Capes de Mathématiques
Préparation à l'écrit

Université Joseph Fourier
Année 2008-2009

Liste des fiches de probabilités

Capes de Mathématiques **Université Joseph Fourier**
Préparation à l'écrit **Année 2008-2009**

Probabilités 1 : Introduction aux espaces probabilisés

« Le degré d'excitation qu'éprouve un joueur en faisant un pari est égal au produit du gain par la probabilité de gagner. » Pascal

Introduction

On fait une expérience dont on ne peut pas déterminer le résultat à l'avance, par exemple on jette une pièce de monnaie en l'air et on s'intéresse au côté sur lequel elle retombe (« pile » ou « face »). Que peut-on dire de plus du résultat de cette expérience ? Si on ne jette la pièce qu'une fois, pas grand'chose. Mais si on la jette un grand nombre de fois, on constate que la fréquence d'apparition de « pile » se stabilise et, pour de nombreuses pièces, que cette fréquence devient proche de 50%. On associe alors au résultat de l'expérience « obtenir pile » la valeur 50%, qu'on appellera la probabilité de l'événement « obtenir pile ».

Lier la probabilité d'un événement à la fréquence d'apparition de cet événement lors d'un grand nombre d'expériences résulte d'un théorème important : la « loi des grands nombres ». En effet, on montre que si F_n désigne la valeur observée de cette fréquence après n lancers et si la pièce est « équilibrée » en un certain sens, alors F_n tend vers 50% avec probabilité 1 quand n tend vers l'infini.

De plus, on peut quantifier ce premier résultat en étudiant les fluctuations de la suite $(F_n)_{n\geqslant 1}$ autour de 50%. Par exemple, on a lancé 1000 fois une pièce de monnaie et on a obtenu 537 piles ; est-ce normal ou doit-on en conclure que la pièce est biaisée ? Le « théorème central limite » répond à cette question en montrant que la vitesse de convergence de la suite aléatoire $(F_n)_n$ vers 50% est en $1/\sqrt{n}$. (Dans l'exemple, d'après le théorème central limite, la probabilité d'obtenir 537 piles ou plus que 537 piles vaut à peu près $0,964\%$, valeur approchée que l'on peut comparer avec la vraie valeur $1,046\%$.)

Le cours commence par définir un espace probabilisé $(\Omega, \mathcal{F}, P)$, les probabilités conditionnelles, les variables aléatoires et leurs lois, et la notion d'indépendance. Comme on l'a vu dans l'exemple précédent, on étudie souvent des phénomènes qui correspondent à la répétition d'une même expérience donc on s'interessera ensuite aux suites de variables aléatoires et on énoncera les deux théorèmes limites fondamentaux mentionnés ci-dessus que sont la loi forte des grands nombres et le théorème central limite.

1 Espaces probabilisés

1.1 Définition

Pour définir un espace probabilisé, on a besoin d'un ensemble Ω appelé **univers**, qui peut représenter l'ensemble des résultats possibles de l'expérience considérée. Une première solution est de choisir pour Ω un ensemble aussi petit que possible, donc la collection des résultats possibles de l'expérience.

Quelques exemples :
- Pour le jet d'une pièce de monnaie, $\Omega = \{0, 1\}$ convient, où 0 représente « pile » et 1 représente « face » (ou vice versa !).
- Pour le jet de n pièces de monnaie ou pour n jets d'une seule pièce de monnaie, $\Omega = \{0, 1\}^n$ convient.
- Pour le lancer d'une fléchette sur une cible, $\Omega =$ un disque du plan euclidien convient.
- Pour la durée de vie d'une ampoule électrique, $\Omega = \mathbb{R}^+$.
- Pour battre un jeu de 52 cartes, $\Omega = \mathfrak{S}_{52}$ convient, où $\mathfrak{S}_k$ désigne l'ensemble des permutations de l'ensemble $\{1, 2, \ldots, k\}$.
- Pour compter le nombre de clients dans une file d'attente : $\Omega = \mathbb{N}$ convient.
- Pour lancer une pièce de monnaie une infinité de fois : $\Omega = \{0, 1\}^{\mathbb{N}^*}$ convient, donc Ω est l'ensemble des suites à valeurs dans $\{0, 1\}$, c'est-à-dire l'ensemble des suites $(x_n)_{n \geqslant 1}$ où $x_n = 1$ si le tirage numéro n donne pile et $x_n = 0$ sinon.

Dans ces exemples, les ensembles Ω sont minimaux et on voit bien que tout ensemble plus « gros » conviendrait aussi. Par exemple, $\Omega = \{0, 1\}^{\mathbb{N}^*}$ convient pour modéliser le jet de n pièces de monnaie pour tout $n \geqslant 1$: il suffit de ne s'intéresser qu'aux n premières coordonnées.

Cette remarque suggère une deuxième solution qui consiste à ne pas se préoccuper de la forme exacte de Ω et, par contre, à spécifier ce que l'observateur peut voir ou non. Pour cela, on introduit la tribu des **événements**[1], souvent notée $\mathcal{F}$, qui correspond à « ce que l'observateur voit ». Les éléments de $\mathcal{F}$ sont donc des parties

[1]Le lecteur attentif aura remarqué que nous utilisons dans ces notes la graphie « événement » et non pas « évènement », qui est pourtant recommandée par le service du Dictionnaire de l'Académie française dans un rapport controversé datant de 1990. À chacun de décider pour ce qui le concerne.

de Ω.

Quelles propriétés une telle classe d'événements doit-elle vérifier ? Si l'observateur peut voir si A est réalisé, il peut voir si A^c l'est. De même si l'observateur peut voir si A est réalisé et si B l'est, il peut voir si $A \cup B$ l'est ou non et si $A \cap B$ l'est ou non.

Enfin, on exige que si $(A_n)_{n \geqslant 0}$ est une suite d'événements, alors leur réunion l'est aussi[2].

Définition 1.1. *Une tribu (ou σ-algèbre) de Ω est une partie $\mathcal{F}$ de $\mathcal{P}(\Omega)$ vérifiant les trois axiomes suivants.*

1. *Ω appartient à $\mathcal{F}$.*
2. *Si A appartient à $\mathcal{F}$, alors $\Omega \setminus A$ appartient à $\mathcal{F}$.*
3. *Si $(A_n)_{n \geqslant 0}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n=0}^{+\infty} A_n$ appartient à $\mathcal{F}$.*

Exemple 1.2. *Si Ω est fini, on choisira souvent $\mathcal{F} = \mathcal{P}(\Omega)$, mais pas toujours. Supposons par exemple que l'on jette un dé mais qu'on ne nous donne que la parité du résultat. La tribu associée sera $\mathcal{F} = \{\emptyset, pair, impair, \Omega\}$ et non pas une classe de 2^6 parties.*

La dernière étape de la construction consiste à mesurer chaque élément A de $\mathcal{F}$ par sa « probabilité », notée $P(A)$.

Définition 1.3. *Une probabilité sur $(\Omega, \mathcal{F})$ est une fonction $P : \mathcal{F} \to [0,1]$ telle que :*

1. *$P(\Omega) = 1$*
2. *Pour toute suite $(A_n)_{n \geqslant 0}$ d'éléments de $\mathcal{F}$ deux à deux disjoints et de réunion $A = \bigcup_{n=0}^{+\infty} A_n$,*

$$P(A) = \sum_{n=0}^{+\infty} P(A_n).$$

*Le triplet $(\Omega, \mathcal{F}, P)$ est appelé un **espace probabilisé** ou un **espace de probabilité**.*

[2] Ce dernier point peut faire (et a fait) débat et il existe des versions « finitistes » des axiomes des probabilités qui l'excluent. Ces versions sortent du cadre du programme et nous n'en parlerons donc pas du tout.

1.2 Conséquences

Quelques propriétés faciles.

1) $P(\emptyset) = 0$;

2) Pour tout événement A, $P(\Omega \setminus A) = 1 - P(A)$;

3) Pour tous événements A et B, si $A \subset B$, $P(A) \leqslant P(B)$;

4) Pour tous événements A et B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

5) Si $(A_n)_{n \geqslant 0}$ est une suite croissante d'événements de réunion A, $P(A)$ est la limite de la suite croissante $(P(A_n))_{n \geqslant 0}$;

6) Si $(A_n)_{n \geqslant 0}$ est une suite décroissante d'événements d'intersection A, $P(A)$ est la limite de la suite décroissante $(P(A_n))_{n \geqslant 0}$.

Preuves

1) $\emptyset$ et $\emptyset$ sont disjoints de réunion $\emptyset$ donc $P(\emptyset) = 2P(\emptyset)$ d'où $P(\emptyset) = 0$.

2) A et A^c sont disjoints et de réunion égale à Ω donc $P(\Omega) = 1 = P(A) + P(A^c)$.

3) A et $B \setminus A$ appartiennent à $\mathcal{F}$ et sont disjoints, de réunion égale à B. On en déduit donc que $P(B) = P(A) + P(B \setminus A)$ d'où $P(B) \geqslant P(A)$.

4) A et $B \setminus A$ sont disjoints de réunion égale à $A \cup B$ d'où $P(A \cup B) = P(A) + P(B \setminus A)$. De plus $B \setminus A$ et $B \cap A$ sont disjoints de réunion égale à B d'où $P(B) = P(B \setminus A) + P(B \cap A)$. On en déduit le résultat.

5) Soit $B_n = A_{n+1} \setminus A_n$. La suite $(B_n)_{n \geqslant 0}$ est une suite d'éléments de $\mathcal{F}$ deux à deux disjoints et, pour tout $n \geqslant 0$, A_{n+1} est la réunion des B_k pour $0 \leqslant k \leqslant n$. Donc $P(A_{n+1}) = \displaystyle\sum_{k=0}^{n} P(B_k)$. Cette série est donc convergente et de somme égale à

$$\sum_{k=0}^{+\infty} P(B_k) = P\left(\bigcup_{k=0}^{+\infty} B_k\right) = P\left(\bigcup_{k=0}^{+\infty} A_k\right).$$

On obtient le résultat sur les suites décroissantes par passage au complémentaire.

Exemple 1.4 (Formule de Poincaré)**.** *Soit $A_1, \ldots, A_n$ des événements. Montrer que*

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k+1} s_k, \quad \textit{avec } s_k = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

1.3 Espaces probabilisés finis

On suppose que Ω est fini avec $\Omega = \{\omega_1, \ldots, \omega_n\}$, et que $\mathcal{F} = \mathcal{P}(\Omega)$.

Ou bien on suppose que $\mathcal{F}$ est finie donc (exercice) engendrée par une partition $(A_i)_{1\leqslant i\leqslant n}$ au sens où

$$\mathcal{F} = \{A_I\,;\, I \subset \{1, 2, \ldots, n\}\}, \quad A_I = \bigcup_{i\in I} A_i.$$

La correspondance entre les deux descriptions est $\{\omega_i\} \to A_i$. On revient désormais à la première description.

Pour i dans $\{1, 2, \ldots, n\}$, on note $p_i = P(\{\omega_i\})$, donc $0 \leqslant p_i \leqslant 1$ et $\displaystyle\sum_{i=1}^{n} p_i = 1$.

Réciproquement, à tout n-uplet $(p_1, \ldots, p_n)$ vérifiant ces conditions, on associe une unique probabilité P sur $(\Omega, \mathcal{F})$ donnée par $P(\{\omega_i\}) = p_i$ pour tout i dans $\{1, 2, \ldots, n\}$. Alors, pour toute partie A de Ω,

$$P(A) = \sum_{i\in I} p_i, \qquad I = \{1 \leqslant i \leqslant n\,;\, \omega_i \in A\}.$$

Premier exemple On jette 3 dés équilibrés n fois de suite. Donc $\Omega = (\{1, 2, \ldots, 6\}^3)^n$ et $P(\{\omega\}) = 1/6^{3n}$ pour tout ω dans Ω.

La probabilité d'obtenir au moins une fois 421 vaut $1 - (1 - 1/36)^n$. Quelle est la probabilité d'obtenir au moins une fois 444 ?

Deuxième exemple Une urne contient 10 boules rouges et 5 boules noires. On tire 4 boules simultanément et on s'intéresse au nombre de boules rouges, donc on pourra choisir l'ensemble $\Omega = \{0, 1, 2, 3, 4\}$ et $\omega = i$ si on obtient i boules rouges et $4 - i$ boules noires.

Par exemple, $P(\{3\}) = \displaystyle\binom{5}{1}\binom{10}{3}\Big/\binom{15}{4} = 40/91$ et la probabilité de tirer au moins une boule rouge vaut $1 - \displaystyle\binom{5}{4}\Big/\binom{15}{4} = 272/273$.

2 Exemples de probabilités

2.1 Probabilité uniforme

Pour tout ω dans Ω de cardinal N, on pose $P(\{\omega\}) = 1/N$.

Alors, $P(A) = (\mathrm{card}\, A)/N$ pour toute partie A (c'est le fameux « nombre de cas favorables sur nombre de cas possibles »). Cette probabilité correspond à la notion de tirage au hasard. Tous les tirages sont équiprobables. Les calculs de telles probabilités sont donc des problèmes de dénombrement.

2.2 Probabilités multinomiales

Soit un ensemble fini U de cardinal N (une urne), et $\Omega = U^n$. Cela correspond à l'expérience de tirer n éléments avec replacement dans U. Le cardinal de Ω est N^n et on suppose tous les tirages équiprobables.

Fixons $U_0 \subset U$ de cardinal N_0. Pour $0 \leqslant k \leqslant n$, on s'intéresse à la probabilité p_k qu'un échantillon de taille n tiré au hasard dans U contienne exactement k éléments de U_0. Alors,

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \qquad p = N_0/N.$$

En effet, $p_k = (\mathrm{card}\, A_k)/N^n$ où A_k est l'ensemble des ω tels que ω_i appartient à U_0 pour exactement k indices i. Pour une partie I donnée, de cardinal k, l'ensemble des ω tels que ω_i appartient à U_0 si et seulement si i appartient à I est de cardinal $N_0^k(N-N_0)^{n-k}$, et il y a $\binom{n}{k}$ parties I de cette sorte, donc $\mathrm{card}\,(A_k) = \binom{n}{k} N_0^k (N-N_0)^{n-k}$.

Par définition, on a $p_0+p_1+\ldots+p_n = 1$ donc $(p_0, p_1, \cdots, p_n)$ définit une probabilité sur $\{0, 1, \ldots, n\}$ appelée probabilité binomiale de paramètre (n, p).

On dit aussi que le nombre d'éléments de U_0 suit une loi binomiale de paramètre (n, p).

L'extension à plus de deux parties est la suivante : soit $U = U_1 \cup \ldots \cup U_r$ une partition de U en r parties avec $r \geqslant 2$ et avec $\mathrm{card}\, U_i = N_i$. Notons $p_i = N_i/N$.

Si $n_1 + \ldots + n_r = n$, la probabilité $p(n_1, \ldots, n_r)$ qu'un échantillon de taille n tiré au hasard dans U contienne exactement n_i éléments de U_i pour chaque $1 \leqslant i \leqslant r$ vaut

$$p(n_1, \ldots, n_r) = \frac{n!}{n_1!\; n_2! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r}.$$

On parle alors de probabilité multinomiale de paramètre $(n, (p_1, \ldots, p_r))$.

2.3 Probabilités hypergéométriques

On considère les ensembles précédents et l'expérience consistant à tirer n éléments dans U **sans replacement**, donc $0 \leqslant n \leqslant N$. On suppose tous les tirages équiprobables.

Soit $\Omega = \mathcal{P}_n(U)$ l'ensemble des parties à n éléments de U. Le cardinal de Ω est $\binom{N}{n}$.

Soit $0 \leqslant k \leqslant n$ un entier tel que $k \leqslant N_0$ et $n - k \leqslant N - N_0$. La probabilité p_k qu'une sous-population de taille n de U contienne exactement k éléments de U_0 vaut

$$p_k = \binom{N_0}{k}\binom{N-N_0}{n-k}\Big/\binom{N}{n}.$$

En effet $p_k = \operatorname{card} A_k/\binom{N}{n}$ où A_k est l'ensemble des éléments ω de Ω (donc ω est une partie de U) tels que $\operatorname{card}(\omega \cap U_0) = k$. De tels ω s'écrivent comme $\omega = \omega_0 \cup \omega_1$ avec ω_0 dans $\mathcal{P}_k(U_0)$ et ω_1 dans $\mathcal{P}_{n-k}(U \setminus U_0)$.

Ceci définit une probabilité sur $\{0, 1, \ldots, n\}$ appelée probabilité hypergéométrique de paramètre (n, N, N_0).

On retrouve la formule $\binom{N}{n} = \sum_{k=0}^{n} \binom{N_0}{k}\binom{N-N_0}{n-k}$, où par convention $\binom{n}{k} = 0$ si $k > n$ ou $k < 0$.

Remarque 2.1. *Si N est grand, les deux tirages ne diffèrent pas beaucoup. En effet, si $N \to \infty$ et $N_0/N \to p$ avec $p \in]0,1[$, alors $p_k \to \binom{n}{k} p^k (1-p)^{n-k}$. En effet,*

$$\frac{\binom{N_0}{k}\binom{N-N_0}{n-k}}{\binom{N}{n}} = \binom{n}{k} \frac{N_0 \cdots (N_0 - k + 1)(N - N_0) \cdots (N - N_0 - n + k + 1)}{N(N-1)\cdots(N-n+1)},$$

donc

$$\frac{\binom{N_0}{k}\binom{N-N_0}{n-k}}{\binom{N}{n}} \sim \binom{n}{k} \frac{N_0^k (N-N_0)^{n-k}}{N^n} \to \binom{n}{k} p^k (1-p)^{n-k}.$$

3 Probabilités conditionnelles et indépendance de deux événements

La notion de probabilité conditionnelle apparaît lorsqu'on connaît un résultat partiel de l'expérience.

Par exemple, supposons qu'une usine construise des objets de types A et B, et que ces objets peuvent ou non être de bonne qualité, propriété qu'on notera Q.

