale au rang d'apparition du premier succès : on reconnaît une loi géométrique, de paramètre $\dfrac{1}{n}$ (on suppose qu'à chaque répétition de l'expérience, les n tirages possibles sont équiprobables), et par conséquent $E(X) = n$.

Chapitre V : Vecteurs aléatoires discrets

V.1 - Loi de probabilité

Définition 5.1 *Soit $(\Omega; \mathcal{A}; P)$ un espace probabilisé.*

Un vecteur aléatoire V est un n-uplet $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ où les X_i sont des variables aléatoires réelles (sur $(\Omega; \mathcal{A}; P)$), c'est-à-dire une application :

$$V : \begin{cases} \Omega & \rightarrow & \mathbb{R}^n \\ \omega & \mapsto & \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} \end{cases}$$

V est appelé vecteur aléatoire discret si les X_i sont des variables aléatoires discrètes.

On considère dans la suite le cas $n = 2$.

Exemple 5.1 *On lance deux fois un dé équilibré. On modélise cette expérience en posant $\Omega = [\![1; 6]\!]^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P la probabilité uniforme. Soient X la v.a.r. égale à la somme des 2 lancers et Y la v.a.r. égale au maximum des 2 lancers. L'application*

$$V : \begin{cases} \Omega & \rightarrow & \mathbb{R}^2 \\ \omega = (\omega_1; \omega_2) & \mapsto & \begin{pmatrix} X(\omega) = \omega_1 + \omega_2 \\ Y(\omega) = \max(\omega_1; \omega_2) \end{pmatrix} \end{cases}$$

est un vecteur aléatoire discret.

Propriété 5.1 *On a directement d'après la définition :*

$$V(\Omega) \subset X(\Omega) \times Y(\Omega)$$

et comme le montre l'exemple précédent, cette inclusion peut être stricte ($(3; 6) \notin V(\Omega)$ et $(3; 6) \in X(\Omega) \times Y(\Omega)$).

On note dans toute la suite $X(\Omega) = \{x_i; i \in I\}$ et $Y(\Omega) = \{y_j; j \in J\}$ où I et J sont des parties de $\mathbb{N}$.

La loi de probabilité de V est par définition l'application :

$$\mu : \left\{ \begin{array}{lcl} \mathcal{B}_{\mathbb{R}^2} & \to & [0;1] \\ B & \mapsto & P(V \in B) \end{array} \right.$$

où $\mathcal{B}_{\mathbb{R}^2}$ est la tribu des boréliens de $\mathbb{R}^2$ (engendrée par exemple par les pavés $[a_1; b_1] \times [a_2; b_2]$). Or cette application μ est entièrement déterminée par la donnée de $V(\Omega)$ et, pour tout couple $(x_i; y_j) \in V(\Omega)$, la donnée de $P((X;Y) = (x_i; y_j))$. L'additivité de P permet alors de calculer $P(V \in B)$ pour tout pavé, et donc pour tout borélien de $\mathbb{R}^2$. On prend donc la définition suivante :

Définition 5.2 *La loi de V, ou loi conjointe du couple $(X;Y)$, est la donnée de :*

1. *$V(\Omega)$, ensemble des valeurs possibles de V.*
2. *$P((X;Y) = (x_i; y_j))$ pour tous les couples $(x_i; y_j)$ de $V(\Omega)$*

On note $p_{i,j} = P((X;Y) = (x_i; y_j))$.
En pratique, on donne tous les $p_{i,j}$ en posant bien sûr $p_{i,j} = 0$ lorsque $(x_i; y_j)$ n'appartient pas à $V(\Omega)$.
Ceci permet de donner la loi, lorsque I et J sont finis (et de cardinal raisonnable !), sous la forme d'un tableau à double entrée.

Exemple 5.2 *On reprend le premier exemple, où on lance deux fois un dé équilibré. On a la loi conjointe suivante :*

X/Y	1	2	3	4	5	6
2	$\frac{1}{36}$	0	0	0	0	0
3	0	$\frac{2}{36}$	0	0	0	0
4	0	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0
5	0	0	$\frac{2}{36}$	$\frac{2}{36}$	0	0
6	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0
7	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
8	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
9	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$
10	0	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$
11	0	0	0	0	0	$\frac{2}{36}$
12	0	0	0	0	0	$\frac{1}{36}$

Exemple de calcul :

$$\begin{aligned} P(X=3 \cap Y=2) &= P(\{1;2\} \cup \{2;1\}) \\ &= P(\{1;2\}) + P(\{2;1\}) \\ &= \frac{2}{36} \end{aligned}$$

V.2 - Lois marginales

Définition 5.3 *Ce sont les lois des variables X et Y.*

Elles se déduisent de la loi conjointe :

$$\begin{aligned} (X = x_i) &= \bigcup_{j \in J} ((X = x_i) \cap (Y = y_j)) \text{ (union disjointe)} \\ \Rightarrow P(X = x_i) &= \sum_{j \in J} P((X = x_i) \cap (Y = y_j)) \end{aligned}$$

Soit, en notant $p_{i,.} = P(X = x_i)$: $p_{i,.} = \sum_{j \in J} p_{i,j}$.
(même raisonnement pour la loi de Y).

Exemple 5.3 *Dans l'exemple précédent, on obtient la loi suivante pour Y :*

$Y(\Omega)$	1	2	3	4	5	6
$p_{.,j}$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

V.3 - Lois conditionnelles

On fixe un événement $(X = x_i)$ (rappel : c'est un abus de notation utilisé fréquemment, il s'agit de $X^{-1}(\{x_i\})$, qui est bien un élément de la tribu $\mathcal{A}$, puisque X est une variable aléatoire, et c'est donc bien un événement...), ou plus généralement un événement $(X \in I)$ où I est un intervalle de $\mathbb{R}$, de probabilité non nulle.
On veut déterminer la loi de Y sachant que $(X = x_i)$ est réalisé.
Par définition, c'est l'application :

$$P_{[X=x_i]} : \left\{ \begin{array}{ccc} \mathcal{B}_{\mathbb{R}} & \to & [0;1] \\ B & \mapsto & \dfrac{P(Y \in B \cap X = x_i)}{P(X = x_i)} \end{array} \right.$$

Comme $Y(\Omega) = \{y_j ; j \in J\}$, cette application est déterminée par la donnée de $P_{[X=x_i]}(Y = y_j), j \in J$, d'où la définition :

Définition 5.4 *La loi conditionnelle de Y sachant $(X = x_i)$ est la donnée de :*

$$\left\{ \left(y_j ; \frac{P(Y = y_j \cap X = x_i)}{P(X = x_i)} \right)_{j \in J} \right\}$$

c'est-à-dire, avec les notations usuelles pour les couples de variables discrètes, la donnée de :

$$\left\{ \left(y_j ; \frac{p_{i,j}}{p_{i,.}} \right)_{j \in J} \right\}$$

On a la même définition pour la loi conditionnelle de X sachant $(Y = y_j)$.

Exemple 5.4 *Avec l'exemple précédent, la loi conditionnelle de X sachant $(Y = 3)$ est :*

$X_{[y=3]}(\Omega)$	4	5	6
$P_{[Y=3]}(X = x_i)$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

V.4 - Indépendance de 2 variables aléatoires réelles

On rappelle la définition vue au chapitre II.

Définition 5.5 *Deux v.a.r. X et Y sont indépendantes si :*

$$\forall(B_1;B_2)\in(\mathcal{B}_{\mathbb{R}})^2, P(X\in B_1\cap Y\in B_2)=P(X\in B_1)P(Y\in B_2)$$

Propriété 5.2 *Lorsque X et Y sont discrètes, la définition est équivalente à :*

$$\forall(i;j)\in I\times J, P(X=x_i\cap Y=y_j)=P(X=x_i)P(Y=y_j)$$

Avec les notations usuelles pour les couples de v.a.r. discrètes, il faut et il suffit donc de vérifier que :

$$\forall(i;j)\in I\times J, p_{i,j}=p_{i,.}p_{.,j}$$

Exemple 5.5 *Les variables X et Y de l'exemple précédent ne sont pas indépendantes puisque $P(X=2\cap Y=2)\neq P(X=2)P(Y=2)$.*

V.5 - Espérance du produit de 2 variables aléatoires

Théorème 5.1 *Soient X et Y deux v.a.r. discrètes indépendantes, admettant une espérance. Alors la v.a.r. XY admet une espérance, et :*

$$E(XY)=E(X)E(Y)$$

Démonstration : on vérifie tout d'abord la sommabilité de la famille $(x_iy_jp_{i,j})_{i\in I,j\in J}$. Soient $I'\subset I$, I' fini, et $J'\subset J$, J' fini.

$$\begin{aligned}\sum_{i\in I',j\in J'}|x_iy_jp_{i,j}| &= \sum_{i\in I'}|x_i|p_{i,.}\sum_{j\in J'}|y_j|p_{.,j}\\ &\leq \sum_{i\in I}|x_i|p_{i,.}\sum_{j\in J}|y_j|p_{.,j}\end{aligned}$$

donc pour toutes parties finies I' et J', la somme est majorée par une constante, d'où la sommabilité de la famille. On peut alors effectuer le calcul :

$$\begin{aligned}E(XY) &= \sum_{i,j}x_iy_jp_{i,j}\\ &= \sum_{i,j}x_iy_jp_{i,.}p_{.,j}\\ &= \left(\sum_i x_ip_{i,.}\right)\left(\sum_j y_jp_{.,j}\right)\\ &= E(X)E(Y)\end{aligned}$$

Remarque : on a $P(XY = x_i y_j) \geq P(X = x_i \cap Y = y_j)$ et l'inégalité peut être stricte. Sur l'exemple traité tout au long de ce chapitre, on a $P(XY = 2 \times 6) = P(X = 2 \cap Y = 6) + P(X = 3 \cap Y = 4)$. Mais comme on somme sur tous les couples de $I \times J$, on calcule bien $E(XY)$.

Attention : la réciproque de ce théorème 5.1 est fausse : l'égalité $E(XY) = E(X)E(Y)$ n'implique pas l'indépendance de X et Y (voir exercice 5.6).

V.6 - Indépendance de n variables aléatoires

On rappelle la définition vue au chapitre II :

Définition 5.6 *Soient X_1, X_2,..., X_n n v.a.r. (définies sur le même espace). Elles sont indépendantes si pour tout ensemble d'indices I inclus dans $[\![1;n]\!]$, on a, pour toute famille de boréliens $(B_i)_{i\in I}$:*

$$P\left(\bigcap_{i\in I}(X_i \in B_i)\right) = \prod_{i\in I} P(X_i \in B_i)$$

Propriété 5.3 *Lorsque les n v.a.r. sont discrètes, il est équivalent de vérifier que c'est vrai sur leur support, c'est-à-dire : pour tout ensemble d'indices I inclus dans $[\![1;n]\!]$, on a, pour toute famille $(B_i)_{i\in I}$ de $(X_i(\Omega))_{i\in I}$*

$$P\left(\bigcap_{i\in I}(X_i \in B_i)\right) = \prod_{i\in I} P(X_i \in B_i)$$

Propriété 5.4 *L'indépendance de n v.a.r. entraîne leur indépendance 2 à 2, et la réciproque est fausse.*

Démonstration : on suppose que X_1,...,X_n sont indépendantes. En appliquant la définition avec $B_3 = B_4 = ... = B_n = \mathbb{R}$, on a :

$$\begin{aligned}\forall (B_1; B_2) \in (\mathcal{B}_{\mathbb{R}})^2, & P(X_1 \in B_1 \cap X_2 \in B_2 \cap X_3 \in \mathbb{R} \cap ... \cap X_n \in \mathbb{R}) \\ &= P(x_1 \in B_1)P(X_2 \in B_2)P(X_3 \in \mathbb{R})...P(X_n \in \mathbb{R})\end{aligned}$$

soit : $P(X_1 \in B_1 \cap X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2)$, et X_1 et X_2 sont donc indépendantes.

Pour le fait que la réciproque est fausse, voir l'exercice 5.5.

Le programme officiel admet le résultat suivant :

Propriété 5.5 *Si X_1,...,X_n sont indépendantes, alors toute fonction de X_1,...,X_p est indépendante de toute fonction de X_{p+1},...,X_n.*

Par exemple, si X et Y sont indépendantes, alors X^2 et Y sont indépendantes.
On peut aussi le voir directement en revenant à la définition de tribu engendrée par une variable aléatoire (définition 2.2) : on remarque que X et Y sont indépendantes si et seulement si tout événement de $\mathcal{A}_X$ est indépendant de tout événement de $\mathcal{A}_Y$. Or $\mathcal{A}_{X^2} \subset \mathcal{A}_X$: en effet, pour tout intervalle $]a;b[$, $(X^2)^{-1}\Big(]a;b[\Big)$ est égal à l'ensemble vide si a et b sont négatifs, à $X^{-1}\Big(]-\sqrt{b};\sqrt{b}[\Big)$ si a est négatif et b positif, et à $X^{-1}\Big(]-\sqrt{b};-\sqrt{a}[\Big) \cup X^{-1}\Big(]\sqrt{a};\sqrt{b}[\Big)$ si a et b sont positifs, donc dans tous les cas, $(X^2)^{-1}\Big(]a;b[\Big) \in \mathcal{A}_X$. Comme les intervalles $]a;b[$ engendrent $\mathcal{B}_{\mathbb{R}}$, on a $\mathcal{A}_{X^2} \subset \mathcal{A}_X$, et donc tout événement de $\mathcal{A}_{X^2}$ est indépendant de tout événement de $\mathcal{A}_Y$.

V.7 - Somme de n variables aléatoires

On considère l'ensemble des v.a.r. discrètes définies sur le même espace $(\Omega;\mathcal{A};P)$, noté F.
Muni de la loi $+$ (addition usuelles de 2 fonctions : $(X+Y)(\omega)$ est égal par définition à $X(\omega)+Y(\omega)$) et de la loi . (loi de composition externe classique : $\forall\lambda \in \mathbb{R}, \lambda.X$ est la fonction définie par : $(\lambda.X)(\omega) = \lambda X(\omega)$), $(F;+;.)$ est un espace vectoriel.
On détaille ici les propriétés des opérateurs E et V, ainsi que celles de la covariance.

Propriété 5.6 *Soit F_1 l'ensemble des éléments de F qui possèdent une espérance :*

1. *F_1 est un sous-espace vectoriel de F.*
2. *E est une forme linéaire sur F_1.*

Ces résultats ont été démontrés au paragraphe 4 du chapitre III (il y avait un cas simple, lorsque Ω est fini ou dénombrable, et le cas général a été traité en écrivant les variables aléatoires comme combinaison de fonctions indicatrices...).

Propriété 5.7 *Soit F_2 l'ensemble des éléments de F qui possèdent une variance. F_2 est un sous-espace vectoriel de F inclus dans F_1.*

Démonstration : soient X et Y deux éléments de F_2.
Il faut montrer que $X+Y$ possède une variance, ce qui équivaut à montrer que $X+Y$ possède un moment d'ordre 2, autrement dit que $(X+Y)^2 = X^2+Y^2+2XY$ possède une espérance. Comme X et Y possèdent une variance, elles possèdent un moment d'ordre 2 : $E(X^2)$ et $E(Y^2)$ existent.

On montre que cela implique que $E(XY)$ existe (en cas de difficulté, c'est l'objet de l'exercice 5.1). On a donc X^2, Y^2 et XY qui possèdent une espérance. Comme F_1 est un sous-espace vectoriel, X^2+Y^2+2XY possède une espérance, et donc $(X+Y)^2$ possède un moment d'ordre 2.
On montre enfin que $F_2 \subset F_1$: on prouve que si X possède un moment d'ordre 2, alors X possède un moment d'ordre 1 (voir l'exercice 3.5).

Définition 5.7 *L'application :*

$$\mathrm{Cov} : \begin{cases} (F_2)^2 & \to \mathbb{R} \\ (X;Y) & \mapsto E(XY) - E(X)E(Y) \end{cases}$$

est bien définie, et est appelée covariance de X et Y.

(On vient de montrer que si X et Y sont deux éléments de F_2, alors XY possède une espérance).

Propriété 5.8 $\forall (X_1;...;X_n) \in (F_2)^n$,

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

Démonstration : procéder par récurrence sur n, en utilisant la linéarité de E.

Un cas particulier très important est celui où les variables X_1,...,X_n sont indépendantes deux à deux. On a alors :

$$\forall i \neq j, E(X_iX_j) = E(X_i)E(X_j)$$

Autrement dit : $\forall i \neq j, \mathrm{Cov}(X_i;X_j) = 0$. D'où la formule suivante :

Propriété 5.9 *Soient X_1,...,X_n des variables appartenant à F_2 qui sont indépendantes deux à deux. Alors :*

$$V\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} V(X_i)$$

Par linéarité de E, on obtient directement une autre expression de la covariance :

Propriété 5.10 $\forall (X;Y) \in (F_2)^2$,

$$\mathrm{Cov}(X;Y) = E\Big((X - E(X))(Y - E(Y))\Big)$$

La covariance est donc l'espérance du produit des deux variables préalablement « centrées ».

V.8 - Covariance et coefficient de corrélation linéaire

On donne ici quelques propriétés d'algèbre bilinéaire, qui seront notamment reprises pour « éclairer » la leçon sur les séries statistiques à deux variables.

Théorème 5.2 *L'application* Cov *est une forme bilinéaire symétrique positive sur l'espace vectoriel* F_2.

Démonstration :

1. Symétrie : on a immédiatement $\mathrm{Cov}(X;Y) = \mathrm{Cov}(Y;X)$.
2. Bilinéarité : il suffit d'utiliser la linéarité de E.
3. Positivité : $\forall X \in F_2, \mathrm{Cov}(X;X) = E\Big((X - E(X))^2\Big) = V(X) \geq 0$.

La covariance n'est pas définie-positive puisqu'on a vu que :

$$V(X) = 0 \Rightarrow X \text{ est constante (presque partout)}$$

On n'a donc pas un produit scalaire sur F_2, et σ n'est donc pas une norme. Si on se restreint au sous-espace vectoriel de F_2 formé par les variables centrées, alors la covariance est un produit scalaire : si $V(X) = 0$, alors X est constante, et comme son espérance est nulle, elle est nulle.

Définition 5.8 *Soient X et Y deux éléments de F_2, de variance strictement positive. On appelle coefficient de corrélation linéaire le réel :*

$$\rho(X;Y) = \frac{\mathrm{Cov}(X;Y)}{\sigma(X)\sigma(Y)}$$

On précisera son interprétation lors du commentaire de la leçon sur les séries statistiques doubles.

Propriété 5.11

$$\forall (X;Y) \in (F_2)^2, |\rho(X;Y)| \leq 1$$

Démonstration : on doit démontrer que $|\mathrm{Cov}(X;Y)| \leq \sigma(X)\sigma(Y)$.
Si on avait un produit scalaire, il s'agirait donc de montrer que la valeur absolue du produit scalaire est inférieure ou égale au produit des normes, ce qui est l'inégalité de Cauchy-Schwarz, que vous devez savoir impérativement démontrer !

On va utiliser la même démonstration ici (elle ne nécessite pas la définie-positivité) : soit $\lambda \in \mathbb{R}$.

$$V(X+\lambda Y) = V(X) + \lambda^2 V(Y) + 2\lambda \operatorname{Cov}(X;Y)$$

Ce trinôme en λ est toujours positif ou nul ($V(X+\lambda Y) \geq 0$), donc son discriminant est négatif ou nul, et comme $\Delta = 4(\operatorname{Cov}(X;Y))^2 - 4V(X)V(Y)$, on en déduit que $(\operatorname{Cov}(X;Y))^2 \leq V(X)V(Y)$.

Remarque : si on centre X et Y, c'est-à-dire si on considère $X^* = X - E(X)$ et $Y^* = Y - E(Y)$, le coefficient de corrélation est inchangé :

$$\rho(X;Y) = \rho(X^*;Y^*)$$

En effet $V(X^*) = V(X - E(X)) = V(X)$ et :

$$\begin{aligned}
\operatorname{Cov}(X^*;Y^*) &= E\Big((X^* - E(X^*))(Y^* - E(Y^*))\Big) \\
&= E(X^*Y^*) \\
&= E\Big((X - E(X))(Y - E(Y))\Big) \\
&= \operatorname{Cov}(X;Y)
\end{aligned}$$

V.9 - Somme de variables de Bernoulli indépendantes

On considère une expérience modélisée par un espace probabilisé $(\Omega; \mathcal{A}; P)$. On s'intéresse à un événement que l'on appellera « succès » noté S, et on note p sa probabilité. On répète n fois cette expérience, et on se place sur $(\Omega^n; \mathcal{A}^{\otimes n})$ que l'on munit de la probabilité produit $P^{\otimes n}$ (ce qui permet de traduire que les répétitions de l'épreuve sont « indépendantes »).
On définit, pour $i \in [\![1;n]\!]$:

$$X_i : \left\{ \begin{array}{ccc} \Omega^n & \to & \{0;1\} \\ (\omega_1;...;\omega_n) & \mapsto & 1 \text{ si } \omega_i \in S, \ 0 \text{ sinon} \end{array} \right.$$

X_i est donc l'indicatrice du succès lors de la $i^{ème}$ épreuve. On a $X_i(\Omega^n) = \{0;1\}$ et :

$$\begin{aligned}
& P^{\otimes n}(X_i = 1) \\
= \ & P^{\otimes n}((\omega_1;...;\omega_{i-1};\omega_i;\omega_{i+1};...;\omega_n) \in (\Omega;...;\Omega;S;\Omega;...;\Omega)) \\
= \ & P(\omega_1 \in \Omega)...P(\omega_{i-1} \in \Omega)P(\omega_i \in S)P(\omega_{i+1} \in \Omega)...P(\omega_n \in \Omega) \\
= \ & 1...1 \times p \times 1...1 = p
\end{aligned}$$

Par conséquent : $X_i \sim \mathcal{B}(p)$.
Les variables X_1,...,X_n sont indépendantes, et on a : $\forall(\alpha_1;...;\alpha_n) \in \{0;1\}^n$,

$$\begin{aligned}
P^{\otimes n}((X_1;...;X_n) = (\alpha_1;...;\alpha_n)) &= P^{\otimes n}(X_1 = \alpha_1)...P^{\otimes n}(X_n = \alpha_n) \\
&= p^k(1-p)^{n-k}
\end{aligned}$$

où k est le nombre de α_i égaux à 1.

On définit $Z = \sum_{i=1}^{n} X_i$, v.a.r. égale au nombre de succès obtenus lors des n réalisations. On a : $Z(\Omega^n) = [\![0;n]\!]$, et :

$$\forall k \in [\![0;n]\!], P^{\otimes n}(Z = k) = p^k(1-p)^{n-k}\binom{n}{k}$$

car il y a $\binom{n}{k}$ n-uplets $(\alpha_1;...;\alpha_n)$ distincts qui comportent k fois 1 et $n-k$ fois 0.
Conclusion : $Z \sim \mathcal{B}(n;p)$, d'où le théorème :

Théorème 5.3 *Soient X_1,...,X_n des v.a.r. définies sur le même espace, indépendantes et qui suivent toutes une loi de Bernoulli de même paramètre p.*
Alors leur somme $Z = \sum_{i=1}^{n} X_i$ suit une loi binomiale $\mathcal{B}(n;p)$.

On retrouve alors facilement l'espérance et la variance pour la loi $\mathcal{B}(n;p)$:

$$E(Z) = E\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} E(X_i) = np$$

Pour la variance, on développe en utilisant l'indépendance des X_i :

$$V(Z) = V\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} V(X_i) = np(1-p)$$

V.10 - Somme de variables binomiales indépendantes

On a la stabilité de la loi binomiale pour la somme au sens suivant :

Théorème 5.4 *Soient X_1,...,X_n des v.a.r. définies sur le même espace, indépendantes et telles que $\forall i \in [\![1;n]\!], X_i \sim \mathcal{B}(n_i;p)$. Alors :*

$$\sum_{i=1}^{n} X_i \sim \mathcal{B}\Big(\sum_{i=1}^{n} n_i;p\Big)$$

Démonstration : on traite le cas $n = 2$ (la démonstration se fait par récurrence en utilisant la même idée).
Comme $X_1 \sim \mathcal{B}(n_1;p)$, elle peut s'écrire comme une somme de n_1 variables

de Bernoulli indépendantes : $X_1 = \sum_{i=1}^{n_1} Y_i$ où $\forall i \in [\![1;n_1]\!], Y_i \sim \mathcal{B}(p)$ et les Y_i sont indépendantes.

On décompose de la même façon X_2 : $X_2 = \sum_{i=n_1+1}^{n_1+n_2} Y_i$ où $\forall i \in [\![n_1+1;n_1+n_2]\!], Y_i \sim \mathcal{B}(p)$ et les Y_i sont indépendantes.

Alors $X_1 + X_2 = \sum_{i=1}^{n_1+n_2} Y_i$ où les $n_1 + n_2$ variables Y_i suivent une loi $\mathcal{B}(p)$ et sont indépendantes. On a donc bien : $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2; p)$.

V.11 - Somme de variables de Poisson indépendantes

On a là aussi une stabilité pour la somme :

Théorème 5.5 *Soient $X_1, \ldots, X_n$ des v.a.r. définies sur le même espace, indépendantes et telles que $\forall i \in [\![1;n]\!], X_i \sim \mathcal{P}(\lambda_i)$. Alors :*

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\Big(\sum_{i=1}^{n} \lambda_i\Big)$$

Démonstration : on traite le cas $n = 2$ (faire la démonstration par récurrence, les calculs se mènent exactement de la même façon). Soient X_1 une v.a.r. qui suit une loi $\mathcal{P}(\lambda_1)$ et X_2 qui suit une loi $\mathcal{P}(\lambda_2)$, avec X_1 et X_2 indépendantes.

On a d'une part $(X_1 + X_2)(\Omega) = \mathbb{N}$, et d'autre part, pour $k \in \mathbb{N}$:

$$\begin{aligned}
P(X_1 + X_2 = k) &= \sum_{i=0}^{k} P(X_1 = i \cap X_2 = k - i) \\
&= \sum_{i=0}^{k} P(X_1 = i) P(X_2 = k - i) \\
&\quad \text{(car } X_1 \text{ et } X_2 \text{ sont indépendantes)} \\
&= \sum_{i=0}^{k} e^{-\lambda_1} \frac{\lambda_1^i}{i!} e^{-\lambda_2} \frac{\lambda_2^{k-i}}{(k-i)!} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^{k} \frac{\begin{pmatrix} k \\ i \end{pmatrix}}{k!} \lambda_1^i \lambda_2^{k-i} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}
\end{aligned}$$

V.12 - Énoncé des exercices

Exercice 5.1 *Soient X et Y deux v.a.r. discrètes admettant une variance.*

1. *Montrer qu'alors XY admet une espérance et que :*

$$E(|XY|) \leq \frac{1}{2}\Big(E(X^2) + E(Y^2)\Big)$$

2. *Montrer que : $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$ (penser à la démonstration de l'inégalité de Cauchy-Schwarz).*

Exercice 5.2 *On lance deux fois une pièce de monnaie équilibrée. Soient X la variable prenant pour valeur le nombre de Pile obtenus moins un, et Y la variable prenant pour valeur le nombre de Pile au deuxième lancer moins le nombre de Pile au premier lancer.*

1. *Déterminer la loi conjointe du couple $(X;Y)$.*
2. *Calculer la covariance du couple.*
3. *Les variables X et Y sont-elles indépendantes ?*

Exercice 5.3 *Soit $(X;Y)$ un couple de variables aléatoires discrètes dont la loi est donnée par le tableau suivant :*

$X\backslash Y$	1	2	3	4
1	0.08	0.04	0.16	0.12
2	0.04	0.02	0.08	0.06
3	0.08	0.04	0.16	0.12

1. *Déterminer les lois marginales du couple et préciser si X et Y sont indépendantes.*
2. *Calculer* $\mathrm{Cov}(X;Y)$.
3. *Déterminer la loi du couple* $(\min(X;Y);\max(X;Y))$.

Exercice 5.4 *On lance trois fois une pièce de monnaie équilibrée; X et Y désignent respectivement le nombre de Face apparues lors des deux premiers lancers et le nombre de Pile apparus lors des deux derniers.*

1. *Déterminer la loi du couple $(X;Y)$ puis les lois marginales de X et Y.*
2. *X et Y sont-elles indépendantes ?*
3. *Calculer* $\mathrm{Cov}(X;Y)$.

Exercice 5.5 *Soient X et Y deux variables indépendantes de support* $\{-1;1\}$ *et distribuées uniformément (poids de 0.5 sur -1 et sur 1).*
On considère $Z = XY$: les variables X, Y et Z sont-elles indépendantes 2 à 2 ? Sont-elles mutuellement indépendantes ?

Exercice 5.6 *Soit X de support $\{-a; -b; b; a\}$ $(0 < b < a)$ distribuée uniformément.*

1. *Donner la loi de X^2 et la loi conjointe du couple $(X; X^2)$.*
2. *Calculer* $\mathrm{Cov}(X; X^2)$.
3. *Les variables X et X^2 sont-elles indépendantes ?*

Exercice 5.7 *Soient X et Y suivant toutes les deux une loi de Bernoulli $\mathcal{B}(p)$. Montrer que :*

$$X \text{ et } Y \text{ sont indépendantes} \Leftrightarrow \mathrm{Cov}(X; Y) = 0$$

Exercice 5.8 *Soit X de loi $\mathcal{B}(n; p)$ et soit Y une v.a.r. à valeurs dans* $\mathbb{N}$ *telle que la loi conditionnelle de Y sachant $(X = k)$ est la loi $\mathcal{B}(k; q)$.*

1. *Montrer que :*

$$\forall \alpha \in [\![0; n]\!], \forall i \in [\![\alpha; n]\!], \binom{i}{\alpha}\binom{n}{i} = \binom{n}{\alpha}\binom{n-\alpha}{i-\alpha}$$

2. *Déterminer la loi de Y.*

Exercice 5.9 *Une particule se déplace sur une droite graduée. A l'instant zéro, la particule est en zéro. A l'issue de chaque instant, elle s'est déplacée d'une unité dans le sens positif avec la probabilité p, ou dans le sens négatif avec la probabilité $1-p$. On note X_n la variable aléatoire prenant pour valeur l'abscisse de la particule à l'issue de l'instant n.*

1. *Déterminer $P(X_n = 0)$.*
2. *Calculer l'espérance et la variance de X_n.*

Indication : on pourra décomposer X_n comme une somme de v.a.r. qui prennent les valeurs +1 et -1.

Exercice 5.10 *Soient A et B deux événements indépendants.*
Les indicatrices $\mathbf{1}_A$ et $\mathbf{1}_B$ sont-elles indépendantes ? Réciproque ?

V.13 - Correction des exercices

Exercice 5.1

1. On note $X(\Omega) = \{x_i; i \in I\}$ et $Y(\Omega) = \{y_j; j \in J\}$, où I et J sont des parties de $\mathbb{N}$. Pour montrer l'existence de $E(XY)$, il faut montrer la sommabilité de la famille $\left(x_i y_j p_{i;j}\right)$. On a l'inégalité :

$$|x_i y_j| \leq \frac{1}{2}\left(x_i^2 + y_j^2\right)$$

Soient I' une partie finie incluse dans I et J' une partie finie incluse dans J :

$$\begin{aligned}
\sum_{i\in I';j\in J'} |x_i y_j| p_{i;j} &\leq \frac{1}{2}\sum_{i\in I';j\in J'} x_i^2 p_{i;j} + \frac{1}{2}\sum_{i\in I';j\in J'} y_j^2 p_{i;j} \\
&\leq \frac{1}{2}\sum_{i\in I'} x_i^2 \sum_{j\in J} p_{i;j} + \frac{1}{2}\sum_{j\in J'} y_j^2 \sum_{i\in I} p_{i;j} \\
&\leq \frac{1}{2}\sum_{i\in I'} x_i^2 p_{i;.} + \frac{1}{2}\sum_{j\in J'} y_j^2 p_{.;j} \\
&\leq \frac{1}{2}\sum_{i\in I} x_i^2 p_{i;.} + \frac{1}{2}\sum_{j\in J} y_j^2 p_{.;j} \\
&\leq \frac{1}{2}\Big(E(X^2) + E(Y^2)\Big)
\end{aligned}$$

La famille est donc bien sommable, et comme l'inégalité est vraie pour toutes les parties finies I' et J', on a :

$$\begin{aligned}
\sum_{i\in I;j\in J} |x_i y_j| p_{i;j} &\leq \frac{1}{2}\Big(E(X^2) + E(Y^2)\Big) \\
\Leftrightarrow E(|XY|) &\leq \frac{1}{2}\Big(E(X^2) + E(Y^2)\Big)
\end{aligned}$$

2. On développe $E((X + \lambda Y)^2)$:

$$E((X + \lambda Y)^2) = E(X^2) + \lambda^2 E(Y^2) + 2\lambda E(XY)$$

On obtient un polynôme de degré 2 en λ (sauf si $E(Y^2) = 0$, auquel cas Y est nulle et l'inégalité est évidente), qui est toujours positif ou nul, donc son discriminant est négatif ou nul :

$$\Delta = (E(XY))^2 - 4E(X^2)E(Y^2) \leq 0$$

D'où :

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

Exercice 5.2

On pose $\Omega = \{P; F\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P la probabilité uniforme.