L'expérience nous permet de connaître $P(A \cap Q)$, estimé par $N_{A\cap Q}/N$ où N_C est le nombre d'objets partageant la propriété C parmi N produits.

On se restreint maintenant à la sous-population des objets de type A et on se demande quelle est la probabilité qu'un objet tiré au hasard dans cette sous-population soit de bonne qualité.

On va donc estimer ceci par $N_{A\cap Q}/N_A$. Or $N_{A\cap Q}/N_A = (N_{A\cap Q}/N)/(N_A/N)$.

On a besoin d'une nouvelle notation pour indiquer qu'on se restreint à cette sous population et on notera $P(Q|A)$, la probabilité de Q sachant A.

On a donc $P(Q|A) \approx N_{A\cap Q}/N_A \approx P(Q \cap A)/P(A)$.

Définition 3.1. *Soit A un événement de probabilité $P(A)$ non nulle. On définit une application $P(\ |A) : \mathcal{F} \to [0,1]$ par*

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

La quantité $P(B|A)$ est la probabilité conditionnelle de B sachant A.

Remarque 3.2. *La relation précédente entraîne*

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

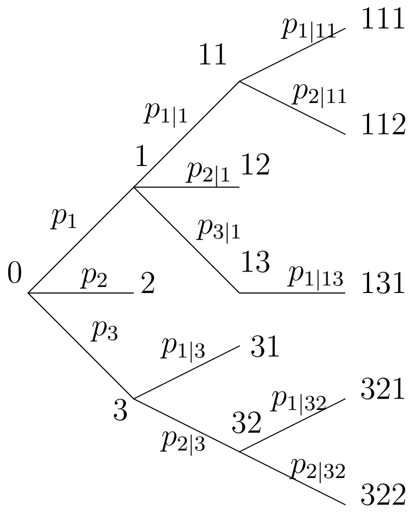
Par récurrence, on en déduit :

Proposition 3.3 (Probabilités conditionnelles en cascade)**.**

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}),$$

à condition de donner au second membre la valeur zéro dès que l'un de ses termes est nul (même si certaines des probabilités conditionnelles y figurant ne sont alors pas définies).

Dans certains problèmes où apparaissent les probabilités conditionnelles, on peut utiliser des arbres. Par exemple, supposons qu'on associe à un individu choisi au

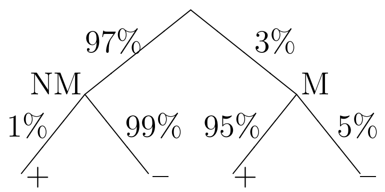


hasard dans une population un « comportement » $(X_1, \ldots, X_n)$, dont la longueur n peut dépendre de l'individu. On considèrera alors l'arbre représenté ici.

On a noté $p_{x_k|x_1 \cdots x_{k-1}}$ la probabilité du comportement $(x_1, \ldots, x_k)$ sachant qu'on a le comportement $(x_1, \ldots, x_{k-1})$.

Considérons par exemple la situation suivante : une maladie atteint 3% d'une population. Un test de dépistage donne les résultats suivants : chez les malades, 95% des tests sont positifs ; chez les non-malades, 99% des tests sont négatifs.

On représentera ces données dans l'arbre suivant :



Les probabilités conditionnelles sont souvent utilisées pour étudier des phénomènes évoluant au cours du temps, par exemple des suites de lancers de pile ou face ou de tirages dans une urne.

Pour un exemple typique, supposons qu'une urne U contient N éléments et que l'on tire r éléments les uns après les autres de la façon suivante : quels que soient les k premiers éléments tirés, au tirage numéro $k+1$, chacun des $N-k$ éléments restants a une probabilité $1/(N-k)$ d'être choisi.

Soit Ω l'ensemble des r-uplets d'éléments distincts de U et notons X_i l'application donnant le résultat du tirage numéro i. Fixons un élément $a = (a_1, a_2, \ldots, a_r)$ de

Ω et, pour $1 \leqslant k \leqslant r$, soit A_k l'événement « on a tiré l'élément a_k au tirage numéro k ». Alors,

$$P(\{a\}) = P(A_1 \cap A_2 \cap \ldots \cap A_r),$$

donc

$$P(\{a\}) = P(A_1)P(A_2|A_1) \cdots P(A_r|A_1 \cap A_2 \cap \ldots \cap A_{r-1}),$$

donc

$$P(\{a\}) = 1/(N(N-1)\cdots(N-r+1)),$$

c'est-à-dire que P est la probabilité uniforme sur Ω comme on pouvait s'en douter.

On peut calculer la probabilité de tirer k éléments dans $U_0 \subset U$ de la même façon, et on retrouve les probabilités hypergéométriques.

Proposition 3.4. *Soit A un événement de probabilité $P(A)$ non nulle. Alors $P(\ |A) : \mathcal{F} \to [0,1]$, $B \mapsto P(B|A)$, est une probabilité sur $(\Omega, \mathcal{F})$.*

La démonstration découle immédiatement de la définition.

Proposition 3.5 (Formule des probabilités totales)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $(A_i)_{i\in I}$ une partition dénombrable de Ω en événements de probabilités $P(A_i)$ strictement positives pour tout $i \in I$. Alors, pour tout événement A,*

$$P(A) = \sum_{i\in I} P(A|A_i)P(A_i).$$

Preuve : La famille $(A \cap A_i)_{i\in I}$ est une partition de A, donc,

$$P(A) = \sum_{i\in I} P(A \cap A_i) = \sum_{i\in I} P(A|A_i)P(A_i).$$

Proposition 3.6 (Formule de Bayes)**.** *Sous les hypothèses de la proposition précédente, pour tout événement A de probabilité $P(A)$ non nulle et pour tout i,*

$$P(A_i|A) = \frac{P(A|A_i)P(A_i)}{\displaystyle\sum_{j\in I} P(A|A_j)P(A_j)}.$$

Preuve : Il suffit d'écrire $P(A_i|A) = \dfrac{P(A|A_i)P(A_i)}{P(A)}$ et d'utiliser la formule des probabilités totales.

Supposons à présent que les événements A et B sont tels que la réalisation de A n'induise rien sur la réalisation de B. Alors, $P(B|A) = P(B)$ donc $P(B \cap A) = P(A)P(B)$.

Définition 3.7. *Deux événements A et B sont indépendants si et seulement si*

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

Remarque 3.8. ***Attention :*** *la notion d'indépendance dépend de la probabilité P considérée !*

Exemple : on dispose de deux pièces de monnaie, l'une équilibrée et l'autre équilibrée ou non. On choisit une des deux pièces au hasard de manière uniforme puis on la lance deux fois. On peut donc toujours utiliser comme espace de probabilité $\Omega = \{\text{pile}, \text{face}\}^2$ et $\mathcal{F} = \mathcal{P}(\Omega)$ mais la probabilité P dépend du biais éventuel de la deuxième pièce.

Soit A l'événement « le premier jet donne face » et B l'événement « le deuxième jet donne face ». Montrer que A et B sont indépendants si et seulement si la deuxième pièce est équilibrée.

Proposition 3.9. *Soient A et B deux événements. Soient $\mathcal{A}$ et $\mathcal{B}$ les tribus engendrées respectivement par A et B, c'est à dire $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ et $\mathcal{B} = \{\emptyset, B, B^c, \Omega\}$. Alors A et B sont indépendants si et seulement si pour tout $C \in \mathcal{A}$ et pour tout $D \in \mathcal{B}$, $P(C \cap D) = P(C)P(D)$.*

Preuve : Par exemple, $P(A \cap B^c)$ vaut

$$P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c).$$

4 Indépendance d'événements

On veut généraliser la notion d'indépendance à un nombre quelconque d'événements.

Définition 4.1. *Les événements $(A_i)_{1 \leqslant i \leqslant n}$ sont indépendants si et seulement si pour tout $I \subset \{1, 2, \ldots, n\}$, on a*

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Rappelons que par convention, un produit sur un ensemble vide d'indices vaut 1 (de même qu'une somme sur un ensemble vide d'indices vaut 0), donc on peut choisir $I = \emptyset$ dans la définition précédente.

Proposition 4.2. *Considérons les propriétés suivantes :*

(i) Les événements $(A_i)_{1\leqslant i\leqslant n}$ *sont indépendants.*

(ii) Pour toute famille $(B_i)_{1\leqslant i\leqslant n}$ *telle que* $B_i \in \{A_i, A_i^c\}$,

$$P\left(\bigcap_{1\leqslant i\leqslant n} B_i\right) = \prod_{1\leqslant i\leqslant n} P(B_i).$$

(iii) Pour toute famille $(B_i)_{1\leqslant i\leqslant n}$ *telle que* $B_i \in \{\emptyset, A_i, A_i^c, \Omega\}$,

$$P\left(\bigcap_{1\leqslant i\leqslant n} B_i\right) = \prod_{1\leqslant i\leqslant n} P(B_i).$$

Les propriétés (i), (ii) et (iii) sont équivalentes.

Preuve Tout d'abord, (iii) implique (i) et (ii) car (i) revient à se restreindre aux $B_i \in \{A_i, \Omega\}$ et (ii) revient à se restreindre aux $B_i \in \{A_i, A_i^c\}$. Ensuite (ii) implique (iii) : si on utilise une fois $\emptyset$ dans (iii), ça marche ; sinon, (iii) s'obtient en sommant 2^a égalités (ii), où a désigne le nombre de fois où (iii) utilise Ω. Reste à montrer que (i) implique (iii).

Supposons que (i) est vraie et (iii) fausse. Pour chaque famille $B = (B_i)_{1\leqslant i\leqslant n}$ qui infirme (iii), notons $n(B)$ le nombre de fois où B utilise A_i^c et non pas A_i ou Ω (puisque dès qu'on utilise $\emptyset$, ça marche). Le cas $n(B) = 0$ est impossible car (i) est vraie, donc $n(B) \geqslant 1$. Supposons que la famille B est telle que $n(B)$ est minimal et soit i un indice tel que $B_i = A_i^c$. Alors la famille B' obtenue en remplaçant A_i^c par A_i dans la famille B est telle que $n(B') = n(B) - 1 < n(B)$ donc B' vérifie (iii). En sommant le fait que (iii) pour B est fausse et le fait que (iii) pour B' est vraie, on obtient que (iii) est fausse pour la famille B'' obtenue à partir de B en remplaçant A_i^c par Ω. Comme $n(B'') = n(B) - 1$, c'est absurde.

Remarque 4.3. ***Attention :*** *le fait que les événements* $(A_i)_{i\in I}$ *sont indépendants entraîne qu'ils sont indépendants deux à deux mais la réciproque est fausse.*

Exemple 4.4. *On lance une pièce deux fois de suite. Soit* A *l'événement « obtenir un face au premier jet »,* B *l'événement « obtenir un face au deuxième jet », et* C *l'événement « obtenir deux résultats différents ».*

Alors $P(A) = P(B) = P(C) = \frac{1}{2}$ *et* $P(A\cap B) = P(A\cap C) = P(B\cap C) = \frac{1}{4}$, *donc* A, B, C *sont indépendants deux à deux. Mais* $A\cap B\cap C = \emptyset$ *donc* $P(A\cap B\cap C) = 0 \neq P(A)P(B)P(C) = \frac{1}{8}$ *donc* (A, B, C) *n'est pas une famille indépendante.*

Question : Pour tout $n \geqslant 2$, trouver n événements non indépendants tels que toute collection de $n-1$ événements parmi ces n est indépendante.

5 Exercices

Espaces de probabilité

1. Premier problème du chevalier de Méré Quel est le plus probable : jouant avec un dé, obtenir au moins une fois 6 en 4 coups, ou bien jouant avec deux dés, obtenir au moins une fois un double 6 en 24 coups ?

2. Second problème du chevalier de Méré Le chevalier de Méré avait posé à Pascal le problème suivant : deux joueurs jouent à un jeu de hasard en plusieurs parties ; celui qui, le premier, a gagné trois parties emporte la totalité de l'enjeu. Si les joueurs doivent arrêter le jeu alors qu'il ne manque au premier joueur, pour l'emporter, qu'une partie, et au second que deux parties, comment doit-on répartir équitablement l'enjeu ?

3. Problème des anniversaires Quelle est la probabilité pour que n personnes prises au hasard aient toutes des jours d'anniversaire différents ?

On supposera que tous les jours de naissance sont équiprobables et on ne tiendra pas compte des années bissextiles.

Déterminer la plus petite valeur de n telle que cette probabilité soit inférieure à 50%.

4. Formule de Poincaré, problème des rencontres a) Montrer que si A_1, $\ldots$, A_n sont n événements d'un même espace probabilisé et si A désigne la réunion de ces n événements, on a :

$$P(A) = \sum_{k=1}^{n} (-1)^{k+1} s_k, \quad s_k = \sum_{i_1 < i_2 < \ldots < i_k} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}).$$

b) On tire sans remise n boules numérotées de 1 à n. Déterminer la probabilité p_n pour qu'il existe un entier k tel que la boule portant le numéro k soit tirée au tirage numéro k.

c) Déterminer la limite de p_n quand n tend vers l'infini.

5. Bridge Calculer la probabilité qu'au bridge, chaque joueur ait un as.

6. Balles et paniers On a r balles et n paniers numérotés de 1 à n.

On répondra aux questions dans les deux cas suivants :

(a) Les r balles sont discernables (par exemple parce qu'elles sont de couleurs différentes).

(b) Les r balles sont indiscernables.

Question 1 : Quel est le nombre de répartitions possibles (un panier peut contenir plusieurs balles) ?

Question 2 : Quelle est la probabilité p_k qu'un panier donné contienne exactement k balles. Étudier la monotonie de la suite $(p_k)_{0 \leqslant k \leqslant r}$.

Question 3 : On suppose que n et r tendent vers l'infini et que r/n tend vers λ. Montrer que chaque terme p_k admet une limite et calculer celle-ci.

7. Problème du scrutin Lors d'un vote opposant deux candidats A et B, A obtient a voix et B obtient b voix. On suppose que $a < b$. Quelle est la probabilité pour qu'au cours du dépouillement, B ait toujours été strictement en tête ?

On pourra représenter le dépouillement par un chemin du plan constitué de segments horizontaux ou verticaux de longueur 1 joignant l'origine au point de coordonnées (a, b) et compter le nombre de chemins situés au dessus de la diagonale.

8. Milieu Soit S l'ensemble des points de l'espace dont les coordonnées x, y et z sont entières et vérifient $0 \leqslant x \leqslant 2$, $0 \leqslant y \leqslant 3$ et $0 \leqslant z \leqslant 4$. Deux points sont choisis au hasard dans S, c'est à dire uniformément dans $\mathcal{P}_2(S)$. Quelle est la probabilité que le milieu du segment qu'ils déterminent appartiennent à S ?

9. Table On dispose $n \geqslant 3$ personnes autour d'une table ronde. Trois personnes distinctes sont choisies au hasard. Calculer la probabilité de l'événement « au moins deux parmi les trois étaient assises l'une à coté de l'autre. »

10. Journal Un responsable de jeux dispose de trois enveloppes de couleurs différentes. Il met de l'argent dans l'une, et du papier journal dans les deux autres. Il fait entrer un joueur et lui fait choisir une enveloppe qu'il garde fermée. Parmi les deux enveloppes restantes, il y en a toujours au moins une qui contient du papier journal. Le responsable ouvre alors une de ces deux enveloppes dont il sait qu'elle contient du papier journal, et propose au joueur de changer l'enveloppe qu'il a en main contre celle qui reste. Le joueur a t-il intérêt à changer ?

Probabilités conditionnelles et indépendance

11. Écrous Dans une usine d'écrous, trois machines A, B et C produisent respectivement 25%, 35% et 40% du total de la production. Elles produisent respectivement 5%, 4% et 2% de pièces défectueuses. Un écrou est tiré au hasard et s'avère défectueux. Quelle est la probabilité qu'il ait été produit par la machine A ? B ? ou C ?

12. Sexe M. Dupont a deux enfants dont une fille, quelle est la probabilité que l'autre soit un garçon ? M. Durand a deux enfants dont le plus jeune est une fille, quelle est la probabilité pour que l'autre soit un garçon ?

13. Le rouge et le noir Une urne contient n boules noires et n boules rouges. On tire deux par deux sans remise, toutes les boules de l'urne. Quelle est la probabilité d'obtenir à chaque tirage deux boules de couleurs différentes ?

14. Urnes On dispose de $N+1$ urnes numérotées de 0 à N. L'urne numéro k contient k boules rouges et $N-k$ boules noires. On tire une des urnes avec équiprobabilité, puis on procède avec cette urne à une série de n tirages avec remise.

a) Calculer la probabilité d'avoir choisi l'urne numéro 1 sachant qu'on a tiré n boules rouges.

b) Calculer la probabilité de tirer n boules rouges.

c) Calculer la probabilité de tirer une boule rouge au tirage $n+1$ sachant qu'on a déjà tiré n boules rouges.

d) Déterminer les limites des probabilités précédentes quand $N \to +\infty$.

15. Un peu d'arithmétique Pour tout entier $n \geqslant 2$ fixé, soit P_n la probabilité uniforme sur l'ensemble $\{1, 2, \ldots, n\}$. Pour tout diviseur m de n désignons par A_m le sous-ensemble de $\{1, 2, \ldots, n\}$ formé des multiples de m.

Montrer que $P_n(A_m) = 1/m$.

Montrer que les A_p où p parcourt les diviseurs premiers de n, sont des événements indépendants dans l'espace probabilisé $(\{1, 2, \ldots, n\}, P_n)$. En déduire que l'ensemble des entiers de $\{1, 2, \ldots, n\}$ premiers avec n a une probabilité $\displaystyle\prod_p \left(1 - \frac{1}{p}\right)$

où p parcourt l'ensemble des diviseurs premiers de n et en déduire le cardinal de cet ensemble. Retrouver ainsi une formule d'Euler.