1. On a les valeurs de X et Y pour chaque événement élémentaire :

ω	$X(\omega)$	$Y(\omega)$
$(P;P)$	1	0
$(P;F)$	0	-1
$(F;P)$	0	1
$(F;F)$	-1	0

D'où la loi du couple $(X;Y)$:

$X\backslash Y$	-1	0	1
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

2. On obtient par un calcul immédiat :

$$\mathrm{Cov}\,(X;Y) = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i y_j p_{i;j} = 0$$

3. Les variables X et Y ne sont pas indépendantes car

$$P(X=-1 \cap Y=-1) = 0 \text{ et } P(X=-1)P(Y=-1) = \frac{1}{16}$$

Exercice 5.3

1. On obtient immédiatement les lois marginales :

$X(\Omega)$	1	2	3
$p_{i;.}$	0.4	0.2	0.4

et

$Y(\Omega)$	1	2	3	4
$p_{.;j}$	0.2	0.1	0.4	0.3

X et Y sont indépendantes (il faut effectuer les 12 vérifications).

2. Par conséquent :

$$\mathrm{Cov}\,(X;Y) = 0$$

3. On obtient facilement :

$\min(X;Y)\backslash\max(X;Y)$	1	2	3	4
1	0.08	0.08	0.24	0.12
2	0	0.02	0.12	0.06
3	0	0	0.16	0.12

Exemple de calcul :

$$\begin{aligned}
& P(\min(X;Y) = 1 \cap \max(X;Y) = 3) \\
= \; & P((X=1 \cap Y=3) \cup (X=3 \cap Y=1)) \\
= \; & P((X=1 \cap Y=3) + P((X=3 \cap Y=1)) \\
= \; & 0.16 + 0.08 \\
= \; & 0.24
\end{aligned}$$

Exercice 5.4

On pose $\Omega = \{P;F\}^3$, $\mathcal{A} = \mathcal{P}(\Omega)$ et P la probabilité uniforme.

1. On a les valeurs de X et Y pour chaque événement élémentaire :

ω	$X(\omega)$	$Y(\omega)$
$(P;P;P)$	0	2
$(P;P;F)$	0	1
$(P;F;P)$	1	1
$(P;F;F)$	1	0
$(F;P;P)$	1	2
$(F;P;F)$	1	1
$(F;F;P)$	2	1
$(F;F;F)$	2	0

D'où la loi du couple $(X;Y)$ et les lois marginales :

$X\backslash Y$	0	1	2	$p_{i;.}$
0	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
2	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
$p_{.;j}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

2. On constate que les variables ne sont pas indépendantes, puisque :

$$P(X = 2 \cap Y = 2) \neq P(X = 2)P(Y = 2)$$

3. Par le calcul, $\operatorname{Cov}(X;Y) = -\frac{1}{4}$.

Exercice 5.5

D'après l'énoncé, X et Y ont pour loi :

$X(\Omega)$	-1	1
	0.5	0.5

et :

$Y(\Omega)$	-1	1
	0.5	0.5

1. L'hypothèse d'indépendance de X et Y permet d'obtenir la loi de Z : on a bien sûr $Z(\Omega) = \{-1;1\}$, et :

$$\begin{aligned} P(Z=1) &= P((X=1 \cap Y=1) \cup (X=-1 \cap Y=-1)) \\ &= P(X=1 \cap Y=1) + P(X=-1 \cap Y=-1) \\ &= P(X=1)P(Y=1) + P(X=-1)P(Y=-1) \\ &= \frac{1}{2} \end{aligned}$$

On obtient de la même façon $P(Z=-1)=\dfrac{1}{2}$.

2. On détermine la loi du couple $(X;Z)$:

$X\backslash Z$	-1	1
-1	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$

Exemple de calcul :

$$\begin{aligned} P(X=-1\cap Z=-1) &= P(X=-1\cap XY=-1)\\ &= P(X=-1\cap Y=1)\\ &= P(X=-1)P(Y=1)\\ &= \frac{1}{4}\end{aligned}$$

L'examen de la loi du couple $(X;Z)$ et des lois de X et de Z permet d'en déduire que X et Z sont indépendantes (4 vérifications). Les calculs sont identiques pour la loi de $(Y;Z)$, et on conclut également qu'elles sont indépendantes. Les variables X, Y et Z sont donc indépendantes deux à deux.
Pour l'indépendance mutuelle, il faudrait en plus que X, Y et Z soient indépendantes, ce qui n'est pas le cas puisque :

$$\begin{aligned} P(X=1\cap Y=1\cap Z=-1) &= 0\\ \text{et } P(X=1)P(Y=1)P(Z=-1) &= \frac{1}{8}\end{aligned}$$

Exercice 5.6
La loi de X est d'après l'énoncé :

$X(\Omega)$	$-a$	$-b$	b	a
	0.25	0.25	0.25	0.25

1. On en déduit la loi de X^2 :

$X^2(\Omega)$	a^2	b^2
	0.5	0.5

D'où la loi du couple $(X;X^2)$:

$X\backslash X^2$	a^2	b^2
$-a$	0.25	0
$-b$	0	0.25
b	0	0.25
a	0.25	0

Exemple de calcul :

$$\begin{aligned}&P(X = a \cap X^2 = a^2)\\&= P(X = a \cap X = -a) + P(X = a \cap X = a) = \frac{1}{4}\end{aligned}$$

2. On obtient la covariance par le calcul, ou directement :

$$\text{Cov}\,(X; X^2) = E(X^3) - E(X)E(X^2) = 0$$

puisque $E(X) = 0$ et $E(X^3) = 0$ (sans calcul, ces deux variables ayant une distribution parfaitement équilibrée autour de 0).

3. Les variables X et X^2 ne sont pas indépendantes :

$$P(X = a \cap X^2 = b) = 0 \text{ et } P(X = a)P(X^2 = b^2) \neq 0$$

Remarque : on a donc deux variables qui ont une covariance nulle et qui ne sont pas indépendantes. La réciproque du théorème 5.1 est donc fausse. Dans l'exercice qui suit, on montre que l'équivalence entre indépendance et nullité de la covariance est vraie pour des variables de Bernoulli.

Exercice 5.7

On a démontré dans le cours que :

$$X \text{ et } Y \text{ indépendantes} \Rightarrow \text{Cov}\,(X; Y) = 0$$

Réciproque : on suppose que $X \rightsquigarrow \mathcal{B}(p)$ et $Y \rightsquigarrow \mathcal{B}(p)$, et que ces deux variables ont une covariance nulle, c'est-à-dire que $E(XY) = E(X)E(Y)$. Dans l'expression de $E(XY)$, il n'y a qu'un terme non nul :

$$E(XY) = \sum_{i=0}^{1}\sum_{j=0}^{1} ijP(X = i \cap Y = j) = P(X = 1 \cap Y = 1)$$

d'où $P(X = 1 \cap Y = 1) = E(X)E(Y) = p^2$. On en déduit :

$$P(X = 1 \cap Y = 0) = P(X = 1) - P(X = 1 \cap Y = 1) = p - p^2$$

puis de la même façon les deux derniers poids de probabilité, ce qui donne comme loi conjointe :

$X\backslash Y$	0	1	
0	$1 - 2p + p^2$	$p - p^2$	$1 - p$
1	$p - p^2$	p^2	p
	$1 - p$	p	1

On vérifie bien que les deux variables sont indépendantes (faire les quatre produits).

Exercice 5.8

1. C'est immédiat par le calcul :

$$\begin{aligned}
\binom{i}{\alpha}\binom{n}{i} &= \frac{i!}{\alpha!(i-\alpha)!}\frac{n!}{i!(n-i)!} \\
&= \frac{n!}{\alpha!(n-\alpha)!}\frac{(n-\alpha)!}{(i-\alpha)!(n-i)!} \\
&= \binom{n}{\alpha}\binom{n-\alpha}{i-\alpha}
\end{aligned}$$

2. Soit $\alpha \in \mathbb{N}$. On développe avec la formule des probabilités totales :

$$\begin{aligned}
P(Y=\alpha) &= \sum_{i=0}^{n} P(Y=\alpha \cap X=i) \\
&= \sum_{i=0}^{n} P_{[X=i]}(Y=\alpha)P(X=i)
\end{aligned}$$

Or la loi de Y sachant $(X=i)$ est la loi $\mathcal{B}(i;q)$, donc :

$$P_{[X=i]}(Y=\alpha) = \begin{cases} 0 & si \quad \alpha > i \\ \binom{i}{\alpha} q^{\alpha}(1-q)^{i-\alpha} & si \quad 0 \leq \alpha \leq i \end{cases}$$

Donc, si $\alpha > n$, $P(Y=\alpha)=0$ et si $\alpha \leq n$:

$$\begin{aligned}
P(Y=\alpha) &= \sum_{i=\alpha}^{n} \binom{i}{\alpha} q^{\alpha}(1-q)^{i-\alpha}\binom{n}{i} p^{i}(1-p)^{n-i} \\
&= \binom{n}{\alpha} q^{\alpha}\sum_{i=\alpha}^{n}\binom{n-\alpha}{i-\alpha}(1-q)^{i-\alpha}p^{i}(1-p)^{n-i}
\end{aligned}$$

En posant $j=i-\alpha$, il vient :

$$\begin{aligned}
P(Y=\alpha) &= \binom{n}{\alpha} q^{\alpha}\sum_{j=0}^{n-\alpha}\binom{n-\alpha}{j}(1-q)^{j}p^{\alpha+j}(1-p)^{n-j-\alpha} \\
&= \binom{n}{\alpha}(qp)^{\alpha}\underbrace{\sum_{j=0}^{n-\alpha}\binom{n-\alpha}{j}(p-qp)^{j}(1-p)^{n-j-\alpha}}_{(p-qp+1-p)^{n-\alpha}} \\
&= \binom{n}{\alpha}(qp)^{\alpha}(1-qp)^{n-\alpha}
\end{aligned}$$

Conclusion : $Y \sim \mathcal{B}(n; qp)$.

Exercice 5.9

1. Pour être en 0 à l'issue de l'instant n, la particule doit avoir fait autant de sauts vers la droite que vers la gauche. Donc si n est impair, $P(X_n = 0) = 0$, et si $n = 2k$:

$$P(X_{2k} = 0) = \left(\begin{array}{c} 2k \\ k \end{array} \right) p^k(1-p)^{n-k}$$

2. $X_n = \sum_{i=1}^{n} Y_i$ où les variables Y_i sont indépendantes de loi :

$Y_i(\Omega)$	-1	1
	$1-p$	p

Donc $E(Y_i) = 2p - 1$ et $V(Y_i) = 4p(1-p)$. Par conséquent :

$$E(X_n) = \sum_{i=1}^{n} E(Y_i) = n(2p-1) \text{ et } V(X_n) = \sum_{i=1}^{n} V(Y_i) = 4np(1-p)$$

Exercice 5.10

Les variables $\mathbf{1}_A$ et $\mathbf{1}_B$ ont pour loi :

$\mathbf{1}_A$	0	1
	$1-P(A)$	$P(A)$

et

$\mathbf{1}_B$	0	1
	$1-P(B)$	$P(B)$

Si A et B sont indépendants, alors $\overline{A}$ et $\overline{B}$ le sont aussi, donc :

$$\begin{aligned} P(\mathbf{1}_A = 0 \cap \mathbf{1}_B = 0) &= P(\overline{A} \cap \overline{B}) \\ &= P(\overline{A})P(\overline{B}) \\ &= P(\mathbf{1}_A = 0)P(\mathbf{1}_B = 0) \end{aligned}$$

On procède de la même façon pour les trois autres vérifications.

Réciproquement, si les variables $\mathbf{1}_A$ et $\mathbf{1}_B$ sont indépendantes, alors :

$$\begin{aligned} P(A \cap B) &= P(\mathbf{1}_A = 1 \cap \mathbf{1}_B = 1) \\ &= P(\mathbf{1}_A = 1)P(\mathbf{1}_B = 1) = P(A)P(B) \end{aligned}$$

et les événements A et B sont donc indépendants.

Chapitre VI : Variables aléatoires à densité

VI.1 - Généralités

Définition 6.1 *Soit X une v.a.r. : on dit que X admet une densité f si sa fonction de répartition F est continue et peut s'écrire sous la forme :*

$$\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^{x} f(t)dt$$

avec :

1. *f est positive.*
2. *f possède un nombre fini de points de discontinuité.*
3. *$\int_{-\infty}^{+\infty} f(t)dt = 1$.*

Remarque : f n'est pas unique. Il suffit de la modifier en un point, et on obtient une autre fonction vérifiant toutes les conditions de la définition (modifier une fonction en un point ne change pas la valeur de l'intégrale). Il est donc plus correct de parler d'une densité de X plutôt que de la densité de X.

F étant une fonction de répartition, elle possède toutes les propriétés générales de ces fonctions (limites, régularité...) énoncées au chapitre II. Dans le cas des variables à densité, F est par définition continue, et possède en plus la propriété suivante, qui relie fonction de répartition et densité :

Propriété 6.1 *En tout point x_0 où f est continue, F est dérivable et :*

$$F'(x_0) = f(x_0)$$

Démonstration : soit x_0 un point où f est continue. On a donc :

$$\forall \epsilon > 0, \exists \alpha > 0, |x - x_0| < \alpha \Rightarrow |f(x) - f(x_0)| < \epsilon$$

Soient $\epsilon > 0$ fixé, et h tel que $|h| < \alpha$.

En remarquant que $f(x_0) = \frac{1}{h}\int_{x_0}^{x_0+h} f(x_0)dx$, on a :

$$\begin{aligned}
\left|\frac{F(x_0+h)-F(x_0)}{h} - f(x_0)\right| &= \left|\frac{1}{h}\int_{x_0}^{x_0+h} f(x) - f(x_0)dx\right| \\
&\leq \frac{1}{h}\int_{x_0}^{x_0+h} |f(x) - f(x_0)|dx \\
&\leq \frac{1}{h}h\epsilon \\
&\leq \epsilon
\end{aligned}$$

On a donc bien $\lim\limits_{h\to 0} \frac{F(x_0+h)-F(x_0)}{h} = f(x_0)$.

Exemple 6.1 *Soit f définie par :*

$$f(t) = \begin{cases} 0 & si \quad t \leq 0 \\ \frac{1}{4\sqrt{t}} & si \quad 0 < t < 1 \\ \frac{1}{2t^2} & si \quad 1 \leq t \end{cases}$$

f vérifie les 3 conditions de la définition : elle est positive et discontinue en 0 et en 1. De plus :

$$\int_a^1 \frac{1}{4\sqrt{t}}dt = \left[\frac{1}{2}\sqrt{t}\right]_a^1 = \frac{1}{2} - \frac{1}{2}\sqrt{a} \xrightarrow[a\to 0]{} \frac{1}{2}$$

Donc $\int_0^1 \frac{1}{4\sqrt{t}}dt$ converge (on le savait sans calcul par le critère de Riemann), et vaut $\frac{1}{2}$.
De même :

$$\int_1^b \frac{1}{2t^2}dt = \left[\frac{-1}{2t}\right]_1^b = \frac{-1}{2b} + \frac{1}{2} \xrightarrow[b\to +\infty]{} \frac{1}{2}$$

Donc $\int_1^{+\infty} \frac{1}{2t^2}dt$ converge (là aussi, on le savait au départ par le critère de Riemann) et vaut $\frac{1}{2}$. Par conséquent, $\int_{-\infty}^{+\infty} f(t)dt$ converge et vaut 1.

On obtient ensuite F :

$$F(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \frac{1}{2}\sqrt{x} & si \quad 0 < x < 1 \\ 1 - \frac{1}{2x} & si \quad 1 \leq x \end{cases}$$

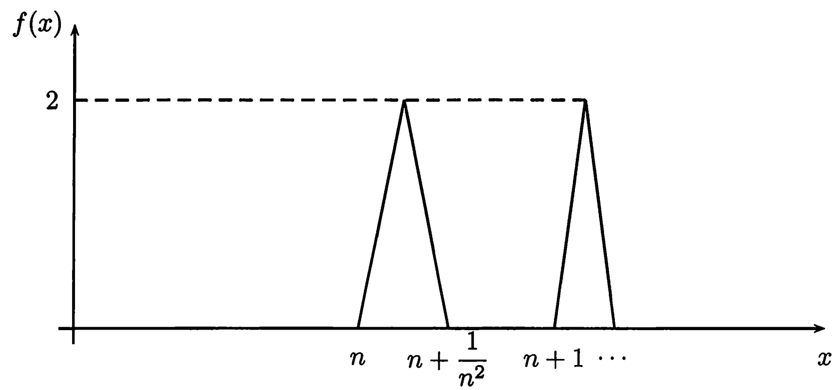
Remarque importante sur les intégrales généralisées :
pour une série $\sum u_n$, on sait que :

$$\sum u_n \text{ converge} \Rightarrow (u_n) \text{ converge et } \lim_{n\to+\infty} u_n = 0$$

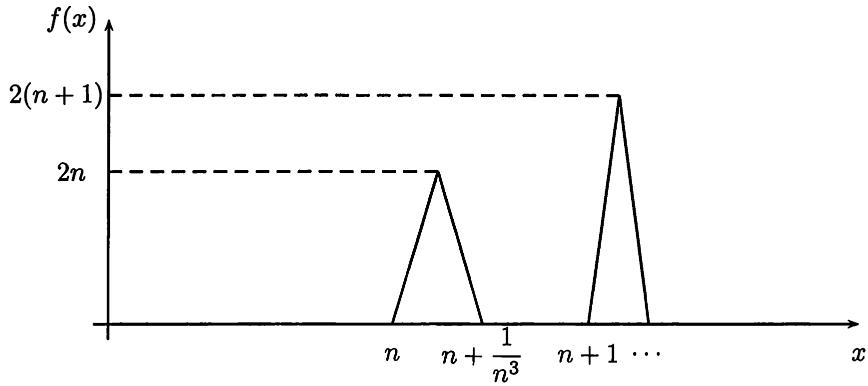
Il suffit d'écrire : $u_n = \sum_{k=0}^{n} u_k - \sum_{k=0}^{n-1} u_k$.
Pour les intégrales généralisées, on n'a pas le même résultat sur le comportement de la fonction à l'infini :
soit $f : [a; +\infty[\to \mathbb{R}$ continue et positive. Alors :

$$\int_a^{+\infty} f(t)dt \textit{ converge} \nRightarrow f(t) \xrightarrow[t\to+\infty]{} 0$$

En effet, la fonction suivante :



ne tend pas vers 0 en l'infini et pourtant son intégrale converge : puisque f est positive, la fonction $x \mapsto \int_0^x f(t)dt$ (définie sur $[0; +\infty[$) est croissante.
De plus, elle est majorée (par $\sum_{n=1}^{+\infty} \frac{1}{n^2}$), donc elle possède une limite en $+\infty$.
Il n'est même pas nécessaire que la fonction soit bornée :



Ce qui est vrai, c'est que si la fonction a une limite, alors cette limite est nulle (démontrez-le, par exemple par l'absurde, en supposant que f a une limite strictement positive).
Si on rajoute comme hypothèse que f est uniformément continue, alors la convergence de son intégrale implique que f possède une limite en $+\infty$ (et cette limite est alors nulle).

Propriété 6.2 *Soit X une v.a.r. de fonction de répartition F, admettant f pour densité. Alors, pour tout $(a; b) \in \mathbb{R}^2$ tel que $a \leq b$,*

$$P(a < X \leq b) = \int_a^b f(t)dt$$

Démonstration :

$$\begin{aligned}
P(a < X \leq b) &= P(X \in]-\infty; b]) - P(X \in]-\infty; a]) \\
&= F(b) - F(a) \\
&= \int_{-\infty}^b f(t)dt - \int_{-\infty}^a f(t)dt \\
&= \int_a^b f(t)dt
\end{aligned}$$

Conséquence : pour une variable à densité, la valeur de la probabilité ne change pas selon que l'on met des inégalités strictes ou larges :

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b)$$

En effet : $\forall a \in \mathbb{R}, P(X = a) = \lim\limits_{t \to a^+} F(t) - F(a) = 0$, et donc par exemple :

$$P(a \leq X \leq b) = P(a < X \leq b) + P(X = a) = P(a < X \leq b)$$

VI.2 - Espérance et variance d'une v.a.r. à densité

On a une analogie entre les cas discrets et continus (variable à densité) : dans le cas discret, les poids de probabilité étaient les $p_i = P(X = x_i)$ (et on devait donc avoir $\sum\limits_{i \in I} p_i = 1$).
Dans le cas continu, l'idée est que les poids de probabilités sont donnés par les valeurs de la densité f (et on demande donc que la somme de ces poids, c'est-à-dire $\displaystyle\int_{-\infty}^{+\infty} f(t)dt$, soit égale à 1).
L'unification de ces deux cas se fait par la théorie de l'intégration au sens de Lebesgue, puisqu'on a dans les deux cas une mesure et on demande

qu'elle soit de masse totale égale à 1.
Sans faire référence à cette théorie de la mesure (clairement hors du programme), on peut garder à l'esprit l'analogie décrite plus haut, et les formules qui suivent, si elles ont été comprises dans le cas discret, seront assez « naturelles ».

Définition 6.2 *Soit X une v.a.r. de densité f. On appelle espérance de X le réel :*

$$E(X) = \int_{-\infty}^{+\infty} tf(t)dt$$

sous réserve de convergence absolue de cette intégrale.

Remarque : c'est l'analogue de l'expression de l'espérance dans le cas discret $E(X) = \sum_{i \in I} x_i p_i$.

Exemple 6.2 *En reprenant la densité f définie par :*

$$f(t) = \begin{cases} 0 & si \quad t \leq 0 \\ \dfrac{1}{4\sqrt{t}} & si \quad 0 < t < 1 \\ \dfrac{1}{2t^2} & si \quad 1 \leq t \end{cases}$$

la v.a.r. X de densité f ne possède pas d'espérance, à cause de la divergence de $\int_1^{+\infty} \frac{1}{2t} dt$.

Exemple 6.3 *Soit X de densité f définie par*

$$\forall t \in \mathbb{R}, f(t) = \frac{1}{\ln 2} \times \frac{1}{t^3 + t} \mathbf{1}_{[1;+\infty[}(t)$$

On peut vérifier que f est une densité : les conditions de positivité et de régularité (discontinuité en $t = 1$) sont clairement vérifiées. Le calcul de l'intégrale se fait par décomposition en éléments simples (on sait qu'elle converge, puisqu'en $+\infty$, $f(t) \sim \frac{1}{t^3 \ln 2}$) :
soit $A \geq 1$:

$$\begin{aligned} \int_1^A \frac{1}{t^3 + t} dt &= \int_1^A \frac{1}{t} - \frac{t}{t^2 + 1} dt \\ &= \left[\ln t - \frac{1}{2} \ln (t^2 + 1) \right]_1^A \\ &= \ln \left(\frac{A}{\sqrt{A^2 + 1}} \right) - \ln \frac{1}{2} \\ &\xrightarrow[A \to +\infty]{} \ln 2 \end{aligned}$$

$E(X)$ existe, puisque $tf(t)$ équivaut à $\dfrac{1}{t^2\ln 2}$ en $+\infty$, et vaut :

$$E(X) = \frac{1}{\ln 2}\int_1^{+\infty} \frac{1}{t^2+1}dt = \frac{1}{\ln 2}\Big[\arctan t\Big]_1^{+\infty} = \frac{\pi}{4\ln 2}$$

Pour l'espérance, on a les mêmes propriétés que dans le cas discret, mais elles sont délicates à démontrer sans faire appel à la théorie de la mesure ($E(X) = \displaystyle\int_\Omega XdP$). Par contre, on n'a plus de structure d'espace vectoriel : la somme de deux variables à densité n'est pas nécessairement une variable à densité (considérer $X - X$ par exemple). On donne donc sans démonstration les deux résultats suivants :

Propriété 6.3 *Linéarité de E.*
Soient X et Y deux v.a.r. à densité admettant une espérance, et soit λ un réel. Alors $X + \lambda Y$ admet une espérance, et :

$$E(X + \lambda Y) = E(X) + \lambda E(Y)$$

Théorème 6.1 *Théorème de transfert.*
Soient X une v.a.r. de densité f et ϕ une fonction de $\mathbb{R}$ dans $\mathbb{R}$ telle que $|\phi|f$ soit intégrable sur $\mathbb{R}$. Alors $\phi(X)$ possède une espérance, et :

$$E(\phi(X)) = \int_{-\infty}^{+\infty} \phi(x)f(x)dx$$

Pour la variance et le moment d'ordre 2, on a aussi les mêmes définitions et les mêmes résultats que dans le cas discret :

Définition 6.3 *Soit X de densité f. On appelle moment d'ordre 2 l'espérance, si elle existe, de la variable X^2.*
C'est donc le réel, d'après le théorème de transfert :

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

sous réserve de convergence de cette intégrale.

Propriété 6.4 *Si X possède un moment d'ordre 2, alors X possède une espérance.*

Démonstration : identique au cas discret (écrire que pour tout réel x, $|x| \leq x^2 + 1$).

Définition 6.4 *Soit X de densité f. On appelle variance de X l'espérance, si elle existe, de la variable $(X - E(X))^2$.*
En vertu du théorème de transfert, c'est le réel :

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx$$

sous réserve de convergence de cette intégrale.

Comme dans le cas discret, l'existence du moment d'ordre 2 équivaut à celle de la variance, et le calcul effectif se fait souvent avec la formule de Koenig :

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

Démonstration analogue au cas discret.

On donne ensuite, comme dans le cas discret, les lois à densité usuelles au programme avec leurs principales caractéristiques.

VI.3 - Loi uniforme sur $[a;b]$

Définition 6.5 *X suit la loi uniforme sur $[a;b]$ si elle a pour densité f définie par :*

$$f(x) = \begin{cases} 0 & si \quad x \notin [a;b] \\ \dfrac{1}{b-a} & si \quad x \in [a;b] \end{cases}$$

On note alors : $X \sim \mathcal{U}([a;b])$.

On obtient sans difficulté :

Propriété 6.5 *Sa fonction de répartition est la fonction :*

$$F(x) = \begin{cases} 0 & si \quad x < a \\ \dfrac{1}{b-a}(x-a) & si \quad a \leq x \leq b \\ 1 & si \quad x > b \end{cases}$$

ainsi que son espérance et sa variance :

Propriété 6.6

$$E(X) = \frac{a+b}{2} \text{ et } V(X) = \frac{(b-a)^2}{12}$$

VI.4 - Loi exponentielle

Définition 6.6 *Soit $\lambda > 0$. X suit la loi exponentielle de paramètre λ si elle a pour densité :*

$$f(x) = \begin{cases} 0 & si \quad x < 0 \\ \lambda e^{-\lambda x} & si \quad x \geq 0 \end{cases}$$

On note alors : $X \sim \mathcal{E}(\lambda)$.

f est bien une densité : on a directement par le calcul la convergence de $\displaystyle\int_0^{+\infty} f(t)dt$ et sa valeur.

Propriété 6.7 *Sa fonction de répartition est :*

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 1 - e^{-\lambda x} & si \quad x \geq 0 \end{cases}$$

En procédant par exemple à des intégrations par parties, on obtient l'espérance et la variance :

Propriété 6.8

$$E(X) = \frac{1}{\lambda} \text{ et } V(X) = \frac{1}{\lambda^2}$$

Cette loi apparaît dans la partie 3 de l'épreuve 1 du CAPES 2001, où on la trouvait dans un contexte usuel : on s'intéressait à l'arrivée de messages dans un réseau informatique. T_1 était le temps d'attente depuis l'instant initial et pour $k \geq 2$, T_k était le temps d'attente du $k^{ième}$ message depuis l'arrivée du $(k-1)^{ième}$. L'hypothèse était que :

$$\forall k \geq 1, T_k \sim \mathcal{E}(\lambda)$$

et que ces variables sont indépendantes. On démontrait dans ce sujet que la loi exponentielle est « sans mémoire », c'est-à-dire qu'elle vérifie :

Propriété 6.9 *Soit $X \sim \mathcal{E}(\lambda)$. Alors :*

$$\forall s \in \mathbb{R}, \forall t \geq 0, P(X > s + t / X > t) = P(X > s)$$

De plus, cette propriété caractérise la loi exponentielle.

Démonstration : voir l'exercice 6.8.

VI.5 - Loi normale, ou loi de (Laplace-)Gauss

C'est une loi fondamentale car elle apparaît comme « loi limite » dans de très nombreuses situations, en vertu du Théorème Central Limite, que l'on abordera dans le chapitre sur les convergences de suites de variables aléatoires.

Définition 6.7 *X suit la loi normale centrée réduite si elle a pour densité f définie par :*

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

On note alors : $X \sim \mathcal{N}(0;1)$.

On peut vérifier que f est bien une densité :
f étant paire, $\displaystyle\int_{-\infty}^{+\infty} f(x)dx$ converge si et seulement si $\displaystyle\int_{0}^{+\infty} f(x)dx$ converge. Or c'est bien le cas puisque $f(x) = \underset{x\to+\infty}{\text{o}}\left(\dfrac{1}{x^2}\right)$.
Pour le calcul de l'intégrale, le problème est que la fonction f ne possède pas de primitive « usuelle » permettant de mener le calcul et on n'y parvient pas avec des intégrations par parties ou des changements de variables.
Attention : f étant continue sur $\mathbb{R}$, elle possède bien des primitives (et même une infinité !).
On donne ici une façon très classique de calculer $\displaystyle\int_{0}^{+\infty} e^{-\frac{x^2}{2}} dx$, qu'il est bon d'avoir vue une fois : l'idée est de passer dans le plan, et d'utiliser les coordonnées polaires.
On pose, pour $a > 0$: $I(a) = \displaystyle\int_{0}^{a} e^{-\frac{x^2}{2}} dx$.
On a, par le théorème de Fubini :

$$(I(a))^2 = \left(\int_0^a e^{-\frac{x^2}{2}} dx\right)\left(\int_0^a e^{-\frac{y^2}{2}} dy\right) = \iint_{[0;a]^2} e^{-\frac{x^2+y^2}{2}} dxdy$$

On ne peut pas passer facilement en coordonnées polaires puisqu'on intègre sur le carré $[0;a] \times [0;a]$. On va intégrer sur le quart de disque en posant :

$$D_a = \left\{(r\cos\theta; r\sin\theta)/r \in [0;a]; \theta \in \left[0;\frac{\pi}{2}\right]\right\}$$

On peut alors calculer, en posant $x = r\cos\theta$ et $y = r\sin\theta$:

$$J(a) = \iint_{D_a} e^{-\frac{x^2+y^2}{2}} dxdy = \iint_{[0;a]\times\left[0;\frac{\pi}{2}\right]} e^{-\frac{r^2}{2}} rdrd\theta$$

En utilisant de nouveau le théorème de Fubini, on obtient :

$$J(a) = \left(\int_0^{\frac{\pi}{2}} d\theta\right)\left(\int_0^a re^{\frac{-r^2}{2}} dr\right) = \frac{\pi}{2}\left[-e^{-\frac{r^2}{2}}\right]_0^a = \frac{\pi}{2}\left(1 - e^{\frac{-a^2}{2}}\right)$$

et donc : $\lim\limits_{a\to+\infty} J(a) = \frac{\pi}{2}$.
On encadre ensuite $(I(a))^2$. On a : $D_a \subset [0;a]^2 \subset D_{a\sqrt{2}}$, et comme la fonction intégrée est positive :

$$J(a) \leq (I(a))^2 \leq J(a\sqrt{2})$$

En passant à la limite, par encadrement, $(I(a))^2 \xrightarrow[a\to+\infty]{} \frac{\pi}{2}$, et on a donc $\int_0^{+\infty} e^{\frac{-x^2}{2}} dx = \sqrt{\frac{\pi}{2}}$.
Conclusion :

$$\int_{-\infty}^{+\infty} e^{\frac{-x^2}{2}} dx = \sqrt{2\pi}$$

Remarque : une autre méthode classique est exposée dans l'exercice 6.6.

La fonction de répartition n'a pas d'expression « explicite » à l'aide des fonctions usuelles.
Elle est donc donnée sous la forme d'une table fournie à la fin de ce chapitre, où sont rassemblées des valeurs approchées de $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

Propriété 6.10

$$E(X) = 0 \textit{ et } V(X) = 1$$

Démonstration : l'existence de $\int_{-\infty}^{+\infty} xf(x)dx$ est assurée par le fait que la fonction $x \mapsto xf(x)$ est continue et qu'en l'infini, $xf(x) = \text{o}\left(\frac{1}{x^2}\right)$. Comme elle est impaire, on a bien :

$$\int_{-\infty}^{+\infty} x\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

Pour la variance, l'existence de $\int_{-\infty}^{+\infty} x^2 f(x)dx$ est assurée pour la même raison et le calcul se fait en intégrant par parties (poser $u = x$ et $v' = xe^{-\frac{x^2}{2}}$).

Remarque : cela explique la notation $\mathcal{N}(0;1)$: 0 est l'espérance de X et 1 est son écart-type.

Définition 6.8 *Soient $m \in \mathbb{R}$ et $\sigma > 0$. X suit une loi normale de paramètre m et σ si elle a pour densité :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

On note alors : $X \sim \mathcal{N}(m;\sigma)$.

On se ramène à la loi normale centrée réduite grâce à la propriété suivante :

Propriété 6.11

$$X \sim \mathcal{N}(m;\sigma) \Leftrightarrow \frac{X-m}{\sigma} \sim \mathcal{N}(0;1)$$

Démonstration : si X suit une loi $\mathcal{N}(m;\sigma)$, alors pour tout réel y :

$$P\Big(\frac{X-m}{\sigma} \leq y\Big) = P(X \leq y\sigma + m) = \int_{-\infty}^{y\sigma+m} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

En posant $t = \dfrac{x-m}{\sigma}$, on obtient :

$$P\Big(\frac{X-m}{\sigma} \leq y\Big) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

La variable $\dfrac{X-m}{\sigma}$ a donc pour densité $t \mapsto \dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ et elle suit donc une loi $\mathcal{N}(0;1)$.
La réciproque se démontre de la même façon.

On en déduit l'espérance et la variance :

Propriété 6.12 *Soit $X \sim \mathcal{N}(m;\sigma)$. Alors $E(X) = m$ et $V(X) = \sigma^2$.*

Démonstration :

$$\begin{aligned}
X \sim \mathcal{N}(m;\sigma) &\Rightarrow \frac{X-m}{\sigma} \sim \mathcal{N}(0;1) \\
&\Rightarrow \left\{ \begin{array}{lcl} E\Big(\dfrac{X-m}{\sigma}\Big) &=& 0 \\ V\Big(\dfrac{X-m}{\sigma}\Big) &=& 1 \end{array} \right. \\
&\Rightarrow \left\{ \begin{array}{lcl} E(X-m) &=& 0 \\ \dfrac{1}{\sigma^2}V(X) &=& 1 \end{array} \right. \\
&\Rightarrow \left\{ \begin{array}{lcl} E(X) &=& m \\ V(X) &=& \sigma^2 \end{array} \right.
\end{aligned}$$

Lorsqu'on considère $\dfrac{X-m}{\sigma}$, on dit que l'on a centré et réduit X.

VI.6 - Énoncé des exercices

Exercice 6.1 *Soit X de loi uniforme sur $\left[-1;\dfrac{2}{3}\right]$.*
Déterminer la loi de X^2.

Exercice 6.2 *1. Soit X une v.a.r. dont la fonction de répartition F est strictement croissante et continue sur $\mathbb{R}$.*
Déterminer la loi de $Y = F \circ X$.

2. Soit X de loi uniforme sur $]0;1[$. Soit F une fonction de répartition. On définit « F^{-1} » par :

$$\forall x \in]0;1[, F^{-1}(x) = \inf\{y/F(y) \geq x\}$$

Préciser pourquoi cette définition a bien un sens, et montrer que la variable aléatoire $Z = F^{-1} \circ X$ a pour fonction de répartition F.

Remarque *: ce résultat est d'une grande importance « pratique » : on peut simuler la fonction de répartition de n'importe quelle variable aléatoire à partir de la loi uniforme sur $[0;1]$; il suffit de générer des nombres au hasard entre 0 et 1, et de composer par « F^{-1} », ce qui est aisé lorsque F est bijective.*

Exercice 6.3 *1. Soit X une v.a.r. de loi $\mathcal{N}(20;2)$. A l'aide de la table de la loi $\mathcal{N}(0;1)$, donner $P(X > 21)$, $P(17 \leq X)$ et $P(18 \leq X < 20.5)$, puis déterminer x tel que $P(20 - x \leq X \leq 20 + x) = 0.95$.*

2. Soit X une v.a.r. de loi $\mathcal{N}(m;\sigma)$. Déterminer $P(|X - m| > 2\sigma)$.

Exercice 6.4 *Soit $a > 0$ fixé et X une v.a.r. de loi $\mathcal{N}(m;\sigma)$.*
Déterminer b pour que $P(b \leq X \leq b + a)$ soit maximale.

Exercice 6.5 *Soit X de loi uniforme sur $\left]-\dfrac{\pi}{2};\dfrac{\pi}{2}\right[$.*

1. Déterminer la loi de $Y = \tan X$. Cette loi est appelée loi de Cauchy.

2. Y possède-t-elle une espérance ?

Exercice 6.6 *Une autre façon de calculer l'intégrale de Gauss $\displaystyle\int_0^{+\infty} e^{-t^2}\,dt$ (pas de probabilités dans cet exercice, mais des outils classiques d'analyse).*
Soit :

$$g : \left\{ \begin{array}{ccc} [0;+\infty[& \to & \mathbb{R} \\ x & \mapsto & \displaystyle\int_0^1 \frac{e^{-(t^2+1)x^2}}{t^2+1}\,dt \end{array} \right.$$

1. *Montrer que g est dérivable, et exprimer g en fonction de f définie par :*

$$\forall x \geq 0, f(x) = \int_0^x e^{-t^2} dt$$

2. *En encadrant $g(x)$, montrer que $g(x) \xrightarrow[x \to +\infty]{} 0$.*
3. *En déduire la limite de f en $+\infty$.*

Exercice 6.7 *Soit f définie sur $\mathbb{R}$ par $f(x) = ae^{-|x|}$.*

1. *Déterminer a pour que f soit une densité de probabilité.*
   Soit X une v.a.r. de densité f : on dit que X suit une loi de Laplace.
2. *Montrer que X a des moments de tout ordre et que :*

$$\forall k \geq 1, E(X^k) = \frac{1}{2}(1 + (-1)^k) \int_0^{+\infty} t^k e^{-t} dt$$

On retrouve donc la très classique fonction Γ :

$$\Gamma : x \mapsto \int_0^{+\infty} t^{x-1} e^{-t} dt$$

3. *Montrer que Γ est définie sur $]0; +\infty[$ et que $\forall x > 0, \Gamma(x+1) = x\Gamma(x)$.*
4. *En déduire $E(X^k)$, et en particulier l'espérance et la variance d'une variable qui suit une loi de Laplace.*

(Dans vos révisions, faites absolument un problème sur la fonction Γ, c'est une très bonne façon de s'entraîner à utiliser des théorèmes et des techniques classiques d'analyse : montrez qu'elle est de classe C^∞ et convexe sur $]0; +\infty[$, déterminez un équivalent en 0^+...).

Exercice 6.8 *Caractérisation de la loi exponentielle par son absence de mémoire.*
Montrer que X suit une loi exponentielle si et seulement si :

$$\forall s \in \mathbb{R}, \forall t \geq 0, P(X > s + t / X > t) = P(X > s)$$

Exercice 6.9 *Soient X_1,...,X_n des v.a.r. indépendantes qui suivent toutes une loi uniforme sur $[0; 1]$. Soit $Y = \max\limits_{1 \leq i \leq n} \{X_i\}$.*
Déterminer la loi de Y.

VI.7 - Correction des exercices

Exercice 6.1

On détermine la fonction de répartition de X^2 : $X(\Omega) = \left[-1; \frac{2}{3}\right[$, donc $X^2(\Omega) = [0; 1]$. On en déduit que si $y \leq 0$, $P(X^2 \leq y) = 0$, et que si $y \geq 1$, $P(X^2 \leq y) = 1$.

Soit F la fonction de répartition de X et soit $y \in [0; 1]$:

$$P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$$

Or on connaît F :

$$F(x) = \begin{cases} 0 & si \quad x < -1 \\ \frac{3}{5}(x+1) & si \quad -1 \leq x \leq \frac{2}{3} \\ 1 & si \quad x > \frac{2}{3} \end{cases}$$

Par conséquent : si $0 \leq y \leq \frac{4}{9}$, $0 \leq \sqrt{y} \leq \frac{2}{3}$ et donc

$$F(\sqrt{y}) - F(-\sqrt{y}) = \frac{3}{5}(\sqrt{y}+1) - \frac{3}{5}(-\sqrt{y}+1) = \frac{6}{5}\sqrt{y}$$

et si $\frac{4}{9} \leq y \leq 1$, $\frac{2}{3} \leq \sqrt{y} \leq 1$, d'où :

$$F(\sqrt{y}) - F(-\sqrt{y}) = 1 - \frac{3}{5}(-\sqrt{y}+1) = \frac{3}{5}\sqrt{y} + \frac{2}{5}$$

Conclusion : X^2 suit une loi dont la fonction de répartition G est définie par :

$$G(y) = \begin{cases} 0 & si \quad y \leq 0 \\ \frac{6}{5}\sqrt{y} & si \quad 0 \leq y \leq \frac{4}{9} \\ \frac{3}{5}\sqrt{y} + \frac{2}{5} & si \quad \frac{4}{9} \leq y \leq 1 \\ 1 & si \quad 1 \leq y \end{cases}$$

Une densité g de X^2 s'obtient alors en dérivant G en tout point où elle est dérivable, et en donnant n'importe quelle valeur à g aux points où G n'est pas dérivable (quels sont ces points ?...).

Exercice 6.2

1. Comme F est une fonction de répartition, elle est à valeurs dans $[0;1]$, donc si $y \leq 0$, alors $P(F \circ X \leq y) = 0$, et si $y \geq 1$, $P(F \circ X \leq y) = 1$.
Soit $y \in [0;1]$:

$$P(Y \leq y) = P(F \circ X \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$$

(F étant continue et strictement croissante, elle est bijective et F^{-1} est strictement croissante). Y a donc pour fonction de répartition :

$$G(y) = \begin{cases} 0 & si \quad y \leq 0 \\ y & si \quad y \in [0;1] \\ 1 & si \quad y \geq 1 \end{cases}$$

et donc $Y \sim \mathcal{U}_{[0;1]}$.

2. On note, pour tout $x \in]0;1[$, $A_x = \{y \in \mathbb{R}/F(y) \geq x\}$.
Comme F est une fonction de répartition, elle tend vers 1 en $+\infty$ et A_x est donc non vide.
Si A_x n'est pas minoré, on peut trouver une suite (u_n) d'éléments de A_x qui tend vers $-\infty$. Or :

$$u_n \in A_x \Rightarrow F(u_n) \geq x$$

Comme F tend vers 0 en $-\infty$, la suite $(F(u_n))$ tend vers 0, et on a, en passant à la limite : $0 \geq x$, ce qui est absurde.
L'ensemble A_x est donc minoré.
Comme l'ensemble A_x est une partie non vide et minorée de $\mathbb{R}$, il possède une borne inférieure.

Remarque : il existe des ensembles dont les parties non vides minorées ne possèdent pas nécessairement de borne inférieure, comme l'ensemble $\mathbb{Q}$: vous pouvez par exemple considérer

$$A = \{x \in \mathbb{Q}/x^2 \leq 2\}$$

et montrer que A est non vide, minorée, et ne possède pas de borne inférieure.

Soient $x \in]0;1[$ et $z \in \mathbb{R}$. On a :

$$F^{-1}(x) \leq z \Leftrightarrow x \leq F(z)$$

En effet, si $x \leq F(z)$, alors $z \in A_x$, et donc $z \geq F^{-1}(x)$.
Réciproquement, si $z \geq F^{-1}(x)$: par définition de la borne inférieure,

il existe une suite décroissante (u_n) d'éléments de A_x qui converge vers $F^{-1}(x)$. On a donc, pour tout n, $F(u_n) \geq x$, et en passant à la limite, comme F est continue à droite : $F(F^{-1}(x)) \geq x$.
Comme $z \geq F^{-1}(x)$, $F(z) \geq F(F^{-1}(x))$, donc $F(z) \geq x$.
On en déduit la fonction de répartition de la v.a.r. $F^{-1} \circ X$:

$$\begin{aligned} P(F^{-1} \circ X \leq z) &= P(X \leq F(z)) \\ &= F(z) \ (\text{car } X \sim \mathcal{U}_{]0;1[}) \end{aligned}$$

$F^{-1} \circ X$ a donc pour fonction de répartition F.

Exercice 6.3
Les résultats numériques s'obtiennent par lecture sur la table qui donne des valeurs approchées de la fonction de répartition de la loi $\mathcal{N}(0;1)$ (et en utilisant la parité de la densité).

1. On note X^* la variable centrée réduite associée à X : $X^* = \dfrac{X-20}{2}$.
   - $P(X > 21) = P\Big(X^* > \dfrac{1}{2}\Big) \simeq 1 - 0.6915.$
   - $P(17 \leq X) = P\Big(-\dfrac{3}{2} \leq X^*\Big) = P\Big(X^* \leq \dfrac{3}{2}\Big) \simeq 0.9332.$
   - On développe :

$$\begin{aligned} P(18 \leq X < 20.5) &= P(-1 \leq X^* < 0.25) \\ &= P(X^* < 0.25) - P(X^* \leq -1) \\ &\simeq 0.5987 - (1 - 0.8413) \simeq 0.44 \end{aligned}$$

   - On obtient x par lecture sur la table :

$$\begin{aligned} P(20 - x < X < 20 + x) &= 0.95 \\ \Leftrightarrow P\Big(-\frac{x}{2} < X^* < \frac{x}{2}\Big) &= 0.95 \\ \Leftrightarrow P\Big(X^* < \frac{x}{2}\Big) &= 0.975 \\ \Leftrightarrow \frac{x}{2} &\simeq 1.96 \end{aligned}$$

2. Le résultat numérique qui suit est à comparer avec la majoration fournie par l'inégalité de Bienaymé-Tchebychev (Chapitre VIII).

$$\begin{aligned} P(|X - m| > 2\sigma) &= P(|X^*| > 2) = 2P(X^* > 2) \\ &\simeq 2(1 - 0.9772) \simeq 0.0456 \end{aligned}$$

Exercice 6.4
Soit g définie sur $\mathbb{R}$ par $g(b) = P(b \leq X \leq b + a)$. On note ϕ la fonction

de répartition de X :

$$\forall x \in \mathbb{R}, \phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}} dt$$

On a donc $g(b) = \phi(b+a) - \phi(b)$. On étudie les variations de g :

$$g'(b) = \frac{1}{\sigma\sqrt{2\pi}} \left(e^{-\frac{(b+a-m)^2}{2\sigma^2}} - e^{-\frac{(b-m)^2}{2\sigma^2}} \right)$$

D'où le signe de g' :

$$\begin{aligned} g'(b) \geq 0 &\Leftrightarrow (b+a-m)^2 \leq (b-m)^2 \\ &\Leftrightarrow a^2 + 2a(b-m) \leq 0 \\ &\Leftrightarrow b \leq m - \frac{a}{2} \end{aligned}$$

Le tableau de variations de g est donc :

b	$-\infty$		$m - \frac{a}{2}$		∞
$g(b)$		$\nearrow$		$\searrow$	

et g est maximale pour $b = m - \frac{a}{2}$.
On pouvait obtenir le résultat graphiquement en traçant la densité de X :

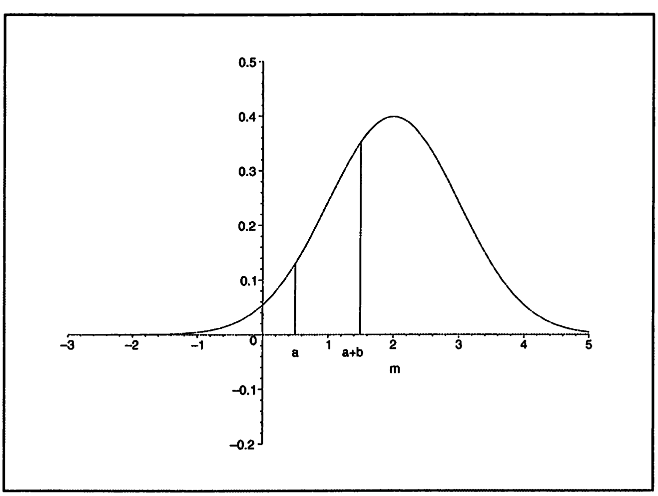


FIGURE 6.1 – Loi normale

Exercice 6.5

La fonction de répartition de X est donnée par :

$$F(x) = \begin{cases} 0 & si \quad x \leq -\dfrac{\pi}{2} \\ \dfrac{1}{\pi}\left(x + \dfrac{\pi}{2}\right) & si \quad -\dfrac{\pi}{2} \leq x \leq \dfrac{\pi}{2} \\ 1 & si \quad x \geq \dfrac{\pi}{2} \end{cases}$$

Soit $y \in \mathbb{R}$:

$$P(Y \leq y) = P(\tan X \leq y) = P(X \leq \arctan y) = \frac{1}{\pi}\left(\arctan y + \frac{\pi}{2}\right)$$

On obtient une densité g de Y en dérivant sa fonction de répartition :

$$\forall y \in \mathbb{R}, g(y) = \frac{1}{\pi(1+y^2)}$$

Y ne possède pas d'espérance car $xg(x) \underset{x \to +\infty}{\sim} \dfrac{1}{\pi x}$, ce qui entraîne la divergence de $\displaystyle\int_1^{+\infty} xg(x)dx$ et donc celle de $\displaystyle\int_{-\infty}^{+\infty} xg(x)dx$.

Remarque : la densité est pourtant une fonction paire, donc la distribution est « équilibrée » autour de 0. On pourrait être tenté de conclure hâtivement que l'espérance est nulle, mais celle-ci n'existe pas.

Exercice 6.6

1. La fonction de deux variables :

$$\phi : \begin{cases} [0;1] \times \mathbb{R}^+ & \to & \mathbb{R} \\ (t;x) & \mapsto & \dfrac{e^{-(t^2+1)x^2}}{t^2+1} \end{cases}$$

admet une dérivée partielle par rapport à x qui est continue.
Le théorème de dérivabilité sous le signe intégral (l'intervalle d'intégration étant le segment $[0;1]$) permet d'en déduire que g est dérivable sur $\mathbb{R}^+$ et que $\forall x \geq 0$,

$$g'(x) = \int_0^1 \frac{\partial \phi}{\partial x}(x;t)dt = \int_0^1 -2xe^{-(t^2+1)x^2}dt = -2xe^{-x^2}\int_0^1 e^{-t^2x^2}dt$$

En effectuant le changement de variables $u = tx$ dans l'intégrale, on fait apparaître f :

$$\forall x \geq 0, g'(x) = -2e^{-x^2}\int_0^x e^{-u^2}du = -2f'(x)f(x)$$

En intégrant : $\exists\, C \in \mathbb{R}, \forall x \geq 0, g(x) = -f^2(x) + C$.
L'évaluation en $x = 0$ donne $C = \dfrac{\pi}{4}$.

Remarque : le fait que deux fonctions ayant des dérivées égales diffèrent d'une constante peut être faux si l'on n'est pas sur un intervalle :

$$h : x \mapsto \begin{cases} 1 & si \quad x \in]0;1[\\ 2 & si \quad x \in]2;3[\end{cases} \quad \text{et } k : x \mapsto \begin{cases} 1 & si \quad x \in]0;1[\\ 3 & si \quad x \in]2;3[\end{cases}$$

sont telles que : $\forall x \in]0;1[\cup]2;3[, h'(x) = k'(x)$. Pourtant, il n'existe pas de constante C telle que :

$$\forall x \in]0;1[\cup]2;3[, h(x) = k(x) + C$$

2. On encadre g : $0 \leq g(x) \leq e^{-x^2} \displaystyle\int_0^1 \frac{1}{1+t^2} dt$, soit :
$$0 \leq g(x) \leq e^{-x^2} \frac{\pi}{4}$$
et par encadrement $\displaystyle\lim_{x \to +\infty} g(x) = 0$.

3. On a donc $\displaystyle\lim_{x \to +\infty} f^2(x) = \frac{\pi}{4}$, c'est-à-dire : $\displaystyle\int_0^{+\infty} e^{-t^2} dt = \sqrt{\frac{\pi}{4}}$.

Exercice 6.7

1. Il faut que $\displaystyle\int_{-\infty}^{+\infty} ae^{-|x|} dx$ converge vers 1. Comme f est continue et que $f(x) = \underset{x \to +\infty}{o}\left(\dfrac{1}{x^2}\right)$, on a l'existence de $\displaystyle\int_0^{+\infty} f(x)dx$. Comme f est paire, on a l'existence de son intégrale sur $\mathbb{R}$, et :
$$\int_{-\infty}^{+\infty} f(x)dx = 2 \int_0^{+\infty} ae^{-x} dx = 2a$$
d'où $a = \dfrac{1}{2}$.

2. $\forall k \geq 1$, la fonction $x \mapsto x^k e^{-|x|}$ est continue sur $\mathbb{R}$ et négligeable en l'infini devant $x \mapsto \dfrac{1}{x^2}$, donc on a l'existence de $\displaystyle\int_{-\infty}^{+\infty} x^k e^{-|x|} dx$.
On calcule le moment d'ordre k :
$$\begin{aligned} E(X^k) &= \int_{-\infty}^{+\infty} \frac{1}{2} t^k e^{-|t|} dt \\ &= \frac{1}{2}\left(\int_{-\infty}^{0} t^k e^t dt + \int_0^{+\infty} t^k e^{-t} dt\right) \\ &= \frac{1}{2}((-1)^k + 1) \int_0^{+\infty} t^k e^{-t} dt \end{aligned}$$

3. Au voisinage de 0, on a $t^{x-1}e^{-t} \sim t^{x-1}$, donc la fonction est intégrable au voisinage de 0 si et seulement si $x - 1 > -1$ (critère de Riemann). En $+\infty$, $t^{x-1}e^{-t}$ est négligeable devant $\frac{1}{t^2}$, donc la fonction est intégrable au voisinage de $+\infty$, et comme elle est continue sur $]0; +\infty[$, on a finalement :
$$\int_0^{+\infty} t^{x-1}e^{-t}dt \text{ existe } \Leftrightarrow x > 0$$
Une intégration par parties donne :
$$\Gamma(x+1) = \int_0^{+\infty} t^x e^{-t}dt = [-t^x e^{-t}]_0^{+\infty} + \int_0^{+\infty} xt^{x-1}e^{-t}dt = x\Gamma(x)$$
4. Comme $\Gamma(1) = 1$, on en déduit que $\forall k \geq 1, \Gamma(k+1) = k!$.
Par conséquent :
$$E(X^k) = \frac{1}{2}(1 + (-1)^k)\Gamma(k+1) = \frac{1}{2}(1 + (-1)^k)k!$$
En particulier $E(X) = 0$ et $V(X) = 2$.

Exercice 6.8

On suppose que X suit une loi exponentielle $\mathcal{E}(\lambda)$. Par définition de la probabilité conditionnelle :
$$\forall s \in \mathbb{R}, \forall t \geq 0, P(X > s+t/X > t) = \frac{P(X > s+t \cap X > t)}{P(X > t)}$$
Si $s \geq 0$, alors :
$$\begin{aligned}\frac{P(X > s+t \cap X > t)}{P(X > t)} &= \frac{P(X > s+t)}{P(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= P(X > s)\end{aligned}$$
Si $s < 0$, alors $X > t \Rightarrow X > s+t$, et donc :
$$P(X > s+t/X > t) = \frac{P(X > t)}{P(X > t)} = 1$$
Comme $P(X > s) = 1$, on a l'égalité souhaitée.
Réciproquement, en notant F la fonction de répartition, l'hypothèse s'écrit, dans le cas où $s \geq 0$: $\frac{1 - F(s+t)}{1 - F(t)} = 1 - F(s)$, soit, en posant $G = 1 - F$:
$$\forall s \geq 0, \forall t \geq 0, G(s+t) = G(s)G(t)$$

On reconnaît l'équation fonctionnelle caractérisant les fonctions exponentielles : $\exists\, a \in \mathbb{R}, \forall t \geq 0, G(t) = e^{at}$.
Comme $\lim\limits_{t\to+\infty} F(t) = 1$, $G(t)$ tend vers 0 et a est strictement négatif. En posant $\lambda = -a > 0$:

$$\forall t \geq 0, F(t) = 1 - e^{-\lambda t}$$

Il reste à montrer que F est nulle sur $]-\infty; 0[$:

$$s < 0 \Rightarrow \frac{P(X > s+t \cap X > t)}{P(X > t)} = \frac{P(X > t)}{P(X > t)} = 1$$

donc $P(X > s) = 1$ et on a bien $F(s) = P(X \leq s) = 0$.
Conclusion : X suit une loi exponentielle.

Exercice 6.9
On a $Y(\Omega) = [0; 1]$. Soit $y \in [0; 1]$. Grâce à l'indépendance des X_i, on obtient :

$$P(Y \leq y) = P\left(\bigcap_{i=1}^{n} X_i \leq y\right) = \prod_{i=1}^{n} P(X_i \leq y) = y^n$$

Y a donc pour fonction de répartition :

$$G(y) = \begin{cases} 0 & si \quad y \leq 0 \\ y^n & si \quad 0 \leq y \leq 1 \\ 1 & si \quad 1 \leq y \end{cases}$$

Loi normale centrée réduite

Le tableau donne avec une précision de 10^{-4} les valeurs de la fonction de répartition de la loi normale centrée réduite : $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt.$

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999

Exemple : $F(1.12) \simeq 0.8686$.

Chapitre VII : Vecteurs aléatoires à densité

VII.1 - Généralités en dimension 2

Définition 7.1 *Un vecteur aléatoire* $V : \left\{ \begin{array}{ccc} \Omega & \rightarrow & \mathbb{R}^2 \\ \omega & \mapsto & (X(\omega); Y(\omega)) \end{array} \right.$ *admet une densité s'il existe une fonction* $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ *positive, dont l'intégrale sur* $\mathbb{R}^2$ *existe et vaut 1, et telle que :*

$$\forall(\alpha;\beta) \in \mathbb{R}^2, P(X \leq \alpha \cap Y \leq \beta) = \int_{-\infty}^{\alpha} \int_{-\infty}^{\beta} f(x;y)dxdy$$

Exemple 7.1 *f définie par :*

$$f(x;y) = \left\{ \begin{array}{ccc} e^{-(x+y)} & si & x \geq 0 \; et \; y \geq 0 \\ 0 & sinon & \end{array} \right.$$

est positive et son intégrale sur $\mathbb{R}^2$ *vaut bien 1 :*

$$\iint_{\mathbb{R}^2} f(x;y)dxdy = \int_0^{+\infty} e^{-x}dx \int_0^{+\infty} e^{-y}dy = 1$$

Exemple 7.2 *Soient $R > 0$ et f définie par :*

$$f(x;y) = \left\{ \begin{array}{ccc} \dfrac{1}{\pi R^2} & si & (x;y) \in D(O;R) \\ 0 & sinon & \end{array} \right.$$

Il est immédiat de vérifier que f est une densité. On dit naturellement d'un couple $(X;Y)$ ayant pour densité cette fonction f qu'il suit la loi uniforme sur le disque $D(0;R)$.

Propriété 7.1 *Soit $(X;Y)$ un couple de densité f. X et Y sont alors des variables à densité, dont une densité pour X est :*

$$f_X : x \mapsto \int_{-\infty}^{+\infty} f(x;y)dy$$

et une densité pour Y est :

$$f_Y : y \mapsto \int_{-\infty}^{+\infty} f(x;y)dx$$

Démonstration : soit $\alpha \in \mathbb{R}$.

$$\begin{aligned} P(X \leq \alpha) &= P(X \leq \alpha \cap Y \in \mathbb{R}) \\ &= \int_{-\infty}^{\alpha} \int_{-\infty}^{+\infty} f(x;y)dxdy \\ &= \int_{-\infty}^{\alpha} \Big(\int_{-\infty}^{+\infty} f(x;y)dy \Big) dx \text{ (Théorème de Fubini)} \end{aligned}$$

Définition 7.2 *Les fonctions f_X et f_Y sont appelées densités marginales du couple $(X;Y)$.*

Exemple 7.3 *Soit $(X;Y)$ de densité f définie par :*

$$\forall (x;y) \in \mathbb{R}^2, f(x;y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

On dit que le vecteur $(X;Y)$ suit une loi normale centrée réduite en dimension 2. X a pour densité :

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy}_{1}$$

On remarque en particulier que $X \sim \mathcal{N}(0;1)$.

VII.2 - Indépendance de 2 variables à densité

On rappelle la définition vue au chapitre II :

Définition 7.3 *Deux v.a.r. X et Y définies sur le même espace probabilisé sont indépendantes si :*

$$\forall (B_1;B_2) \in (\mathcal{B}_{\mathbb{R}})^2, P(X \in B_1 \cap Y \in B_2) = P(X \in B_1)P(Y \in B_2)$$

La tribu des boréliens de $\mathbb{R}$ étant engendrée par $\{]-\infty;a], a \in \mathbb{R}\}$, il est équivalent de vérifier que :

$$\forall (\alpha;\beta) \in \mathbb{R}^2, P(X \leq \alpha \cap Y \leq \beta) = P(X \leq \alpha)P(Y \leq \beta)$$

Comme dans le cas discret, on a une caractérisation plus simple :

Théorème 7.1 *Si X et Y sont indépendantes, de densité f_X et f_Y, alors le couple $(X;Y)$ a pour densité $f = f_X f_Y$.*
Réciproquement, soit $V = (X;Y)$ de densité f. Si f se factorise sous la forme : $\forall (x;y) \in \mathbb{R}^2, f(x;y) = g(x)h(y)$, où g et h sont positives et intégrables sur $\mathbb{R}$, alors X et Y sont indépendantes et :

$$\exists \lambda > 0, f_X = \lambda g \text{ et } f_Y = \frac{1}{\lambda} h$$

Démonstration : soient X et Y indépendantes, de densité f_X et f_Y :

$$\begin{aligned}\forall(\alpha;\beta)\in\mathbb{R}^2, P(X\leq\alpha\cap Y\leq\beta) &= P(X\leq\alpha)P(Y\leq\beta)\\ &= \int_{-\infty}^{\alpha} f_X(x)dx\int_{-\infty}^{\beta} f_Y(y)dy\\ &= \iint_{]-\infty,\alpha]\times]-\infty,\beta]} f_X(x)f_Y(y)dxdy\end{aligned}$$

Réciproquement, si $(X;Y)$ a une densité f qui s'écrit : $\forall(x;y)\in\mathbb{R}^2$, $f(x;y)=g(x)h(y)$, où g et h sont positives et intégrables sur $\mathbb{R}$, alors :

$$\forall x\in\mathbb{R}, f_X(x)=\int_{\mathbb{R}} f(x;y)dy = g(x)\underbrace{\int_{\mathbb{R}} h(y)dy}_{\lambda}$$

De plus :

$$\iint_{\mathbb{R}^2} f(x;y)dxdy = 1 \Rightarrow \int_{\mathbb{R}} h(y)dy\int_{\mathbb{R}} g(x)dx = 1$$

Donc λ est bien strictement positif et $\displaystyle\int_{\mathbb{R}} g(x)dx=\frac{1}{\lambda}$.

On a alors $\displaystyle f_Y(y)=\int_{\mathbb{R}} f(x;y)dx = h(y)\int_{\mathbb{R}} g(x)dx=\frac{1}{\lambda}h(y)$. Les variables X et Y sont indépendantes puisque :

$$\begin{aligned}\forall(\alpha;\beta)\in\mathbb{R}^2, P(X\leq\alpha\cap Y\leq\beta) &= \int_{-\infty}^{\alpha}\int_{-\infty}^{\beta} f(x;y)dxdy\\ &= \int_{-\infty}^{\alpha} f_X(x)dx\int_{-\infty}^{\beta} f_Y(y)dy\\ &= P(X\leq\alpha)P(Y\leq\beta)\end{aligned}$$

Remarque : bien voir l'analogie avec le cas discret, où l'on avait la « factorisation » $p_{i,j}=p_{i,\cdot}p_{\cdot,j}$.