16. Un peu de génétique Les gènes (unités d'hérédité) se présentent dans les cas les plus simples en paires et sous deux formes, appelées allèles, et notées A et a. Chaque individu possède un des trois génotypes AA, aa (qui sont homozygotes) et Aa (qui est hétérozygote).

Chaque individu reçoit au hasard un gène de chacun des génotypes de ses parents, chacun des gènes de ces génotypes ayant la probabilité $\frac{1}{2}$ de passer à l'enfant. Par exemple, deux parents Aa donneront à leur enfant le génotype AA avec probabilité $\frac{1}{4}$, Aa avec probabilité $\frac{1}{2}$ et aa avec probabilité $\frac{1}{4}$.

Les génotypes des parents de la génération 0 sont supposés indépendants. La probabilité qu'un parent de la génération 0 ait AA (respectivement Aa, respectivement aa) comme génotype est notée u (resp. $2v$, resp. w), donc $u+2v+w = 1$ et $(u, 2v, w)$ est appelé la fréquence des génotypes.

On note $p = u + v$ et $q = v + w$ la fréquence des allèles A et a.

a) Montrer qu'un individu de la première génération appartient à l'un des génotypes AA, Aa, aa avec les probabilités respectives $u_1 = p^2$, $2v_1 = 2pq$ et $w_1 = q^2$. En conclure que passé la première génération, la loi de probabilité u, v, w des génotypes des individus d'une population doit vérifier sous les hypothèses précédentes la relation de Hardy-Weinberg : $v^2 = uw$. Montrer que dans ce cas, $u = p^2$, $2v = 2pq$ et $w = q^2$.

b) On considère dans les questions suivantes que $u = p^2$, $2v = 2pq$ et $w = q^2$.

Calculer la probabilité que le génotype d'un individu de la première génération soit Aa sachant que son frère a le même génotype.

c) Calculer pour i et j dans $\mathcal{G} = \{AA, Aa, aa\}$, la probabilité $p_{i,j}$ que le génotype d'un enfant soit j sachant que le génotype de son père est i. On note P la matrice associée, indexée par $\mathcal{G} \times \mathcal{G}$.

d) Montrer que la probabilité que le génotype d'un individu de la seconde génération soit j sachant que le génotype de son grand-père paternel est i est donnée par

$$p^{(2)}_{i,j} = \sum_{k \in \mathcal{G}} p_{i,k} p_{k,j}$$

Plus généralement, montrer que la probabilité que le génotype d'un individu de la génération n soit j sachant que le génotype de son ancêtre masculin de la génération 0 était i est donnée par le coeficient (i, j) de la matrice P^n. Calculer P^n.

17. Urne de Pólya Soient $a \geqslant 0$, $b \geqslant 0$ et $c \geqslant 0$ des entiers avec $a+b \geqslant 1$. Une urne contient a boules noires et b boules rouges. Si on tire une boule, on remet dans l'urne c boules de la couleur de la boule tirée. (Le cas du tirage avec remise simple est donnée par $c = 1$ et celui du tirage sans remise par $c = 0$).

a) Calculer la probabilité qu'au deuxième tirage, on tire une boule noire.

b) Calculer la probabilité qu'au troisième tirage, on tire une boule noire.

c) On note X_i la variable aléatoire valant 1 si on tire une boule noire au tirage numéro i et 0 sinon. Que représente la variable aléatoire $S_i = X_1 + \cdots + X_i$?

En utilisant la formule des probabilités totales montrer que

$$P(X_{i+1} = 1) = (E[S_i] + a)/(a + b + i).$$

En déduire $P(X_i = 1)$.

18. Trois Anne, Barbara et Cléo jouent une suite de parties en suivant la règle suivante : chaque partie oppose deux joueuses et la perdante de la partie s'efface à la fin de la partie pour laisser la place aux deux autres joueuses. Les parties sont supposées indépendantes et chaque joueuse a la même probabilité à chaque fois de gagner. Le jeu s'arrête dès qu'une joueuse a gagné deux parties consécutives ; cette joueuse est alors la gagnante du jeu.

Déterminer la probabilité de chacune des trois joueuses de gagner si Anne et Barbara jouent la première partie. Pour tout entier n, calculer la probabilité que le jeu s'arrête au bout de n parties exactement.

19. QCM Dans un QCM, il y a m réponses possibles. Un candidat a une probabilité p de connaître la réponse à une question prise au hasard parmi un ensemble fini de questions.

Sachant que le candidat a répondu juste à la question, quelle est la probabilité qu'il connaisse effectivement la réponse ? On suppose qu'un candidat qui ne connaît pas la réponse, répond au hasard et donc que les m réponses sont équiprobables.

Probabilités 2 : Variables aléatoires discrètes

« Il est souvent assez difficile de faire des prédictions, surtout en ce qui concerne le futur. » Apocryphe, parfois attribué à Niels Bohr

On convient que « dénombrable » signifie « fini » ou « dénombrable infini. »

La première partie du chapitre est consacrée à des rappels, sans doute en partie indigestes, sur ce qu'il est licite et illicite de faire avec des sommes infinies ; rien de spécifiquement probabiliste donc, mais des résultats à connaître en particulier pour faire des probabilités. On introduit ensuite les variables aléatoires, leur loi, leurs moments et la propriété d'indépendance, le tout dans le cas discret.

1 Familles sommables

On rappelle que toute partie A de $\mathbb{R}$ non vide et majorée admet une borne supérieure, notée $\sup A$; c'est le plus petit des majorants de A. Si A n'est pas majorée, on note $\sup A = +\infty$.

On se donne un ensemble dénombrable I et une famille $(x_i)_{i\in I}$ de nombres réels **positifs**.

Définition 1.1. *Soit $A \subset \mathbb{R}^+$ l'ensemble des sommes finies $\sum_{i\in J} x_i$ avec J partie finie de I. On pose $\sum_{i\in I} x_i := \sup A$. Si A est borné, on dit que la famille $(x_i)_{i\in I}$ est sommable.*

La somme $\sum_{i\in I} x_i$ est donc un élément de $\mathbb{R}^+ \cup \{+\infty\}$.

Dans le cas particulier où I est vide, on pose $\sum_{i\in I} x_i := 0$, ce qui correspond au supremum de $\emptyset$ dans $\mathbb{R}^+$.

Si $(y_i)_{i\in I}$ est une seconde famille de réels positifs telle que pour tout i dans I, $x_i \leqslant y_i$, alors $\sum_{i\in I} x_i \leqslant \sum_{i\in I} y_i$. Cas particulier : si $J \subset I$, alors $\sum_{i\in J} x_i \leqslant \sum_{i\in I} x_i$.

On explique maintenant comment calculer la valeur d'une somme infinie à partir de sommes finies.

Proposition 1.2. *Soit $(J_n)_{n\geqslant 0}$ une famille croissante de parties de I, finies ou infinies, dont la réunion est égale à I. Alors*

$$\sum_{i\in I} x_i = \lim_{n\to+\infty} \sum_{i\in J_n} x_i.$$

Il suffit donc de faire le calcul pour **une** suite croissante $(J_n)_n$ de réunion I, et le résultat sera toujours le même.

Preuve : On doit montrer que $S = \ell$ où S désigne la somme sur I tout entier et ℓ la limite du membre de droite. Pour tout n, $\sum\limits_{i\in J_n} x_i \leqslant S$. Comme la suite de terme général $\sum\limits_{i\in J_n} x_i$ est croissante, en passant à la limite on obtient $\ell \leqslant S$.

Dans l'autre sens, soit $J \subset I$ une partie finie. À tout élément j de J on peut associer un entier $n(j)$ tel que $j \in J_{n(j)}$. En posant $N = \max_{j\in J} n(j)$, on obtient $J \subset J_N$ et

$$\sum_{i\in J} x_i \leqslant \sum_{i\in J_N} x_i \leqslant \ell,$$

ce qui termine la preuve.

Remarque 1.3. *Si $I = \mathbb{N}$, on obtient* $\sum\limits_{i\in I} x_i = \sum\limits_{n=0}^{+\infty} x_n = \lim\limits_{n\to+\infty} \sum\limits_{k=0}^{n} x_n$.

Désormais, $(x_i)_{i\in I}$ désigne une famille de nombres réels **pas forcément positifs**. Pour tout nombre réel x, on note x^+ sa partie positive et x^- sa partie négative, définies par

$$x^+ := \max(x, 0), \quad x^- := \max(-x, 0),$$

ou bien, ce qui est équivalent, par les deux relations

$$x = x^+ - x^-, \quad |x| = x^+ + x^-.$$

On remarquera que $x^+ \geqslant 0$ et $x^- \geqslant 0$ pour tout nombre réel x.

Définition 1.4. *La famille $(x_i)_{i\in I}$ est sommable si et seulement si la famille $(|x_i|)_{i\in I}$ d'éléments de $\mathbb{R}^+$ est sommable. Alors, comme $x_i^+ \leqslant |x_i|$ et $x_i^- \leqslant |x_i|$, les familles $(x_i^+)_{i\in I}$ et $(x_i^-)_{i\in I}$ sont également sommables. On peut donc poser*

$$\sum_{i\in I} x_i := \sum_{i\in I} x_i^+ - \sum_{i\in I} x_i^-.$$

Si les deux séries $\sum_{i\in I} x_i^+$ et $\sum_{i\in I} x_i^-$ n'étaient pas convergentes, la différence de leurs sommes n'aurait tout simplement pas de sens.

Proposition 1.5. *Soit $(x_i)_{i\in I}$ une famille sommable et $(J_n)_{n\geq 0}$ une famille croissante de parties de I, finies ou infinies, dont la réunion est égale à I. Alors*

$$\sum_{i\in I} x_i = \lim_{n\to+\infty} \sum_{i\in J_n} x_i.$$

Preuve : Comme chaque famille $(x_i)_{i\in J_n}$ est sommable,

$$\sum_{i\in J_n} x_i = \sum_{i\in J_n} x_i^+ - \sum_{i\in J_n} x_i^-,$$

et il suffit de passer à la limite en utilisant la proposition 1.2. Fin de la preuve.

Remarque 1.6. *Si $I = \mathbb{N}$, la famille $(x_n)_{n\in\mathbb{N}}$ est sommable si et seulement si la série $\sum_{n=0}^{+\infty} x_n$ est absolument convergente.*

En appliquant ce qui précède à des ensembles J_n qui sont finis, on obtient

$$\left|\sum_{i\in J_n} x_i\right| \leqslant \sum_{i\in J_n} x_i^+ + \sum_{i\in J_n} x_i^- = \sum_{i\in J_n} |x_i|.$$

D'où le résultat suivant :

Proposition 1.7. *Si la famille $(x_i)_{i\in I}$ est sommable, alors*

$$\left|\sum_{i\in I} x_i\right| \leqslant \sum_{i\in I} |x_i|.$$

Voici d'autres propriétés.

Proposition 1.8 (Linéarité)**.** *Soit $(x_i)_{i\in I}$ et $(y_i)_{i\in I}$ deux familles sommables. Soit a et b deux nombres réels et, pour tout $i \in I$, posons $z_i := ax_i + by_i$. Alors la famille $(z_i)_{i\in I}$ est sommable et*

$$\sum_{i\in I} z_i = a\sum_{i\in I} x_i + b\sum_{i\in I} y_i.$$

Preuve : Si J est une partie finie de I,

$$\sum_{i\in J}|z_i| \leqslant |a|\sum_{i\in J}|x_i| + |b|\sum_{i\in J}|y_i| \leqslant |a|\sum_{i\in I}|x_i| + |b|\sum_{i\in I}|y_i|,$$

donc la famille $(z_i)_{i\in I}$ est sommable. Si $(J_n)_{n\geq 0}$ est une suite croissante de parties finies de I de réunion égale à I,

$$\sum_{i\in J_n} z_i = a\sum_{i\in J_n} x_i + b\sum_{i\in J_n} y_i,$$

donc on obtient le résultat par passage à la limite. Fin de la preuve.

Proposition 1.9 (Sommation par paquets)**.** *Soit $(I_k)_{k\in K}$ une partition de I, donc les ensembles I_k sont deux à deux disjoints et leur réunion est égale à I. Soit $(x_i)_{i\in I}$ une famille sommable. Alors,*
1) Pour tout k dans K, la famille $(x_i)_{i\in I_k}$ est sommable.
2) Soit $S_k = \sum_{i\in I_k} x_i$. La famille $(S_k)_{k\in K}$ est sommable et

$$\sum_{k\in K} S_k := \sum_{i\in I} x_i.$$

Réciproquement si pour tout $k \in K$, la famille $(x_i)_{i\in I_k}$ est sommable et si la famille $(T_k)_{k\in K}$ est sommable où on a noté, pour chaque k, $T_k := \sum_{i\in I_k}|x_i|$, alors la famille $(x_i)_{i\in I}$ est sommable.

Preuve : 1) Si $J \subset I_k$ est fini,

$$\sum_{i\in J}|x_i| \leqslant \sum_{i\in I}|x_i|,$$

donc $\sum_{i\in I_k}|x_i| \leqslant \sum_{i\in I}|x_i|$, qui est fini.
2) Pour tout $k \in K$, introduisons une suite croissante $(I_{k,n})_{n\geqslant 0}$ de parties finies de I_k dont la réunion est I_k. On pose $S_{k,n} = \sum_{i\in I_{k,n}} x_i$. On sait par la proposition 1.2 que $S_{k,n}$ tend vers S_k quand n tend vers l'infini.
Pour tout $F \subset K$ fini et pour tout n,

$$\sum_{k\in F}|S_{k,n}| \leqslant \sum_{k\in F}\sum_{i\in I_{k,n}}|x_i| = \sum_{i\in F_n}|x_i| \leqslant \sum_{i\in I}|x_i|,$$

où F_n désigne la réunion des $I_{k,n}$ pour $k \in F$. En faisant tendre n vers $+\infty$,

$$\sum_{k\in F} |S_k| \leqslant \sum_{i\in I} |x_i|, \quad \sum_{k\in K} |S_k| \leqslant \sum_{i\in I} |x_i|,$$

donc la famille $(S_k)_{k\in K}$ est sommable. De plus,

$$\sum_{k\in F} S_{k,n} = \sum_{k\in F} \sum_{i\in I_{k,n}} x_i = \sum_{i\in F_n} x_i.$$

Or la suite $(F_n)_{n\geq 0}$ de parties de I est croissante et sa réunion vaut la réunion des I_k pour $k \in F$, notée I_F, donc

$$\sum_{k\in F} S_k = \sum_{i\in I_F} x_i.$$

En considérant maintenant une suite croissante de parties finies F de K, on obtient le résultat par la proposition 1.2.

Pour la réciproque, si J est une partie finie de I, $J \cap I_k$ est une partie finie de I_k donc $\displaystyle\sum_{i\in J\cap I_k} |x_i| \leqslant T_k$. De plus, l'ensemble K_J des indices k dans K tels que $J \cap I_k$ n'est pas vide est fini, donc $\displaystyle\sum_{i\in J} |x_i| = \sum_{k\in K_J} \sum_{i\in J\cap I_k} |x_i| \leqslant \sum_{k\in K} T_k$, qui est fini. Fin de la preuve.

Corollaire 1.1 (Théorème de Fubini). *Soient $(x_{i,j})_{(i,j)\in I\times J}$ une famille de nombres réels indexée par deux ensembles dénombrables I et J.*
1) Supposons que, pour tout (i,j), $x_{i,j} \geqslant 0$. Alors

$$\sum_{(i,j)\in I\times J} x_{i,j} = \sum_{i\in I} \left(\sum_{j\in J} x_{i,j} \right) = \sum_{j\in J} \left(\sum_{i\in I} x_{i,j} \right), \tag{1}$$

avec la convention que, pour tout réel x, $x + (+\infty)$ et $(+\infty) + (+\infty)$ valent $+\infty$.
2) La famille $(x_{i,j})_{(i,j)\in I\times J}$ est sommable si et seulement si, pour tout $i \in I$, $\displaystyle\sum_{j\in J} |x_{i,j}|$ est fini et si $\displaystyle\sum_{i\in I} \left(\sum_{j\in J} |x_{i,j}| \right)$ est fini. L'égalité (1) est alors valide.

Preuve : 1) Si $\displaystyle\sum_{(i,j)\in I\times J} x_{i,j}$ est finie, il suffit d'appliquer la sommation par paquets à la partition $I \times J = \displaystyle\bigcup_{i\in I} (\{i\} \times J)$.

Si $\sum_{(i,j)\in I\times J} x_{i,j}$ est infinie, soit il existe $i \in I$ tel que $\sum_{j\in J} x_{i,j}$ est infinie et dans ce cas l'égalité est vraie, soit pour tout $i \in I$, $\sum_{j\in J} x_{i,j}$ est finie mais alors le théorème de sommation par paquets entraîne que $\sum_{i\in I}\left(\sum_{j\in J} x_{i,j}\right)$ est infinie.
2) Sommation par paquets pour la partition $I \times J = \bigcup_{i\in I}(\{i\} \times J)$. Fin de la preuve.

2 Variables aléatoires

2.1 Flèches

Définition 2.1. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité.*
Une variable aléatoire réelle est une application $X : \Omega \to \mathbb{R}$ telle que, pour tout nombre réel x, l'ensemble $\{\omega \in \Omega\,;\, X(\omega) \leqslant x\}$ appartient à $\mathcal{F}$.
Une variable aléatoire vectorielle est une application $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ telle que, pour tout i, X_i est une variable aléatoire réelle.
On dit que la variable aléatoire X est discrète quand $X(\Omega)$ est dénombrable.

Quelques remarques et notations.

Si $\mathcal{F} = \mathcal{P}(\Omega)$, toute application de Ω dans $\mathbb{R}$ ou $\mathbb{R}^d$ est une variable aléatoire.

Si Ω est dénombrable, toute variable aléatoire est discrète.

On admettra le résultat suivant :

Proposition 2.2. *Si $f : \mathbb{R}^d \to \mathbb{R}^n$ est continue et si $X : \Omega \to \mathbb{R}^d$ est une variable aléatoire, alors $f \circ X$ noté $f(X)$ est une variable aléatoire, à valeurs dans $\mathbb{R}^n$. Par exemple une somme et un produit de variables aléatoires sont des variables aléatoires.*

Pour une variable aléatoire X on notera $\{X = x\}$, $\{X \leqslant x\}$ et $\{X < x\}$ les ensembles $\{\omega \in \Omega\,;\, X(\omega) = x\}$, $\{\omega \in \Omega\,;\, X(\omega) \leqslant x\}$ et $\{\omega \in \Omega\,;\, X(\omega) < x\}$. De façon générale, on notera $\{X \in F\}$ l'ensemble $\{\omega \in \Omega\,;\, X(\omega) \in F\}$.

Quelques propriétés basées sur ces ensembles.

Proposition 2.3. *1) Soit X une application de Ω dans $\mathbb{R}$ telle que $X(\Omega)$ est dénombrable. Alors X est une variable aléatoire si et seulement si $\{X = x\} \in \mathcal{F}$ pour tout $x \in X(\Omega)$.*

2) Si X est une variable aléatoire discrète et $f : \mathbb{R}^d \to \mathbb{R}^n$ une application quelconque, alors $f(X)$ est une variable aléatoire discrète.

Pour la partie 1), on remarque que, pour tout x, $\{X = x\}$ est l'intersection dénombrable de $\{X \leqslant x\}$ et des $\{X \leqslant x - 1/n\}^c$ pour $n \geqslant 1$, et $\{X \leqslant x\}$ est la réunion dénombrable des $\{X = y\}$ pour tous les y dans $X(\Omega)$ tels que $y \leqslant x$.

On admet la partie 2).

Attention ! Dans le cas général non discret, le fait que $\{X = x\}$ soit mesurable pour tout x n'implique pas que X est une variable aléatoire.

2.2 Fonctions de répartition

Définition 2.4. *Soit X une variable aléatoire réelle. On appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \to [0,1]$ définie pour tout réel x par*

$$F_X(x) = P(X \leqslant x).$$

Proposition 2.5. *1) F_X est une fonction croissante et continue à droite.*
2) La limite de F_X en $-\infty$ vaut 0.
3) La limite de F_X en $+\infty$ vaut 1.
4) Pour tout réel x, $F_X(x-) = P(X < x)$.
5) Pour tout réel x, $P(X = x) = F_X(x) - F_X(x-)$.

D'après 5), si l'on connaît la fonction de répartition d'une variable aléatoire à valeurs entières, on connaît $P(X = n)$ pour tout entier n.

Preuve de la proposition :

1) Si $x \leqslant y$, $\{X \leqslant x\} \subset \{X \leqslant y\}$. Par conséquent, $P(X \leqslant x) \leqslant P(X \leqslant y)$.

Pour le reste de la preuve, soit x un nombre réel, $(x_n)_{n\geqslant 0}$ une suite de nombres réels et $A_n := \{X \leqslant x_n\}$.

Si $(x_n)_{n\geqslant 0}$ décroît vers x, la suite $(A_n)_{n\geqslant 0}$ est décroissante et $\{X \leqslant x\}$ est l'intersection des A_n, donc $P(X \leqslant x) = \lim_n P(X \leqslant x_n)$.

2) Si $(x_n)_{n\geqslant 0}$ décroît vers $-\infty$, la suite $(A_n)_{n\geqslant 0}$ est décroissante et l'intersection des A_n est vide, donc $\lim_n P(X \leqslant x_n) = 0$.

3) Si $(x_n)_{n\geqslant 0}$ croît vers $+\infty$, la suite $(A_n)_{n\geqslant 0}$ est croissante et la réunion des A_n est Ω tout entier, donc $\lim_n P(X \leqslant x_n) = 1$.

4) Si $(x_n)_{n\geqslant 0}$ croît strictement vers x, la suite $(A_n)_{n\geqslant 0}$ est croissante et $\{X < x\}$ est la réunion des A_n, donc $P(X < x) = \lim_n P(X \leqslant x_n)$.

5) $P(X = x) = P(X \leqslant x) - P(X < x)$. Fin de la preuve.

2.3 Variables aléatoires et lois discrètes

Définition 2.6. *Soit X une variable aléatoire discrète à valeurs dans $\mathbb{R}^d$ définie sur $(\Omega, \mathcal{F}, P)$. La loi de X (sous P) est la probabilité sur $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$, notée P_X et définie par la relation $P_X(B) = P(X \in B)$ pour tout $B \subset \mathbb{R}^d$.*

On note parfois aussi $X(P)$ la loi P_X de X.

Si $X(\Omega) = \{x_i\,;\, i \in I\}$ avec I dénombrable, la probabilité P_X est caractérisée par la donnée des nombres $P_X(\{x_i\}) = P(X = x_i)$.

En effet, pour $B \subset \mathbb{R}^d$, $P_X(B) = \displaystyle\sum_{i \in I(B)} P_X(\{x_i\})$ où $I(B)$ désigne l'ensemble des i dans I tels que x_i appartient à B, et cette somme est bien définie.

Si B est une partie de $\mathbb{R}^d$ qui contient $X(\Omega)$, alors la restriction de P_X à $(B, \mathcal{P}(B))$ est une probabilité sur $(B, \mathcal{P}(B))$.

Exemple 2.7. *On lance deux dés et on s'intéresse au plus grand nombre obtenu. On peut considérer pour cela $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $P(\{(i, j)\}) = 1/36$. Soit $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$ définie par $X(i, j) = \max(i, j)$. La loi de X est une probabilité sur $\{1, 2, 3, 4, 5, 6\}$ d'ecrite par les nombres $p_i = P(X = i)$, avec*

$$p_1 = 1/36, \quad p_2 = 3/36, \quad p_3 = 5/36, \quad p_4 = 7/36, \quad p_5 = 9/36, \quad p_6 = 11/36.$$

On vérifie que $p_1 + p_2 + \cdots + p_6 = 1$.

Proposition 2.8. *Soit X une variable aléatoire discrète et $f : \mathbb{R}^d \to \mathbb{R}^n$. La loi de $Y = f(X)$ est donnée par*

$$P_Y(\{y\}) = \sum_{i \in I(y)} P_X(\{x_i\}), \qquad I(y) := \{i \in I\,;\, f(x_i) = y\}.$$

Exemple 2.9. *On reprend l'exemple 2.7 et on considère la variable aléatoire Y qui vaut $Y = 1$ si $X = 1$ ou $X = 6$ et $Y = -1$ sinon. Ceci modélise le jeu suivant : on gagne 1 € si on obtient au moins un 6 ou un double 1 et on perd 1 € sinon.*

3 Moments et indépendance

3.1 Espérance

Définition 3.1. *Soit X une variable aléatoire discrète réelle définie sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $X(\Omega) = \{x_i\,;\, i \in I\}$ avec I dénombrable. Pour tout i dans I, on*

note $p_i = P_X(\{x_i\}) = P(X = x_i)$.
Si la famille $(x_i p_i)_{i\in I}$ *est sommable, on dit que* X *admet une espérance, notée* $E[X]$ *et définie par*

$$E[X] := \sum_{i\in I} x_i p_i.$$

Quelques remarques.

Si $X(\Omega)$ est fini, X admet toujours une espérance.

Si X ne prend que des valeurs positives alors $E[X] \geqslant 0$.

Si X est une variable aléatoire constante égale à x, alors $E[X] = x$.

Si a et b sont des nombres réels et si X admet une espérance, alors $aX + b$ aussi et $E[aX + b] = aE[X] + b$. Plus géneralement, on dispose du résultat suivant.

Proposition 3.2 (Formule fondamentale)**.** *Soit* X *une variable aléatoire discrète à valeurs dans* $\mathbb{R}^d$ *définie sur* $(\Omega, \mathcal{F}, P)$. *Notons* $X(\Omega) = \{x_i\,;\, i \in I\}$ *avec* I *dénombrable, l'ensemble des valeurs prises par* X *et, pour tout* i *dans* I, $p_i = P_X(\{x_i\}) = P(X = x_i)$.
Soit $f : \mathbb{R}^d \to \mathbb{R}$ *et notons* $f(X(\Omega)) = \{y_j\,;\, j \in J\}$ *avec* J *dénombrable. Enfin, soit* $Y = f(X)$.

Si la famille $(f(x_i)p_i)_{i\in I}$ *est sommable, alors* Y *admet une espérance et*

$$E[Y] = E[f(X)] = \sum_{i\in I} f(x_i)p_i = \sum_{j\in J} y_j\, P_Y(\{y_j\}).$$

Preuve : Notons $q_j = P_Y(\{y_j\}) = \displaystyle\sum_{i\in I(j)} p_i$ avec $I(j) := \{i \in I\,;\, f(x_i) = y_j\}$.

Il faut d'abord montrer que la somme $\displaystyle\sum_{j\in J} |y_j| q_j$ converge. Pour cela, on utilise la formule de Fubini pour la suite $(a_{i,j})_{i,j}$ définie par

$$a_{i,j} = f(x_i)p_i \text{ si } f(x_i) = y_j, \qquad a_{i,j} = 0 \text{ sinon.}$$

Alors, pour tout $i \in I$, $\displaystyle\sum_{j\in J} |a_{i,j}| = |f(x_i)|p_i$. Le théorème de Fubini permet de conclure.

Remarque : Si Ω est dénombrable, toute variable aléatoire X sur Ω est discrète et si X admet une espérance,

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

Pour le voir, il suffit d'appliquer la proposition précédente à $f = X$ et à l'application identité de Ω dans Ω.

Proposition 3.3 (et définition)**.** *Soient X et Y deux variables aléatoires réelles discrètes. On note $X(\Omega) = \{x_i\,;\, i \in I\}$ et $Y(\Omega) = \{y_j\,;\, j \in J\}$ avec I et J dénombrables.*
La loi de (X,Y) est donnée par $p_{i,j} = P((X,Y) = (x_i, y_j))$. Alors les lois de X et Y sont appelées les lois marginales de (X,Y) et sont données par

$$p_i = P(X = x_i) = \sum_{j \in J} p_{i,j}, \qquad q_j = P(Y = y_j) = \sum_{i \in I} p_{i,j}.$$

Preuve : Les événements $\{Y = y_j\}$ qui sont non vides forment une partition de Ω, donc, pour tout i fixé, les événements $\{X = x_i, Y = y_j\}$ qui sont non vides forment une partition de $\{X = x_i\}$. On en déduit

$$P(X = x_i) = \sum_{j \in J} P(\{X = x_i\} \cap \{Y = y_j\}),$$

et on a la même chose pour l'autre formule. Fin de la preuve.

Exemple 3.4. *On lance deux dés. Soit X le minimum des deux nombres obtenus et Y leur maximum. On peut représenter la loi de (X,Y) dans le tableau suivant :*

$y \backslash x$	1	2	3	4	5	6	$P(Y = y)$
1	1/36	0	0	0	0	0	1/36
2	2/36	1/36	0	0	0	0	3/36
3	2/36	2/36	1/36	0	0	0	5/36
4	2/36	2/36	2/36	1/36	0	0	7/36
5	2/36	2/36	2/36	2/36	1/36	0	9/36
6	2/36	2/36	2/36	2/36	2/36	1/36	11/36
$P(X = x)$	11/36	9/36	7/36	5/36	3/36	1/36	

Proposition 3.5. *Soit X et Y deux variables aléatoires réelles discrètes admettant des espérances. Alors $X + Y$ admet une espérance et*

$$E[X + Y] = E[X] + E[Y].$$

Preuve : Gardons les notations de la proposition précédente. On utilise la formule fondamentale avec la variable aléatoire (X,Y) qui prend les valeurs (x_i, y_j) quand i décrit I et j décrit J, et avec la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $(x,y) \mapsto x + y$.

Pour tout $i \in I$, $\sum_{j\in J} |x_i| p_{i,j} = |x_i| p_i$ et la somme $\sum_{i\in I} |x_i| p_i$ converge car X admet une espérance, donc par le théorème de Fubini, la famille $(x_i p_{i,j})_{(i,j)\in I\times J}$ est sommable. Il en est de même pour la famille $(y_j p_{i,j})_{(i,j)\in I\times J}$.

Donc par linéarité la famille $((x_i + y_j) p_{i,j})_{(i,j)\in I\times J}$ est sommable et sa somme vaut

$$\sum_{(i,j)\in I\times J} x_i p_{i,j} + \sum_{(i,j)\in I\times J} y_j p_{i,j} = \sum_{i\in I} x_i p_i + \sum_{j\in J} y_j q_j = E[X] + E[Y].$$

Fin de la preuve.

Corollaire 3.1. *Soit X et Y deux variables aléatoires discrètes telles que pour tout $\omega \in \Omega$, $X(\omega) \leqslant Y(\omega)$. Si X et Y admettent des espérances alors $E[X] \leqslant E[Y]$.*

Il suffit de voir que la variable aléatoire $Y - X$ ne prend que des valeurs positives et que $E[Y - X] = E[Y] - E[X]$.

Si X et Y ne prennent que des valeurs positives, si $X \leqslant Y$ et si Y admet une espérance, alors X aussi.

3.2 Variance et covariance

Définition 3.6. *La variable aléatoire X est de carré intégrable si $E[X^2]$ est fini. On a alors*

$$E[X^2] = \sum_{i\in I} x_i^2 p_i.$$

Lemme 3.1. *Si X est de carré intégrable et si $E[X^2] = 0$, alors $X = 0$ partout.*

Preuve laissée au lecteur.

Proposition 3.7 (Inégalité de Cauchy-Schwarz)**.** *Soit X et Y deux variables aléatoires réelles discrètes de carré intégrable. Alors la variable aléatoire XY admet une espérance et*

$$|E[XY]| \leqslant E[X^2]^{1/2} E[Y^2]^{1/2}.$$

De plus, $X + Y$ est de carré intégrable.

Preuve : Pour tout $\omega \in \Omega$, $2|X(\omega)Y(\omega)| \leqslant X^2(\omega) + Y^2(\omega)$. On en déduit par un résultat précédent que la variable aléatoire XY admet une espérance et que, pour tout réel t, la variable aléatoire $(X + tY)^2 = X^2 + 2tXY + t^2Y^2$ aussi. Comme $(X + tY)^2$ est positive partout, $E[(X + tY)^2] \geqslant 0$.

On développe et on traduit le fait que le polynome en t obtenu est toujours positif, ce qui nous donne le résultat. Fin de la preuve.

Preuve alternative : Une autre méthode consiste à considérer la variable aléatoire

$$Z := E[Y^2]^{1/2}X - E[X^2]^{1/2}Y.$$

Comme $Z^2 \leqslant 2E[Y^2]X^2 + 2E[X^2]Y^2$, Z est de carré intégrable et $E[Z^2] \geqslant 0$. En développant Z^2 et en réarrangeant les termes, on obtient

$$2E[Y^2]E[X^2] \geqslant 2E[X^2]^{1/2}E[Y^2]^{1/2}E[XY].$$

Si on peut diviser les deux membres par $2E[X^2]^{1/2}E[Y^2]^{1/2}$, on obtient

$$E[XY] \leqslant E[X^2]^{1/2}E[Y^2]^{1/2},$$

et en considérant $T := E[Y^2]^{1/2}X + E[X^2]^{1/2}Y$, on démontre l'autre inégalité.

Sinon, cela signifie que $E[X^2] = 0$ ou $E[Y^2] = 0$, donc que $X = 0$ partout ou $Y = 0$ partout, auquel cas tout fonctionne. Cette méthode fournit aussi les cas d'égalité puisque $E[Z^2] = 0$ ou $E[T^2] = 0$ implique que $Z = 0$ partout ou $T = 0$ partout, donc X et Y sont proportionnelles. Fin de la preuve alternative.

Proposition 3.8 (et définition)**.** *Soit X une variable aléatoire discrète de carré intégrable. Alors X admet une espérance et*

$$|E[X]| \leqslant E[|X|] \leqslant E[X^2]^{1/2}.$$

On définit alors la variance de X, notée $\text{var}(X)$ *ou* $\sigma^2(X)$*, et l'écart type de X, noté* $\sigma(X)$*, par*

$$\text{var}(X) = E[X^2] - E[X]^2 = E[(X - E[X])^2], \quad \sigma(X) = \sqrt{\text{var}(X)}.$$

Pour tout réel x, $E[(X - x)^2] = \text{var}(X) + (x - E[X])^2$ donc l'espérance est aussi la valeur de x qui minimise $E[(X-x)^2]$, et la valeur minimale obtenue est la variance.

Preuve : La première inégalité est une conséquence de la proposition précédente appliquée aux variables aléatoires X et $Y = 1$. Par ailleurs,

$$E[(X - x)^2] = E[X^2] - 2xE[X] + x^2 = E[X^2] - E[X]^2 + (E[X] - x)^2,$$

ce qui termine la preuve.

Proposition 3.9 (et définition)**.** *Soient X et Y deux variables aléatoires réelles discrètes de carré intégrable. On définit la covariance de X et Y, notée* $\text{cov}(X, Y)$*, par*

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

On a alors

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

Preuve : Pour le premier point il suffit de développer le produit et d'utiliser le résultat sur l'espérance d'une somme et le fait que si t est un réel et X une variable aléatoire discrète admettant une espérance, alors $E[tX] = tE[X]$.
Pour le second point, on développe $(X + Y)^2 = X^2 + Y^2 + 2XY$, chacun de ces trois termes admet une espérance. Fin de la preuve.

Remarque : une variance est toujours positive ou nulle mais une covariance peut être positive ou nulle ou négative.