Exemple 7.4 *Soit $V=(X;Y)$ de densité f définie par :*

$$f(x;y)=\begin{cases} e^{-(x+y)} & si \quad x\geq 0 \text{ et } y\geq 0\\ 0 & sinon\end{cases}$$

X et Y sont indépendantes, car f s'écrit :

$$\forall(x;y)\in\mathbb{R}^2, f(x;y)=(e^{-x}\mathbf{1}_{[0;+\infty[}(x))(e^{-y}\mathbf{1}_{[0;+\infty[}(y))$$

et $x \mapsto e^{-x}\mathbf{1}_{[0;+\infty[}(x)$ est bien positive et intégrable sur $\mathbb{R}$.
Comme son intégrale sur $\mathbb{R}$ vaut 1, on a les densités marginales :

$$f_X(x) = e^{-x}\mathbf{1}_{[0;+\infty[}(x) = f_Y(x)$$

$(\lambda = 1)$

La suite de ce chapitre ne concerne que les candidats à l'Agrégation Interne. Néanmoins, elle peut intéresser les candidats au CAPES, puisqu'elle fait intervenir des outils classiques, comme le produit de convolution que l'on retrouvait dans l'épreuve 1 du CAPES 2001.

VII.3 - Densité d'une somme de 2 variables à densité

Théorème 7.2 *Soient X et Y deux variables de densité respectives f et g.*
Si X et Y sont indépendantes, alors $X+Y$ est une variable aléatoire à densité, dont une densité est :

$$h : t \mapsto \int_{-\infty}^{+\infty} f(x)g(t-x)dt$$

c'est-à-dire : $h = f * g$, produit de convolution de f et g.

Démonstration : comme X et Y sont indépendantes, le couple $(X;Y)$ a une densité qui s'écrit $\phi(x;y) = f(x)g(y)$. Soit $\alpha \in \mathbb{R}$. On intègre la densité sur le domaine $\Delta = \{(x;y) \in \mathbb{R}^2 / x+y \leq \alpha\}$:

$$P(X+Y \leq \alpha) = \iint_\Delta \phi(x;y)dxdy = \int_{-\infty}^{+\infty}\Big(\int_{-\infty}^{\alpha-x} g(y)dy\Big)f(x)dx$$

On pose $t = x+y$:

$$\begin{aligned} P(X+Y \leq \alpha) &= \int_{-\infty}^{+\infty}\Big(\int_{-\infty}^{\alpha} g(t-x)dt\Big)f(x)dx \\ &= \int_{-\infty}^{\alpha}\underbrace{\Big(\int_{-\infty}^{+\infty} g(t-x)f(x)dx\Big)}_{h(t)}dt \end{aligned}$$

VII.4 - Vecteur aléatoire à densité en dimension p

On étend la définition vue dans le cas $p = 2$.

Définition 7.4 *Un vecteur aléatoire $V = (X_1;\cdots;X_p)$ admet une densité s'il existe une fonction $f : \mathbb{R}^p \to \mathbb{R}$ positive, dont l'intégrale sur $\mathbb{R}^p$ existe et vaut 1, et telle que : pour tout $(\alpha_1;\cdots;\alpha_p) \in \mathbb{R}^p$,*

$$P\Big(\bigcap_{i=1}^{p} X_i \leq \alpha_i\Big) = \int_{-\infty}^{\alpha_1}\cdots\int_{-\infty}^{\alpha_p} f(x_1;\cdots;x_p)dx_1\cdots dx_p$$

Le théorème de transfert se généralise également :

Théorème 7.3 *Soit $X = (X_1; \cdots; X_p)$ un vecteur aléatoire de densité f. Soient $g : \mathbb{R}^p \to \mathbb{R}$ continue, et Δ un domaine « géométriquement simple » de $\mathbb{R}^p$, tels que $|g|\mathbf{1}_\Delta f$ soit intégrable sur $\mathbb{R}^p$.*
Alors $(g\mathbf{1}_\Delta)(X)$ est une variable aléatoire réelle, d'espérance :

$$E((g\mathbf{1}_\Delta)(X)) = \int_\Delta g(x_1; \cdots; x_p) f(x_1; \cdots; x_p) dx_1 \cdots dx_p$$

Ce théorème est admis.

Exemple 7.5 *Soit $(X;Y)$ un vecteur de densité f. En appliquant le théorème précédent avec $\Delta = \mathbb{R}^2$ et $g : (x;y) \to x$, on a que, si $(x;y) \to xf(x;y)$ est intégrable sur $\mathbb{R}^2$:*

$$E(X) = \iint_{\mathbb{R}^2} xf(x;y)dxdy$$

VII.5 - Indépendance de p variables à densité

On rappelle la définition générale, pour des v.a.r. quelconques :

Définition 7.5 *Les v.a.r. $X_1, \cdots, X_p$ sont indépendantes si pour tout ensemble d'indices $I \subset [\![1; \cdots; n]\!]$, on a, pour tout $(B_i)_{i\in I} \in (\mathcal{B}_\mathbb{R})^{\mathrm{Card}\, I}$,*

$$P\left(\bigcap_{i\in I} X^{-1}\left(B_i\right)\right) = \prod_{i\in I} P\left(X^{-1}\left(B_i\right)\right)$$

Comme dans le cas de deux variables, cette condition équivaut à :

$$\forall I \subset [\![1; \cdots; n]\!],\ \forall(\alpha_i)_{i\in I} \in \mathbb{R}^{\mathrm{Card}\, I}, P\left(\bigcap_{i\in I}(X_i \leq \alpha_i)\right) = \prod_{i\in I} P(X_i \leq \alpha_i)$$

Le théorème de « factorisation » vu pour un couple se généralise aussi :

Théorème 7.4 *Si $X_1, \cdots, X_p$ sont indépendantes de densités respectives $f_{X_1}, \cdots, f_{X_p}$, alors le vecteur $V = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ a pour densité $\prod_{i=1}^{p} f_{X_i}$.*

Réciproquement, si $V = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ a une densité f qui s'écrit :

$$\forall(x_1; \cdots; x_p) \in \mathbb{R}^p, f(x_1; \cdots; x_p) = f_1(x_1) \cdots f_p(x_p)$$

où les f_i sont positives et intégrables sur $\mathbb{R}$, alors $X_1, \cdots, X_p$ sont indépendantes, et :

$$\exists(\lambda_1; \cdots; \lambda_p) \in (\mathbb{R}_+^*)^p \text{ tels que } \prod_{i=1}^{p} \lambda_i = 1 \text{ et } \forall i, f_{X_i} = \lambda_i f_i$$

Démonstration : analogue au cas de deux variables.

VII.6 - Covariance et coefficient de corrélation

Propriété 7.2 *Soit $(X; Y)$ un vecteur de densité f.*
Si $(x; y) \mapsto |xy| f(x; y)$ est intégrable sur $\mathbb{R}^2$, alors XY est une v.a.r. d'espérance :

$$E(XY) = \iint_{\mathbb{R}^2} xyf(x; y)dxdy$$

Démonstration : on applique le théorème de transfert avec $g(x; y) = xy$.

Définition 7.6 *Soit $(X; Y)$ un couple de v.a.r. de densité f, tel que X et Y admettent une espérance. On appelle covariance de $(X; Y)$ le réel :*

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

sous réserve d'existence de $E(XY)$.
D'après le théorème de transfert, c'est le réel :

$$\mathrm{Cov}(X; Y) = \iint_{\mathbb{R}^2} xyf(x; y)dxdy - \iint_{\mathbb{R}^2} xf(x; y)dxdy \iint_{\mathbb{R}^2} yf(x; y)dxdy$$

sous réserve d'intégrabilité des fonctions en jeu.

On a le même résultat que dans le cas discret :

Propriété 7.3 *Si X et Y sont indépendantes, alors $\mathrm{Cov}(X; Y) = 0$, et la réciproque est fausse.*

Démonstration : l'indépendance de X et Y entraîne que $f = f_X f_Y$, d'où :

$$\begin{aligned} E(XY) &= \iint_{\mathbb{R}^2} xyf(x; y)dxdy \\ &= \iint_{\mathbb{R}^2} xyf_X(x)f_Y(y)dxdy \\ &= \int_{\mathbb{R}} xf_X(x)dx \int_{\mathbb{R}} yf_Y(y)dy \\ &= E(X)E(Y) \end{aligned}$$

La nullité de la covariance n'entraîne pas l'indépendance : voir exercice 7.3.

Définition 7.7 *Soient X et Y possédant une variance (strictement positive). On appelle coefficient de corrélation de X et Y le réel :*

$$\rho(X;Y) = \frac{\mathrm{Cov}(X;Y)}{\sigma(X)\sigma(Y)}$$

VII.7 - Espérance et variance d'une somme de p variables à densité

En général, la somme de p variables à densité n'est pas une variable à densité (prendre $X + (-X)$). Le calcul de l'espérance et de la variance se fait de la même façon que dans le cas discret : $E\Big(\sum_{i=1}^{p} X_i\Big) = \sum_{i=1}^{p} E(X_i)$ et $V\Big(\sum_{i=1}^{p} X_i\Big) = \sum_{i=1}^{p} V(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i; X_j)$.

VII.8 - Loi normale en dimension 2

Définition 7.8 *Le vecteur $V = (X;Y)$ suit une loi normale centrée s'il a une densité f définie par :*

$$\forall (x;y) \in \mathbb{R}^2, f(x;y) = \frac{1}{\alpha} e^{-\frac{1}{2}Q(x;y)}$$

où Q est une forme quadratique définie-positive sur $\mathbb{R}^2$, et $\alpha = \displaystyle\iint_{\mathbb{R}^2} e^{-\frac{1}{2}Q(x;y)}dxdy$. On dit aussi que V est un vecteur normal, ou gaussien (centré).

Remarque : la définie-positivité de Q permet d'avoir l'existence de l'intégrale sur $\mathbb{R}^2$ de f.

Q étant une forme quadratique sur $\mathbb{R}^2$, elle s'écrit :

$$Q(x;y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2bxy + cy^2$$

On rappelle que Q est définie-positive si et seulement si $a > 0$ et $ac - b^2 > 0$. On peut le voir en opérant une réduction de Gauss :

$$Q(x;y) = ax^2 + 2bxy + cy^2 = a\Big(x + \frac{b}{a}y\Big)^2 + \frac{ac - b^2}{a}y^2$$

On peut aussi dire que $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ est symétrique réelle, donc diagonalisable : elle est semblable à $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, et Q est définie-positive si

et seulement si $\lambda_1 > 0$ et $\lambda_2 > 0$.
Or :

$$\begin{aligned}
\begin{cases} \lambda_1 > 0 \\ \lambda_2 > 0 \end{cases} &\Leftrightarrow \begin{cases} \lambda_1\lambda_2 > 0 \\ \lambda_1 + \lambda_2 > 0 \end{cases} \Leftrightarrow \begin{cases} \det A > 0 \\ \operatorname{tr} A > 0 \end{cases} \\
&\Leftrightarrow \begin{cases} ac - b^2 > 0 \\ a + c > 0 \end{cases} \Leftrightarrow \begin{cases} ac - b^2 > 0 \\ a > 0 \end{cases}
\end{aligned}$$

Calcul de α : on a vu au chapitre précédent que :

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx = \sigma\sqrt{2\pi}$$

ce qui permet d'obtenir :

$$\begin{aligned}
\alpha &= \iint_{\mathbb{R}^2} e^{-\frac{1}{2}a(x+\frac{b}{a}y)^2} e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2} dxdy \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} e^{-\frac{1}{2}a(x+\frac{b}{a}y)^2} dx \right) e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2} dy \\
&= \frac{\sqrt{2\pi}}{\sqrt{a}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2} dy \\
&= \frac{2\pi}{\sqrt{a}} \sqrt{\frac{a}{ac-b^2}} \\
&= \frac{2\pi}{\sqrt{ac-b^2}}
\end{aligned}$$

Propriété 7.4 *On a, avec les notations précédentes :*

$$E(X) = E(Y) = 0 \text{ et } \operatorname{Cov}(X;Y) = \frac{-b}{ac-b^2}$$

Démonstration : on utilise aussi les calculs d'intégrales faits pour la loi normale (chapitre précédent) :

$$\begin{aligned}
\int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx &= m \\
\text{et } \int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx &= \sigma^2 + m^2
\end{aligned}$$

($E(T)$ et $E(T^2)$) pour une v.a.r. T qui suit une loi $\mathcal{N}(m;\sigma)$).

Calcul de $E(X)$:

$$\begin{aligned}
E(X) &= \iint_{\mathbb{R}^2} xf(x;y)dxdy \\
&= \frac{1}{\alpha}\int_{-\infty}^{+\infty}\Big(\int_{-\infty}^{+\infty} xe^{-\frac{1}{2}a(x+\frac{b}{a}y)^2}dx\Big)e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2}dy \\
&= \frac{1}{\alpha\sqrt{a}}\int_{-\infty}^{+\infty}\sqrt{2\pi}\Big(\frac{-b}{a}\Big)ye^{-\frac{1}{2}\frac{ac-b^2}{a}y^2}dy \\
&= 0 \text{ (imparité de la fonction à intégrer)}
\end{aligned}$$

Calcul de $\mathrm{Cov}(X;Y)$:

$$\begin{aligned}
E(XY) &= \iint_{\mathbb{R}^2} xyf(x;y)dxdy \\
&= \frac{1}{\alpha}\int_{-\infty}^{+\infty}\Big(\int_{-\infty}^{+\infty} xe^{-\frac{1}{2}a(x+\frac{b}{a}y)^2}dx\Big)ye^{-\frac{1}{2}\frac{ac-b^2}{a}y^2}dy \\
&= \frac{1}{\alpha}\int_{-\infty}^{+\infty}\frac{\sqrt{2\pi}}{\sqrt{a}}\Big(\frac{-b}{a}\Big)y^2e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2}dy \\
&= \frac{1}{\alpha}\frac{\sqrt{2\pi}}{\sqrt{a}}\Big(\frac{-b}{a}\Big)\sqrt{2\pi}\Big(\frac{a}{ac-b^2}\Big)^{\frac{3}{2}} \\
&= \frac{-b}{ac-b^2}
\end{aligned}$$

Propriété 7.5 *Les composantes X et Y du vecteur $V = (X;Y)$ suivent des lois normales.*

Démonstration : on calcule la densité de X et de Y :

$$\begin{aligned}
\forall y \in \mathbb{R}, f_Y(y) &= \int_{-\infty}^{+\infty} f(x;y)dx \\
&= \frac{1}{\alpha}e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}a(x+\frac{b}{a}y)^2}dx \\
&= \frac{\sqrt{ac-b^2}}{2\pi}\frac{\sqrt{2\pi}}{\sqrt{a}}e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2} \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{ac-b^2}{a}}e^{-\frac{1}{2}\frac{ac-b^2}{a}y^2}
\end{aligned}$$

Donc : $Y \sim \mathcal{N}(0;\sigma_Y)$ avec $\sigma_Y = \sqrt{\dfrac{a}{ac-b^2}}$. On obtient de la même façon : $X \sim \mathcal{N}(0;\sigma_X)$ avec $\sigma_X = \sqrt{\dfrac{c}{ac-b^2}}$.

Propriété 7.6 *Soit $V = (X;Y)$ un vecteur normal. Alors on a l'équivalence : X et Y sont indépendantes si et seulement si* $\mathrm{Cov}(X;Y) = 0$.

Démonstration : on a vu que l'indépendance entraîne la nullité de la covariance (c'est toujours vrai).
Réciproquement :

$$\begin{aligned}
\mathrm{Cov}(X;Y) = 0 \ &\Rightarrow \ b = 0 \\
&\Rightarrow \ \begin{cases} f_X(x) = \dfrac{1}{\sqrt{2\pi}}\sqrt{a}e^{-\frac{1}{2}ax^2} \\ f_Y(y) = \dfrac{1}{\sqrt{2\pi}}\sqrt{c}e^{-\frac{1}{2}cx^2} \end{cases} \\
&\Rightarrow \ f_X(x)f_Y(y) = \frac{1}{2\pi}\sqrt{ac}e^{-\frac{1}{2}(ax^2+cy^2)} \\
&\Rightarrow \ f_X(x)f_Y(y) = f(x;y)
\end{aligned}$$

Définition 7.9 *Soit $V = (X;Y)$ un vecteur normal. On appelle matrice de variances-covariances la matrice Σ :*

$$\Sigma = \begin{pmatrix} V(X) & \mathrm{Cov}(X;Y) \\ \mathrm{Cov}(X;Y) & V(Y) \end{pmatrix}$$

D'après les calculs qui viennent d'être faits, on a :

$$\Sigma = \frac{1}{ac-b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$$

On a donc $\Sigma = A^{-1}$, d'où la propriété :

Propriété 7.7 *$V = (X;Y)$ suit une loi normale si et seulement s'il admet une densité f qui s'écrit :*

$$\forall (x;y) \in \mathbb{R}^2, f(x;y) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\Sigma^{-1}\begin{pmatrix} x \\ y \end{pmatrix}}$$

où Σ est la matrice de variances-covariances de V.

Par translation, on définit la loi normale générale (non nécessairement centrée) :

Définition 7.10 *Soit $V = (X;Y)$ d'espérance $m = (E(X);E(Y))$.
V suit une loi normale si et seulement si $V - m$ suit une loi normale centrée.*

La matrice de variances-covariances de V est la même que celle de $V - m$ et un changement de variables donne une densité de V :

Propriété 7.8 *Un vecteur aléatoire V d'espérance $m = (m_1; m_2)$ et de matrice de variances-covariances Σ suit une loi normale si et seulement si une densité de V est f définie par :*

$$\forall u \in \mathbb{R}^2, f(u) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}\ {}^t(u-m)\Sigma^{-1}(u-m)}$$

On a montré que les composantes X et Y d'un vecteur normal sont indépendantes si et seulement si $\mathrm{Cov}(X; Y) = 0$, ce qui revient à dire :

Propriété 7.9 *Soit $V = (X; Y)$ un vecteur normal de matrice de variances-covariances Σ. Les v.a.r. X et Y sont indépendantes si et seulement si Σ est une matrice diagonale.*

VII.9 - Loi normale en dimension p

On généralise ce qui vient d'être vu :

Définition 7.11 *Le vecteur $V = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ suit une loi normale (centrée) s'il a une densité f qui s'écrit :*

$$\forall (x_1; \cdots; x_p) \in \mathbb{R}^p, f(x_1; \cdots; x_p) = \frac{1}{\alpha} e^{-\frac{1}{2}Q(x_1;\cdots;x_p)}$$

où Q est une forme quadratique définie-positive sur $\mathbb{R}^p$, et

$$\alpha = \int \cdots \int_{\mathbb{R}^p} e^{-\frac{1}{2}Q(x_1;\cdots;x_p)} dx_1 \cdots dx_p$$

Propriété 7.10 *Soit $V = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$, tel que les composantes X_i possèdent un moment d'ordre 2. En notant Σ sa matrice de variances-covariances définie par :*

$$\forall (i; j) \in [\![1; \cdots; p]\!]^2, \Sigma_{i;j} = \mathrm{Cov}(X_i; X_j)$$

on a : V suit une loi normale (centrée) si et seulement s'il a une densité qui peut s'écrire :

$$\forall u \in \mathbb{R}^p, f(u) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{\sqrt{\det \Sigma}} e^{-\frac{1}{2}\ {}^t u \Sigma^{-1} u}$$

De manière générale, V suit une loi normale si $V - m$ suit une loi normale centrée (où $m = (E(X_1); \cdots ; E(X_p)))$, ce qui équivaut à dire que V a une densité qui s'écrit :

$$\forall u \in \mathbb{R}^p, f(u) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{\sqrt{\det \Sigma}} e^{-\frac{1}{2}\ {}^t(u-m)\Sigma^{-1}(u-m)}$$

et on note alors : $X \sim \mathcal{N}(m; \Sigma)$.

Remarque : si $p = 1$, on retrouve bien la fonction $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}$, puisque $\Sigma = (\sigma^2)$. Par cohérence avec la notation en dimension p, on note parfois $\mathcal{N}(m; \sigma^2)$ au lieu de $\mathcal{N}(m; \sigma)$.

Les résultats démontrés dans le cas $p = 2$ se généralisent : si V suit une loi normale, alors ses composantes suivent une loi normale, et elles sont indépendantes si et seulement si Σ est diagonale.

VII.10 - Énoncé des exercices

Exercice 7.1 *Dans le carré unité, on choisit au hasard un point* $(X; Y)$*, c'est-à-dire :*

$$\forall S \subset [0; 1]^2, P((X; Y) \in S) = aire(S)$$

1. *Donner une densité de* $(X; Y)$.
2. *Déterminer les lois de* X *et* Y.
3. X *et* Y *sont-elles indépendantes ?*

Exercice 7.2 *Soit le polynôme* $P(X) = X^2 - 2AX + B$. *On suppose que* A *et* B *sont des v.a.r. indépendantes qui suivent la loi uniforme sur* $[0; 1]$.

1. *Quelle est la probabilité que* P *possède 2 racines réelles distinctes ?*
2. *Quelle est la probabilité que* P *possède une racine double ?*
3. *Quelle est la probabilité que* P *possède 2 racines complexes (et non réelles) ?*

Exercice 7.3 *Soit* $(X; Y)$ *de loi uniforme sur* $D(0; R)$*, disque de centre* 0 *et de rayon* R $(R > 0)$.

1. *Donner une densité de* $(X; Y)$.
2. *Déterminer les densités marginales de* X *et* Y.
3. X *et* Y *sont-elles indépendantes ?*

4. *Calculer* $\mathrm{Cov}(X;Y)$.
5. *Soit* $U = \sqrt{X^2+Y^2}$. *Déterminer la fonction de répartition, une densité et l'espérance de* U.

Exercice 7.4 *Soit* f *définie par :*

$$f(x;y) = \begin{cases} ae^{-(x+y)} & si \quad 0 \leq x \leq y \\ 0 & sinon \end{cases}$$

1. *Déterminer* a *pour que* f *soit la densité d'un vecteur aléatoire* $V = (X;Y)$.
2. *Déterminer les densités marginales.*
3. *Les composantes* X *et* Y *sont-elles indépendantes ?*

Exercice 7.5 *Soient* X *et* Y *deux v.a.r. qui suivent une loi normale.*

1. *Leur somme* $Z = X+Y$ *suit-elle nécessairement une loi normale ?*
2. *En supposant en plus* $\mathrm{Cov}(X;Y) = 0$, Z *suit-elle une loi normale ?*
3. *Le vecteur* $V = (X;Y)$ *est-il nécessairement gaussien ?*

Exercice 7.6 *Soient* X *et* Y *deux v.a.r. à densité. Le vecteur* $V = (X;Y)$ *est-il nécessairement un vecteur à densité ?*

VII.11 - Correction des exercices

Exercice 7.1

1. f est la densité uniforme sur la carré unité $C = [0;1]^2$:

$$f(x;y) = \begin{cases} 0 & si \quad (x;y) \notin C \\ 1 & si \quad (x;y) \in C \end{cases}$$

2. Loi de X :

$$f_X(x) = \int_{\mathbb{R}} f(x;y)dy = \begin{cases} \displaystyle\int_0^1 1dy = 1 & si \quad x \in [0;1] \\ 0 & si \quad x \notin [0;1] \end{cases}$$

Même calcul pour f_Y.

3. On a $\forall (x;y) \in \mathbb{R}^2, f_X(x)f_Y(y) = f(x;y)$, et les variables X et Y sont donc indépendantes.

Exercice 7.2

Comme A et B sont indépendantes, le vecteur $(A;B)$ a pour densité $f = f_A f_B = \mathbf{1}_{[0;1]^2}$.

1. P possède 2 racines réelles distinctes si et seulement si $A^2 - B > 0$. Soit $\Delta_1 = \{(a;b) \in [0;1]^2 / a^2 - b > 0\}$:

$$P(A^2 - B > 0) = \iint_{\Delta_1} 1 dxdy = \int_0^1 \Big(\int_0^{x^2} dy\Big) dx = \frac{1}{3}$$

2. On note $\Delta_2 = \{(a;b) \in [0;1]^2 / a^2 - b = 0\}$ et on obtient $P(A^2 - B = 0) = 0$ (l'intégrale de f sur Δ_2 est nulle).
3. On passe à l'événement contraire :

$$P(A^2 - B < 0) = 1 - P(A^2 - B \geq 0) = \frac{2}{3}$$

Exercice 7.3

1. $(X;Y)$ suit la loi uniforme sur $D(O;R)$:

$$f(x;y) = \begin{cases} 0 & si \quad (x;y) \notin D(0;R) \\ \dfrac{1}{\pi R^2} & si \quad (x;y) \in D(0;R) \end{cases}$$

2. Densité de X : $f_X(x) = \displaystyle\int_{\mathbb{R}} f(x;y)dy$, donc f_X est nulle si $x \notin]-R;R[$, et si $x \in]-R;R[$:

$$f_X(x) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2} dy = \frac{2}{\pi R^2}\sqrt{R^2 - x^2}$$

Même calcul pour f_Y.

3. X et Y ne sont pas indépendantes puisque $f_X f_Y \neq f$.
On peut aussi justifier la non indépendance en remarquant que :

$$P\Big(X > \frac{R}{\sqrt{2}}\Big)P\Big(Y > \frac{R}{\sqrt{2}}\Big) \neq 0 \text{ et } P\Big(X > \frac{R}{\sqrt{2}} \cap Y > \frac{R}{\sqrt{2}}\Big) = 0$$

4. Il est clair que $E(X) = E(Y) = 0$. On calcule $E(XY)$ en passant en coordonnées polaires :

$$E(XY) = \iint_{D(0;R)} xy \frac{1}{\pi R^2} dxdy = \frac{1}{\pi R^2} \int_0^R \int_0^{2\pi} r^2 \cos\theta \sin\theta r dr d\theta = 0$$

Par conséquent $\mathrm{Cov}(X;Y) = 0$.

Remarque : on a donc un exemple de variables dont la covariance est nulle, et qui ne sont pas indépendantes.

5. Comme U est à valeurs dans $[0;R]$, si $u \leq 0$ alors $P(U \leq u) = 0$, et si $u \geq R$, $P(U \leq u) = 1$. Soit $u \in [0;R]$:

$$P(U \leq u) = \iint_{D(0;u)} f(x;y)dxdy = \frac{u^2}{R^2}$$

La fonction de répartition F est donc :

$$F(u) = \begin{cases} 0 & si \quad u \leq 0 \\ \dfrac{u^2}{R^2} & si \quad 0 \leq u \leq R \\ 1 & si \quad u \geq R \end{cases}$$

On obtient une densité f en dérivant F (et en donnant une valeur arbitraire à $f(R)$). On peut choisir $f : u \mapsto \frac{2u}{R^2}\mathbf{1}_{[0;R]}(u)$, et on en déduit l'espérance :

$$E(U) = \int_{\mathbb{R}} uf(u)du = \int_0^R \frac{2u^2}{R^2}du = \frac{2}{3}R$$

Exercice 7.4

1. Soit $T = \{(x;y) \in \mathbb{R}^2 / 0 \leq x \leq y\}$:

$$\begin{aligned} \iint_{\mathbb{R}^2} f(x;y)dxdy &= a\iint_T e^{-(x+y)}dxdy \\ &= a\int_0^{+\infty}\Big(\int_0^y e^{-x}dx\Big)e^{-y}dy \\ &= a\int_0^{+\infty}(1-e^{-y})e^{-y}dy \\ &= \frac{1}{2}a \end{aligned}$$

On pose donc : $a = 2$.

2. Densité de X : $f_X(x) = \int_{\mathbb{R}} f(x;y)dy$ donc si $x < 0$, $f_X(x) = 0$, et si $x \geq 0$, $f_X(x) = \int_x^{+\infty} 2e^{-x}e^{-y}dy = 2e^{-2x}$.

   Densité de Y : $f_Y(y) = \int_{\mathbb{R}} f(x;y)dx$ donc f_Y est nulle si $y < 0$, et si $y \geq 0$, $f_Y(y) = \int_0^y 2e^{-x}e^{-y}dx = 2(1-e^{-y})e^{-y}$.

3. Les variables X et Y ne sont pas indépendantes puisque $f_X f_Y \neq f$.

Exercice 7.5

1. Il suffit de prendre X de loi $\mathcal{N}(0;1)$ et $Y = -X$ pour constater que la réponse est négative.
2. La réponse est non. Un contre-exemple classique est le suivant : on considère deux v.a.r. indépendantes X et T telles que $X \sim \mathcal{N}(0;1)$ et T de loi discrète :

$T(\Omega)$	-1	1
$P(T = t_i)$	$\frac{1}{2}$	$\frac{1}{2}$

On pose ensuite $Y = XT$: $Y \sim \mathcal{N}(0;1)$. En effet, soit $y \in \mathbb{R}$:

$$\begin{aligned}
P(Y \leq y) &= P(XT \leq y) \\
&= P_{(T=-1)}(XT \leq y)P(T=-1) \\
&\quad + P_{(T=1)}(XT \leq y)P(T=-1) \\
&= \frac{1}{2}P(-X \leq y) + \frac{1}{2}P(X \leq y) \\
&= \frac{1}{2}(P(X \geq -y) + P(X \leq y)) \\
&= P(X \leq y)
\end{aligned}$$

De plus, l'indépendance de X et T entraîne celle de X^2 et T, donc :

$$\mathrm{Cov}(X;Y) = E(XY) = E(X^2T) = E(X^2)E(T) = 0$$

Mais $X + Y$ ne suit pas une loi normale car si c'était le cas, on aurait $P(X + Y = 0) = 0$ (variable à densité). Or ici :

$$\begin{aligned}
P(X+Y=0) &= P(X(1+T)=0) \\
&= P((X=0) \cup (T=-1)) \\
&= P(X=0) + P(T=-1) - P(X=0)P(T=-1) \\
&= \frac{1}{2}
\end{aligned}$$

3. Là aussi, la réponse est non : Soit $V = (X;Y)$ de densité f définie par

$$\forall (x;y) \in \mathbb{R}^2, f(x;y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)} + axy\mathbf{1}_{[-1;1]\times[-1;1]}(x;y)$$

où a est tel que f soit positive sur $\mathbb{R}^2$.
Il est clair que V n'est pas gaussien. Or ses composantes le sont :

$$f_X(x) = \int_{\mathbb{R}} f(x;y)dy = \frac{1}{2\pi}e^{-\frac{1}{2}x^2}\int_{\mathbb{R}} e^{-\frac{1}{2}y^2}dy + \int_{-1}^{1} xydy = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

donc $X \sim \mathcal{N}(0;1)$ (même résultat pour Y).

Exercice 7.6
Soit $(X;Y)$ de densité f. On note δ la droite d'équation $x+y=0$.

$$P(Y=-X)=P(X+Y=0)=\iint_{\delta} f(x;y)dxdy=0$$

On en déduit que le vecteur $(X;-X)$ ne peut pas admettre de densité car si c'était le cas, on aurait $P(-X=-X)=0$. Or on a évidemment $P(-X=-X)=1$.

Chapitre VIII : Suites de variables aléatoires

Les v.a.r. considérées sont soit discrètes, soit à densité.

VIII.1 - Inégalité de Bienaymé-Tchebychev

Théorème 8.1 *Soit X une v.a.r. possédant une espérance m et une variance σ^2. Alors :*

$$\forall a > 0, P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}$$

Démonstration : elle repose sur le fait que pour un événement A, on a $P(A) = E(\mathbf{1}_A)$ (puisque la variable aléatoire $\mathbf{1}_A$ suit une loi de Bernoulli de paramètre $P(A)$). Soit A l'événement : $A = \{\omega / |X - m| \geq a\}$.
Soient Y et Z les variables aléatoires définies par $Y = |X - m|$ et $Z = a\mathbf{1}_A$.
On compare Y et Z :
Si $\omega \in A$, alors $Y(\omega) = |X(\omega) - m|$, donc $Y(\omega) \geq a$, et comme $Z(\omega) = a$, on a : $Y(\omega) \geq Z(\omega)$.
Si $\omega \notin A$, alors $Z(\omega) = 0$ et comme $Y(\omega)$ est toujours positif : Y$(\omega) \geq Z(\omega)$. D'ou finalement :

$$\forall \omega \in \Omega, Y(\omega) \geq Z(\omega)$$

Par conséquent, $Y^2 \geq Z^2$, et donc $E(Y^2) \geq E(Z^2)$. Or $E(Y^2) = \sigma^2$ et $E(Z^2) = a^2 P(A)$, d'où $P(A) \leq \dfrac{\sigma^2}{a^2}$.