3.3 Variables aléatoires indépendantes

Définition 3.10. *Les variables aléatoires discrètes X_k, $1 \leqslant k \leqslant n$, à valeurs dans E_1,...,E_n respectivement, sont indépendantes si et seulement si, pour tout $(x_1, \ldots, x_n)$ dans $E_1 \times \ldots \times E_n$,*

$$P(\forall 1 \leqslant k \leqslant n,\ X_k = x_k) = \prod_{k=1}^{n} P(X_k = x_k).$$

Proposition 3.11. *Les variables aléatoires discrètes $(X_1, \ldots, X_n)$ à valeurs respectivement dans E_1,...,E_n sont indépendantes si et seulement si pour toutes fonctions $f_k : E_k \to \mathbb{R}$ positives,*

$$E\left[\prod_{k=1}^{n} f_k(X_k)\right] = \prod_{k=1}^{n} E[f_k(X_k)].$$

Dans ce cas, si on se donne, pour tout k, une fonction $g_k : E_k \to \mathbb{R}$ telle que $g_k(X_k)$ est intégrable et si on pose $Z := g_1(X_1) \cdots g_n(X_n)$, alors Z est intégrable et

$$E[Z] = \prod_{k=1}^{n} E[g_k(X_k)].$$

Preuve : Supposons pour simplifier que $n = 2$ et appliquons la proposition 3.2 avec $f(x, y) = f_1(x)f_2(y)$. Notons x_i pour i dans I les valeurs prises par X_1, y_j pour j dans J les valeurs prises par X_2. Notons $p_i := P(X_1 = x_i)$, $q_j := P(X_2 = y_j)$ et $p_{i,j} := P((X_1, X_2) = (x_i, y_j))$. Par hypothèse, $p_{i,j} = p_i q_j$, donc $E[f(X_1, X_2)]$ vaut

$$\sum_{i,j} f_1(x_i)f_2(y_j)p_{i,j} = \sum_{i} f_1(x_i)p_i \sum_{j} f_2(y_j)q_j = E[f_1(X_1)]E[f_2(X_2)].$$

La réciproque s'obtient en considérant les applications $f_1 := \mathbf{1}_{\{x_i\}}$ et $f_2 = \mathbf{1}_{\{y_j\}}$.

Pour démontrer le dernier résultat, on applique ce qui précède à $|g_k|$. On en déduit que

$$E[|Z|] = \prod_{k=1}^{n} E[|g_k(X_k)|],$$

et on réutilise la proposition 3.2 de la même façon que précédemment. Fin de la preuve.

Proposition 3.12. *Soit X et Y deux variables aléatoires réelles discrètes de carré intégrable. On suppose que X et Y sont indépendantes. Alors,*

$$E[XY] = E[X]E[Y], \quad \operatorname{cov}(X, Y) = 0, \quad \operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

Preuve : On applique la proposition précédente.

Remarque 3.13. *Attention : la réciproque au résultat précédent est fausse, on peut trouver des variables aléatoires X et Y telles que* $\operatorname{cov}(X, Y) = 0$ *sans que X et Y soient indépendantes.*

Par exemple, soit X une variable aléatoire de loi $P(X = 1) = P(X = -1) = \frac{1}{2}$, Z une variable aléatoire indépendante de X et de loi $P(Z = -1) = \frac{2}{3}$ et $P(Z = 2) = \frac{1}{3}$, $Y = XZ$.

Montrer que $E[X] = E[Y] = E[XY] = 0$ donc X et Y ne sont pas corrélées, mais que $P(X = 1, Y = -2) = 0$, et en conclure que X et Y ne sont pas indépendantes.

3.4 Fonctions génératrices

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$.

La distribution de X est une loi de probabilité sur $\mathbb{N}$ donc elle est entièrement déterminée par les quantités $P(X = n)$ pour n dans $\mathbb{N}$. En effet, pour tout $A \subset \mathbb{N}$,

$$P(X \in A) = \sum_{n \in A} P(X = n).$$

Donc X admet une espérance si et seulement si $\sum_{n} nP(X = n)$ converge, auquel cas

$$E[X] = \sum_{n \geqslant 0} nP(X = n).$$

De même, X est de carré intgrable si et seulement si $\sum_{n} n^2 P(X = n)$ converge, auquel cas

$$E[X^2] = \sum_{n \geqslant 0} n^2 P(X = n).$$

Si f est une application positive ou telle que $\sum_n |f(n)|P(X=n)$ converge, alors

$$E[f(X)] = \sum_{n\geqslant 0} f(n)P(X=n).$$

Définition 3.14. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}\cup\{+\infty\}$. La fonction génératrice de X est la fonction $g_X : [0,1] \to [0,1]$ définie par*

$$g_X(s) := \sum_{n=0}^{+\infty} s^n P(X=n) = E[s^X].$$

Proposition 3.15. *Toute fonction génératrice vérifie les propriétés suivantes.*

1. *La fonction g_X est convexe, croissante sur $[0,1]$ et indéfiniment dérivable sur $[0,1[$.*
2. *La fonction g_X détermine la loi de X.*
3. *Si on suppose que X est à valeurs dans $\mathbb{N}$, X admet un moment d'ordre p si et seulement si g_X admet une dérivée à gauche d'ordre p en 1. Dans ce cas, pour $p=1$, $E[X] = g_X'(1-)$ et pour $p \geqslant 1$,*

$$E[X(X-1)\cdots(X-p+1)] = g_X^{(p)}(1-).$$

4. *Si X et Y sont indépendantes, alors $g_{X+Y} = g_X g_Y$.*

Preuve : La fonction g_X est donnée par une série entière de rayon de convergence au moins égal à 1, car $\sum_n P(X=n)$ converge. C'est donc une fonction indéfiniment dérivable sur $[0,1[$, et

$$g_X^{(p)}(s) = \sum_{n=0}^{+\infty} n(n-1)\cdots(n-p+1)s^{n-p}P(X=n),$$

donc

$$g_X^{(p)}(s) = E[X(X-1)\cdots(X-p+1)s^{X-p}].$$

En particulier, $g_X^{(n)}(0) = n!\,P(X=n)$. On en déduit facilement les points 1. et 2.

Démontrons le point 3. pour $p=1$. On suppose donc que X est à valeurs dans $\mathbb{N}$ et on pose $g = g_X$. Alors,

$$h(s) := \frac{g(1)-g(s)}{1-s} = \sum_{n=0}^{+\infty} \frac{1-s^n}{1-s} P(X=n).$$

Supposons que X est intégrable. On va utiliser le fait que, pour tout entier $n \geqslant 0$ et pour tout s dans $[0,1[$,

$$0 \leqslant \frac{1-s^n}{1-s} \leqslant n.$$

Par conséquent,

$$h(s) \leqslant \sum_{n=0}^{+\infty} nP(X=n) = E[X].$$

Dans l'autre sens, pour tout entier $N \geqslant 0$,

$$h(s) \geqslant \sum_{n=0}^{N} \frac{1-s^n}{1-s} P(X=n),$$

et la fraction $(1-s^n)/(1-s)$ tend vers n quand s tend vers 1. Pour N fixé, on ne manipule que des sommes finies, donc

$$\liminf_{s \to 1} h(s) \geqslant \sum_{n=0}^{N} nP(X=n).$$

Cette minoration est vraie pour tout $N \geqslant 0$, donc

$$\liminf_{s \to 1} h(s) \geqslant \sup_N \sum_{n=0}^{N} nP(X=n) = \sum_{n=0}^{+\infty} nP(X=n) = E[X],$$

donc $h(s)$ tend vers $E[X]$ quand s tend vers 1, c'est-à-dire que la dérivée à gauche en 1 de g existe et vaut $E[X]$, cqfd.

Réciproquement, supposons que g admet une dérivée à gauche en 1. Alors, pour tout $N \geqslant 0$,

$$\sum_{n=0}^{N} nP(X=n) = \lim_{s \to 1} \sum_{n=0}^{N} \frac{1-s^n}{1-s} P(X=n) \leqslant \lim_{s \to 1} \frac{1-g(s)}{1-s} = g'(1-),$$

donc X est intégrable et peut on utiliser la partie précédente de la preuve pour conclure.

Finalement, g est dérivable à gauche en 1 si et seulement si $\sum_n nP(X=n)$ converge si et seulement si X est intégrable.

Le résultat général se prouve de la même manière.

Pour le point 4., pour tout s dans $[0,1]$,

$$E[s^{X+Y}] = E[s^X s^Y] = E[s^X]E[s^Y],$$

grâce à la proposition 3.11 et au fait que la fonction de $\mathbb{N}$ dans $\mathbb{R}$ définie par $x \mapsto s^x$ est bornée pour tout s dans $[0,1]$. Fin de la preuve.

Remarque 3.16. *Attention, la réciproque du point 4 est fausse. Exemple : soit* (X, Y) *dont la loi est donnée dans le tableau suivant :*

$x \backslash y$	*0*	*1*	*2*	$P(X = x)$
0	*1/9*	*0*	*2/9*	*1/3*
1	*2/9*	*1/9*	*0*	*1/3*
2	*0*	*2/9*	*1/9*	*1/3*
$P(Y = y)$	*1/3*	*1/3*	*1/3*	

On voit que X *et* Y *ne sont pas indépendantes puisque* $P(X = 0, Y = 1)$ *est nul et* $P(X = 0)$ *et* $P(Y = 1)$ *ne le sont pas. Dans cet exemple, la loi de* $X + Y$ *est donnée par*

k	*0*	*1*	*2*	*3*	*4*
$P(X + Y = k)$	*1/9*	*2/9*	*1/3*	*2/9*	*1/9*

On vérifie que $g_X(s) = g_Y(s) = \frac{1}{3}(1 + s + s^2)$ *et* $g_{X+Y} = g_X g_Y$.

4 Exemples fondamentaux de lois discrètes

4.1 Loi de Dirac δ_x

Paramètre x réel (par exemple) : $X = x$ partout.

Donc $E[X] = x$, $\mathrm{var}(X) = 0$, et $g_X(s) = s^x$.

Si $\mathrm{var}(X) = 0$, on est dans ce cas.

Modèle : expérience dont le résultat est certain.

4.2 Loi de Bernoulli $b(p)$

Paramètre p avec $0 \leqslant p \leqslant 1$: X à valeurs dans $\{0, 1\}$ et $P(X = 1) = p = 1 - P(X = 0)$.

Donc $E[X] = p$, $\mathrm{var}(X) = p(1 - p)$, et $g_X(s) = 1 - p + ps$.

Modèle : jeu de pile ou face.

4.3 Loi uniforme $U(A)$

Paramètre A un ensemble fini de cardinal $n \geqslant 1$: X est à valeurs dans E et pour tout x dans A, $P(X = x) = 1/n$.

Modèle : expérience avec n résultats possibles, si on ne dispose d'aucune information supplémentaire, ou bien si la situation est entièrement symétrique.

4.4 Loi binomiale $B(n, p)$

Paramètres (n, p) avec $0 \leqslant p \leqslant 1$ réel et $n \geqslant 0$ entier : X est à valeurs dans $\{0, 1, \ldots, n\}$ et, pour tout $0 \leqslant k \leqslant n$, $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Donc $g_X(s) = (1 - p + ps)^n$.

Modèle : nombre de piles parmi n résultats d'un jeu de pile ou face.

Proposition 4.1. *Soit $(X_1, \ldots, X_n)$ n variables aléatoires de Bernoulli indépendantes de paramètre p. Alors $X = X_1 + \cdots + X_n$ suit une loi binomiale de paramètre (n, p).*

Preuve : On peut utiliser les fonctions génératrices. Grâce à l'indépendance,

$$g_X(s) = \prod_k g_{X_k}(s) = (1 - p + ps)^n.$$

Comme la fonction génératrice caractérise la loi, le résultat suit.

Donc $E[X] = np$ et $\mathrm{var}(X) = np(1 - p)$.

Proposition 4.2. *Si n tend vers l'infini et p tend vers 0 de sorte que np tende vers a avec a strictement positif et fini, alors, pour tout $k \geqslant 0$,*

$$P(X_n = k) \to \frac{a^k}{k!} \mathrm{e}^{-a}.$$

4.5 Loi de Poisson $P(a)$

Paramètre a strictement positif : X à valeurs dans $\mathbb{N}$ et $P(X = k) = \dfrac{a^k}{k!} \mathrm{e}^{-a}$ pour tout $k \geqslant 0$.

Donc $g_X(s) = \mathrm{e}^{-a(1-s)}$, $E[X] = a$ et $\mathrm{var}(X) = a$.

Modèle : d'après la proposition 4.2, loi des événements rares.

4.6 Loi géométrique $G(p)$

Paramètre p dans $[0,1]$: X est à valeurs dans $\mathbb{N}$ et $P(X = n) = p(1-p)^n$ pour tout $n \geqslant 0$.

Donc $g_X(s) = p/(1-(1-p)s)$, $E[X] = (1-p)/p$ et $\mathrm{var}(X) = (1-p)/p^2$.

Modèle : instant du premier succès dans une suite d'expériences indépendantes chacune avec probabilité p de succès.

Proposition 4.3. *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre p (c'est-à-dire que pour tout $n \geqslant 0$, les variables aléatoires $(X_0, \ldots, X_n)$ sont indépendantes). Soit X la variable aléatoire définie par*

$$X = \min\{n \geqslant 0 \,|\, X_n = 1\}.$$

Alors X est finie avec probabilité 1 et de loi géométrique de paramètre p.

Preuve : L'événement $\{X = n\}$ vaut l'intersection de $\{X_n = 1\}$ et des n événements $\{X_k = 0\}$ pour $0 \leqslant k \leqslant n-1$. Par indépendance, $P(X = n) = p(1-p)^n$.

5 Autres exemples de lois discrètes

5.1 Loi hypergéométrique $H(n, N, N_1)$

Paramètres (n, N, N_1) avec $0 \leqslant n \leqslant N$ et $0 \leqslant N_1 \leqslant N$: X est à valeurs dans $\{0, 1, \ldots, n\}$ et pour tout $0 \leqslant k \leqslant n$,

$$P(X = k) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}},$$

avec la convention que $\displaystyle\binom{i}{j} = 0$ si $j < 0$ ou si $j > i$.

Modèle : on effectue n tirages sans remise dans un ensemble A de cardinal N ; le nombre d'objets de $A_1 \subset A$ tirés suit cette loi si le cardinal de A est N et celui de A_1 est N_1.

Proposition 5.1. *Soit X une variable aléatoire de loi hypergéométrique $H(n, N, N_1)$. Alors,*

$$E[X] = nN_1/N, \qquad \mathrm{var}(X) = \frac{nN_1(N-N_1)(N-n)}{N^2(N-1)}.$$

Preuve : On utilise le fait que $k\binom{N_1}{k} = N_1\binom{N_1-1}{k-1}$. Ainsi,

$$k\,P(X=k) = \frac{k\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}} = \frac{N_1\binom{N_1-1}{k-1}\binom{N-N_1}{(n-1)-(k-1)}}{N\binom{N-1}{n-1}/n},$$

donc

$$k\,P(X=k) = (nN_1/N)P(Y=k-1),$$

si Y suit une loi hypergéométrique de paramètres $(n-1, N-1, N_1-1)$. En sommant sur k et en utilisant le fait que la somme des $P(Y=k-1)$ vaut 1, on trouve la valeur de $E[X]$.

En appliquant le même principe à $P(Y=k-1)$, on obtient

$$k(k-1)\,P(X=k) = (nN_1/N)(k-1)P(Y=k-1),$$

et

$$(k-1)P(Y=k-1) = ((n-1)(N_1-1)/(N-1))P(Z=k-2),$$

si Z suit une loi hypergéométrique de paramètres $(n-2, N-2, N_1-2)$, donc

$$E[X(X-1)] = \frac{n(n-1)N_1(N_1-1)}{N(N-1)},$$

ce qui permet de trouver la valeur de

$$\mathrm{var}(X) = E[X(X-1)] + E[X] - E[X]^2.$$

Fin de la preuve.

Proposition 5.2. *Si N et N_1 tendent vers l'infini de sorte que N_1/N converge vers p, alors $P(X=k) \to \binom{n}{k}p^k(1-p)^{n-k}$.*

Preuve laissée en exercice.

Ceci montre qu'un tirage sans remise équivaut à un tirage avec remise si le cardinal de l'ensemble est grand et le rapport N_1/N à peu près constant.

Remarque 5.3. *Soit X_i la variable aléatoire valant $X_i = 1$ si au tirage numéro i, on obtient un objet de A_1 et $X_i = 0$ sinon. Alors $X = X_1 + \cdots + X_n$ et les variables aléatoires X_i suivent des lois de Bernoulli. Quels sont leurs paramètres ? Puisque $E[X] = nN_1/N$ et $E[X] = E[X_1] + \cdots + E[X_n]$, on voit que $P(X_i = 1) = N_1/N$ pour tout i.*

Mais attention, les variables aléatoires X_i ne sont pas indépendantes ; sinon la loi de leur somme X serait binomiale $(n, N_1/N)$.

5.2 Loi multinomiale $M(n; p_1, \ldots, p_r)$

Paramètres $n \geqslant 0$ entier et $(p_1, \ldots, p_r)$ avec p_k dans $[0, 1]$ et $p_1 + \cdots + p_r = 1$: alors $X = (X_1 \ldots, X_r)$ est à valeurs dans l'ensemble des r-uplets $(k_1, \ldots, k_r)$ dans $\mathbb{N}^r$ tels que $k_1 + \cdots + k_r = n$ et, pour tout tel $(k_1, \ldots, k_r)$,

$$P(X = (k_1, \ldots, k_r)) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} \ldots p_r^{k_r}, \quad \binom{n}{k_1, \ldots, k_r} = \frac{n!}{k_1! \ldots k_r!}.$$

Quand $r = 2$, (X_1, X_2) suit la loi de $(Y, n - Y)$ avec Y de loi binomiale (n, p_1). Plus généralement :

Proposition 5.4. *Si X suit la loi $M(n; p_1, \ldots, p_r)$, les lois marginales des X_i sont les lois binomiales (n, p_i) pour $1 \leqslant i \leqslant r$.*

Preuve : On peut faire le calcul directement en remarquant que $\{X_i = k\}$ est la réunion des $\{X = (k_1, \ldots, k_r)\}$ sur tous les r-uplets $(k_1, \ldots, k_r)$ tels que $k_i = k$. C'est lourd. On peut aussi utiliser les fonctions génératrices. Pour tout $(s_1, \ldots, s_r)$ dans $[0, 1]^r$,

$$E[s_1^{X_1} \cdots s_r^{X_r}] = \sum_{(k_1, \ldots, k_r)} \binom{n}{k_1, \ldots, k_r} (p_1 s_1)^{k_1} \ldots (p_r s_r)^{k_r}.$$

La somme du membre de droite vaudrait 1 si $p_1 s_1 + \cdots + p_r s_r$ valait 1 car ce serait la masse totale d'une distribution. Par homogénéité, on en déduit

$$E[s_1^{X_1} \cdots s_r^{X_r}] = (p_1 s_1 + \cdots + p_r s_r)^n.$$

Il reste à choisir $s_j = 1$ pour tout $j \neq i$ et à remarquer que la somme des p_j pour $j \neq i$ vaut $1 - p_i$ pour obtenir

$$E[s_i^{X_i}] = (1 - p_i + p_i s_i)^n.$$

C'est la fonction génératrice de la loi binomiale (n, p_i). cqfd.

Exercice : Calculer la loi de (X_1, X_2).

5.3 Loi binomiale négative $NB(n, p)$

Paramètres $n \geqslant 1$ et $0 \leqslant p \leqslant 1$: $P(X = k) = \binom{k-1}{n-1} p^n (1 - p)^{k-n}$ pour tout $0 \leqslant n \leqslant k$ (attention à l'ordre des deux indices).

Construction : rang du succès numéro n dans une suite d'épreuves indépendantes, chacune avec probabilité de succès p.

Donc $NB(n,p)$ est la loi de la somme de n variables aléatoires indépendantes de loi $G(p)$ et $NB(1,p) = G(p)$.

Donc $g_X(s) = p^n/(1-(1-p)s)^n$, $E[X] = n/p$ et $\text{var}(X) = n(1-p)/p^2$.

6 Exercices

1. Espérance discrète Soit X une variable aléatoire à valeurs dans $\{0, 1, ..., n\}$. Montrer que :

$$E(X) = \sum_{k=1}^{n} P(X \geqslant k).$$

2. Loi sans mémoire Soit T une variable aléatoire à valeurs dans $\mathbb{N}$ telle que pour tout $n \geqslant 0$ et $k \geqslant 0$:

$$P(T \geqslant n + k \,|\, T \geqslant n) = P(T \geqslant k).$$

Déterminer la loi de T.

3. Boules colorées Une urne contient N boules dont N_1 portent le numéro 1, N_2 portent le numéro 2, ..., et N_k portent le numéro k. On fait un tirage de n boules avec remise. Soit X_i le nombre de boules tirées qui portent le numéro i, et $X = (X_1, \ldots, X_k)$.

a) Donner la loi de X.

b) Donner la loi de X_i pour $1 \leqslant i \leqslant k$.

c) Donner la loi de (X_i, X_j) pour $i \neq j$.

Facultatif Traiter les mêmes questions pour un tirage sans remise.

4. Instants On effectue une suite infinie de lancers indépendants d'une pièce de monnaie. Le résultat du lancer numéro n est une variable aléatoire X_n qui vaut 1 si l'on a obtenu « pile », ce qui arrive avec une probabilité p, et qui vaut 0 si l'on a obtenu « face », ce qui arrive avec une probabilité $1 - p$.

On note N_1, N_2, etc., les numéros des lancers successifs où l'on a obtenu « pile », c'est-à-dire

$$N_1 = \inf\{n \geqslant 1 \,;\, X_n = 1\}, \qquad N_{k+1} = \inf\{n \geqslant N_k + 1 \,;\, X_n = 1\}, \ k \geqslant 1.$$

a) Donner la loi de N_1 et montrer que N_1 est fini avec probabilité 1.

b) Donner la loi de (N_1, N_2). Montrer que les variables aléatoires N_1 et $N_2 - N_1$ sont indépendantes et de même loi.

c) Donner la loi de N_2. Donner la loi de N_1 sachant $N_2 = n_2$.

d) Donner la loi de N_k pour tout $k \geqslant 3$.

5. Lois de Bernoulli Soit X et Y deux variables aléatoires de lois de Bernoulli, pas forcément de même loi. On suppose que $\text{cov}(X, Y) = 0$. Montrer que X et Y sont indépendantes.

6. Lois binomiales Soit n et m deux entiers strictement positifs et $0 \leqslant p \leqslant 1$. Si X suit une loi binomiale de paramètre (n, p), si Y suit une loi binomiale de paramètre (m, p) et si X et Y sont indépendantes, calculer la loi de $Z := X + Y$. En déduire la formule suivante : pour tout $0 \leqslant k \leqslant n + m$,

$$\sum_{i=0}^{n} \binom{n}{i} \binom{m}{k-i} = \binom{n+m}{k}.$$

7. Les boîtes d'allumettes de Banach On prétend que Stefan Banach avait toujours une boîte de N allumettes dans chacune de ses poches. Quand il voulait une allumette, il choisissait une de ses deux poches au hasard, et ses choix successifs constituaient une suite de tirages de Bernoulli avec paramètre $p = \frac{1}{2}$.

On considère le moment où Banach découvre pour la première fois qu'une boîte est vide. À ce moment, l'autre boîte contient X_N allumettes. Donner la loi de X_N.

Facultatif Montrer que $E[X_N] = 2(2N+1)u_N - 1$ où $u_N = P(X_N = 0)$. Montrer en utilisant la formule de Stirling que, quand N tend vers l'infini, $E[X_N]$ est équivalent à $2\sqrt{N/\pi}$ (énoncé à vérifier).

8. Méthode du maximum observé Une urne contient N balles numérotées de 1 à N. On effectue n tirages avec remise. Soit X le plus grand nombre tiré lors des n tirages.

a) Donner la fonction de répartition de X.

b) Donner la loi de X.

c) Calculer $E[X]$ et donner un équivalent de $E[X]$ quand $N \to +\infty$

Remarque Les statisticiens utilisent cette méthode du maximum observé pour estimer le paramètre inconnu N. Prolongement : comparer l'efficacité de cette méthode avec l'estimateur basé sur la moyenne des numéros observés pendant n tirages.

9. Clés Un homme possède n clés et veut ouvrir une porte. Une seule parmi les clés dont il dispose ouvre la porte. Il essaie les clés au hasard. Trouver l'espérance et la variance du nombre d'essais nécessaires si :

a) Les clés qui ne marchent pas sont remises avec les autres.

b) Les clés qui ne marchent pas sont mises de côté.

10. Poisson(s) Soient X et Y deux variables aléatoires indépendantes de lois de Poisson de paramètre a et b.

a) Déterminer la loi de la variable aléatoire $S = X + Y$.

b) Déterminer, pour tout couple (n, k) d'entiers naturels, la probabilité conditionnelle $P(X = k \,|\, S = n)$.

c) (Facultatif) Soit $r \geqslant 1$ un entier et X_k des variables aléatoires indépendantes de lois de Poisson de paramètres respectifs a_k. Donner la loi conditionnelle de $(X_1, ..., X_r)$ sachant $\{X_1 + ... + X_r + X_{r+1} = n\}$, pour tout $n \geqslant 0$.

11. Tirages biaisés On dispose de n pièces numérotées de 1 à n et biaisées de façon que la probabilité d'obtenir pile avec la pièce i vaut $1/(2i + 1)$.

On lance les n pièces et on veut calculer la probabilité d'obtenir un nombre impair de piles.

Soit X_i la variable aléatoire qui vaut $X_i = 1$ si la pièce i est tombée sur pile et $X_i = 0$ sinon, et $X = X_1 + \cdots + X_n$.

a) Préciser ce que représente la variable aléatoire X.

b) Calculer la fonction génératrice φ de X.

c) Montrer que la probabilité cherchée vaut $\frac{1}{2}(\varphi(1)+\varphi(-1))$ et calculer cette valeur.

12. Lois géométriques Soit X et Y deux variables aléatoires indépendantes de lois géométriques de paramètres respectifs a et b. Soit $Z = \min(X, Y)$ et $U = |X - Y|$.

a) Déterminer $P(X \geqslant n)$ pour tout $n \geqslant 0$. Déterminer $P(Z \geqslant n)$. Préciser la loi de Z.

b) Calculer $P(U = 0)$. Déterminer la loi de U.

c) Montrer que Z et U sont indépendantes.

13. Régression linéaire Soit X et Y deux variables aléatoires discrètes de variances non nulles.

On pose
$$\varrho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}, \quad \overline{X} = X - E[X], \quad \overline{Y} = Y - E[Y].$$

1) Montrer que $|\operatorname{cov}(X,Y)| = |E[\overline{X}\,\overline{Y}]| \leqslant \sigma_X \sigma_Y$.

En déduire que $-1 \leqslant \varrho_{X,Y} \leqslant 1$.

2) Montrer que $|\varrho_{X,Y}| = 1$ si et seulement sil existe a non nul et b tels que $P(Y = aX + b) = 1$.

On pourra calculer $E[(\overline{Y} + t\overline{X})^2]$.

3) Préciser la valeur de $\varrho_{X,Y}$ si X et Y sont indépendantes.

4) On cherche la meilleure approximation de Y comme fonction affine de X au sens des moindres carrés, c'est-à-dire que l'on cherche les valeurs de a et b qui minimisent $E[(aX + b - Y)^2]$. Notons $\Phi(a,b) = E[(aX + b - Y)^2]$

Montrer que $\Phi(a,b) = E[(\overline{Y} - a\overline{X})^2] + (E[Y] - (aE[X] + b))^2$.

En déduire que le couple (a_0, b_0) qui minimise Φ vaut
$$a_0 = \varrho_{X,Y}\sigma_Y/\sigma_X, \qquad b_0 = E[Y] - a_0 E[X].$$

On appelle la droite d'équation $y = a_0 x + b_0$ la droite de régression linéaire de Y en X.

5) On suppose que (X,Y) suit la loi uniforme sur un ensemble de cardinal n, c'est-à-dire qu'il existe n points (x_i, y_i) dans le plan tels que $P(X = x_i, Y = y_i) = 1/n$ pour tout $1 \leqslant i \leqslant n$.

Déterminer la droite de régression linéaire de Y en X dans ce cas.

14. Loi jointe On effectue une suite infinie de lancers indépendants d'un dé équilibré. On numérote les lancers à partir de 1. On définit X comme le numéro du premier lancer qui donne 6, et Y comme le nombre de 5 obtenus avant d'obtenir le premier 6.

Déterminer la loi du couple (X,Y).
Déterminer la loi conditionnelle de Y sachant l'événement $\{X = n\}$.
Déterminer la loi absolue de Y.

Capes de Mathématiques **Université Joseph Fourier**
Préparation à l'écrit **Année 2008-2009**

Probabilités 3 : Variables aléatoires densitables

« *Les questions les plus importantes de la vie ne sont en effet, pour la plupart, que des problèmes de probabilité.* » *Pierre-Simon Laplace*

1 Description

La fonction de répartition $F_X : \mathbb{R} \to [0,1]$ d'une variable aléatoire X à valeurs réelles est définie, pour tout réel x, par $F_X(x) = P(X \leqslant x)$, et elle caractérise la loi de X. La fonction F_X est croissante, continue à droite et à valeurs dans $[0,1]$, de plus $F_X(x)$ tend vers 0 quand x tend vers $-\infty$ et $F_X(x)$ tend vers 1 quand x tend vers $+\infty$. Réciproquement, toute fonction vérifiant ces propriétés (c'est-à-dire toute fonction à valeurs dans $[0,1]$, croissante, continue à droite, de limites 0 en $-\infty$ et 1 en $+\infty$) est la fonction de répartition d'une variable aléatoire.

1.1 Densité

Définition 1.1. *On dit que la variable aléatoire $X : \Omega \to \mathbb{R}$ admet la loi de densité f si la fonction $f : \mathbb{R} \to \mathbb{R}^+$ est continue par morceaux et si, pour tout réel x,*

$$F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

Par conséquent, si f est la densité de la loi d'une variable aléatoire, $f \geqslant 0$ et

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1.$$

Si F_X est continue et admet une dérivée sauf en un nombre fini de points, alors la loi de X admet une densité donnée par la dérivée de F_X là où cette dérivée existe et par n'importe quelle valeur là où cette dérivée n'existe pas.

Remarque 1.2. *En toute généralité, la loi d'une variable aléatoire X à valeurs réelles comporte trois parties : une partie absolument continue, décrite par une densité f comme ci-dessus ; une partie atomique, correspondant à une mesure discrète et décrite dans le chapitre précédent par des nombres p_x ; et une troisième partie, appelée partie singulière et plus difficile à décrire, qui est en fait simplement la partie qui n'est ni atomique, ni densitable. Cette décomposition est unique. Quand la loi comporte seulement une partie absolument continue et une partie atomique (ce qui sera toujours le cas dans ce cours), on peut utiliser la notation*

$$P_X(\mathrm{d}x) = f_X(x)\,\mathrm{d}x + \sum_x p_x\,\delta_x.$$

Cette notation signifie que, pour tout B,

$$P(X \in B) = \int_B f_X(x)\,\mathrm{d}x + \sum_{x\in B} p_x.$$

Même si on en parlera peu dans la suite, de tels « mélanges » apparaissent naturellement.

Exemple/exercice : soit X une variable aléatoire de loi uniforme sur l'intervalle $[0,4]$ (voir la définition plus bas) et $Y = \max(1, \min(X,2))$. Décrire la loi de Y.

1.2 Espérance

On revient au cas densitable.

Définition 1.3. *Quand l'intégrale $\displaystyle\int_{-\infty}^{+\infty} |x|\,f(x)\,\mathrm{d}x$ converge, on dit que la variable aléatoire X de densité f est intégrable (ou admet une espérance). Dans ce cas, on pose*

$$E(X) = \int_{-\infty}^{+\infty} x\,f(x)\,\mathrm{d}x.$$

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

La formule de changement de variables ci-dessous permet de calculer des espérances de fonctions de variables aléatoires densitables. C'est l'analogue de la « formule fondamentale » du chapitre 2 pour les variables aléatoires discrètes.

Proposition 1.4. *Soit X une variable aléatoire densitable de densité f_X et soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction. On pose $Y = \varphi(X)$. Alors la variable aléatoire Y est* *intégrable si et seulement si l'intégrale de la fonction $|\varphi| f_X$ sur $\mathbb{R}$ converge. Dans ce cas,*

$$E(Y) = \int_{-\infty}^{+\infty} \varphi(x)\, f_X(x)\, \mathrm{d}x.$$

Exemple 1.5. *Soit $Y = aX + b$ avec $a \neq 0$. Alors Y est intégrable si et seulement si X l'est et dans ce cas, $E(Y) = aE(X) + b$. Pour tout y,*

$$F_Y(y) = F_X((y-b)/a) \text{ si } a > 0, \quad F_Y(y) = 1 - F_X((y-b)/a) \text{ si } a < 0.$$

Donc Y admet une densité f_Y donnée par $f_Y(y) = f_X((y-b)/a)/|a|$.

L'espérance est un opérateur linéaire : si X et Y sont des variables aléatoires réelles intégrables et a et b des nombres réels,

$$E(aX + bY) = aE(X) + bE(Y).$$

L'espérance est un opérateur positif : si $X \geqslant 0$, alors $E(X) \geqslant 0$. Par conséquent, l'espérance préserve l'ordre, c'est-à-dire que si $X \leqslant Y$, alors $E(X) \leqslant E(Y)$.

Définition 1.6. *Soit X une variable aléatoire continue et p un réel strictement positif. On dit que X admet un moment d'ordre p si $E(|X|^p)$ est fini. Alors $E(X^p)$ existe et on l'appelle le moment d'ordre p de X.*

Le cas $p = 2$ est important et la section suivante lui est consacrée.

1.3 Deuxième moment

Définition 1.7. *On dit que la variable aléatoire X est de carré intégrable si et seulement si X^2 est intégrable si et seulement si l'intégrale*

$$\int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x$$

est finie. Alors,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx.$$

Voici un rappel.

Proposition 1.8 (Inégalité de Cauchy-Schwarz)**.** *Soit f et g deux fonctions telles que f^2 et g^2 sont d'intégrale finie sur $\mathbb{R}$. Alors fg est d'intégrale finie sur $\mathbb{R}$ et*

$$\left(\int_{-\infty}^{+\infty} f(x)g(x)\mathrm{d}x \right)^2 \leqslant \int_{-\infty}^{+\infty} f(x)^2 \mathrm{d}x \cdot \int_{-\infty}^{+\infty} g(x)^2 \mathrm{d}x.$$

Preuve (rappel) : une méthode consiste à considérer $h = af \pm bg$ pour deux réels quelconques a et b. Alors $h^2 \geqslant 0$ donc l'intégrale de h^2 est positive ou nulle. En notant $I(\varphi)$ l'intégrale d'une fonction intégrable φ et en élevant au carré, on obtient

$$4a^2b^2I(fg)^2 \leqslant [a^2I(f^2) + b^2I(g^2)]^2.$$

Choisissons $a = \sqrt{I(g^2)}$ et $b = \sqrt{I(f^2)}$. Il vient

$$I(f^2)I(g^2)I(fg)^2 \leqslant I(f^2)^2I(g^2)^2.$$

Si $I(f^2)$ et $I(g^2)$ sont strictement positifs, on a fini. Sinon, supposons par exemple que $I(f^2) = 0$. Donc $f = 0$ partout, $fg = 0$ partout et $I(fg)^2 = 0$ donc l'inégalité est vraie là aussi.

Une autre méthode, peut-être plus classique mais à mon avis moins bonne, consiste à remarquer que pour tout réel t, l'intégrale du carré de $f + tg$ est positive. Le discriminant du polynôme du second degré en t correspondant est donc négatif ou nul, ce qui fournit le résultat.

Ces deux preuves sont à connaître.