Remarque : l'inégalité a surtout un intérêt théorique (voir preuve de la loi faible des grands nombres), et elle est vraie pour toute variable ayant une espérance et une variance.
Numériquement, la majoration est souvent médiocre.
Par exemple, si $X \sim \mathcal{N}(m; \sigma)$, l'inégalité de Bienaymé-Tchebychev fournit la majoration :

$$P(|X - m| \geq 2\sigma) \leq \frac{1}{4}$$

alors que :

$$\begin{aligned}
P(|X - m| \geq 2\sigma) &= P(|X^*| \geq 2) \text{ où } X^* = \frac{X - m}{\sigma} \sim \mathcal{N}(0; 1) \\
&= P(X^* \leq -2) + P(X^* \geq 2) \\
&= 2(1 - P(X^* \leq 2)) \\
&\simeq 0.0455 \text{ (lecture sur la table de la loi } \mathcal{N}(0; 1))
\end{aligned}$$

VIII.2 - Convergence en probabilité

Définition 8.1 *Soit* (X_n) *une suite de v.a.r. et* X *une v.a.r. définies sur le même espace* $(\Omega; \mathcal{A}; P)$.
On dit que (X_n) *converge vers* X *en probabilité si :*

$$\forall \alpha > 0, P(|X_n - X| > \alpha) \xrightarrow[n \to +\infty]{} 0$$

La définition exprime donc que la probabilité d'avoir un écart entre X_n et X supérieur à α tend vers 0 quand $n \to +\infty$.
L'inégalité de Bienaymé-Tchebychev permet alors d'obtenir la loi faible des grands nombres :

Théorème 8.2 *Soit* (X_n) *une suite de v.a.r. définies sur le même espace* $(\Omega; \mathcal{A}; P)$*, indépendantes, et admettant la même espérance* m *et la même variance* σ^2*. On définit leur moyenne* Z_n *:* $\forall n \in \mathbb{N}^*, Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i$*. Alors :*

$$Z_n \xrightarrow[n \to +\infty]{} m \text{ en probabilité}$$

Démonstration : d'une part,

$$E(Z_n) = E\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E(Xi) = m$$

et d'autre part, grâce à l'indépendance des X_i :

$$V(Z_n) = V\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}V\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

On applique l'inégalité de Bienaymé-Tchebychev à Z_n :

$$\forall \alpha > 0, P(|Z_n - m| > \alpha) \leq \frac{\frac{\sigma^2}{n}}{\alpha^2} \xrightarrow[n \to +\infty]{} 0$$

Remarque : le théorème n'impose pas que les variables X_n aient la même loi, mais seulement qu'elles aient la même espérance et la même variance (et qu'elles soient indépendantes).
La loi faible des grands nombres permet d'obtenir, dans le cas de variables de Bernoulli, le théorème de Bernoulli :

Théorème 8.3 *On considère des variables de Bernoulli indépendantes, associées à la répétition de la même expérience aléatoire (pour laquelle on s'intéresse à un événement, appelé succès, noté S) :*

$$X_n = \begin{cases} 0 & si \quad \text{à la } n^{ième} \text{ épreuve, on a } \overline{S} \\ 1 & si \quad \text{à la } n^{ième} \text{ épreuve, on a } S \end{cases}$$

En notant $p = P(S)$, on a $\forall n \in \mathbb{N}^*, X_n \sim \mathcal{B}(p)$, et :

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{} p \text{ en probabilité}$$

Démonstration : on est dans les conditions du théorème de la loi faible des grands nombres.

$\sum_{i=1}^{n} X_i$ compte le nombre de succès obtenus au cours des n premières épreuves.
Le nombre moyen de succès obtenus au cours des n premières épreuves tend donc à se stabiliser autour de p, ce qui permet en pratique d'affecter la valeur p à $P(S)$.

VIII.3 - Convergence en loi

Définition 8.2 *Soient (X_n) une suite de v.a.r. et X une v.a.r., définies sur le même espace $(\Omega; \mathcal{A}; P)$. On note, pour tout n, F_n la fonction de répartition de X_n, et F la fonction de répartition de X. On dit que (X_n) converge en loi vers X si :*

$$F_n(x) \xrightarrow[n\to+\infty]{} F(x)$$

en tout point de continuité de F.

Remarque : la définition n'impose pas que l'écart entre X_n et X tende vers 0. Elle ne concerne que les lois de ces variables, exprimant la convergence simple de la suite de fonctions F_n vers la fonction F (en tout point de continuité de F).
Les deux modes de convergence ne sont pas équivalents. On a simplement :

Propriété 8.1

$$X_n \xrightarrow[n\to+\infty]{} X \text{ en probabilité} \Rightarrow X_n \xrightarrow[n\to+\infty]{} X \text{ en loi}$$

et la réciproque est fausse.

On admet que la convergence en probabilité entraîne la convergence en loi. Pour voir que la réciproque est fausse, on peut considérer une variable X qui suit une loi $\mathcal{N}(0;1)$, et une suite de variables (X_n) définie par : $\forall n, X_n = -X$. Par symétrie de la densité de X, on a pour tout réel x, $P(X_n \leq x) = P(X \geq -x) = P(X \leq x)$, donc X_n suit une loi $\mathcal{N}(0;1)$, et (X_n) converge en loi vers X. Cependant, il n'y a pas convergence en probabilité : soit $\epsilon > 0$. $P(|X_n - X| > \epsilon) = P\Big(|X| > \dfrac{\epsilon}{2}\Big) > 0$, donc $P(|X_n - X| > \epsilon)$ ne tend pas vers 0 lorsque $n \to +\infty$.

Théorème 8.4 *Approximation de la loi hypergéométrique par la loi binomiale.*
Soit (X_N) une suite de variables aléatoires telle que pour tout $N \geq 1$, $X_N \sim \mathcal{H}(N;n;p)$. Alors la suite (X_N) converge en loi vers une variable aléatoire X qui suit une loi binomiale $\mathcal{B}(n;p)$.

L'interprétation est que, lorsque le nombre de boules est très important, la loi du nombre de succès avec des tirages sans remise tend vers la loi du nombre de succès avec des tirages avec remise.

Démonstration : $X_N \sim \mathcal{N}(N;n;p)$ si :
1. $X_N(\Omega) = [\![\max(0; n - N(1-p)); \min(n; Np)]\!]$.
2. $\forall k \in X_N(\Omega), P(X_N = k) = \dfrac{\begin{pmatrix} Np \\ k \end{pmatrix}\begin{pmatrix} N(1-p) \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$.

Pour N assez grand, on a donc $X_N(\Omega) = [\![0;n]\!]$. Soit $k \in [\![0;n]\!]$:

$$\begin{aligned}
&\frac{\begin{pmatrix} Np \\ k \end{pmatrix}\begin{pmatrix} N(1-p) \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
&= \frac{(Np)!}{k!(Np-k)!} \times \frac{(N(1-p))!}{(n-k)!(N(1-p)-n+k)!} \times \frac{n!(N-n)!}{N!} \\
&= \begin{pmatrix} n \\ k \end{pmatrix} \frac{(Np)!}{(Np-k)!} \times \frac{(N(1-p))!}{(N(1-p)-(n-k))!} \times \frac{(N-n)!}{N!}
\end{aligned}$$

Or $\forall \alpha \in \mathbb{N}, \dfrac{q!}{(q-\alpha)!} = q(q-1)\cdots(q-\alpha+1) \underset{q \to +\infty}{\sim} q^\alpha$, donc :

$$\begin{aligned}
P(X_N = k) &\underset{N \to +\infty}{\sim} \begin{pmatrix} n \\ k \end{pmatrix} \frac{(Np)^k (N(1-p))^{n-k}}{N^n} \\
&= \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}
\end{aligned}$$

c'est-à-dire : $\forall k \in [\![0;n]\!], P(X_N = k) \xrightarrow[N\to+\infty]{} P(X = k)$, où $X \sim B(n;p)$. On en déduit la convergence simple de la suite de fonctions F_N vers F : si $x < 0$, alors $F_N(x) = 0 = F(x)$, et si $x > n$, $F_N(x) = 1 = F(x)$. Pour un réel x de $[0;n]\backslash[\![0;n]\!]$:

$$F_N(x) = \sum_{k=0}^{[x]} P(X_N = k) \xrightarrow[N\to+\infty]{} \sum_{k=0}^{[x]} P(X = k) = F(x)$$

($[x]$ désigne la partie entière de x). La suite (X_N) converge donc en loi vers X.

Théorème 8.5 *Approximation de la loi binomiale par la loi de Poisson.* *Soit (X_n) une suite de v.a.r. telle que $\forall n \geq 1, X_n \sim \mathcal{B}\left(n;\frac{\lambda}{n}\right)$, λ étant un réel strictement positif fixé.*
Alors la suite (X_n) converge en loi vers une variable aléatoire X qui suit une loi de Poisson $\mathcal{P}(\lambda)$.

La suite de variables binomiales $\mathcal{B}(n;p)$ doit donc être telle que np est constant ($np = \lambda$).
Dans la pratique, pour remplacer une loi binomiale par une loi de Poisson, on prend comme condition $n \geq 30$ et p assez faible (tel que $np \leq 5$).

Démonstration : soit $k \in \mathbb{N}$.

$$\begin{aligned}
\forall n \geq k, P(X_n = k) &= \binom{n}{k} p^k(1-p)^{n-k} \\
&= \binom{n}{k} (\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k} \\
&= \frac{1}{k!}\frac{n!}{(n-k)!}\frac{\lambda^k}{n^k}e^{(n-k)\ln(1-\frac{\lambda}{n})} \\
&= \frac{\lambda^k}{k!}\frac{n!}{(n-k)!}\frac{1}{n^k}e^{(n-k)(-\frac{\lambda}{n}+o(\frac{1}{n}))} \\
&\xrightarrow[n\to+\infty]{} \frac{\lambda^k}{k!}e^{-\lambda} \\
&= P(X = k)
\end{aligned}$$

On montre ensuite que (F_n) converge simplement vers F exactement de la même façon que précédemment.

Le théorème qui suit explique l'importance de la loi de Gauss :

Théorème 8.6 *Théorème central limite (TCL).*
Soit (X_n) une suite de v.a.r. indépendantes et de même loi, admettant une espérance m et une variance σ^2.
Soit $S_n = \sum_{i=1}^{n} X_i$ (on a donc $E(S_n) = nm$ et $V(S_n) = n\sigma^2$).
Alors la variable S_n centrée et réduite converge en loi vers une v.a.r. X qui suit une loi $\mathcal{N}(0;1)$, soit :

$$\forall x \in \mathbb{R}, P\left(\frac{S_n - nm}{\sqrt{n}\sigma} \leq x\right) \xrightarrow[n \to +\infty]{} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Remarque : on n'a pas besoin de connaître la loi des X_n.

La démonstration du TCL est tout à fait hors programme. Elle s'appuie sur la notion de fonction caractéristique d'une variable aléatoire :

$$\phi(u) = E\left(e^{-iuX}\right) = \int_{-\infty}^{+\infty} e^{-iux} f(x) dx$$

Exemple 8.1 *Soit (X_i) une suite de variables de Bernoulli $\mathcal{B}(p)$, indépendantes.*
La variable $S_n = \sum_{i=1}^{n} X_i$ suit une loi binomiale $\mathcal{B}(n;p)$.
Alors la variable centrée réduite $S_n^* = \dfrac{S_n - np}{\sqrt{np(1-p)}}$ converge en loi vers X qui suit une loi $\mathcal{N}(0;1)$.
La loi $\mathcal{N}(0;1)$ est donc une approximation de la loi binomiale centrée réduite.
Dans la pratique, on prend comme condition $n \geq 30$ et $np(1-p) \geq 9$ pour remplacer une loi binomiale centrée réduite par une loi normale $\mathcal{N}(0;1)$.
Ce résultat est connu sous le nom de théorème de De Moivre-Laplace. Il a été démontré en 1733 (sans le TCL et les fonctions caractéristiques !) par A. De Moivre. Cette preuve « historique » est très longue, et s'appuie notamment sur une approximation de $n!$.

Remarque : S_n étant discrète, S_n^* l'est aussi, et on a donc une suite de v.a.r. discrètes qui converge vers une v.a.r. continue.

VIII.4 - Convergence presque sûre

Cette partie ne concerne que les candidats à l'Agrégation Interne.

Définition 8.3 *Soient une v.a.r. X et une suite (X_n) de v.a.r. définies sur un même espace probabilisé $(\Omega; \mathcal{A}; P)$. On dit que la suite (X_n) converge* *presque sûrement vers X si*

$$P(\{\omega \in \Omega / \lim_{n \to +\infty} X_n(\omega) = X(\omega)\}) = 1$$

Exemple 8.2 *Soit (X_n) une suite de variables aléatoires définies sur le même espace $(\mathbb{R}; \mathcal{B}_{\mathbb{R}}; P)$ par*

$$X_n(\omega) = \begin{cases} n & si \quad \omega \in \left[0; \frac{1}{n}\right] \\ 0 & sinon \end{cases}$$

On a $X_n(0) = n \to +\infty$, et si $\omega \neq 0$, il existe un entier N tel que si $n \geq N$, alors $\omega \notin \left[0; \frac{1}{n}\right]$, donc $X_n(\omega) \to 0$. Par conséquent, (X_n) converge presque sûrement vers 0.

Théorème 8.7 *Loi forte des grands nombres.*
Soit (X_n) une suite de v.a.r. indépendantes, de même loi, ayant une variance finie, et d'espérance m. On pose pour tout $n \geq 1$, $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors la suite (Z_n) converge presque sûrement vers m.

Ce théorème est admis.

VIII.5 - Énoncé des exercices

Exercice 8.1 *On désire tester un dé afin de savoir s'il n'est éventuellement pas pipé.*
Combien de lancers de ce dé doit-on effectuer pour pouvoir affirmer, avec un risque d'erreur inférieur à 5 %, que la fréquence d'apparition du 6 au cours de ces lancers diffère de $\frac{1}{6}$ d'au plus 0.01 ?

Exercice 8.2 *Soient (X_n) une suite de v.a.r. et X une v.a.r. telles que :*

$$X(\Omega) = \{\alpha_1; \cdots; \alpha_n\} \text{ et } \forall n \in \mathbb{N}, X_n(\Omega) = X(\Omega)$$

Quel est le lien entre les deux propositions :

i (X_n) *converge en loi vers* X.

ii $E(X_n) \xrightarrow[n \to +\infty]{} E(X)$.

Exercice 8.3 *Démonstration à l'aide des probabilités du Théorème de Weierstrass-Stone.*
Il s'agit du théorème suivant : toute fonction continue sur $[0;1]$ *(plus généralement sur un compact* $[a;b]$*) est limite uniforme d'une suite de polynômes.*
Autrement dit, il existe une suite (P_n) *d'éléments de* $\mathbb{R}[X]$ *telle que :*

$$\sup_{x\in[0;1]} |P_n(x) - f(x)| \xrightarrow[n\to+\infty]{} 0$$

On va démontrer que la suite de polynômes (P_n) *définie par :*

$$\forall n \geq 1, P_n(x) = \sum_{k=0}^{n} \binom{n}{k} x^k(1-x)^{n-k} f(\frac{k}{n})$$

convient (ces polynômes sont appelés polynômes de Bernstein). On se place sur un espace probabilisé $(\Omega; \mathcal{A}; P)$*. Pour tout* $x \in [0;1]$*, on considère une suite de v.a.r.* (X_i) *indépendantes et de même loi de Bernoulli de paramètre* x*.*

1. *Soit* $S_n = \sum_{i=1}^{n} X_i$*. Quelle est la loi de* S_n *?*
2. *Montrer que* $E\Big(f\Big(\dfrac{S_n}{n}\Big)\Big) = P_n(x)$*.*
3. *Comme* f *est continue sur le compact* $[0;1]$*, elle est uniformément continue (Théorème de Heine, démonstration par l'absurde...) d'où :*

   $$\forall \epsilon > 0, \exists \alpha > 0, |x-y| \leq \alpha \Rightarrow |f(x) - f(y)| < \epsilon$$

   On pose alors $I = \left\{k / \left|\dfrac{k}{n} - x\right| < \alpha\right\}$*. Calculer* $P\left(\left|\dfrac{S_n}{n} - x\right| \geq \alpha\right)$*, et montrer que :*

   $$|P_n(x) - f(x)| \leq \epsilon + 2MP\left(\left|\frac{S_n}{n} - x\right| \geq \alpha\right)$$

   où $M = \sup_{x\in[0;1]} \{|f(x)|\}$*.*
4. *A l'aide de l'inégalité de Bienaymé-Tchebychev, majorer* $P\left(\left|\dfrac{S_n}{n} - x\right| \geq \alpha\right)$*, et conclure.*

Exercice 8.4 *Inégalité de Markov.*
Soit X *une v.a.r. à valeurs positives, admettant une espérance. Montrer que :*

$$\forall t > 0, P(X \geq t) \leq \frac{E(X)}{t}$$

Indication : s'inspirer de la preuve de l'inégalité de Bienaymé-Tchebychev.

Exercice 8.5 *Soient (X_i) une suite de variables indépendantes de même loi de Poisson $\mathcal{P}(1)$.*

1. *Quelle est la loi de $S_n = \sum_{i=1}^{n} X_i$?*
2. *En utilisant le T.C.L., déterminer $\lim_{n\to+\infty} P(S_n^* \leq 0)$, où S_n^* est la variable centrée réduite associée à S_n : $S_n^* = \dfrac{S_n - E(S_n)}{\sigma(S_n)}$.*
3. *En déduire un équivalent de $\sum_{k=0}^{n} \dfrac{n^k}{k!}$.*

Exercice 8.6 *Montrer que :*

$$\forall x > 0, \int_0^x e^{-\frac{t^2}{2}} dt \geq \sqrt{\frac{\pi}{2}}\Big(1 - \frac{1}{x^2}\Big)$$

VIII.6 - Correction des exercices

Exercice 8.1

On pose, pour $i \in [\![1; n]\!]$:

$$X_i = \begin{cases} 1 & \text{si on obtient 6 au } i^{i\grave{e}me} \text{ lancer} \\ 0 & \text{sinon} \end{cases}$$

(L'univers est $[\![1; 6]\!]^n$, on prend comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et on choisit pour P la probabilité produit, qui permet d'avoir l'indépendance des X_i). Les v.a.r. X_i suivent toutes une loi de Bernoulli $\mathcal{B}(p)$, où p est la probabilité d'obtenir un 6 lorsqu'on lance une fois le dé, et elles sont indépendantes.

La v.a.r. $\sum_{i=1}^{n} X_i$ compte le nombre de succès (nombre de fois où l'on a obtenu 6), donc elle suit une loi binomiale $\mathcal{B}(n; p)$. Pour obtenir la fréquence d'apparition du 6, on pose naturellement :

$$F_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

On calcule son espérance et sa variance :

$$E(F_n) = p \text{ et } V(F_n) = \frac{1}{n^2}V\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

On suppose le dé non pipé et on applique l'inégalité de Bienaymé-Tchebychev :

$$P\left(\left|F_n - \frac{1}{6}\right| \geq 0.01\right) \leq \frac{5}{36n \times 0.01^2}$$

Il suffit donc que $\dfrac{5}{36n \times 0.01^2} \leq 0.05$, soit $n \geq 27778$.

Remarque : on peut aussi penser à utiliser le TCL qui permet d'approcher une loi binomiale par une loi normale : on a ici $nF_n \sim \mathcal{B}(n;p)$, et donc $\dfrac{nF_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to+\infty]{} Y$, où $Y \sim \mathcal{N}(0;1)$ (convergence au sens de la convergence en loi). D'où :

$$\begin{aligned} P\left(\left|F_n - \frac{1}{6}\right| \geq 0.01\right) &= P\left(\frac{|nF_n - \frac{n}{6}|}{\sqrt{n\frac{5}{36}}} \geq \frac{0.01\sqrt{n}}{\sqrt{\frac{5}{36}}}\right) \\ &\xrightarrow[n\to+\infty]{} P\left(|Y| \geq \frac{0.01\sqrt{36n}}{\sqrt{5}}\right) \end{aligned}$$

Il suffit donc (moyennant cette approximation) que n soit tel que $P\left(|Y| \geq \dfrac{0.01\sqrt{36n}}{\sqrt{5}}\right) \leq 0.05$. En utilisant la table de la loi normale, ceci implique $\dfrac{0.01\sqrt{36n}}{\sqrt{5}} \geq 1.96$, soit $n \geq 5336$.

Exercice 8.2

On a $i \to ii$: $E(X_n) = \sum_{k=1}^{p} P(X_n = \alpha_k)\alpha_k$. Or $P(X_n = \alpha_k) \xrightarrow[n\to+\infty]{} P(X = \alpha_k)$.

En effet, $P(X_n = \alpha_k) = F_n(x) - F_n(y)$, où x et y sont deux réels fixés tels que $\alpha_{k-1} < y < \alpha_k < x < \alpha_{k+1}$ (ou $y < \alpha_1 < x < \alpha_2$ si $k = 1$ et $\alpha_{p-1} < y < \alpha_p < x$ si $k = p$), et par définition, la fonction F_n converge simplement vers la fonction F.

En passant à la limite, on a bien :

$$E(X_n) \xrightarrow[n\to+\infty]{} \sum_{k=1}^{p} P(X = \alpha_k)\alpha_k = E(X)$$

La réciproque $ii \Rightarrow i$ est fausse :

il suffit de prendre les X_n de même loi , et X telle que :

α_k	1	2	3
$P(X_n = \alpha_k)$	0.1	0.8	0.1
$P(X = \alpha_k)$	0.2	0.6	0.2

On a $E(X_n) = E(X)$ et (X_n) ne converge pas en loi vers X.

Exercice 8.3

1. $S_n \sim \mathcal{B}(n;x)$.

2. On a par définition de l'espérance : $E(S_n) = \sum_{k=0}^{n} k \begin{pmatrix} n \\ k \end{pmatrix} x^k(1-x)^{n-k}$.

   On applique le théorème de transfert :

$$E\Big(f\Big(\frac{S_n}{n}\Big)\Big) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big) \begin{pmatrix} n \\ k \end{pmatrix} x^k(1-x)^{n-k}$$

3. la variable $\frac{S_n}{n}$ a pour univers image $\frac{S_n}{n}(\Omega) = \Big\{0; \frac{1}{n}; \frac{2}{n}; \cdots; 1\Big\}$. Par conséquent :

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - x\right| \geq \alpha\right) &= \sum_{k \notin I} P\Big(\frac{S_n}{n} = \frac{k}{n}\Big) \\ &= \sum_{k \notin I} P(S_n = k) \\ &= \sum_{k \notin I} \begin{pmatrix} n \\ k \end{pmatrix} x^k(1-x)^{n-k} \end{aligned}$$

   En remarquant que $f(x) = \sum_{k=0}^{n} f(x) \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$, on a :

$$\begin{aligned} |P_n(x) - f(x)| &= \left|\sum_{k=0}^{n} \Big(f\Big(\frac{k}{n}\Big) - f(x)\Big) \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}\right| \\ &\leq \sum_{k \in I} \left|f\Big(\frac{k}{n}\Big) - f(x)\right| \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \\ &\quad + \sum_{k \notin I} \left|f\Big(\frac{k}{n}\Big) - f(x)\right| \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \\ &\leq \sum_{k \in I} \epsilon \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \\ &\quad + 2M \sum_{k \notin I} \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \\ &\leq \epsilon + 2MP\left(\left|\frac{S_n}{n} - x\right| \geq \alpha\right) \end{aligned}$$

4. Comme $E\left(\frac{S_n}{n}\right) = x$ et $V\left(\frac{S_n}{n}\right) = \frac{x(1-x)}{n}$, on a par l'inégalité de Bienaymé-Tchebychev :

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - x\right| \geq \alpha\right) &\leq \frac{x(1-x)}{n\alpha^2} \\ &\leq \frac{1}{n\alpha^2} \end{aligned}$$

Conclusion : $|P_n(x) - f(x)| \leq \epsilon + \frac{2M}{n\alpha^2}$, et donc si n est assez grand : $|P_n(x) - f(x)| \leq 2\epsilon$.

Exercice 8.4

Soit A l'événement : $A = \{\omega / X(\omega) \geq t\}$.
On va utiliser, comme dans la preuve de l'inégalité de Bienaymé-Tchebychev, que $P(A) = E(\mathbf{1}_A)$.
Soit $Z = t\mathbf{1}_A$. Si $\omega \in A$, $Z(\omega) = t$ et $X(\omega) \geq t$, donc $X(\omega) \geq Z(\omega)$.
Si $\omega \notin A$, $Z(\omega) = 0$, et comme X est à valeurs positives, $X(\omega) \geq Z(\omega)$.
On en conclut que : $\forall \omega, X(\omega) \geq Z(\omega)$. Par conséquent, $E(X) \geq E(Z)$, et comme $E(Z) = tE(\mathbf{1}_A) = tP(A)$, on a bien : $E(X) \geq tP(A)$.

Exercice 8.5

1. Comme on l'a démontré au chapitre IV, $S_n \sim \mathcal{P}(n)$.
2. On a donc $E(S_n) = n$ et $V(S_n) = n$. Le T.C.L. s'applique directement à la variable centrée réduite S_n^* :

$$P(S_n^* \leq 0) \xrightarrow[n\to+\infty]{} \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{2}$$

3. Comme $P(S_n^* \leq 0) = P(S_n \leq n) = \sum_{k=0}^{n} P(S_n = k) = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$, on obtient un équivalent : $\sum_{k=0}^{n} \frac{n^k}{k!} \sim_{n\to+\infty} \frac{1}{2} e^n$.

Exercice 8.6

Le membre de droite fait bien sûr penser à la loi normale. Soient $X \sim \mathcal{N}(0;1)$ et x un réel positif fixé. On a :

$$\int_0^x e^{-\frac{t^2}{2}} dt = \sqrt{2\pi} P(0 \leq X \leq x)$$

Par parité de la densité de la loi de X, $P(0 \leq X \leq x) = \dfrac{1}{2}P(-x \leq X \leq x)$.
L'inégalité de Bienaymé-Tchebychev donne alors :

$$P(|X| \geq x) \leq \frac{1}{x^2}$$

D'où $P(|X| \leq x) = 1 - P(|X| \geq x) \geq 1 - \dfrac{1}{x^2}$.
Conclusion :

$$\int_0^x e^{-\frac{t^2}{2}} = \sqrt{\frac{\pi}{2}}P(|X| \leq x) \geq \sqrt{\frac{\pi}{2}}\Big(1 - \frac{1}{x^2}\Big)$$

Chapitre IX : Leçons d'oral (CAPES Externe)

Ce chapitre ne concerne que les candidats au CAPES externe, les leçons de probabilité à l'Agrégation Interne étant différentes dans leur intitulé, leur contenu et leur esprit.

Les leçons qui suivent ne sont pas des exposés type : elles sont souvent trop longues pour être présentées dans le temps imparti, car on a voulu donner toutes les preuves, même les plus élémentaires, et essayer de suggérer plusieurs illustrations. A chaque candidat de construire sa leçon suivant ses goûts et son niveau.

On trouvera en particulier dans la plupart des leçons des commentaires, qui invitent à être attentif à des points particuliers, ou à cadrer l'exposé et permettre au candidat de se préparer à l'entretien avec le jury.

Enfin, un complément est consacré à la leçon sur les séries statistiques doubles.

IX.1 - Utilisation d'arbres, de tableaux, de diagrammes pour des exemples de dénombrement. Dénombrement des arrangements et des permutations.

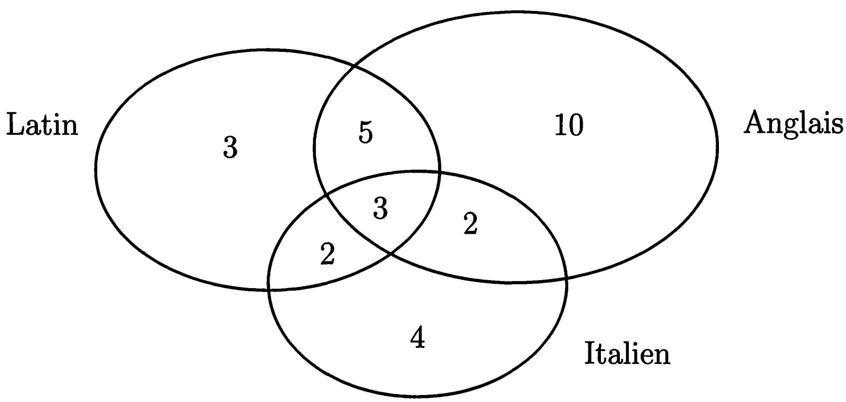
IX.1.1) Utilisation de diagrammes et de tableaux

Utilisation de tableaux : on lance 2 fois un dé équilibré, dont les faces portent les numéros : 1, 1, 2, 2, 3 et 3. On fait la somme des deux numéros obtenus. Quelle est la somme que l'on a le plus de chance d'obtenir ?

	1	2	3
1	2	3	4
2	3	4	5
3	4	5	6

La somme égale à 4 est donc la plus probable.

Utilisation de diagrammes : dans une classe, chaque élève étudie au moins une langue : latin, italien, anglais. 3 étudient les 3 langues, 8 étudient le latin et l'anglais, 5 le latin et l'italien, 5 l'anglais et l'italien, 20 l'anglais, 13 le latin et 11 l'italien. Combien n'étudient que le latin ? Combien la classe comporte-t-elle d'élèves ?



Il y a donc 3 élèves qui n'étudient que le latin, et 29 élèves dans la classe. On utilise, pour construire le diagramme, la relation : $\text{Card}(A \cup B) = \text{Card}\, A + \text{Card}\, B - \text{Card}(A \cap B)$, formule qui découle du « principe de la somme » :

Théorème 9.1 *Si A et B sont deux ensembles finis d'union disjointe, alors $A \cup B$ est fini et :* $\text{Card}(A \cup B) = \text{Card}\, A + \text{Card}\, B$.

Démonstration : soient $n = \text{Card}\, A$ et $p = \text{Card}\, B$. Il existe par définition deux applications $\phi_1 : A \to \{1; \cdots; n\}$ et $\phi_2 : B \to \{1; \cdots; p\}$ bijectives.

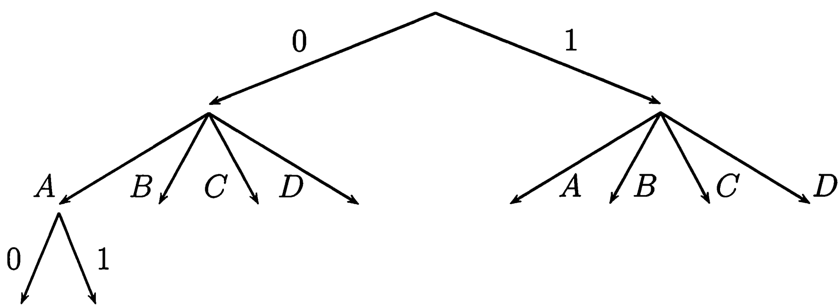
On construit une bijection de $A \cup B$ vers $\{1; \cdots; n+p\}$ en faisant « glisser » la numérotation ϕ_2 :

$$\phi : \left\{ \begin{array}{ccc} A \cup B & \to & \{1; \cdots; n+p\} \\ x & \mapsto & \left\{ \begin{array}{ll} \phi_1(x) & si \quad x \in A \\ n + \phi_2(x) & si \quad x \in B \end{array} \right. \end{array} \right.$$

est bijective, et donc : $\mathrm{Card}(A \cup B) = n + p$.

IX.1.2) Utilisation d'arbres

Exemple 9.1 *Le code d'entrée d'un immeuble est composé d'un chiffre égal à 0 ou 1, puis d'une lettre choisie parmi $\{A; B; C; D\}$, puis d'un chiffre égal à 0 ou 1. Combien y a-t-il de codes possibles ?*
On raisonne sur l'arbre suivant :



Un arbre est donc une succession d'étapes $E_1, \cdots, E_p$. L'étape 1 consiste à choisir un élément parmi n_1 éléments possibles. Ensuite, chaque étape E_k consiste à choisir un élément pour chacun des éléments choisis en E_{k-1} : on suppose que le nombre de possibilités est le même pour chacun des éléments choisis en E_{k-1}, et on le note n_k. Alors, le nombre total de possibilités (de « branches de l'arbre ») est $n_1 n_2 \cdots n_p$: c'est le « principe du produit ».
Le nombre de codes possibles est donc : $2 \times 4 \times 2 = 16$.

Mathématiquement, on se donne p ensembles finis $A_1, \cdots, A_p$ et on veut calculer le nombre d'éléments $(a_1; \cdots; a_p) \in A_1 \times \cdots \times A_p$.

Définition 9.1 *Un arbre est un produit cartésien d'ensembles finis $A_1 \times \cdots \times A_p$; une branche est un p-uplet $(a_1; \cdots; a_p) \in A_1 \times \cdots \times A_p$.*

Le « principe du produit » est donc le théorème suivant :

Théorème 9.2 *Si $A_1, \ldots, A_p$ sont des ensembles finis, alors $A_1 \times \cdots \times A_p$ est un ensemble fini et* $\mathrm{Card}(A_1 \times \cdots \times A_p) = \displaystyle\prod_{i=1}^{p} \mathrm{Card}\, A_i$.