Application :
Soit f une fonction positive ou nulle. En écrivant $|x|f(x)$ comme le produit $|x|\sqrt{f(x)} \times \sqrt{f(x)}$, on obtient

$$\left(\int_{-\infty}^{+\infty} |x|f(x)\mathrm{d}x\right)^2 \leqslant \int_{-\infty}^{+\infty} x^2f(x)\mathrm{d}x \cdot \int_{-\infty}^{+\infty} f(x)\mathrm{d}x.$$

On en déduit le résultat suivant.

Proposition 1.9 (et définition). *Si X est de carré intégrable, X admet une espérance et*

$$E[X]^2 \leqslant E[|X|]^2 \leqslant E[X^2].$$

On définit alors la variance $\mathrm{var}(X)$ *et l'écart type $\sigma(X)$ de X par*

$$\mathrm{var}(X) = E[(X - E[X])^2) = E[X^2] - E[X]^2, \quad \sigma(X) = \sqrt{\mathrm{var}(X)}.$$

On note souvent $\mathrm{var}(X) = \sigma^2(X)$.

Proposition 1.10 (et définition). *1) Si X est de carré intégrable, alors $Y = aX+b$ aussi, pour tous a et b, et* $\mathrm{var}(Y) = a^2\mathrm{var}(X)$.
2) Pour toute variable aléatoire X de carré intégrable, $\mathrm{var}(X) \geqslant 0$ *et* $\mathrm{var}(X) = 0$ *si et seulement si X est constante.*
3) Pour toutes variables aléatoires X et Y de carré intégrable,

$$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y),$$

où la covariance $\text{cov}(X,Y)$ *de* X *et* Y *est définie, si* X *et* Y *sont de carré intégrable, par la formule*

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

Preuve de 1) : $E[Y] = aE[X] + b$ et $Y^2 = a^2X^2 + 2abX + b^2$ donc, toujours par linéarité, $E[Y^2]$ vaut $a^2E[X^2] + 2abE[X] + b^2$ et en effectuant la soustraction, on obtient le résultat.

1.4 Fonction caractéristique

Comme son nom l'indique, la fonction caractéristique fournit une façon de décrire entièrement la loi d'une variable aléatoire. La fonction caractéristique joue dans le cas général le rôle de la fonction génératrice dans le cas discret entier.

Définition 1.11. *On appelle fonction caractéristique de* X *la transformée de Fourier de la loi de* X*, c'est-à-dire la fonction* $\varphi_X : \mathbb{R} \to \mathbb{C}$ *définie pour tout réel* x*, par*

$$\varphi_X(x) = E[\mathrm{e}^{\mathrm{i}xX}].$$

La fonction caractéristique φ_X existe toujours puisque $|\mathrm{e}^{\mathrm{i}xX}| \leqslant 1$ est une fonction bornée de X.

Proposition 1.12. *La transformée de Fourier caractérise la loi : si les fonctions* φ_X *et* φ_Y *sont égales, alors les lois de* X *et de* Y *sont égales.*

Si $X \geqslant 0$ partout, on peut aussi considérer la transformée de Laplace λ_X de (la loi de) X, définie pour tout réel positif x, par

$$\lambda_X(x) = E[\mathrm{e}^{-xX}].$$

La transformée de Laplace λ_X existe toujours si $X \geqslant 0$ partout puisque, dans ce cas, $\mathrm{e}^{-xX} \leqslant 1$ est une fonction bornée de X.

La transformée de Laplace caractérise la loi : si X et Y sont des variables aléatoires positives et si $\lambda_X = \lambda_Y$, alors la loi de X et la loi de Y sont égales.

La notation φ_X est canonique, la notation λ_X ne l'est pas.

2 Familles de variables aléatoires

On considère maintenant un couple de variables aléatoires (X,Y). Sa fonction de répartition est définie par

$$F_{(X,Y)}(x,y) = P(X \leqslant x, Y \leqslant y).$$

Sa fonction caractéristique est définie par

$$\varphi_{(X,Y)}(x,y) = E[\mathrm{e}^{\mathrm{i}(xX+yY)}].$$

Les lois marginales sont les lois de X et de Y. Il est important de réaliser que la loi de X et la loi de Y ne déterminent pas la loi de (X,Y).

Vérifions ceci sur deux exemples.

Exercice 2.1. *Pour chaque couple de variables aléatoires (X,Y) ci-dessous, calculer la loi de X et la loi de Y.*

1) Soit $0 \leqslant t \leqslant 1$. La loi de (X,Y) est

$$\tfrac{1}{2}(1-t)(\delta_{(0,0)} + \delta_{(1,1)}) + \tfrac{1}{2}t(\delta_{(0,1)} + \delta_{(1,0)}).$$

2) La loi de (X,Y) est uniforme sur D, où D est le sous-ensemble du carré $[0,1]^2$ formé des points (x,y) tels que

$$x - \tfrac{1}{2} \leqslant y \leqslant x \quad \text{ou} \quad y \geqslant x + \tfrac{1}{2}.$$

Ensuite, décrire une loi plus simple d'un couple (X',Y') de variables aléatoires telles que X et X' ont la même loi et Y et Y' ont la même loi.

La façon la plus simple de spécifier la loi d'un couple à partir des marginales est d'imposer l'indépendance.

Définition 2.2. *On dit que deux événements A et A' sont indépendants si et seulement si $P(A \cap A') = P(A)P(A')$. On dit que deux variables aléatoires X et Y sont indépendantes si et seulement si, pour tous B et B', $\{X \in B\}$ et $\{Y \in B'\}$ sont des événements indépendants. En d'autres termes, on demande que*

$$P(X \in B, Y \in B') = P(X \in B)P(Y \in B').$$

Proposition 2.3. *Les variables aléatoires X et Y sont indépendantes :*

1. *si et seulement si, pour tous réels x et y, $F_{(X,Y)}(x,y) = F_X(x)F_Y(y)$,*
2. *si et seulement si, pour toutes fonctions h et g bornées,*

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)],$$

3. *si et seulement si, pour tous réels x et y, $\varphi_{(X,Y)}(x,y) = \varphi_X(x)\varphi_Y(y)$.*

Dans le cas densitable, X et Y sont indépendantes si et seulement si, pour tout couple (x,y),

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y).$$

Dans le cas discret, X et Y sont indépendantes si et seulement si, pour tout couple (x,y),

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Définition 2.4. *On dit qu'un vecteur* $(X_k)_{1\leqslant k\leqslant n}$ *de taille* $n \geqslant 2$ *est indépendant si et seulement si, pour tous* B_k,

$$P(\forall 1 \leqslant k \leqslant n,\, X_k \in B_k) = \prod_{k=1}^{n} P(X_k \in B_k).$$

Si le vecteur $(X_k)_{1\leqslant k\leqslant n}$ est indépendant, les variables aléatoires X_k sont indépendantes deux à deux pour $1 \leqslant k \leqslant n$. La réciproque est fausse, comme le montre l'exemple ci-dessous.

Exemple 2.5 (Important)**.** *Soit* X *et* Y *deux variables indépendantes de Bernoulli (voir plus bas) de loi* $\frac{1}{2}(\delta_1 + \delta_{-1})$, *et soit* $Z = XY$. *Calculer la loi de* (X, Y, Z) *et les lois de* (X, Y), (X, Z) *et* (Y, Z). *Préciser si le triplet* (X, Y, Z) *est indépendant. Préciser si les couples* (X, Y), (X, Z) *et* (Y, Z) *sont indépendants.*

Si X et Y sont des variables aléatoires indépendantes et de carré intégrable, alors $\mathrm{cov}(X, Y) = 0$. La réciproque est fausse, comme le montre l'exemple ci-dessous.

Exemple 2.6. *Soit* X *une variable aléatoire de loi* $\frac{1}{3}(\delta_{-1}+\delta_0+\delta_1)$ *et soit* $Y = X^2$. *Montrer que* $XY = X$. *En déduire que les variables aléatoires* X *et* Y *ne sont pas indépendantes mais que* $\mathrm{cov}(X, Y) = 0$.

Définition 2.7 (Coefficient de corrélation)**.** *Toujours dans le cas de carré intégrable, on note*

$$\varrho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}.$$

Donc $-1 \leqslant \varrho(X, Y) \leqslant 1$. Lorsque $|\varrho(X, Y)|$ est proche de 1, on dit que les variables aléatoires sont fortement corrélées.

Cas vectoriel : loi conjointe et marginales sans changement.

Dans le cas de carré intégrable, la matrice de covariance C vaut

$$C_{i,i} = \mathrm{var}(X_i), \quad C_{i,j} = \mathrm{cov}(X_i, X_j),\ i \neq j.$$

Exercice 2.8. *On peut prolonger l'exemple 2.5 ci-dessus comme suit.*

1) Soit X *une variable aléatoire ne prenant que les valeurs* $+1$ *et* -1.

Montrer que $P(X = 1) = \frac{1}{2}(E(X) + 1)$ *et* $P(X = 0) = \frac{1}{2}(1 - E(X))$.

2) Soit X *et* Y *deux variables aléatoires ne prenant que les valeurs* $+1$ *et* -1.

Montrer que X *et* Y *sont indépendantes si et seulement si* $E(XY) = E(X)E(Y)$ *si et seulement si* $P(X = Y = 1) = P(X = 1)P(Y = 1)$.

3) On fixe désormais $n \geqslant 2$, on note $[[n]] = \{1, 2, \ldots, n\}$ et on se donne n variables aléatoires $(X_i)_{i \in [[n]]}$ de loi symétrique $\frac{1}{2}(\delta_{-1} + \delta_1)$. Pour toute partie non vide I de $[[n]]$, on note $X_I = \prod_{i \in I} X_i$.

Montrer que les variables aléatoires $(X_i)_{i \in [[n]]}$ sont indépendantes si et seulement si $E(X_I) = 0$ pour toute partie non vide I de $[[n]]$ si et seulement si $P(\forall i \in I,\, X_i = 1) = \prod_{i \in I} P(X_i = 1)$ pour toute partie non vide I de $[[n]]$.

4) On suppose désormais de plus que les variables aléatoires $(X_i)_{i \in [[n]]}$ sont indépendantes et on note $Y = X_{[[n]]}$ et $Y_i = X_{[[n]] \setminus \{i\}}$ pour tout i dans $[[n]]$.

Montrer que $Y_i = Y X_i$ pour tout i. Calculer le produit $Y_1 Y_2 \cdots Y_n$ (on séparera les cas n pair et n impair).

Montrer que X_I est une variable aléatoire de loi $\frac{1}{2}(\delta_{-1} + \delta_1)$ pour toute partie non vide I de $[[n]]$.

Montrer que X_I et X_J sont indépendantes pour toutes parties I et J distinctes de $[[n]]$.

Calculer $P(Y_1 = \cdots = Y_i = 1)$ pour tout i dans $[[n]]$, $i \neq n$, puis $P(Y_1 = \cdots = Y_n = 1)$. Pour le cas $i = n$, on séparera les cas n pair et n impair.

5) Déduire de tout ceci que toute sous-famille de $(Y_i)_{i \in [[n]]}$ de cardinal au plus $n - 1$ est indépendante, que $(Y_i)_{i \in [[n]]}$ est indépendante si n est pair mais non indépendante si n est impair.

3 Exemples de lois à densité

3.1 Loi uniforme $U(I)$

Paramètre : I intervalle d'intérieur non vide. Densité $|I|^{-1}\mathbf{1}_I$.

Si $I = (a, b)$ avec $a < b$, espérance $(a + b)/2$, variance : $(b - a)^2/12$. Fonction de répartition $F(x) = 0$ si $x \leqslant a$, $F(x) = (x - a)/(b - a)$ si $a \leqslant x \leqslant b$ et $F(x) = 1$ si $x \geqslant b$.

Si la loi de X est $U([0, 1])$ (uniforme standard) et si $A : x \mapsto ax + b$, $a \neq 0$, est une application affine, alors la loi de $A(X) = aX + b$ est uniforme sur $A([0, 1])$ qui vaut $[b, a + b]$ si $a > 0$ et $[a + b, b]$ si $a < 0$.

3.2 Loi exponentielle $\mathcal{E}(a)$

Paramètre : a réel strictement positif. Densité $a\,\mathrm{e}^{-ax}\,\mathbf{1}_{x\geqslant 0}$.

Espérance $1/a$, variance $1/a^2$. Fonction de répartition $F(x) = 0$ si $x \leqslant 0$ et $F(x) = 1 - \mathrm{e}^{-ax}$ si $x \geqslant 0$.

Si la loi de X est $\mathcal{E}(1)$ (exponentielle standard) et si a est un réel positif, alors la loi de aX est $\mathcal{E}(1/a)$.

Construction : horloges. Ou $-\log U$, U uniforme standard.

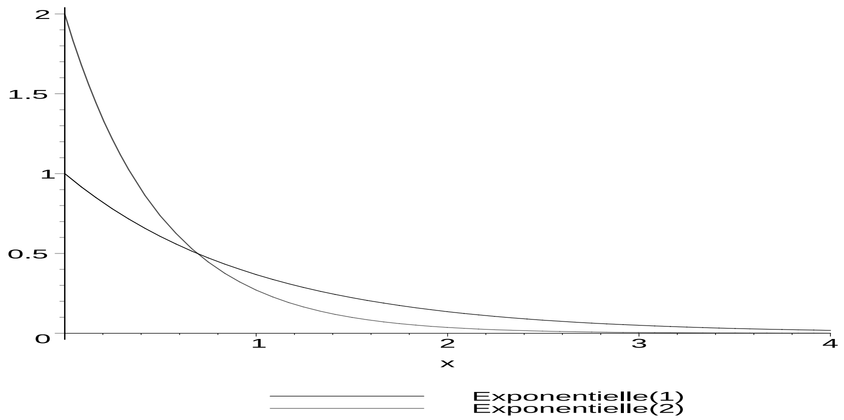


Figure 1. Densité de lois exponentielles

Les lois exponentielles sont les seules lois à valeurs dans $\mathbb{R}^+$ « sans mémoire », c'est-à-dire telles que, pour tous réels positifs s et t,

$$P(X \geqslant s + t | X \geqslant t) = P(X \geqslant s).$$

3.3 Loi normale (ou gaussienne) $\mathcal{N}(m, v)$

Paramètres : m réel et v réel strictement positif. Densité $\mathrm{e}^{-(x-m)^2/(2v)}/\sqrt{2\pi v}$.

Espérance m, variance v.

Si la loi de X est $\mathcal{N}(0, 1)$ (gaussienne standard) et si a et b sont des réels, alors la loi de $aX + b$ est $\mathcal{N}(b, a^2)$.

Par convention, $\mathcal{N}(m, 0)$ est une masse de Dirac en m.

La somme de variables aléatoires indépendantes de loi $\mathcal{N}(m_k, v_k)$ suit la loi $\mathcal{N}(m, v)$ avec

$$m = \sum_k m_k, \qquad v = \sum_k v_k.$$

Construction : loi des erreurs.

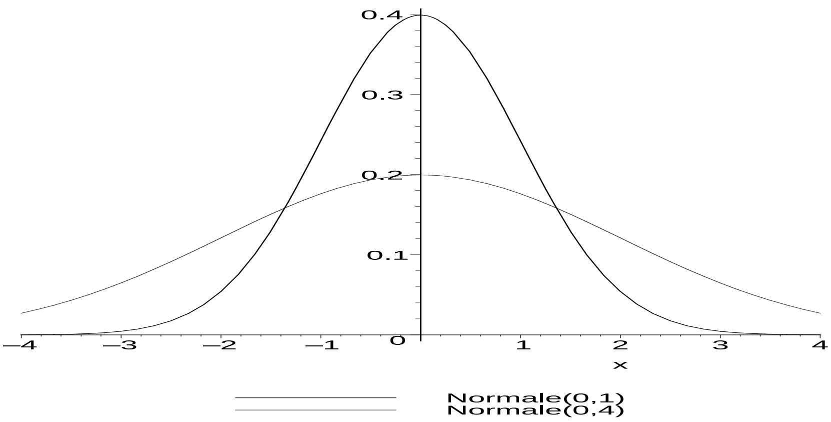


Figure 2. Densité de lois normales

3.4 Loi de Cauchy $C(m, a)$

Paramètres : a strictement positif et m réel. Densité $(a/\pi)/((x-m)^2 + a^2)$.

Pas d'espérance ni de variance car les intégrales divergent. La médiane vaut m.

Si la loi de X est $C(0,1)$ (loi de Cauchy standard) et si a et b sont des réels, alors la loi de $aX + b$ est $C(b, a^2)$.

Par convention, $C(m, 0)$ est une masse de Dirac en m.

La somme de variables aléatoires indépendantes de loi $C(m_k, a_k)$ suit la loi $C(m, a)$ avec $m = \sum_k m_k$ et $a = \sum_k a_k$.

Construction : la mouche. Ou bien le rapport de deux gaussiennes standard.

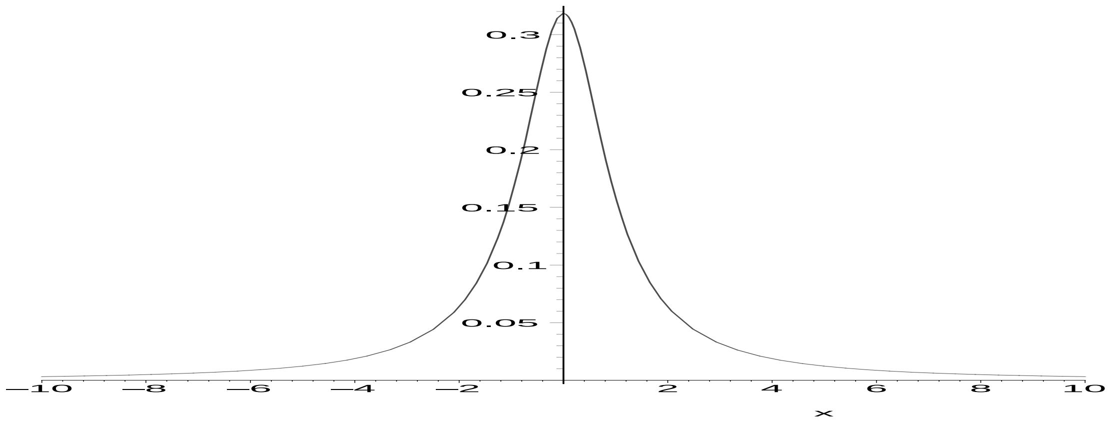


Figure 3. Densité de la loi de Cauchy

3.5 Autres lois

3.5.1 Loi gamma $G(a, b)$

Paramètres : a et b strictement positifs. Densité

$$b^a \Gamma(a)^{-1} x^{a-1} \mathrm{e}^{-bx} \mathbf{1}_{x>0}.$$

Espérance a/b, variance a/b^2.