Démonstration : on commence par montrer que si A et B sont deux ensembles finis, alors $A \times B$ est fini et $\mathrm{Card}(A \times B) = \mathrm{Card}\, A \times \mathrm{Card}\, B$.
On note $n = \mathrm{Card}\, A$, $p = \mathrm{Card}\, B$, et on procède par récurrence sur p.
Si $p = 1$: comme $\mathrm{Card}\, A = n$, il existe une bijection $\phi_1 : A \to \{1; \cdots; n\}$.
L'application :

$$\phi : \left\{ \begin{array}{ccc} A \times B & \to & \{1; \cdots; n\} \\ (a; b) & \mapsto & \phi_1(a) \end{array} \right.$$

est une bijection, et donc : $\mathrm{Card}(A \times B) = n$.
Hérédité : on note $A = \{a_1; \cdots; a_n\}$ et $B = \{b_1; \cdots; b_{p+1}\}$.
On décompose $A \times B$ en union de deux ensembles disjoints et on utilise l'hypothèse de récurrence : $A \times B = (A \times \{b_1; \cdots; b_p\}) \cup (A \times \{b_{p+1}\})$ et donc $\mathrm{Card}(A \times B) = np + n = n(p+1)$.
Le théorème est donc démontré pour deux ensembles. On étend ensuite le résultat à un nombre quelconque d'ensembles par récurrence.

Définition 9.2 *Soit $p \geq 1$. On appelle p-liste d'un ensemble A tout élément $(a_1; \cdots; a_p)$ de A^p, c'est-à-dire tout p-uplet d'éléments de A.*

Exemple 9.2 *On jette 4 fois un dé à 6 faces : $A = \{1; \cdots; 6\}$, et $(1; 2; 2; 2)$ est une 4-liste de A.*

Propriété 9.1 *Si A est fini, alors l'ensemble des p-listes d'éléments de A est fini et a pour cardinal $(\mathrm{Card}\, A)^p$.*

Démonstration : on utilise le théorème sur le cardinal du produit cartésien où l'on pose $A_1 = \cdots = A_p = A$.

Exemple 9.3 *Pour les 4 lancers de dés, on a donc 6^4 résultats possibles.*

Propriété 9.2 *Une p-liste d'éléments de A s'identifie à une application : $\{1; \cdots; p\} \to A$.*

Démonstration : l'application

$$\phi : \left\{ \begin{array}{ccc} A^p & \to & A^{\{1; \cdots; p\}} \\ (a_1; \cdots; a_p) & \mapsto & f : \left\{ \begin{array}{ccc} \{1; \cdots; p\} & \to & A \\ i & \mapsto & a_i \end{array} \right. \end{array} \right.$$

est bijective. On en déduit la propriété suivante :

Propriété 9.3 *Si A et B sont deux ensembles finis, alors le nombre d'applications de B vers A est $(\mathrm{Card}\, A)^{\mathrm{Card}\, B}$.*

IX.1.3) Arrangements

Définition 9.3 *Soit A un ensemble non vide. On appelle arrangement de p éléments de A toute p-liste $(a_1; \cdots ; a_p)$ d'éléments de A deux à deux distincts.*

Comme tout arrangement est une p-liste, on a que, si A est fini, l'ensemble des arrangements est également fini.

Notation : on note $\mathcal{A}_n^p$ l'ensemble des arrangements de p éléments d'un ensemble de cardinal n, et $A_n^p = \operatorname{Card} \mathcal{A}_n^p$.

Théorème 9.3

$$A_n^p = \begin{cases} 0 & si \quad p > n \\ \dfrac{n!}{(n-p)!} & si \quad p \leq n \end{cases}$$

Démonstration : si $p > n$, c'est évident puisque les éléments doivent être deux à deux distincts.
Si $p \leq n$: on procède par récurrence finie sur p.
Si $p = 1$, $A_n^1 = \operatorname{Card} A = n$.
Hérédité :

$$\phi : \begin{cases} \mathcal{A}_n^{p+1} & \to & \mathcal{A}_n^p \times \mathcal{A}_{n-p}^1 \\ (a_1; \cdots ; a_p; a_{p+1}) & \mapsto & (a_1; \cdots ; a_p) \times (a_{p+1}) \end{cases}$$

est bijective, et par conséquent :

$$A_n^{p+1} = A_n^p A_{n-p}^1 = \frac{n!}{(n-p)!} \frac{(n-p)!}{(n-p-1)!} = \frac{n!}{(n-(p+1))!}$$

Propriété 9.4 *Tout arrangement de p éléments de A s'identifie à une injection de $\{1; \cdots ; p\}$ vers A.*

Démonstration : on reprend la même idée que pour les p-listes.

$$\phi : \begin{cases} \mathcal{A}_n^p & \to & Inj(\{1; \cdots\} \to A) \\ (a_1; \cdots ; a_p) & \mapsto & f : \begin{cases} \{1; \cdots ; p\} & \to & A \\ i & \mapsto & a_i \end{cases} \end{cases}$$

est une bijection.
L'application f est bien injective car $f(i) = f(j) \Rightarrow a_i = a_j \Rightarrow i = j$ (puisque $a_1, \cdots , a_p$ sont deux à deux distincts).

Exemple 9.4 *Le nombre de mots de 3 lettres distinctes (ayant un sens ou non...) est* $A_{26}^3 = 26 \times 25 \times 24$.

Exemple 9.5 *Le problème des anniversaires. On considère un groupe de n personnes et on veut déterminer la probabilité qu'au moins deux personnes fêtent leur anniversaire en même temps.*
On suppose que toutes les années comportent 365 jours, et que les 365 jours de naissance possibles sont équiprobables...(pour la modélisation probabiliste, on pose $\Omega = [\![1;365]\!]^n$, $\mathcal{A} = \mathcal{P}(\Omega)$ *et* P *la probabilité uniforme).*
Soit A *: « au moins deux personnes fêtent leur anniversaire en même temps ».*

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{\mathrm{Card}\,\overline{A}}{\mathrm{Card}\,\Omega} = 1 - \frac{A_{365}^n}{365^n}$$

($\simeq 0.5$ *pour* $n = 23$*...).*

IX.1.4) Permutations

Définition 9.4 *Soit* A *un ensemble de cardinal* n. *On appelle permutation tout arrangement des* n *éléments de* A.

De l'étude des arrangements, on déduit immédiatement que le nombre de permutations d'un ensemble A de cardinal n est $n!$, et que toute permutation de A s'identifie à une bijection : $\{1; \cdots; n\} \to A$.

Exemple 9.6 *Nombre d'anagrammes de « admis » (*$5!$*), puis de « admissibles » (*$\frac{11!}{2!3!}$*).*

Exemple 9.7 *Dans le cas où* $A = \{1; \cdots; n\}$, *on note* S_n *l'ensemble des bijections de* $\{1; \cdots; n\}$ *dans lui-même. On a que* $(S_n; \circ)$ *est un groupe fini de cardinal* $n!$.

Commentaire : si on présente cet exemple, il faut avoir quelques idées sur les propriétés de $(S_n; \circ)$: non commutativité si $n \geq 3$, engendrement par exemple par les transpositions, et savoir décrire ses éléments au moins pour $n = 2$ et $n = 3$...

Exemple 9.8 *L'ensemble des isométries qui conservent un triangle équilatéral est isomorphe à* S_3 *(on ramène le problème à la permutation des 3 sommets, voir exposés de géométrie affine) et comporte donc 6 éléments : l'identité, 3 symétries axiales (qui correspondent aux 3 transpositions) et 2 rotations (qui correspondent aux 2 3-cycles).*

IX.2 - Coefficients binomiaux, dénombrement des combinaisons, formule du binôme. Applications.

IX.2.1) Combinaisons

Définition 9.5 *Soit E un ensemble de cardinal n. On appelle combinaison de k éléments de E toute partie de E de cardinal k.*

Notation : on note $\begin{pmatrix} n \\ k \end{pmatrix}$ le nombre de combinaisons de k éléments de E.

Théorème 9.4 *Pour tout entier n,*

$$\begin{pmatrix} n \\ k \end{pmatrix} = \begin{cases} 0 & si \quad k > n \\ \dfrac{n!}{k!(n-k)!} & si \quad k \leq n \end{cases}$$

Démonstration : si $k > n$, c'est évident.
Si $k \leq n$, on procède par récurrence sur n : soit la propriété P_n :

$$\forall k \leq n, \begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$$

Pour $n = 0$, on a bien $\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 1$, la seule partie de l'ensemble vide étant l'ensemble vide.
On montre ensuite que : $\forall n \geq 0, P_n \Rightarrow P_{n+1}$.
Pour $k = 0$, il est clair que $\begin{pmatrix} n+1 \\ 0 \end{pmatrix} = 1$.
Pour $k \geq 1$: soit $E = \{a_1; \cdots; a_{n+1}\}$ un ensemble de cardinal $n + 1$.
L'ensemble des parties à k éléments de E s'écrit comme l'union disjointe de l'ensemble des parties à k éléments qui contiennent a_{n+1} et de l'ensemble des parties à k éléments qui ne contiennent pas a_{n+1}.
Soit A une partie (à k éléments) ne contenant pas a_{n+1} : A est une partie de $\{a_1; \cdots; a_n\}$ et on a donc $\begin{pmatrix} n \\ k \end{pmatrix}$ parties de ce type.
Soit B une partie (à k éléments) contenant a_{n+1} : $B = B' \cup \{a_{n+1}\}$ où B' est une partie de $k - 1$ éléments de $\{a_1; \cdots; a_n\}$, et on a par conséquent $\begin{pmatrix} n \\ k-1 \end{pmatrix}$ parties de ce type.
On a donc $\begin{pmatrix} n+1 \\ k \end{pmatrix} = \begin{pmatrix} n \\ k \end{pmatrix} + \begin{pmatrix} n \\ k-1 \end{pmatrix}$, et en utlisant l'hypothèse de

récurrence :

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

Propriété 9.5 *Les coefficients binomiaux vérifient :*

1. $\forall n \geq 0, \binom{n}{0} = \binom{n}{n} = 1$

et $\forall 1 \leq k \leq n$ *:*

2. $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$

3. $\binom{n}{k} = \binom{n}{n-k}$

4. $k\binom{n}{k} = n\binom{n-1}{k-1}$

Démonstration : ces résultats s'obtiennent soit par le calcul direct (aucune difficulté), soit par le dénombrement :

1. C'est évident.
2. Le résultat a été démontré précédemment (lors de la preuve de l'expression de $\binom{n}{k}$).
3. On définit $\phi : \left\{\begin{array}{ccc} \mathcal{P}(E) & \to & \mathcal{P}(E) \\ A & \mapsto & \overline{A} \end{array}\right.$. ϕ est bijective, car $\phi \circ \phi = Id$ (on dit que ϕ est une involution), et si on note $\mathcal{P}_i$ l'ensemble des parties de E de cardinal i, on a $\phi(\mathcal{P}_k) = \mathcal{P}_{n-k}$. Par conséquent $\mathrm{Card}\,\mathcal{P}_k = \mathrm{Card}\,\mathcal{P}_{n-k}$.
4. On dénombre les couples $(a; A)$ où A est une partie de E de cardinal k, et a un élément de A.
   Si on fixe A, on a k éléments a possibles, d'où $k\binom{n}{k}$ couples.
   Si on fixe a, on a $\binom{n-1}{k-1}$ parties A possibles, d'où $n\binom{n-1}{k-1}$ couples.

La relation $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$ permet de construire ligne après ligne le triangle de Pascal qui donne les valeurs des $\binom{n}{k}$:

$n\backslash k$	0	1	2	3
0	1	0	0	0
1	1	1	0	0
2	1	2	1	0
3	1	3	3	1

Exemple 9.9 *Le nombre de tirages possibles au Loto est* $\begin{pmatrix} 49 \\ 6 \end{pmatrix} = 13983816$.

Exemple 9.10 *Étant donnés deux entiers a et b, le nombre de façons d'aller de $(0;0)$ à $(a;b)$ en se déplaçant d'une unité vers le haut ou vers la droite à chaque fois est :* $\begin{pmatrix} a+b \\ a \end{pmatrix}$.

IX.2.2) Formule du binôme

Théorème 9.5 *Soit $n \in \mathbb{N}$. $\forall (a;b) \in \mathbb{C}^2, (a+b)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} a^k b^{n-k}$.*

Démonstration : on procède par récurrence sur n. Pour $n = 0$, c'est évident, et on montre l'hérédité en développant :

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)(a+b)^n \\
&= (a+b)\sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} a^k b^{n-k} \\
&= \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} a^k b^{n+1-k}
\end{aligned}$$

Après un changement d'indice dans la première somme :

$$\begin{aligned}
(a+b)^{n+1} &= \sum_{k=1}^{n+1}\binom{n}{k-1}a^kb^{n+1-k}+\sum_{k=0}^{n}\binom{n}{k}a^kb^{n+1-k}\\
&= \sum_{k=1}^{n}\underbrace{\left(\binom{n}{k-1}+\binom{n}{k}\right)}_{\binom{n+1}{k}}a^kb^{n-k}+\underbrace{\binom{n}{n}}_{\binom{n+1}{n+1}}a^{n+1}\\
&\quad+\underbrace{\binom{n}{0}}_{\binom{n+1}{0}}b^{n+1}\\
&= \sum_{k=0}^{n+1}\binom{n+1}{k}a^kb^{n+1-k}
\end{aligned}$$

Remarque : la formule est vraie plus généralement dans un anneau commutatif. Si l'anneau n'est pas commutatif, elle est vraie pour deux éléments a et b qui commutent.

Les résultats qui suivent sont une conséquence directe de la formule du binôme :

Propriété 9.6

1. $\displaystyle\sum_{k=0}^{n}\binom{n}{k}=2^n.$

2. $\displaystyle\sum_{k=0}^{n}(-1)^k\binom{n}{k}=0.$

3. $\displaystyle\sum_{k\ pairs,k=0}^{n}\binom{n}{k}=\sum_{k\ impairs,k=0}^{n}\binom{n}{k}=2^{n-1}.$

4. $\displaystyle\sum_{k=0}^{n}k\binom{n}{k}=n2^{n-1}.$

Démonstration :

1. Développer $(1+1)^n$.

Remarque : le nombre de parties d'une ensemble E de cardinal n est donc 2^n, ce qui se montre directement avec la bijection :

$$\begin{cases} \mathcal{P}(E) & \to & \{0;1\}^E \\ A & \mapsto & \mathbf{1}_A \end{cases}$$

2. Développer $(1-1)^n$.

   Remarque : de la même façon, on peut trouver la somme sur les k multiples de 3 de $\binom{n}{k}$ en développant $(1+j)^n$.

3. Séparer la somme précédente en distinguant k pair et k impair.
4. Développer $(1+x)^n$, puis dériver et évaluer en $x=1$.

IX.2.3) Applications

- **Formule de Vandermonde :**

$$\forall\, 0 \leq p \leq n+m, \sum_{k=0}^{p} \binom{n}{k}\binom{m}{p-k} = \binom{n+m}{p}$$

Démonstration : soient E de cardinal n et F de cardinal m deux ensembles disjoints. Dénombrer les parties à p éléments de $E \cup F$.

Remarque : on peut aussi chercher le coefficient de x^p dans le développement de $(x+1)^n(x+1)^m$.
En particulier, $\displaystyle\sum_{k=0}^{n}\left(\binom{n}{k}\right)^2 = \binom{2n}{n}$.

- **Linéarisation de** $(\cos x)^n$, $(\sin x)^n$ **:**

Par exemple :

$$\begin{aligned}
(\cos x)^3 &= \left(\frac{e^{ix}+e^{-ix}}{2}\right)^3 \\
&= \frac{1}{8}\sum_{k=0}^{3}\binom{3}{k} e^{ikx}e^{-i(3-k)x} \\
&= \frac{1}{8}(e^{-i3x}+3e^{-ix}+3e^{ix}+e^{i3x}) \\
&= \frac{1}{4}\cos 3x + \frac{3}{4}\cos x
\end{aligned}$$

Intérêt : par exemple, calculs de primitives.

– **Théorème de Fermat :**

Théorème. *Soit p premier. Alors :*

$$\forall a \in \mathbb{N}, a^p \equiv a[p]$$

Démonstration : récurrence sur a.
Si $a = 0$, le résultat est clair. Pour l'hérédité, on développe par la formule du binôme :

$$(a+1)^p = \sum_{k=0}^{p} \begin{pmatrix} p \\ k \end{pmatrix} a^k$$

Or, si $k \in [\![1; p-1]\!]$, p divise $\begin{pmatrix} p \\ k \end{pmatrix}$. En effet, $k! \begin{pmatrix} p \\ k \end{pmatrix} = p(p-1)\cdots$ $(p-k+1)$ donc p divise $k! \begin{pmatrix} p \\ k \end{pmatrix}$. Comme p est premier avec tous les entiers de 1 à k, on a, par le théorème de Gauss, que p divise $\begin{pmatrix} p \\ k \end{pmatrix}$.
Par conséquent :

$$\begin{aligned} (a+1)^p &\equiv a^p + 1[p] \\ &\equiv a + 1[p] \end{aligned}$$

– **Inégalité de Bernoulli :**

$$\forall n \in \mathbb{N}, \forall x \geq 0, (1+x)^n \geq 1 + nx$$

Démonstration : développer.

– **Calcul de $S_\alpha = \sum_{k=1}^{n} k^\alpha, \alpha \in \mathbb{N}$:**

On peut obtenir les S_α de proche en proche grâce à la formule du binôme.
On connaît :

$$S_1 = \sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

$(S_1 + S_1 = (1 + \cdots + n) + (n + \cdots + 1) = n(n+1)$, puisqu'on a n fois le terme $n+1)$.
Calcul de S_2 :

$$\begin{aligned} (n+1)^3 &= n^3 + 3n^2 + 3n + 1 \\ n^3 &= (n-1)^3 + 3(n-1)^2 + 3(n-1) + 1 \\ &\vdots \\ 2^3 &= 1^3 + 3 \times 1^2 + 3 \times 1 + 1 \end{aligned}$$

En additionnant ces n équations :

$$(n+1)^3 \quad = \quad 1 + 3S_2 + 3S_1 + n$$

Il reste à remplacer S_1 par $\dfrac{n(n+1)}{2}$ pour obtenir :

$$S_2 = \frac{n(n+1)(2n+1)}{6}$$

Pour S_3, développer $(n+1)^4 \cdots$ (résultat : $S_3 = \dfrac{n^2(n+1)^2}{4}$).

IX.3 - Description mathématique d'une expérience aléatoire : ensemble des événements élémentaires, événements, probabilité (on se limitera au cas où l'ensemble des événements élémentaires est fini).

Commentaire : il faut tirer toutes les conséquences du fait que Ω est supposé fini :
on prendra toujours comme tribu $\mathcal{P}(\Omega)$ (il est donc inutile de parler de tribu) et on appellera événement toute partie de Ω.
Il n'y a aucun passage à la limite à évoquer : $\sum p_i$ comporte un nombre fini de termes, et pour l'additivité, il n'y a pas d'union infinie puisqu'on n'a qu'un nombre fini d'événements.

Il s'agit de décrire une expérience aléatoire, donc c'est vous qui choisissez pour chacun de vos exemples un Ω cohérent et installez une probabilité P (justifiée par exemple par les indications de votre énoncé : dé équilibré..., ou par la loi faible des grands nombres...).

IX.3.1) Expérience aléatoire, univers

Définition 9.6 *On appelle expérience aléatoire une expérience sur un système dont le résultat n'est pas connu d'avance et peut varier si on répète cette expérience.*

Exemple 9.11 *Jeter une pièce de monnaie, lancer des dés, prélever des boules dans une urne...*

Le résultat, par hypothèse unique, de la réalisation de l'expérience aléatoire est noté ω.

Définition 9.7 *On appelle univers l'ensemble des résultats possibles. Il est noté Ω.*

Exemple 9.12 *On effectue deux jets successifs d'un dé : $\Omega = [\![1;6]\!]^2$.*

La difficulté vient du fait qu'il est possible, pour une même expérience aléatoire, de définir plusieurs univers, suivant ce que l'on entend par le terme « résultat possible ». Par exemple, pour une expérience aléatoire consistant à prélever une boule dans une urne contenant 2 boules rouges et 3 boules noires, on peut considérer qu'un résultat possible est l'une des 5 boules ($\Omega = \{R_1; R_2; N_1; N_2; N_3\}$) ou une couleur ($\Omega = \{N; R\}$).

Notation : on note $\Omega = \{\omega_1; \cdots; \omega_n\}$ dans la suite.

IX.3.2) Événements

Définition 9.8 *On appelle événement toute partie de Ω, et événement élémentaire tout singleton de Ω.*

Exemple 9.13 *On lance deux fois un dé à six faces. On pose $\Omega = [\![1; 6]\!]^2$.*
On considère A : « la somme obtenue est supérieure ou égale à 11 ».
On a : $A = \{(5; 6); (6; 5); (6; 6)\}$ et A est un événement.
On considère B : « la somme obtenue est divisible par 3 et par 4 ».
On a : $B = \{(6; 6)\}$ et B est un événement élémentaire.

Commentaire : ω_i est un résultat possible (ou éventualité) et $\{\omega_i\}$ est un événement élémentaire.

On utilise la correspondance suivante entre opérations sur les événements et langage ensembliste :

Terminologie probabiliste	Terminologie ensembliste	Notation
Événement certain	Ensemble entier	Ω
Événement impossible	Ensemble vide	$\emptyset$
Événement élémentaire	Singleton	$\{\omega\}$
Événement contraire de A	Complémentaire de A	$\overline{A}$
A ou B	Réunion de A et B	$A \cup B$
A et B	Intersection de A et B	$A \cap B$
A implique B	A inclus dans B	$A \subset B$
A et B incompatibles	A et B disjoints	$A \cap B = \emptyset$
ω réalise A	ω appartient à A	$\omega \in A$

IX.3.3) Probabilité

Définition 9.9 *On appelle probabilité sur $(\Omega; \mathcal{P}(\Omega))$ toute application P : $\mathcal{P}(\Omega) \to \mathbb{R}$ telle que :*

1. *$\forall A \in \mathcal{P}(\Omega), P(A) \geq 0$ (positivité)*
2. *$P(\Omega) = 1$ (totalité)*
3. *$\forall A$ et B disjoints, $P(A \cup B) = P(A) + P(B)$ (additivité)*

Commentaire : P est bien définie sur $\mathcal{P}(\Omega)$ et non sur Ω, et donc $P(\omega_i)$ n'a aucun sens : il faut être attentif à bien écrire $P(\{\omega_i\})$.

Propriété 9.7 *Toute probabilité P vérifie :*

1. $P(\emptyset) = 0$
2. $\forall A \in \mathcal{P}(\Omega), P(\overline{A}) = 1 - P(A)$
3. $A \subset B \Rightarrow P(A) \leq P(B)$
4. $\forall A \in \mathcal{P}(\Omega), P(A) \leq 1$
5. *Si $A_1, \cdots, A_k$ sont 2 à 2 incompatibles, alors :*

$$P(A_1 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i)$$

Démonstration : ces propriétés découlent immédiatement de la définition :

1. $P(\emptyset \cup \Omega) = P(\Omega) + P(\emptyset)$ et $\emptyset \cup \Omega = \Omega$.
2. $\Omega = A \cup \overline{A} \Rightarrow P(\Omega) = P(A) + P(\overline{A})$.
3. $B = A \cup (B \cap \overline{A})$ et l'union étant disjointe :
   $P(B) = P(A) + P(B \cap \overline{A})$, et $P(B \cap \overline{A}) \geq 0$.
4. $A \subset \Omega \Rightarrow P(A) \leq P(\Omega)$.
5. Immédiat par récurrence.

Notation : on note, pour $i \in [\![1; n]\!]$, $p_i = P(\{\omega_i\})$.

Propriété 9.8 *P est entièrement déterminée par la donnée de $(p_i)_{i=1,\cdots,n}$.*

Démonstration : par additivité de P, $\forall A \in \mathcal{P}(\Omega), A \neq \emptyset, P(A) = \sum_{i/\omega_i \in A} p_i$.

Remarque : on a nécessairement $\forall i \in [\![1; n]\!], p_i \geq 0$ et $\sum_{i=1}^{n} p_i = 1$.

Propriété 9.9 *Réciproquement, étant donnés $p_1, \cdots, p_n$ n réels positifs de somme 1, l'application P définie par :*

$$\forall A \in \mathcal{P}(\Omega), A \neq \emptyset, P(A) = \sum_{i/\omega_i \in A} p_i \text{ et } P(\emptyset) = 0$$

est une probabilité sur $(\Omega; \mathcal{P}(\Omega))$.

Démonstration : on vérifie les 3 axiomes.
Positivité : $(\forall i \in [\![1; n]\!], p_i \geq 0) \Rightarrow P(A) \geq 0$.

Totalité : $P(\Omega) = \sum_{i=1}^{n} p_i = 1$.

Additivité : Soient A et B d'intersection vide.

$$P(A \cup B) = \sum_{i/\omega_i \in A \cup B} p_i = \sum_{i/\omega_i \in A} p_i + \sum_{i/\omega_i \in B} p_i = P(A) + P(B)$$

Propriété 9.10 *Formule de Poincaré. Soient A_1 et A_2 deux événements. Alors $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.*

Démonstration :

$$\begin{aligned}
A_1 \cup A_2 &= A_1 \cup (A_2 \cap \overline{A}_1) \text{ (union disjointe)} \\
\Rightarrow P(A_1 \cup A_2) &= P(A_1) + P(A_2 \cap \overline{A}_1) \\
\text{or : } A_2 &= (A_2 \cap A_1) \cup (A_2 \cap \overline{A}_1) \text{ (union disjointe)} \\
\Rightarrow P(A_2) &= P(A_2 \cap A_1) + P(A_2 \cap \overline{A}_1) \\
\text{d'où : } P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2)
\end{aligned}$$

IX.3.4) Cas particulier : la probabilité uniforme

Définition 9.10 *On appelle probabilité uniforme la probabilité P qui vérifie :*

$$P(\{\omega_1\}) = \cdots = P(\{\omega_n\})$$

Remarque : puisque la somme doit être égale à 1, on a évidemment :

$$\forall i \in [\![1; n]\!], P(\{\omega_i\}) = \frac{1}{n}$$

Propriété 9.11 *Si P est la probabilité uniforme, alors :*

$$\forall A \in \mathcal{P}(\Omega), P(A) = \frac{\operatorname{Card} A}{\operatorname{Card} \Omega}$$

Démonstration : on a bien $P(\emptyset) = 0$ et si $A \neq \emptyset$:

$$P(A) = \sum_{i/\omega_i \in A} P(\{\omega_i\}) = \sum_{i/\omega_i \in A} \frac{1}{n} = \frac{1}{n} \operatorname{Card} A$$

Commentaire : il n'y a pas que la probabilité uniforme sur $\mathcal{P}(\Omega)$. Ce sont des considérations relatives à l'expérience (boules indiscernables, dé équilibré) qui conduisent souvent à choisir P uniforme, mais on n'est pas toujours en situation d'équiprobabilité (voir exemple 9.15).

IX.3.5) Exemples

Exemple 9.14 *On tire simultanément 5 cartes parmi 32 au hasard. Quelle est la probabilité d'avoir 5 cœurs ?*
Soit E l'ensemble des 32 cartes. On pose comme univers :

$$\Omega = \{M \in \mathcal{P}(E) / \operatorname{Card} M = 5\}$$

On choisit pour P la probabilité uniforme, et on a, en notant C l'événement « on a obtenu 5 cœurs » :

$$P(C) = \frac{\operatorname{Card} C}{\operatorname{Card} \Omega} = \frac{\begin{pmatrix} 8 \\ 5 \end{pmatrix}}{\begin{pmatrix} 32 \\ 5 \end{pmatrix}}$$

Exemple 9.15 *Une urne U_1 contient une boule rouge (notée R_1) et une urne U_2 contient une boule rouge (notée R_2) et une boule noire (notée N_1) indiscernables au toucher. On choisit une urne au hasard, puis on prélève une boule dans l'urne choisie. Quelle est la probabilité d'obtenir une boule rouge ?*
Comme on a deux expériences aléatoires successives (choix d'une urne, puis choix d'une boule), l'univers va être un ensemble de couples. On pose ici :

$$\Omega = \{(U_1; R_1); (U_2; N_1); (U_2; R_2)\}$$

Choix de P : on note $p_1 = P(\{(U_1; R_1)\})$, $p_2 = P(\{(U_2; N_1)\})$ et $p_3 = P(\{(U_2; R_2)\})$.
D'après l'énoncé : $p_1 = p_2 + p_3$ (la probabilité de choisir U_1 est la même que celle de choisir U_2) et $p_2 = p_3$ (boules « indiscernables au toucher » dans U_2).
De plus : $p_1 + p_2 + p_3 = 1$, donc on obtient : $p_1 = \frac{1}{2}$ et $p_2 = p_3 = \frac{1}{4}$.

Commentaire *: P n'est donc pas uniforme !*

On a alors, en notant R l'événement « on a obtenu une boule rouge » :

$$P(R) = P(\{(U_1; R_1)\} \cup \{(U_2; R_2)\}) = p_1 + p_3 = \frac{3}{4}$$

Exemple 9.16 *Problème posé à Pascal par le chevalier de Méré.*
« Qu'est ce qui est le plus probable : sortir au moins un 6 en lançant 4 fois un dé ou au moins un double 6 en lançant 24 fois deux dés ? »
Premier cas : *4 lancers successifs d'un dé (équilibré).*

On répète 4 fois la même expérience aléatoire et on pose donc $\Omega = [\![1;6]\!]^4$*.*
On prend comme probabilité P *la probabilité uniforme.*
Soit A *l'événement « le résultat obtenu comporte au moins un 6 ».*
On raisonne sur l'événement contraire :

$$P(\overline{A}) = \frac{\mathrm{Card}(\overline{A})}{\mathrm{Card}(\Omega)} = \frac{5^4}{6^4}$$

On a donc :

$$P(A) = 1 - \frac{5^4}{6^4} \simeq 0.518$$

Deuxième cas *: 24 lancers successifs de deux dés (équilibrés).*
On répète 24 fois la même expérience aléatoire et on pose donc $\Omega = ([\![1;6]\!]^2)^{24}$*. On prend comme probabilité* P *la probabilité uniforme.*
Soit A *l'événement « le résultat obtenu comporte au moins un double 6 ».*
On raisonne encore sur l'événement contraire :

$$P(\overline{A}) = \frac{\mathrm{Card}(\overline{A})}{\mathrm{Card}(\Omega)} = \frac{35^{24}}{36^{24}}$$

On a donc :

$$P(A) = 1 - \frac{35^{24}}{36^{24}} \simeq 0.491$$

Exemple 9.17 *Problème posé à Galilée par le Duc de Toscane.*
« Pourquoi, quand on effectue trois lancers de dé, obtient-on plus souvent la somme 10 que la somme 9, bien que chacune soit obtenue de 6 manières différentes ? »
On répète 3 fois la même expérience aléatoire et on pose donc $\Omega = [\![1;6]\!]^3$*.*
On prend comme probabilité P *la probabilité uniforme.*
On note S_{10} *« la somme des 3 éléments du triplet vaut 10 » et* S_9 *« la somme des 3 éléments du triplet vaut 9 ».*
S_{10} *est composé des éléments* $(1;3;6), (1;4;5), (2;3;5), (2;4;4), (2;6;2)$ *et* $(3;3;4)$*, plus tous les triplets obtenus en permutant les éléments de ces 6 triplets. Pour chacun des triplets* $(1;3;6), (1;4;5)$ *et* $(2;3;5)$*, on a 3! permutations, et pour chacun des triplets* $(2,4,4), (2;6;2)$ *et* $(3;3;4)$*, on a 3 permutations.*
Par conséquent :

$$\mathrm{Card}(S_{10}) = 3 \times (3!) + 3 \times 3 = 27$$

En dénombrant les éléments de S_9 *de la même façon, on obtient :*

$$\mathrm{Card}(S_9) = 3 \times (3!) + 2 \times 3 + 1 = 25$$

On en conclut que $P(S_9) < P(S_{10})$*.*

Exemple 9.18 *Le problème des anniversaires. On considère un groupe de n personnes et on veut déterminer la probabilité qu'au moins deux personnes fêtent leur anniversaire en même temps. On suppose que toutes les années comportent 365 jours, et que les 365 jours de naissance possibles sont équiprobables...*
On pose $\Omega = [\![1;365]\!]^n$, et on choisit pour P la probabilité uniforme.
Soit A : « au moins deux personnes fêtent leur anniversaire en même temps ».

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{\operatorname{Card}\overline{A}}{\operatorname{Card}\Omega} = 1 - \frac{A_{365}^n}{365^n}$$

($\simeq 0.5$ pour $n = 23$...).

IX.4 - Probabilité conditionnelle ; indépendance de 2 événements (on se limitera au cas où l'ensemble d'épreuves est fini). Applications à des calculs de probabilité.

Commentaire : on se place sur un espace probabilisé $(\Omega; \mathcal{P}(\Omega); P)$.

IX.4.1) Probabilité conditionnelle

Exemple 9.19 *Soit une population de 100 individus : 40 hommes, dont 10 sont malades et 60 femmes, dont 20 sont malades. On choisit une personne au hasard. On pose donc $\Omega = \{\omega_1; \cdots; \omega_{100}\}$ et on prend comme probabilité la probabilité uniforme.*
On suppose que l'individu choisi est une femme. Cet événement étant supposé réalisé, la probabilité que l'individu soit malade est alors :

$$\frac{20}{60} = \frac{\frac{20}{100}}{\frac{60}{100}} = \frac{\frac{\mathrm{Card}(M \cap F)}{\mathrm{Card}\,\Omega}}{\frac{\mathrm{Card}\,F}{\mathrm{Card}\,\Omega}} = \frac{P(M \cap F)}{P(F)}.$$

Ceci motive la définition suivante :

Définition 9.11 *Soit B un événement de probabilité non nulle. On appelle probabilité conditionnelle à B, ou probabilité sachant B, associée à P l'application :*

$$P_B : \left\{ \begin{array}{l} \mathcal{P}(\Omega) \to \mathbb{R} \\ A \mapsto \dfrac{P(A \cap B)}{P(B)} \end{array} \right.$$

Notation : on note aussi $P(A/B)$ pour $P_B(A)$.

Commentaire : on sera alors attentif au fait que c'est une simple notation, et que A/B n'est pas un événement.
On utilise dans la suite indifféremment l'une de ces deux notations.

Propriété 9.12 *L'application P_B est une probabilité sur $(\Omega; \mathcal{P}(\Omega))$.*

Démonstration : on prouve que P_B vérifie les 3 axiomes :
Positivité : P étant positive, P_B l'est aussi.
Totalité : $P_B(\Omega) = \dfrac{P(\Omega \cap B)}{P(B)} = 1.$

Additivité : si A_1 et A_2 sont disjoints :

$$\begin{aligned} P_B(A_1 \cup A_2) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \\ &= \frac{P((A_1 \cap B) \cup (B \cap A_2))}{P(B)} \\ &= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} \\ &= P_B(A_1) + P_B(A_2) \end{aligned}$$

Propriété 9.13 *Soit B un événement de probabilité non nulle. Alors :*

$$\forall A \in \mathcal{P}(\Omega), P(A \cap B) = P_B(A)P(B)$$

IX.4.2) Formule des probabilités totales et formule de Bayes

Définition 9.12 *On appelle système complet d'événements toute famille d'événements $(A_i)_{i=1,\cdots,n}$ telle que :*

1. $\forall i \in [\![1;n]\!], A_i \neq \emptyset$.
2. $\forall (i;j) \in [\![1;n]\!]^2, i \neq j \Rightarrow A_i \cap A_j = \emptyset$.
3. $\displaystyle\bigcup_{i=1}^{n} A_i = \Omega$.

Propriété 9.14 *Soit $(A_i)_{i=1,\cdots,n}$ un système complet d'événements. Alors on a la formule des probabilités totales :*

$$\forall A \in \mathcal{P}(\Omega), P(A) = \sum_{i=1}^{n} P(A/A_i)P(A_i)$$

(sous réserve que chaque A_i ait une probabilité non nulle, afin que les probabilités conditionnelles soient bien définies).

Démonstration : comme $(A_i)_{i=1,\cdots,n}$ est un système complet, A s'écrit $A = A \cap \left(\displaystyle\bigcup_{i=1}^{n} A_i\right)$ et on a donc la décomposition de A sous forme d'union disjointe :

Remarque : on écrit souvent la formule des probabilités totales relativement au système complet $\{A_1; \overline{A}_1\}$:

$$P(A) = P(A/A_1)P(A_1) + P(A/\overline{A}_1)P(\overline{A}_1)$$

Propriété 9.15 *Soient A et B deux événements de probabilité non nulle. Alors on peut écrire :*

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \quad \textit{(formule de Bayes)}$$

Démonstration : $P(A\cap B) = P(A/B)P(B)$ et $P(B\cap A) = P(B/A)P(A)$, puis $P(A \cap B) = P(B \cap A)$.

Cas particulier : lorsqu'on a un système complet $(A_i)_{i=1,\cdots,n}$, on obtient :

$$\begin{aligned} \forall k \in [\![1; n]\!], P(A_k/A) &= \frac{P(A/A_k)P(A_k)}{P(A)} \\ &= \frac{P(A/A_k)P(A_k)}{\sum_{i=1}^{n} P(A/A_i)P(A_i)} \end{aligned}$$

Dans le cas d'un système complet $\{A_1; \overline{A}_1\}$:

$$P(A_1/A) = \frac{P(A/A_1)P(A_1)}{P(A/A_1)P(A_1) + P(A/\overline{A}_1)P(\overline{A}_1)}$$

Exemple 9.20 *On tire successivement et sans remise 2 boules d'une urne U composée de 2 boules noires et 3 boules rouges (indiscernables au toucher).*
Quelle est la probabilité que la $2^{\text{ème}}$ boule soit rouge ?
Quelle est la probabilité que la $1^{\text{ère}}$ soit rouge sachant que la $2^{\text{ème}}$ est noire ?
On pose : $\Omega = \{(a; b) \in U^2 / a \neq b\}$ et on choisit pour P la probabilité uniforme. On note R_i l'événement « la $i^{\text{ème}}$ boule est rouge » et N_i « la $i^{\text{ème}}$ boule est noire ».
Par la formule des probabilités totales :

$$P(R_2) = P(R_2/R_1)P(R_1) + P(R_2/N_1)P(N_1) = \frac{2}{4} \times \frac{3}{5} + \frac{3}{4} \times \frac{2}{5} = \frac{3}{5}$$

Par la formule de Bayes : $P(R_1/N_2) = \dfrac{P(N_2/R_1)P(R_1)}{P(N_2)} = \dfrac{\frac{2}{4} \times \frac{3}{5}}{\frac{2}{5}} = \dfrac{3}{4}.$

IX.4.3) Indépendance de deux événements

On traduit le fait que la réalisation de B n'a pas « d'influence » sur celle de A, c'est-à-dire, dans le cas où $P(B) \neq 0$: $P_B(A) = P(A)$. On a alors $P(A \cap B) = P(A)P(B)$ et c'est ce que l'on peut prendre comme définition :

Définition 9.13 *Deux événements A et B sont indépendants si $P(A \cap B) = P(A)P(B)$.*

Exemple 9.21 *On jette un dé équilibré, expérience que l'on modélise en posant $\Omega = [\![1;6]\!]$, et en prenant pour P la probabilité uniforme.*
On considère les événements suivants : $A = \{2;4;6\}$, $B = \{5;6\}$ et $C = \{5\}$.
On a immédiatement : $P(A) = \dfrac{1}{2}$, $P(B) = \dfrac{1}{3}$, $P(C) = \dfrac{1}{6}$, $P(A \cap B) = \dfrac{1}{6}$, $P(A \cap C) = 0$ et $P(B \cap C) = \dfrac{1}{6}$.
On en conclut que A et B sont indépendants, mais que A et C ainsi que B et C ne le sont pas.

Commentaire : ne pas confondre incompatibilité $((A \cap B) = \emptyset)$ et indépendance $(P(A \cap B) = P(A)P(B))$.
En particulier, on notera bien que la notion d'indépendance, à la différence de celle d'incompatibilité, dépend de la probabilité P : en toute rigueur, on devrait dire A et B sont P-indépendants.

Propriété 9.16

$$\begin{aligned} A \text{ et } B \text{ indépendants} &\Leftrightarrow \overline{A} \text{ et } B \text{ indépendants} \\ &\Leftrightarrow A \text{ et } \overline{B} \text{ indépendants} \\ &\Leftrightarrow \overline{A} \text{ et } \overline{B} \text{ indépendants} \end{aligned}$$

Démonstration : si A et B sont indépendants, alors :

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

donc $\overline{A}$ et B sont indépendants.
On en déduit de la même façon les autres implications.

Remarque : la relation d'indépendance n'est pas « transitive », comme on le voit sur l'exemple suivant :

Exemple 9.22 *On reprend l'exemple précédent et on considère l'événement $D = \{1;3;5\}$. On a : $P(A) = \dfrac{1}{2}$, $P(B) = \dfrac{1}{3}$, $P(D) = \dfrac{1}{2}$, $P(A \cap B) =$*

$\frac{1}{6}$, $P(A \cap D) = 0$ *et* $P(B \cap D) = \frac{1}{6}$.
Donc A et B sont indépendants, B et D sont indépendants, mais A et D ne sont pas indépendants.

IX.4.4) Applications

Exercice 9.1 *Événement indépendant de tous les autres. Montrer que A est indépendant de tous les événements B de $\mathcal{P}(\Omega)$ si et seulement si $P(A) = 0$ ou $P(A) = 1$.*
On suppose que : $\forall B \in \mathcal{P}(\Omega), P(A \cap B) = P(A)P(B)$.
Pour $B = A$, on a : $P(A) = (P(A))^2$, d'où la condition nécessaire.
Réciproquement : si $P(A) = 0$, alors : $\forall B \in \mathcal{P}(\Omega), (A \cap B) \subset A$, et donc $P(A \cap B) = 0 = P(A)P(B)$, et A et B sont indépendants. Si $P(A) = 1$, alors $P(\overline{A}) = 0$, et donc $\overline{A}$ et B sont indépendants, ce qui entraîne que A et B sont indépendants.

Exercice 9.2 *Fiabilité de test : on considère une population où une personne sur 100 est malade. On dispose d'un test qui permet de détecter la maladie avec une probabilité de $\frac{8}{10}$ et de reconnaître un sujet sain avec une probabilité de $\frac{9}{10}$.*
Quelle est la probabilité qu'une personne reconnue malade par le test le soit effectivement ?
En notant les événements M : « la personne est malade », et T : « le test annonce que la personne est malade », on a :

$$P(T/M) = \frac{8}{10} \text{ et } P(\overline{T}/\overline{M}) = \frac{9}{10}$$

On applique la formule de Bayes :

$$\begin{aligned} P(M/T) &= \frac{P(T/M)P(M)}{P(T/M)P(M) + P(T/\overline{M})P(\overline{M})} \\ &= \frac{\frac{8}{10} \times \frac{1}{100}}{\frac{8}{10} \times \frac{1}{100} + \frac{1}{10} \times \frac{99}{100}} \\ &= \frac{8}{107} \end{aligned}$$

Exercice 9.3 *Un électron peut être chaque seconde dans deux états : A ou B. S'il est dans l'état A, la probabilité qu'il passe à l'état B la seconde*

suivante est 0.6. S'il est dans l'état B, la probabilité qu'il reste dans l'état B la seconde suivante est 0.2. On suppose qu'il est dans l'état A à l'instant initial.
Quelle est la probabilité qu'il y soit toujours dans deux minutes ?
Soient les événements A_n : « l'électron est dans l'état A à la $n^{ième}$ seconde » et B_n : « l'électron est dans l'état B à la $n^{ième}$ seconde ».
On note $a_n = P(A_n)$, et on cherche donc a_{120}.
Par la formule des probabilités totales :

$$\begin{aligned} a_{n+1} &= P(A_{n+1}/A_n)P(A_n) + P(A_{n+1}/B_n)P(B_n) \\ &= 0.4a_n + 0.8(1 - a_n) \\ &= -0.4a_n + 0.8 \end{aligned}$$

On obtient donc une suite arithmético-géométrique, qu'il faut savoir traiter :
on cherche le point fixe : $l = -0.4l + 0.8 \Rightarrow l = \frac{4}{7}$. La suite $\left(a_n - \frac{4}{7}\right)$ est géométrique de raison -0.4. Comme $a_0 = 1$, on obtient $a_n = \frac{3}{7}(-0.4)^n + \frac{4}{7}$, et en particulier a_{120}.

Commentaire : il s'agit de suites usuelles en probabilité, appelées « Chaînes de Markov », en dimension 1 dans l'exercice proposé.

Exercice 9.4 *Formule d'Euler.*
Voir l'exercice 1.13

IX.5 - Variable aléatoire à valeurs réelles dont l'ensemble des valeurs est fini. Loi de probabilité. Espérance mathématique, variance. Exemples.

Commentaire : on se place sur un espace probabilisé $(\Omega; \mathcal{P}(\Omega); P)$. Comme $X(\Omega)$ est fini, la v.a.r. X admet des moments de tous ordres (pas de problème d'existence de l'espérance et de la variance).
La tribu étant $\mathcal{P}(\Omega)$, on a nécessairement que : $\forall B \in \mathcal{B}_{\mathbb{R}}, X^{-1}(B) \in \mathcal{P}(\Omega)$.
Attention : le fait que $X(\Omega)$ soit fini n'implique pas que Ω soit fini (prendre par exemple $\Omega = \mathbb{R}$ et $X = \mathbf{1}_{[0;+\infty[}$).

IX.5.1) Variable aléatoire réelle

Définition 9.14 *On appelle variable aléatoire réelle toute application* $X : \Omega \to \mathbb{R}$.

Exemple 9.23 *On jette deux fois un dé équilibré, on pose* $\Omega = [\![1;6]\!]^2$, *et* P *la probabilité uniforme.*

$$X : \omega = (a;b) \mapsto a+b \text{ (somme des deux scores obtenus)}$$

$$Y : \omega = (a;b) \mapsto \begin{cases} 0 & \text{si } a \text{ et } b \text{ impairs} \\ 1 & \text{si } a \text{ et } b \text{ pairs} \\ 2 & \text{sinon} \end{cases}$$

X *et* Y *sont des variables aléatoires réelles.*

Propriété 9.17 *L'ensemble des v.a.r. sur* $(\Omega; \mathcal{P}(\Omega); P)$ *a une structure d'algèbre pour les lois usuelles : la somme de deux v.a.r., une v.a.r. multipliée par un réel et le produit de deux v.a.r. sont encore des v.a.r.*

Démonstration : les applications obtenues sont bien des applications de Ω vers $\mathbb{R}$.

Notation : on note dans la suite $X(\Omega) = \{x_1; \cdots; x_n\}$.

Exemple 9.24 *On lance un dé équilibré, on pose* $\Omega = [\![1;6]\!]$, *et on prend pour* P *la probabilité uniforme. On réalise un gain nul si l'on obtient 1, un gain de 1 € si l'on obtient 2, 3 ou 4 et de 2 € si le résultat est 5 ou 6. On note* X *la v.a.r. égale au gain obtenu :* $X(\Omega) = \{0; 1; 2\}$.

IX.5.2) Loi de probabilité

Définition 9.15 *La loi de probabilité de X est l'application :*

$$P' : \left\{ \begin{array}{ccc} \mathcal{P}(X(\Omega)) & \to & \mathbb{R} \\ A & \mapsto & P(X^{-1}(A)) \end{array} \right.$$

Commentaire : on rappelle que par définition $X^{-1}(A) = \{\omega \in \Omega / X(\omega) \in A\}$. Pas question de « bijection réciproque » ici...

Notation : $P(X^{-1}(A))$ sera souvent noté $P(X \in A)$.

Théorème 9.6 *P' est une probabilité sur l'espace probabilisable formé par le couple $(X(\Omega); \mathcal{P}(X(\Omega)))$.*

Démonstration : P' est positive puisque P est positive.
Totalité : $P'(X(\Omega)) = P(X^{-1}(X(\Omega))) = P(\Omega) = 1$.
Additivité : soient A et B deux parties disjointes de $X(\Omega)$:

$$P'(A \cup B) = P(X^{-1}(A \cup B)) = P(X^{-1}(A) \cup X^{-1}(B))$$

et comme $X^{-1}(A)$ et $X^{-1}(B)$ sont disjointes :

$$P'(A \cup B) = P(X^{-1}(A)) + P(X^{-1}(B)) = P'(A) + P'(B)$$

En pratique, on se contente de donner les x_i et les valeurs de $P(X = x_i)$, puisqu'on a :

Propriété 9.18 *La loi de probabilité P' est entièrement déterminée par ses valeurs sur les $\{x_i\}_{i=1\cdots n}$.*

Démonstration : comme P' est additive, pour tout $A \in \mathcal{P}(X(\Omega))$,

$$P'(A) = P'\left(\bigcup_{i/x_i \in A} \{x_i\} \right) = \sum_{i/x_i \in A} P'(\{x_i\}) = \sum_{i/x_i \in A} P(X = x_i)$$

Exemple 9.25 *Une urne contient 3 boules rouges et 4 boules noires. On extrait successivement avec remise 2 boules de l'urne.*
On pose $\Omega = \{R_1; R_2; R_3; N_1; N_2; N_3; N_4\}^2$ et on prend P uniforme. On mise au départ 10 € et on gagne 8 € par boule rouge obtenue. Soit X la v.a.r. prenant pour valeur le gain final. X a pour loi :

$X(\Omega)$	-10	-2	6
$P(X = x_i)$	$\frac{16}{49}$	$\frac{24}{49}$	$\frac{9}{49}$

Propriété 9.19 *Puisque P vérifie l'axiome de totalité, on a évidemment :*

$$\sum_{i=1}^{n} P(X = x_i) = 1$$

Démonstration : il suffit d'écrire que les événements $X^{-1}(\{x_i\}), i = 1 \cdots n$, sont disjoints et que leur réunion est égale à Ω.

Commentaire : ne pas confondre la probabilité P, qui a été installée sur l'espace probabilisable $(\Omega; \mathcal{A})$ et la probabilité P', qui est la loi de X. Par exemple, pour le jeu avec mise initiale de 10 €, P est uniforme et P' ne l'est pas.

IX.5.3) Espérance mathématique

Il s'agit de la moyenne des valeurs de X pondérées par leur probabilité d'apparition :

Définition 9.16 *Soit X une v.a.r. discrète. On appelle espérance de X le réel, noté $E(X)$:*

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

$E(X)$ est donc le barycentre de la famille de points pondérés $(x_i; P(X = x_i))_{i=1 \cdots n}$.

Exemple 9.26 *En reprenant le jeu des 10 €,*

$$E(X) = (-10) \times \frac{16}{49} + (-2) \times \frac{24}{49} + 6 \times \frac{9}{49} = -\frac{22}{7}$$

Commentaire : $E(X)$ n'a aucune raison d'être égale à l'un des x_i !

Propriété 9.20 *On suppose les x_i ordonnés : $x_1 \leq x_2 \leq \cdots \leq x_n$. Alors : $x_1 \leq E(X) \leq x_n$.*

Démonstration : pour tout $i \in [\![1; n]\!], x_1 \leq x_i \leq x_n$, ce qui entraîne que $\sum_{i=1}^{n} p_i x_1 \leq \sum_{i=1}^{n} p_i x_i \leq \sum_{i=1}^{n} p_i x_n$, c'est-à-dire $x_1 \leq E(X) \leq x_n$.

Théorème 9.7 *E est une forme linéaire sur l'espace vectoriel des v.a.r. définies sur $(\Omega; \mathcal{P}(\Omega); P)$: si X et Y sont deux v.a.r., si λ est un réel, alors :*

$$E(X + \lambda Y) = E(X) + \lambda E(Y)$$

Démonstration : on note $X(\Omega) = \{x_1; \cdots ; x_n\}$ et $Y(\Omega) = \{y_1; \cdots ; y_m\}$. Soit $Z = X + Y$. On note $Z(\Omega) = \{z_1; \cdots ; z_p\}$: $E(Z) = \sum_{k=1}^{p} z_k P(Z = z_k)$. Or on a :

$$P(Z = z_k) = P(X + Y = z_k) = \sum_{(i;j)/x_i+y_j=z_k} P(X = x_i \cap Y = y_j)$$

d'où :

$$\begin{aligned}
E(Z) &= \sum_{k=1}^{p} \sum_{(i;j)/x_i+y_j=z_k} z_k P(X = x_i \cap Y = y_j) \\
&= \sum_{(i;j)} (x_i + y_j) P(X = x_i \cap Y = y_j) \\
&= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} P(X = x_i \cap Y = y_j) + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} P(X = x_i \cap Y = y_j) \\
&= \sum_{i=1}^{n} x_i P(X = x_i) + \sum_{j=1}^{m} y_j P(Y = y_j) \\
&= E(X) + E(Y)
\end{aligned}$$

De plus, on a bien $E(\lambda X) = \sum_{i=1}^{n} \lambda x_i P(\lambda X = \lambda x_i) = \lambda \sum_{i=1}^{n} x_i P(X = x_i)$.

IX.5.4) Variance et écart-type

Définition 9.17 *On appelle variance de X l'espérance de la variable $(X - E(X))^2$:*

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

Théorème 9.8

$$V(X) = \sum_{i=1}^{n} (x_i - E(X))^2 p_i$$

On mesure la somme des carrés des écarts entre les valeurs de X et la moyenne, pondérés par les poids de probabilité. C'est donc un indicateur de « dispersion » autour de $E(X)$.

Démonstration : on note, pour tout i, $A_i = \{\omega / X(\omega) = x_i\}$, et on a

la décomposition suivante : $(X - E(X))^2 = \sum_{i=1}^{n}(x_i - E(X))^2 \mathbf{1}_{A_i}$. En effet, pour tout $\omega \in \Omega$, les fonctions de droite et de gauche sont égales.
On en déduit, grâce à la linéarité de l'espérance :

$$\begin{aligned} E((X - E(x))^2) &= \sum_{i=1}^{n}(x_i - E(X))^2 E(\mathbf{1}_{A_i}) \\ &= \sum_{i=1}^{n}(x_i - E(X))^2 P(X = x_i) \end{aligned}$$

Commentaire : les deux résultats précédents (linéarité de E et expression de $V(X)$) nécessitent une démonstration. Ils n'ont rien d'évident puisqu'en appliquant juste la définition de E, on a seulement :

$$E(X + Y) = \sum(x_i + y_j)P(X + Y = x_i + y_j)$$

et :

$$E\Big((X - E(X))^2\Big) = \sum(x_i - E(X))^2 P\Big((X - E(X))^2 = (x_i - E(X))^2\Big)$$

Remarque : si Ω est fini, ces résultats sont immédiats, puisqu'on montre que dans ce cas $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$.

Propriété 9.21 *On a $V(X) \geq 0$, et $V(X) = 0 \Leftrightarrow X$ est constante.*

Démonstration : si $V(X) = 0$ alors $\forall i, (x_i - E(X))^2 P(X = x_i) = 0$. On a donc que tous les x_i sont égaux (à $E(X)$).

Le calcul pratique de la variance se fait souvent avec la formule de Koenig-Huygens :

Théorème 9.9 *Formule de Koenig.*

$$V(X) = \sum_{i=1}^{n} x_i^2 P(X = x_i) - (E(X))^2$$

Démonstration : on développe :

$$\begin{aligned} V(X) &= \sum_{i=1}^{n}(x_i - E(X))^2 p_i \\ &= \sum_{i=1}^{n} x_i^2 p_i - 2E(X)\sum_{i=1}^{n} x_i p_i + E(X)^2 \sum_{i=1}^{n} p_i \\ &= \sum_{i=1}^{n} x_i^2 P(X = x_i) - (E(X))^2 \end{aligned}$$

Exemple 9.27 *Pour le jeu des 10 €, on obtient :*

$$V(X) = (-10)^2 \times \frac{16}{49} + (-2)^2 \times \frac{24}{49} + 6^2 \times \frac{9}{49} - \left(\frac{-22}{7}\right)^2 \simeq 31.35$$

Propriété 9.22 *Soit $(a;b) \in \mathbb{R}^2$. Alors $V(aX+b) = a^2V(X)$.*

Démonstration : il suffit d'utiliser la linéarité de E :

$$V(aX+b) = E(a^2(X-E(X))^2) = a^2E(X-E(X))^2 = a^2V(X)$$

Définition 9.18 *On appelle écart-type de X le réel $\sigma(X) = \sqrt{V(X)}$.*

IX.5.5) Exemples

Exemple 9.28 *Soit X de loi uniforme sur $[\![1;n]\!]$. Vérifier que :*

$$E(X) = \frac{n+1}{2} \text{ et } V(X) = \frac{n^2-1}{12}$$

Exemple 9.29 *Une urne est composée de 4 boules (indiscernables au toucher) portant les numéros suivants : 0, 1, 1 et 2 . On tire une poignée de n boules et on appelle X_n le produit des numéros obtenus.*
Déterminer la loi de chaque variable X_i. Combien vaut-il mieux tirer de boules ?
Soient $E = \{0;1;1;2\}$ et $n \in [\![1;4]\!]$. Pour modéliser l'expérience, on pose comme univers $\Omega_n = \{A \subset \mathcal{P}(E) / \operatorname{Card} A = n\}$ et, les poignées étant équiprobables (pour n fixé), on choisit pour probabilité P_n la probabilité uniforme. Le calcul des poids de probabilité se réduit à du dénombrement puisque P_n est uniforme : $\operatorname{Card}\Omega_n = \binom{4}{n}$, et on obtient :

$X_1(\Omega_1)$	0	1	2
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$X_2(\Omega_2)$	0	1	2
	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$

$X_3(\Omega_3)$	0	2
	$\frac{5}{6}$	$\frac{1}{6}$

$X_4(\Omega_4)$	0
	1

On compare les espérances : $E(X_1) = 1$, $E(X_2) = \frac{5}{6}$, $E(X_3) = \frac{1}{3}$, et $E(X_4) = 0$. Il vaut donc mieux prélever une seule boule.

Commentaire : pour cette leçon, il peut être très enrichissant de garder à l'esprit la loi faible des grands nombres qui peut permettre de justifier certains choix de probabilités (voir chapitre VIII). Elle donne aussi une interprétation de l'espérance : la moyenne des valeurs prises par X, lorsque l'on répète l'expérience aléatoire, tend vers $E(X)$. Surtout, ne pas interpréter l'espérance comme la valeur la plus probable de X : il est clair que la variable X_2 du dernier exemple ne vaudra jamais $\frac{5}{6}$!

IX.6 - Schéma de Bernoulli et loi binomiale. Exemples.

IX.6.1) Loi de Bernoulli

Définition 9.19 *Soit $p \in [0;1]$. X suit la loi de Bernoulli de paramètre p si :*

1. $X(\Omega) = \{0;1\}$

2. $P(X=0) = 1-p$ et $P(X=1) = p$

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

Contexte usuel : on considère un espace probabilisé $(\Omega; \mathcal{A}; P)$ et on s'intéresse à un événement S, appelé « succès ».
On définit alors une v.a.r.

$$X : \omega \mapsto \begin{cases} 1 \text{ si } \omega \in S \\ 0 \text{ sinon} \end{cases}$$

X est donc simplement l'indicatrice du succès S, et si on note $p = P(S)$, X suit une loi de Bernoulli de paramètre p.

Propriété 9.23 *$E(X) = p$ et $V(X) = p(1-p)$*

Remarque : les calculs des moments de X sont facilités par le fait que :

$$\forall n \geq 1, X^n = X$$

IX.6.2) Schéma de Bernoulli

On considère une expérience aléatoire ayant 2 issues : succès S ou échec $\overline{S}$.
Une expérience aléatoire ainsi modélisée est appelée épreuve de Bernoulli : on a posé $\Omega = \{S; \overline{S}\}$, et, si on note p la probabilité du succès, on définit une probabilité P_1 par $P_1(S) = p$ (et donc $P_1(\overline{S})) = 1-p$).

Définition 9.20 *On appelle schéma de Bernoulli une suite finie de répétitions indépendantes d'une même épreuve de Bernoulli.*

Soit n le nombre de répétitions de l'épreuve de Bernoulli. Le schéma de Bernoulli a pour univers : $\Omega_n = \{S; \overline{S}\}^n$. On définit ensuite une probabilité P_n sur $\mathcal{P}(\Omega_n)$: comme Ω_n est fini, il suffit de définir P_n sur les singletons (événements élémentaires).
Soit $(\alpha_1; \cdots; \alpha_n) \in \{S; \overline{S}\}^n$. On pose :

$$P_n(\{(\alpha_1; \cdots; \alpha_n)\}) = P_1(\{\alpha_1\}) \cdots P_1(\{\alpha_n\})$$

Soit $k = \text{Card}\{\alpha_i/\alpha_i = S\}$. Alors :

$$P_n(\{(\alpha_1;\cdots;\alpha_n)\}) = p^k(1-p)^{n-k}$$

Commentaire : la probabilité P_n s'appelle la probabilité produit, et choisir cette probabilité permet de traduire que les répétitions de l'épreuve de Bernoulli se font de manière « indépendantes », au sens suivant : soit S_i l'événement « on a obtenu un succès au cours de la $i^{ième}$ épreuve ».

$$\begin{aligned} P_n(S_i) &= P_n((\alpha_1;\cdots;\alpha_n) \in \Omega \times \cdots \times \Omega \times S \times \Omega \cdots \times \Omega) \\ &= P_1(\Omega)\cdots P_1(\Omega)P_1(S)P_1(\Omega)\cdots P_1(\Omega) \\ &= p \end{aligned}$$

Si $i \neq j$:

$$\begin{aligned} P_n(S_i \cap S_j) &= P_n((\alpha_1;\cdots;\alpha_n) \in \Omega \times \cdots \\ &\quad \times\Omega \times S \times \Omega \cdots \times \Omega \times S \times \Omega \cdots \times \Omega) \\ &\quad \text{(où les 2 } S \text{ apparaissent en position } i \text{ et } j) \\ &= P_1(\Omega)\cdots P_1(\Omega)P_1(S)P_1(\Omega)\cdots P_1(\Omega)P_1(S)P_1(\Omega)\cdots P_1(\Omega) \\ &= p^2 \\ &= P_n(S_i)P_n(S_j) \end{aligned}$$

Donc S_i et S_j sont indépendants. P_n s'appelle la probabilité produit.

On définit ensuite, pour $i \in [\![1;n]\!]$:

$$X_i : \left\{ \begin{array}{ccc} \Omega_n & \to & \{0;1\} \\ (\alpha_1;\cdots;\alpha_n) & \mapsto & \left\{ \begin{array}{lll} 0 & si & \alpha_i = \overline{S} \\ 1 & si & \alpha_i = S \end{array} \right. \end{array} \right.$$

On a donc des variables de Bernoulli :

$$\forall i \in [\![1;n]\!], X_i \sim \mathcal{B}(p)$$

Théorème 9.10 *Les v.a.r. X_i sont 2 à 2 indépendantes.*

Démonstration : « $X_i = 1$ » est l'événement S_i et « $X_i = 0$ » est l'événement $\overline{S_i}$, et on a vu que S_i et S_j sont indépendants. Par conséquent :

$$\begin{aligned} P_n(X_i = 1 \cap X_j = 1) &= P_n(S_i \cap S_j) = P_n(S_i)P_n(S_j) \\ &= P_n(X_i = 1)P_n(X_j = 1) \end{aligned}$$

De plus, l'indépendance de S_i et S_j entraîne celle de S_i et de $\overline{S_j}$, celle de $\overline{S_i}$ et $\overline{S_j}$ ainsi que celle de $\overline{S_i}$ et S_j, ce qui permet d'effectuer immédiatement les trois autres vérifications.

IX.6.3) Loi binomiale

C'est la loi du nombre de succès obtenus au cours du schéma de Bernoulli, et on pose donc naturellement :

Définition 9.21 *On dit que Z_n suit la loi binomiale $\mathcal{B}(n;p)$ si*

$$Z_n = \sum_{i=1}^{n} X_i$$

où les X_i suivent toutes la même loi $\mathcal{B}(p)$ et sont 2 à 2 indépendantes.

Propriété 9.24 *On a $Z_n(\Omega_n) = [\![0;n]\!]$, et :*

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

Démonstration : soit $k \in [\![0;n]\!]$:

$$P(Z_n = k) = P\Big(\sum_{i=1}^{n} X_i = k\Big)$$

L'événement $\sum\limits_{i=1}^{n} X_i = k$ est l'union disjointe des événements élémentaires $\{(\alpha_1;\cdots;\alpha_n)\}$ où k éléments α_i sont égaux à 1 et $n-k$ éléments α_i sont égaux à 0.
Chacun de ces événements élémentaires a donc pour probabilité :

$$\begin{aligned} P_n(\{(\alpha_1;\cdots;\alpha_n)\}) &= P_1(\{\alpha_1\})\cdots P_1(\{\alpha_n)\}) \\ &= p^k(1-p)^{n-k} \end{aligned}$$

et ces événements élémentaires sont au nombre de $\binom{n}{k}$, d'où par additivité de P_n :

$$P_n(Z_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Exemple 9.30 *On lance 10 fois un dé équilibré.*
Quelle est la probabilité d'obtenir pour 4 lancers un résultat supérieur ou égal à 5 ?
L'épreuve de Bernoulli est ici : lancer un dé équilibré, on appelle succès l'événement S : « le résultat est supérieur ou égal à 5 », on a $p = \frac{1}{3}$.

Le schéma de Bernoulli est de répéter $n = 10$ fois l'épreuve de Bernoulli.

Commentaire : *la répéter de manière « indépendante » signifie que l'on munit l'espace probabilisable $(\Omega_{10}; \mathcal{P}(\Omega_{10}))$, où $\Omega_{10} = \{S; \overline{S}\}^{10}$ de la probabilité produit P_{10}...*

On a donc :

$$Z_{10} \sim \mathcal{B}\Big(10; \frac{1}{3}\Big) \text{ et } P(Z_{10} = 4) = \begin{pmatrix} 10 \\ 4 \end{pmatrix} \Big(\frac{1}{3}\Big)^4 \Big(\frac{2}{3}\Big)^6 \simeq 0.23$$

Combien de fois faut-il lancer le dé pour que la probabilité d'avoir au moins un succès soit supérieure ou égale à 0.95 ?

$$\begin{array}{rcl} P_n(Z_n \geq 1) & \geq & 0.95 \\ \Rightarrow 1 - P_n(Z_n = 0) & \geq & 0.95 \\ \Rightarrow \Big(\dfrac{2}{3}\Big)^n & \leq & 0.05 \\ \Rightarrow n & \geq & \dfrac{\ln 0.05}{\ln \dfrac{2}{3}} \simeq 7.39 \end{array}$$

Donc il faut au moins 8 lancers.

Propriété 9.25 *Si $Z_n \sim \mathcal{B}(n; p)$, alors :*

$$E(Z_n) = np \text{ et } V(Z_n) = np(1-p)$$

Démonstration : on utilise la linéarité de l'espérance :

$$E(Z_n) = E\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} E(X_i) = np$$

et on développe la variance :

$$V(Z_n) = V\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i; X_j)$$

Comme les X_i sont indépendantes, les covariances sont nulles et donc :

$$V(Z_n) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Exemple 9.31 *En reprenant l'exemple précédent, $E(Z_{10}) = \dfrac{10}{3}$.*

Remarque : pour n fixé, la variance est maximale pour $p = \frac{1}{2}$, ce qui est cohérent puisque cela correspond à l'équiprobabilité entre le succès et l'échec.

IX.7 - Séries statistiques à deux variables numériques. Nuage de points associé. Ajustement affine par la méthode des moindres carrés. Droites de régression. Applications. L'exposé pourra être illustré par un ou des exemples faisant appel à l'utilisation d'une calculatrice.

IX.7.1) Série statistique à deux variables numériques

On considère une population Ω d'individus $\omega_1, \cdots, \omega_n$.
On définit sur cette population deux caractères numériques, c'est-à-dire deux applications :

$$X : \left\{ \begin{array}{ccc} \Omega & \to & \mathbb{R} \\ \omega_i & \mapsto & X(\omega_i) \end{array} \right. \text{ et } Y : \left\{ \begin{array}{ccc} \Omega & \to & \mathbb{R} \\ \omega_i & \mapsto & Y(\omega_i) \end{array} \right.$$

On note $X(\omega_i) = x_i$ et $Y(\omega_i) = y_i$.

Exemple 9.32 *Ω est l'ensemble des candidats ayant composé aux deux épreuves, X associe à chaque candidat sa note à la première épreuve et Y sa note à la deuxième.*

Le but est de déterminer s'il existe une « relation » entre X et Y.
On considère donc les couples $(x_i; y_i)_{i=1\cdots n}$. Comme ils ne sont pas nécessairement distincts, on regroupe les couples identiques en leur affectant un poids p_i égal à la fréquence d'apparition du couple :

Définition 9.22 *On appelle série statistique double la donnée de N couples pondérés $((x_i; y_i); p_i)_{i=1\cdots N}$, où les p_i sont des réels strictement positifs de somme égale à 1.*

Exemple 9.33 *On considère 5 candidats, qui ont obtenu les notes suivantes :*

x_i	10	8	11	7
y_i	8	8	9	6
p_i	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Dans un repère orthonormé $(O; \vec{i}; \vec{j})$, on place les points $M_i(x_i; y_i)$ et entre parenthèses leur poids p_i : on obtient un nuage de points pondérés.
Paramètres de la série :
moyennes : $\overline{X} = \sum_{i=1}^{N} p_i x_i$ et $\overline{Y} = \sum_{i=1}^{N} p_i y_i$.

Variances : $V(X) = \sum_{i=1}^{N} p_i(x_i - \overline{X})^2$ et $V(Y) = \sum_{i=1}^{N} p_i(y_i - \overline{Y})^2$.

Covariance : $\text{Cov}\,(X;Y) = \sum_{i=1}^{N} p_i(x_i - \overline{X})(y_i - \overline{Y})$.

On vérifie facilement en développant que :

$$V(X) = \sum_{i=1}^{N} x_i^2 p_i - (\overline{X})^2 \text{ (idem pour } Y)$$

$$\text{Cov}\,(X;Y) = \sum_{i=1}^{N} p_i x_i y_i - (\overline{X} \times \overline{Y})$$

Définition 9.23 *Le point $G(\overline{X};\overline{Y})$ est appelé point moyen du nuage.*

C'est donc le barycentre du système de points pondérés $(M_i;p_i)$.

IX.7.2) Ajustement affine et droites de régression

On souhaite déterminer la « meilleure » droite Δ passant par le nuage de points.
On suppose dans la suite que X n'est pas constante (sinon, le nuage est formé de points alignés verticalement) ainsi que Y (le nuage étant alors horizontal). Pour donner un sens au terme « meilleure » droite, il faut se donner un critère : on cherche la droite qui minimise :

$$\sum_{i=1}^{N} p_i M_i H_i^2$$

où H_i est le projeté de M_i sur Δ parallèlement à l'axe $[Oy)$. On veut donc minimiser la fonction

$$\phi(a;b) = \sum_{i=1}^{N} p_i(y_i - (ax_i + b))^2$$

Méthode : on met ϕ sous forme canonique par rapport à b, puis par rapport à a.

$$\begin{aligned}
\phi(a;b) &= \sum_{i=1}^{N} p_i b^2 - 2b\sum_{i=1}^{N} p_i(y_i - ax_i) + \sum_{i=1}^{N} p_i(y_i - ax_i)^2 \\
&= b^2 - 2b(\overline{Y} - a\overline{X}) + \sum_{i=1}^{N} p_i y_i^2 + a^2\sum_{i=1}^{N} p_i x_i^2 - 2a\sum_{i=1}^{N} p_i x_i y_i \\
&= (b - (\overline{Y} - a\overline{X}))^2 - (\overline{Y} - a\overline{X})^2 + \sum_{i=1}^{N} p_i y_i^2 + a^2\sum_{i=1}^{N} p_i x_i^2 \\
&\quad -2a\sum_{i=1}^{N} p_i x_i y_i \\
&= (b - (\overline{Y} - a\overline{X}))^2 - \overline{Y}^2 - a^2\overline{X}^2 + 2a\overline{X} \times \overline{Y} + \sum_{i=1}^{N} p_i y_i^2 \\
&\quad +a^2\sum_{i=1}^{N} p_i x_i^2 - 2a\sum_{i=1}^{N} p_i x_i y_i \\
&= (b - (\overline{Y} - a\overline{X}))^2 + V(Y) + a^2V(X) - 2a\,\mathrm{Cov}\,(X;Y) \\
&= (b - (\overline{Y} - a\overline{X}))^2 + V(X)\Big(a - \frac{\mathrm{Cov}\,(X;Y)}{V(X)}\Big)^2 - \frac{\mathrm{Cov}\,(X;Y)^2}{V(X)} \\
&\quad +V(Y)
\end{aligned}$$

ϕ est minimale lorsque les deux carrés sont nuls, d'où le théorème :

Théorème 9.11 *Il existe une unique droite $\Delta : y = ax + b$ qui minimise $\sum_{i=1}^{N} p_i M_i H_i^2$. Elle a pour coefficients :*

$$\begin{aligned}
a &= \frac{\mathrm{Cov}\,(X;Y)}{V(X)} \\
b &= \overline{Y} - a\overline{X}
\end{aligned}$$

Remarque : Δ passe par le point moyen $G(\overline{X};\overline{Y})$ (ce qui n'était pas imposé au départ).

Le minimum de ϕ vaut :

$$V(Y)\Big(1 - \frac{\mathrm{Cov}\,(X;Y)^2}{V(X)V(Y)}\Big)$$

Définition 9.24 *On appelle coefficient de corrélation linéaire :*

$$r = \frac{\mathrm{Cov}\,(X;Y)}{\sigma(X)\sigma(Y)}$$

Propriété 9.26 $-1 \leq r \leq 1$ *et* $|r| = 1$ *si et seulement si les points sont alignés.*

Démonstration : le minimum de ϕ vaut $V(Y)(1-r^2)$. ϕ étant positive, son minimum l'est aussi, et donc $r^2 \leq 1$. De plus, $r^2 = 1$ si et seulement si le minimum de ϕ est nul, auquel cas les points sont tous sur Δ $(\forall i, y_i - (ax_i + b) = 0)$.

Remarque : la démonstration peut se faire aussi en invoquant l'inégalité de Cauchy-Schwarz (développer Cov $(X + \lambda Y; X + \lambda Y)$), et son cas d'égalité.

Le coefficient r mesure donc la « qualité » de la droite Δ obtenue : plus $|r|$ est proche de 1, plus l'ajustement est bon. On fixe souvent la barre à 0.7 (voir l'interprétation en termes d'angle dans la partie consacrée à des compléments sur cette leçon).

Définition 9.25 Δ *est appelée droite de régression de* Y *en* X.

Avec les données numériques de l'exemple, on obtient $\Delta : y = 0.54x + 3.07$, et $r = 0.81$.

Définition 9.26 *Si on échange les rôles de* X *et* Y*, on obtient la droite de régression de* X *en* Y *:* $\Delta' : x = a'y + b'$*, où*

$$\begin{aligned} a' &= \frac{\mathrm{Cov}(X;Y)}{V(Y)} \\ b' &= \overline{X} - a'\overline{Y} \end{aligned}$$

Le coefficient de corrélation est toujours :

$$r = \frac{\mathrm{Cov}(X;Y)}{\sigma(X)\sigma(Y)}$$

(expression symétrique en X, Y*)*

Propriété 9.27 $0 \leq aa' \leq 1$ *et :*

$$aa' = 1 \Leftrightarrow \text{ les points sont alignés} \Leftrightarrow \Delta = \Delta'$$

Démonstration : $aa' = \dfrac{\mathrm{Cov}(X;Y)^2}{V(X)V(Y)} = r^2$.

IX.7.3) Applications

Si la relation affine entre X et Y est « insuffisante » ($|r|$ trop faible), on peut envisager d'autres relations fonctionnelles :
ajustement par une fonction puissance : $y = ax^b$. On a alors $\ln y = b\ln x + \ln a$, et on cherche un ajustement affine entre $\ln X$ et $\ln Y$.
Ajustement par une fonction exponentielle : $y = ae^{bx}$. On a $\ln y = bx + \ln a$, et on cherche un ajustement affine entre X et $\ln Y$.

Dans le cadre des séries chronologiques (où l'une des deux variables est le temps), cela permet d'obtenir une courbe de tendance, premier pas en vue de faire de la prévision.

IX.8 - Compléments sur les statistiques doubles.

Le but est de donner un éclairage sur cette leçon, grâce à l'algèbre bilinéaire.
On considère une population pondérée finie $(\omega_i; p_i)_{i=1\cdots n}$ telle que pour tout i, $p_i \in]0;1[$, et $\sum_{i=1}^{n} p_i = 1$. On définit deux caractères numériques sur cette population :

$$X : \left\{ \begin{array}{ccc} \Omega & \rightarrow & \mathbb{R} \\ \omega_i & \mapsto & x_i \end{array} \right. \text{ et } Y : \left\{ \begin{array}{ccc} \Omega & \rightarrow & \mathbb{R} \\ \omega_i & \mapsto & y_i \end{array} \right.$$

On peut les représenter dans $\mathbb{R}^2$, en se donnant un repère orthonormé et on a alors un nuage de n points pondérés $M_i(x_i; y_i)$. On travaille avec la norme euclidienne, et le but est de trouver la droite Δ qui minimise :

$$\sum_{i=1}^{n} p_i M_i P_i^2$$

où P_i est le projeté de M_i sur la droite Δ parallèlement à l'axe $[oy)$.
Si on note $y = ax + b$ l'équation de Δ, on veut donc déterminer :

$$\inf_{(a;b)\in\mathbb{R}^2} \left\{ \sum_{i=1}^{n} p_i (y_i - ax_i - b)^2 \right\}$$

C'est le point de vue à développer dans la leçon.

On peut aussi travailler dans $\mathbb{R}^{\Omega}$, ensemble des applications de Ω dans $\mathbb{R}$. X et Y sont alors représentées par :

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} et \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

et on identifie $\mathbb{R}^n$ et $\mathbb{R}^{\Omega}$. On définit l'application ϕ sur $(\mathbb{R}^n)^2$ par :

$$\begin{aligned} \phi(X;Y) &= \sum_{i=1}^{n} p_i x_i y_i \\ &= E(XY) \end{aligned}$$

C'est clairement une forme bilinéaire symétrique définie-positive (car pour tout i, $p_i > 0$), c'est-à-dire un produit scalaire. Le problème de minimisation s'écrit alors :

$$\begin{aligned} \inf_{(a;b)\in\mathbb{R}^2} \left\{ \sum_{i=1}^{n} p_i (y_i - ax_i - b)^2 \right\} &= \inf_{(a;b)\in\mathbb{R}^2} \{E((Y - aX - b)^2)\} \\ &= \inf_{Z\in Vect(1;X)} \{E((Y - Z)^2)\} \end{aligned}$$

On suppose X non constante, de sorte que $(1; X)$ est une famille libre de $\mathbb{R}^{\Omega}$.
Comme $(\mathbb{R}^{\Omega}; \phi)$ est un espace euclidien, l'inf recherché est un min, atteint pour $Z = \tilde{Y}$, projeté orthogonal de Y sur $F = Vect(1; X)$. On peut l'obtenir en orthonormalisant la base $(1; X)$ ou en caractérisant directement $\tilde{Y}$ par :

$$\begin{cases} \tilde{Y} \in Vect(1; X) \\ Y - \tilde{Y} \perp 1 \\ Y - \tilde{Y} \perp X \end{cases}$$

On a donc :

$$\begin{cases} \exists (a; b) \in \mathbb{R}^2, \tilde{Y} = aX + b \\ E((Y - \tilde{Y})1) = 0 \\ E((Y - \tilde{Y})X) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \exists (a; b) \in \mathbb{R}^2, \tilde{Y} = aX + b \\ E(Y) = E(\tilde{Y}) \\ E(YX) = E(\tilde{Y}X) \end{cases}$$

La deuxième équation donne :

$$\begin{aligned} & E(Y) = E(aX + b) \\ \Rightarrow\ & E(Y) = aE(X) + b \\ \Rightarrow\ & b = E(Y) - aE(X) \end{aligned}$$

On remarque donc que Δ passe par le barycentre $G(E(X); E(Y))$.
On reporte la valeur de b dans la troisième équation .

$$\begin{aligned} & E(aX^2 + bX) = E(XY) \\ \Rightarrow\ & aE(X^2) + bE(X) = E(XY) \\ \Rightarrow\ & aE(X^2) + E(Y)E(X) - aE(X)^2 = E(XY) \\ \Rightarrow\ & a = \frac{\mathrm{Cov}(X; Y)}{V(X)} \end{aligned}$$

On en déduit la valeur du min, c'est-à-dire la distance de Y à $Vect(1; X)$ (au sens de ϕ !) :

$$\begin{aligned} E((Y - \tilde{Y})^2) &= E(Y^2) - E(\tilde{Y}^2) \text{ (Théorème de Pythagore)} \\ &= E(Y^2) - E(a^2X^2 + b^2 + 2abX) \\ &= E(Y^2) - a^2E(X^2) - E(Y)^2 - a^2E(X)^2 \\ &\quad + 2aE(X)E(Y) - 2aE(X)E(Y) + 2a^2E(X)^2 \\ &= V(Y) - a^2V(X) \\ &= V(Y)\Big(1 - \frac{\mathrm{Cov}(X; Y)^2}{V(X)V(Y)}\Big) \end{aligned}$$

Interprétation du coefficient de corrélation linéaire $r = \dfrac{\mathrm{Cov}(X;Y)}{\sigma(X)\sigma(Y)}$. En notant $X^* = X - E(X)$ et $Y^* = Y - E(Y)$ les variables X et Y centrées réduites, on a :

$$\mathrm{Cov}(X;Y) = E(X - E(X))(Y - E(Y)) = \phi(X^*;Y^*)$$

et $V(X) = E((X - E(X))^2) = \phi(X^*;X^*)$, d'où :

$$r = \frac{\phi(X^*;Y^*)}{\sqrt{\phi(X^*;X^*)}\sqrt{\phi(Y^*;Y^*)}}$$

C'est donc le produit scalaire des vecteurs X^* et Y^* divisé par le produit des normes. Comme on est dans un espace euclidien, c'est donc le cosinus de l'angle formé par ces deux vecteurs :

$$r = \cos(X^*;Y^*)$$

Lorsqu'on donne comme borne $r \geq 0.707$, on considère donc que l'on a un bon ajustement si l'angle formé par $(X^*;Y^*)$ est inférieur à $\dfrac{\pi}{4}$.
Cet angle est celui que forment dans $\mathbb{R}^n$ les vecteurs

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} et \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

après avoir centré (retranché $E(X)$ et $E(Y)$), c'est-à-dire en se plaçant dans un repère de centre $G(E(X);E(Y))$.

Il est évidemment hors de question de développer tout ceci le jour de l'Oral, mais avoir vu cette interprétation peut être enrichissant pour la discussion avec le jury !

Annexe 1 : tribu des boréliens

Définition 9.27 *Étant donné un ensemble Ω, on dit qu'une famille de parties de Ω est une tribu si elle contient l'ensemble vide, si elle est stable par passage au complémentaire et par union dénombrable.*

Une intersection de tribus est donc encore une tribu, ce qui permet de définir la tribu engendrée par une famille de parties de Ω :

Définition 9.28 *Soit $\mathcal{A}$ une famille de parties de Ω. On appelle tribu engendrée par $\mathcal{A}$ l'intersection de toutes les tribus qui contiennent $\mathcal{A}$. On la note dans la suite $\mathcal{T}(\mathcal{A})$.*

Propriété 9.28 *$\mathcal{T}(\mathcal{A})$ est la plus petite tribu (au sens de l'inclusion) qui contient $\mathcal{A}$.*

Démonstration : il est clair que $\mathcal{T}(\mathcal{A})$ est une tribu et qu'elle contient $\mathcal{A}$. Si $\mathcal{B}$ est une tribu qui contient $\mathcal{A}$, alors $\mathcal{B}$ contient $\mathcal{T}(\mathcal{A})$ (puisque $\mathcal{T}(\mathcal{A})$ est l'intersection des tribus contenant $\mathcal{A}$).

On considère le cas où $\Omega = \mathbb{R}$:

Définition 9.29 *On appelle tribu borélienne de $\mathbb{R}$ la tribu engendrée par l'ensemble $\mathcal{O}$ des ouverts de $\mathbb{R}$, et on la note $\mathcal{B}_{\mathbb{R}}$: $\mathcal{B}_{\mathbb{R}} = \mathcal{T}(\mathcal{O})$.*
Les éléments de $\mathcal{B}_{\mathbb{R}}$ sont appelés les boréliens de $\mathbb{R}$.

La tribu $\mathcal{B}_{\mathbb{R}}$ contient en particulier les fermés de $\mathbb{R}$ (puisqu'elle est stable par passage au complémentaire) ; elle contient aussi les singletons (écrire $\{a\} = \bigcap_{n\geq 1} \left]a - \frac{1}{n}; a + \frac{1}{n}\right[$), et comme elle est stable par union finie ou dénombrable, elle contient les parties finies ou dénombrables de $\mathbb{R}$, donc $\mathbb{N}$, $\mathbb{Z}$ et $\mathbb{Q}$ sont des boréliens. Les intervalles sont aussi des boréliens ($[a; b[=]a; b[\cup \{a\}$, $[a; +\infty[= \bigcup_{n\geq a} [a; n[$), et on peut montrer qu'ils engendrent $\mathcal{B}_{\mathbb{R}}$. On peut se limiter aux intervalles du type $]a; b[$:

Propriété 9.29 *$\mathcal{B}_{\mathbb{R}}$ est engendré par l'ensemble des intervalles ouverts $]x; y[$, $x \in \mathbb{R}$, $y \in \mathbb{R}$.*

Démonstration : soit $\mathcal{T}$ la tribu engendrée par ces intervalles. Comme $]x; y[$ est un ouvert, on a immédiatement $\mathcal{T} \subset \mathcal{B}_{\mathbb{R}}$.
On montre que $\mathcal{B}_{\mathbb{R}} \subset \mathcal{T}$: soit O un ouvert de $\mathbb{R}$. On note $B(x; r)$ la boule ouverte de centre x et de rayon r (dans le cas présent, c'est l'intervalle

$]x-r;x+r[$, mais la démonstration est identique dans le cas de $\mathbb{R}^n$), et $\mathbb{Q}$ étant dénombrable, on note $\mathbb{Q} = \{q_m, m \in \mathbb{N}\}$.

Soit $x \in O$. Comme O est ouvert, il existe un entier n tel que $B\left(x; \frac{1}{2^n}\right) \subset O$. De plus, comme $\mathbb{Q}$ est dense dans $\mathbb{R}$, il existe un entier m tel que

$$|x - q_m| \leq \frac{1}{2^{n+1}}$$

On a donc $B\left(q_m; \frac{1}{2^{n+1}}\right) \subset O$, et O s'écrit alors comme une union dénombrable de boules ouvertes :

$$O = \bigcup_{(m;n)/B\left(q_m;\frac{1}{2^{n+1}}\right)\subset O} B\left(q_m; \frac{1}{2^{n+1}}\right)$$

Tout ouvert O appartient donc à la tribu $\mathcal{T}$, et comme les ouverts engendrent $\mathcal{B}_{\mathbb{R}}$, on a $\mathcal{B}_{\mathbb{R}} \subset \mathcal{T}$.

On peut démontrer ensuite qu'une probabilité μ sur $(\mathbb{R}; \mathcal{B}_{\mathbb{R}})$ est entièrement déterminée dès qu'on la connaît sur un ensemble de parties qui engendre $\mathcal{B}_{\mathbb{R}}$. μ est donc déterminée par ses valeurs sur l'ensemble des intervalles $]-\infty; a], a \in \mathbb{R}$, et donc la loi d'une variable aléatoire est entièrement caractérisée par la donnée de sa fonction de répartition, puisque $F(x) = \mu(]-\infty; x])$.

Pour définir une probabilité sur $\mathbb{R}$, on peut se demander pourquoi on ne prend pas comme tribu $\mathcal{P}(\mathbb{R})$ au lieu de $\mathcal{B}_{\mathbb{R}}$: le problème est qu'il n'est pas possible de définir une application sur $\mathcal{P}(\mathbb{R})$ qui respecte les trois axiomes (additivité, totalité, positivité). Il faut une tribu plus petite : La tribu des boréliens convient.

Remarque : il n'est pas très simple de montrer que $\mathcal{B}_{\mathbb{R}}$ est strictement incluse dans $\mathcal{P}(\mathbb{R})$. La construction de parties de $\mathbb{R}$ non boréliennes fait notamment appel à l'axiome du choix.

Annexe 2 : familles sommables

On considère une famille $(a_i)_{i\in I}$ de réels. On veut donner un sens à $\sum_{i\in I} a_i$: si I est une partie (infinie) de $\mathbb{N}$, cette somme peut dépendre de l'ordre dans lequel on prend les a_i, et si I est un ensemble sur lequel on n'a pas d'ordre naturel (par exemple si I est une partie de $\mathbb{N}^2$ ou de $\mathbb{Q}$), il n'y a pas de façon « canonique » de considérer des sommes partielles.

Définition 9.30 *Soit $(a_i)_{i\in I}$ une famille de réels positifs. On dit que cette famille est sommable si l'ensemble :*

$$\left\{\sum_{i\in J} a_i / J \text{ est une partie finie de } I\right\}$$

est majoré.
Si c'est le cas, la borne supérieure de cet ensemble est appelée somme de la famille $(a_i)_{i\in I}$, et on la note $\sum_{i\in I} a_i$.

Dans le cas où I est égal à $\mathbb{N}$ (ou à $[\![n_0;+\infty[\![$) :

Propriété 9.30 *Soit $(a_i)_{i\in I}$ une suite de réels positifs : la famille $(a_i)_{i\in I}$ est sommable si et seulement si la série de terme général a_i converge.*
On a dans ce cas :

$$\sum_{i\in I} a_i = \sum_{i=0}^{+\infty} a_i$$

On considère ensuite une famille $(a_i)_{i\in I}$ de réels quelconques. On pose, pour tout $i \in I$, $a_i^+ = \sup(a_i;0)$ et $a_i^- = -\inf(a_i;0)$. Les éléments a_i^+ et a_i^- sont donc positifs et $a_i = a_i^+ - a_i^-$.

Définition 9.31 1. *La famille $(a_i)_{i\in I}$ est sommable si les familles $(a_i^+)_{i\in I}$ et $(a_i^-)_{i\in I}$ sont sommables, et on pose dans ce cas :*

$$\sum_{i\in I} a_i = \sum_{i\in I} a_i^+ - \sum_{i\in I} a_i^-$$

2. *La famille $(a_i)_{i\in I}$ est absolument sommable si la famille $(|a_i|)_{i\in I}$ est sommable.*

On a le résultat remarquable suivant :

Propriété 9.31 *La famille $(a_i)_{i\in I}$ est sommable si et seulement si elle est absolument sommable, et lorsque $I = \mathbb{N}$, les trois propositions suivantes sont équivalentes :*

i La famille $(a_i)_{i\in\mathbb{N}}$ est sommable.

ii La série $\sum_{i\geq 0} a_i$ est commutativement convergente, c'est-à-dire que pour toute permutation σ de $\mathbb{N}$, la série $\sum_{i\geq 0} a_{\sigma(i)}$ converge.

iii La série $\sum_{i\geq 0} a_i$ est absolument convergente.

Si l'une de ces trois conditions est réalisée, alors, pour toute permutation σ de $\mathbb{N}$:

$$\sum_{i\in\mathbb{N}} a_i = \sum_{i=0}^{+\infty} a_i = \sum_{i=0}^{+\infty} a_{\sigma(i)}$$

En probabilités, lorsque l'on considère une variable aléatoire discrète dont l'univers image $X(\Omega) = \{x_i, i \in I\}$ est dénombrable, on veut définir l'espérance de X indépendamment de la bijection choisie entre $X(\Omega)$ et $\mathbb{N}$. On demande donc la sommabilité de la famille $(x_i p_i)_{i\in I}$, et pas seulement la convergence de la série.

Exemple 9.34 *Soit X une variable aléatoire dont la loi est donnée par $X(\Omega) = \{(1)^n(n+1), n \in \mathbb{N}^*\}$ et, pour tout $n \in \mathbb{N}^*$, $P(X = n) = \dfrac{1}{n(n+1)}$.*

La série de terme général $nP(X = n) = \dfrac{(-1)^n}{n}$ n'étant pas absolument convergente, l'espérance de X n'existe pas. On peut aussi dire que la famille $\left(\dfrac{(-1)^n}{n}\right)_{n\in\mathbb{N}^*}$ n'est pas sommable puisque la famille $\left(\dfrac{1}{n}\right)_{n\geq 1}$ (c'est-à-dire la famille des a_i^+) ne l'est pas.

Cependant, la série $\sum_{n\geq 1} \dfrac{(-1)^n}{n}$ converge vers $-\ln 2$ (application classique de la formule de Taylor avec reste intégral : développer $x \mapsto \ln(1 + x)$ en $x_0 = 1$). On pourrait donc vouloir poser $E(X) = -\ln 2$, mais on a privilégié une bijection entre $X(\Omega)$ et $\mathbb{N}^*$. Si on définit

$$\sigma \left\{ \begin{array}{ccc} \mathbb{N}^* & \to & X(\Omega) \\ 3k & \mapsto & 2k+1 \\ 3k+1 & \mapsto & -2(2k+1) \\ 3k+2 & \mapsto & -2(2k+2) \end{array} \right.$$

c'est-à-dire la numérotation $1 \mapsto 2$, $2 \mapsto -4$, $3 \mapsto 3$, $4 \mapsto -6$, $5 \mapsto -8$, $6 \mapsto 5$, *et c., on a alors la série :*

$$-1 - \frac{1}{3} + \frac{1}{2} - \frac{1}{5} - \frac{1}{7} + \frac{1}{4} \dots$$

c'est-à-dire la série de terme général

$$\sum_{k \geq 1} \left(-\frac{1}{2(2k+1)} - \frac{1}{2(2k+2)} + \frac{1}{2k+1} \right)$$

Cette série est convergente, puisque son terme général est équivalent à $\dfrac{1}{4k^2}$, *mais elle a pour somme :*

$$\begin{aligned}
\sum_{k=1}^{+\infty} \left(-\frac{1}{2(2k+1)} - \frac{1}{2(2k+2)} + \frac{1}{2k+1} \right) &= \sum_{k=1}^{+\infty} \left(\frac{1}{2k+1} - \frac{1}{2k+2} \right) \times \frac{1}{2} \\
&= \frac{1}{2} \sum_{k=1}^{+\infty} \frac{(-1)^k}{k} \\
&= -\frac{\ln 2}{2}
\end{aligned}$$

La variable aurait alors pour espérance $-\dfrac{\ln 2}{2}$ *!*

Exemple 9.35 *Soient* p *et* q *deux réels appartenant à* $]0;1[$. *Considérons une variable aléatoire* X *dont la loi est donnée par* $X(\Omega) = \mathbb{N}^2$, *et, pour tout couple d'entiers* $(i;j)$, $P(X = (i;j)) = (1-q)q^i(1-p)p^j$. *Soit* I *une partie finie de* $\mathbb{N}^2$. *Il existe deux entiers* M *et* N *tels que* $I \subset [\![0;M]\!] \times [\![0;N]\!]$ *et on a la majoration :*

$$\sum_{(i;j) \in I} p_{i,j} \leq (1-q)(1-p) \sum_{i=0}^{M} \sum_{j=0}^{N} q^i p^j \leq 1$$

La famille est donc sommable, et sa somme vaut 1 :

$$\sum_{(i;j) \in I} p_{i,j} = (1-q)(1-p) \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} q^i p^j = 1$$

Cet ouvrage a été achevé d'imprimer en août 2010 par Normandie Roto Impression s.a.s., 61250 Lonrai
N° d'impression : 102872 - Dépôt légal : septembre 2010 - *Imprimé en France*

Cette collection regroupe des ouvrages variés dont le but est de compléter la formation scientifique des candidats aux concours d'Agrégation et de CAPES de Mathématiques, et éventuellement de leur donner une préparation spécifique à une épreuve ou un type d'épreuve.

Cet ouvrage est le fruit de la participation régulière de l'auteur aux jurys des concours ainsi qu'à la préparation des candidats.

Il propose aux candidats préparant le CAPES de Mathématiques un cours de probabilités correspondant au programme de l'écrit, ainsi que des exercices, tous corrigés, sur chaque chapitre. Les leçons d'oral de dénombrements et de probabilités y sont également traitées, et commentées.

Il s'adresse aussi aux candidats à l'agrégation interne, dont le programme du concours comprend celui du CAPES et quelques points spéciques (signalés comme tels) qui sont détaillés.

Ce livre ne nécessite aucun pré-requis en probabilités et, conformément aux programmes, les probabilités ne sont pas abordées avec le point de vue de la théorie de la mesure.

Jérôme Escoffier, professeur agrégé en classes préparatoires au lycée Saint-Joseph d'Avignon, est membre du jury du CAPES et participe à la préparation au CAPES de l'université d'Avignon.

9 782729 858506

www.editions-ellipses.fr