Construction : pour tout entier $n \geqslant 1$, la somme de n variables aléatoires indépendantes de loi $\mathcal{E}(b)$ suit la loi $G(n, b)$.

3.5.2 Loi beta $\mathcal{B}(a, b)$

Paramètres : a et b strictement positifs. Densité

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \mathbf{1}_{0<x<1}.$$

Espérance $a/(a+b)$, variance $ab/((a+b)^2(a+b+1))$.

Construction : pour tous entiers $n \geqslant 1$ et $m \geqslant 1$, si on considère $n+m$ variables aléatoires indépendantes de loi $\mathcal{E}(b)$, le rapport de la somme des n premières sur la somme totale suit la loi $\mathcal{B}(n, m)$.

4 Exercices

1. Marginales Soit (X, Y) un couple de variables aléatoires de densité

$$f(x, y) = \sqrt{x/y}\, \mathbf{1}_D(x, y), \qquad D = \{(x, y) \,|\, 0 < y \leqslant x \leqslant 1\}.$$

Vérifier que f est bien une densité sur $\mathbb{R}^2$.

Déterminer les lois marginales de X et Y. Préciser si les variables aléatoires X et Y sont indépendantes.

Calculer $P(Y < X/2)$.

Réaliser le couple (X, Y) à l'aide de deux variables aléatoires indépendantes à valeurs dans $[0, 1]$.

2. Marginales encore Soit (X, Y) un couple de variables aléatoires de densité

$$f(x, y) = c\,\mathrm{e}^{-x} y^{-2}\,\mathbf{1}_{x>0}\,\mathbf{1}_{y>1}.$$

Calculer la constante c.

Déterminer les lois marginales de X et Y. Préciser si les variables aléatoires X et Y sont indépendantes.

Calculer $P(Y > 2, X < 1)$ et $P(XY > 1)$.

3. Durées de vie On fixe a strictement positif. Une usine fabrique des ampoules dont la durée de vie T mesurée en heures vérifie $P(T > t) = \mathrm{e}^{-at}$ pour tout t positif.

Préciser la loi de T. Calculer la durée de vie moyenne de ces ampoules et l'écart type associé.

On considère n ampoules dont les durées de vie $(T_k)_{1\leqslant k\leqslant n}$ sont indépendantes et de même loi que T. On note $U = \min\{T_k \,|\, 1 \leqslant k \leqslant n\}$ le premier instant où au moins une des ampoules cesse de fonctionner et $V = \max\{T_k \,|\, 1 \leqslant k \leqslant n\}$ le premier instant où toutes les ampoules ont cessé de fonctionner.

Préciser les lois de U et de V.

4. Jacobien Soient X et Y deux variables aléatoires indépendantes de même loi de densité la fonction f définie par $f(x) = x^{-2}\mathbf{1}_{x\geqslant 1}$. Soient $U = XY$ et $V = X/Y$.

Calculer la loi du couple (U, V) et ses lois marginales. Préciser si les variables U et V sont indépendantes.

5. Exponentielles Soit X et Y deux variables indépendantes de lois exponentielles de paramètre respectif a et b strictement positifs. On pose $U = \min(X, Y)$, $V = \max(X, Y)$ et $W = V - U$.

Calculer $P(U = X)$ et montrer que U et W sont indépendantes.

6. Marginales enfin Soit a un nombre réel strictement positif et (X, Y) un couple aléatoire de densité

$$f(x, y) = ce^{-ay}\,\mathbf{1}_D(x, y), \quad D = \{(x, y) \in \mathbb{R}^2 \,|\, 0 < x < y\}.$$

Préciser la valeur de c.

On pose $Z = X/Y$. Déterminer la loi du vecteur (Z, Y). Préciser si les variables Z et Y sont indépendantes. Déterminer leurs lois respectives.

Capes de Mathématiques **Université Joseph Fourier**
Préparation à l'écrit **Année 2008-2009**

Probabilités 4 : Théorèmes limites

« *Mathématiques : dessèchent le coeur.* »
Gustave Flaubert, Dictionnaire des idées reçues

1 Théorèmes limites

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs réelles $X_n : \Omega \to \mathbb{R}$. On rappelle que ces variables aléatoires sont indépendantes si et seulement si, pour tout $n \geqslant 1$ et tous $x_i \leqslant y_i$,

$$P(\forall 1 \leqslant i \leqslant n,\, x_i \leqslant X_i \leqslant y_i) = \prod_{i=1}^{n} P(x_i \leqslant X_i \leqslant y_i).$$

Théorème 1.1 (Loi forte des grands nombres)**.** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi admettant une espérance m. On note $S_n := X_1 + \cdots + X_n$. Alors il existe un événement A de probabilité 1 tel que pour tout $\omega \in A$, $S_n(\omega)/n \longrightarrow m$. On dit que S_n/n converge vers m presque sûrement.*

Corollaire 1.1 (Loi faible des grands nombres)**.** *Sous les mêmes hypothèses, pour tout t strictement positif,*

$$P(nm - nt \leqslant S_n \leqslant nm + nt) \longrightarrow 1.$$

On dit que S_n/n converge vers m en probabilité.

Si on répète une même expérience un grand nombre de fois de manière identique et que l'on regarde le nombre de fois où un résultat R apparaît, la loi forte des grands nombres montre que la fréquence empirique d'apparition de R tend vers la probabilité de R quand le nombre d'expériences tend vers l'infini. Par exemple, si on lance une pièce équilibrée un grand nombre de fois, la suite des fréquences relatives des piles que l'on obtient tend avec probabilité 1 vers 50%. On peut se demander quelle est l'erreur faite quand on remplace directement la fréquence obtenue après n expériences par 50%. Le théorème 1.3 ci-dessous répond d'une certaine manière à cette question.

Nous commençons par quelques rappels sur les variables aléatoires gaussiennes. Une variable aléatoire X suit la loi $\mathcal{N}(m, v)$, donc est gaussienne (ou normale) d'espérance m et de variance v, si X admet pour densité la fonction f définie sur $\mathbb{R}$ par

$$f(x) := \frac{1}{\sqrt{2\pi v}}\,\mathrm{e}^{-(x-m)^2/(2v)}.$$

Quelques conséquences.

La transformée de Fourier d'une variable aléatoire X de loi $\mathcal{N}(m, v)$ vaut, en tout t réel,

$$\varphi_X(t) := E[\mathrm{e}^{\mathrm{i}tX}] = \mathrm{e}^{\mathrm{i}mt}\mathrm{e}^{-vt^2/2}.$$

Pour tous réels a et b, si la loi de X est $\mathcal{N}(m, v)$, alors la loi de $aX + b$ est $\mathcal{N}(am + b, a^2 v)$. (On rappelle que par convention $\mathcal{N}(m, 0)$ est la masse de Dirac en m.)

Si les variables aléatoires réelles X_k sont gaussiennes et indépendantes, alors pour tous nombres réels a_k, la variable aléatoire $\sum_k a_k X_k$ suit une loi gaussienne. L'hypothèse d'indépendance est importante, comme le montre le contrexemple ci-dessous.

Exercice 1.2. Soit T de loi de Bernoulli $b(p)$, donc $P(T = 1) = p$ et $P(T = 0) = 1 - p$. Soit X de loi gaussienne $\mathcal{N}(0, 1)$ et indépendante de T. Soit $Y = TX$ et $Z = X + Y$. Montrer que la loi de Y est gaussienne mais pas celle de Z.

Indication : calculer $P(Z = 0)$.

Nous sommes maintenant en mesure d'énoncer le théorème annoncé.

Théorème 1.3 (Théorème central limite)**.** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi de carré intégrable, d'espérance m et de variance v. On note $S_n = X_1 + \cdots + X_n$. Alors, pour tous x et y tels que $x \leqslant y$,*

$$P(nm + x\sqrt{nv} \leqslant S_n \leqslant nm + y\sqrt{nv}) \longrightarrow P(x \leqslant Z \leqslant y),$$

où Z désigne une variable aléatoire de loi gaussienne centrée réduite, donc

$$P(x \leqslant Z \leqslant y) = \frac{1}{\sqrt{2\pi}} \int_x^y \mathrm{e}^{-z^2/2}\mathrm{d}z.$$

2 Exemples d'application

Soit Z une variable aléatoire de loi gaussienne centrée réduite.

Binomiale Si X_n suit une loi binomiale $B(n,p)$, alors, quand n est grand,

$$\frac{X_n - np}{\sqrt{np(1-p)}} \approx Z.$$

En effet, la loi de X_n est celle d'une somme de n variables aléatoires indépendantes de loi de Bernoulli $b(p)$, de moyenne p et de variance $p(1-p)$.

Poisson Si X_a suit une loi de Poisson $\mathcal{P}(a)$, alors, quand a est grand,

$$\frac{X_a - a}{\sqrt{a}} \approx Z.$$

En effet, si $a = n$ est entier, la loi de X_n est celle d'une somme de n variables aléatoires indépendantes de loi de Poisson $\mathcal{P}(1)$, de moyenne 1 et de variance 1 et, si a n'est pas entier, on peut se ramener au cas entier (preuve omise).

Usage En pratique, les statisticiens observent les règles d'usage suivantes.

1. Si $a \geqslant 10$, on peut approcher la loi de Poisson $\mathcal{P}(a)$ par la loi gaussienne $\mathcal{N}(a,a)$.
2. Si $np \geqslant 10$ et $n(1-p) \geqslant 10$, on peut approcher la loi binomiale $B(n,p)$ par la loi gaussienne $\mathcal{N}(np, np(1-p))$.
3. Si $np < 10$ et $p \leqslant \frac{1}{10}$, on peut approcher la loi binomiale $B(n,p)$ par la loi de Poisson $\mathcal{P}(np)$.
4. Si $n(1-p) < 10$ et $p \geqslant \frac{9}{10}$, on peut approcher la loi binomiale $B(n,p)$ par la loi de $n - X$ avec X de loi de Poisson $\mathcal{P}(n(1-p))$.

Fumeurs Une proportion inconnue $0 < p < 1$ d'une certaine population est constituée de fumeurs. On utilise un tirage avec remise dans cette population pour déterminer le paramètre p. Un candidat naturel pour p est la proportion F_n de fumeurs tirés au cours de n tirages. On désire trouver p avec une erreur plus petite que $t := 0,5\%$, donc on souhaite assurer la réalisation de l'événement $A_n := \{p - t \leqslant F_n \leqslant p + t\}$. On veut déterminer combien de tirages sont nécessaires.

Tout d'abord, il faut réaliser qu'aucun nombre de tirages ne garantit que A_n est réalisé avec certitude : il est possible par exemple de ne tirer que des fumeurs pendant les n premiers tirages, auquel cas $F_n = 1$, mais ceci se produit avec probabilité p^n donc cette éventualité devient très improbable si n devient grand. On peut seulement rendre A_n très probable, donc on se contente de fixer un niveau de confiance a, par exemple $a = 95\%$, et de chercher un entier n tel que $P(A_n) \geqslant a$.

Notons $X_i = 1$ si la personne tiré au tirage numéro i est fumeur et $X_i = 0$ sinon. La suite $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes et de même loi de Bernoulli de paramètre p. Soit $S_n := X_1 + \cdots + X_n$ le nombre de fumeurs tirés jusqu'au temps n, et

$$T_n := \frac{S_n - np}{\sqrt{np(1-p)}}.$$

On a donc $S_n = nF_n$ et on cherche n tel que $P(A_n) \geqslant a$. Or,

$$A_n = \left\{|T_n| < t\sqrt{n/(p(1-p))}\right\} \supset \left\{|T_n| < 2t\sqrt{n}\right\},$$

car $\sqrt{p(1-p)} \leqslant 1/2$. Soit Z une variable aléatoire gaussienne centrée réduite et z_a un nombre réel tel que $P(|Z| \geqslant z_a) = a$.

La façon usuelle d'utiliser le théorème central limite est de décréter que $2t\sqrt{n} = z_a$ doit suffire, donc de choisir $n = z_a^2/(4t^2)$.

Si $a = 95\%$ on trouve grâce à une table de la distribution normale standard que $z_a = 1,96$. Finalement, $n = 40000$ convient.

Remarque : en toute rigueur, le théorème central limite est valide pour un niveau d'erreur fixé et quand n tend vers l'infini ; ici, on l'utilise quand n et le taux d'erreur $2t\sqrt{n}$ tendent tous les deux vers l'infini en même temps, donc en dehors de son domaine théorique de validité, comme le montre l'exercice très simple ci-dessous.

Exercice : exhiber une famille $(x_{n,k})_{n,k}$ de nombres réels tels que, pour tout k fixé, $x_{n,k} \longrightarrow 0$ quand $n \to \infty$, et pourtant $x_{n,n}$ ne tend pas vers 0 quand $n \to \infty$.

3 Exercices

1. Arrondis Un ordinateur effectue les calculs avec 9 chiffres significatifs et arrondit les résultats théoriques à cette précision.

On suppose qu'il effectue 10^6 opérations élémentaires, que les erreurs d'arrondi à chaque opération sont indépendantes et de loi uniforme sur l'intervalle $[-\frac{1}{2}10^{-9}, \frac{1}{2}10^{-9}]$, et que l'erreur sur le résultat final est la somme des erreurs commises à chaque opération.

On demande d'évaluer la probabilité pour que l'erreur finale soit en valeur absolue inférieure à $\frac{1}{2}10^{-6}$.

On indique que $P(0 \leqslant Z \leqslant \sqrt{3}) \simeq 46\%$ pour une variable aléatoire Z gaussienne centrée réduite.

2. Une preuve probabiliste de la formule de Stirling Soit Z une variable aléatoire de loi gaussienne centrée réduite et $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de loi de Poisson de paramètre 1. On pose $S_n = \sum_{k=1}^{n} X_k$ et $T_n = (S_n - n)/\sqrt{n}$.

a) Déterminer la loi de S_n, son espérance et sa variance.

b) Montrer que pour tout $a \leqslant b$, $P(a \leqslant T_n \leqslant b)$ tend vers $P(a \leqslant Z \leqslant b)$.

c) En déduire que $\mathrm{e}^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \longrightarrow \frac{1}{2}$.

d) Montrer que $\int_0^{\infty} P(T_n \geqslant x)\,\mathrm{d}x \longrightarrow \int_0^{\infty} P(Z \geqslant x)\,\mathrm{d}x$ et calculer cette limite.

e) Montrer que $\int_0^{\infty} P(T_n \geqslant x)\,\mathrm{d}x = \mathrm{e}^{-n} n^n \sqrt{n}/n!$.

f) En déduire la formule de Stirling $n! \sim n^n \mathrm{e}^{-n}\sqrt{2\pi n}$.

3. Processus de comptage Pour tout temps $t \geqslant 0$, N_t désigne le nombre de clients qui arrivent dans un service pendant l'intervalle de temps $]0, t]$

a) Que représente, pour t et h strictement positifs, la quantité $N_{t+h} - N_t$?

On suppose qu'il existe un nombre réel a strictement positif tel que, quand $h \to 0$, pour tout t, $P(N_{t+h} - N_t = 1) = ah + o(h)$ et $P(N_{t+h} - N_t \geqslant 2) = o(h)$, et que si $t < s$, les variables aléatoires $N_s - N_t$ et N_t sont indépendantes.

b) Soit $p_n(t) = P(N_t = n)$ et $q_n(t, h) = P(N_{t+h} - N_t = n)$. On suppose que p_n est une fonction dérivable de t.

Montrer que $q_0(t, h) = 1 - ah + o(h)$, $q_1(t, h) = ah + o(h)$ et $q_n(t, h) = o(h)$ pour $n \geqslant 2$.

Montrer que $p_0(t + h) = q_0(t, h)p_0(t)$.

En déduire que $p_0'(t) = -ap_0(t)$.

c) Montrer que, pour $n \geqslant 1$, $p_n(t + h) = \sum_{k=0}^{n} p_k(t)q_{n-k}(t, h)$.

En déduire que $p_n'(t) = -a\,(p_n(t) + p_{n-1}(t))$.

d) Montrer que pout tout $n \geqslant 0$, $p_n(t) = \mathrm{e}^{-at}(at)^n/n!$.

e) Soit T_1 l'instant où arrive le premier client. Montrer que T_1 suit une loi exponentielle de paramètre a.

4. Partie entière Soit X une variable aléatoire de loi uniforme sur $[0,1]$, Y la partie entière de $10X$ et Z la partie entière de $100X - 10Y$.

Préciser ce que représentent Y et Z.

Trouver les lois de (Y, Z), de Y et de Z. Préciser si Y et Z sont indépendantes. Proposer une généralisation de ces résultats et en déduire une façon de simuler une variable aléatoire de loi uniforme sur $[0,1]$ à partir d'une suite de variables aléatoires indépendantes de loi uniforme sur $\{0, 1, \dots, 9\}$.

Déduire de tout ceci une façon de simuler une variable aléatoire de loi uniforme sur $[0,1]$ à partir d'une suite de variables aléatoires indépendantes de loi de Bernoulli $\frac{1}{2}(\delta_0 + \delta_1)$.

Capes de Mathématiques **Université Joseph Fourier**
Préparation à l'écrit **Année 2008-2009**

Probabilités 5 : Applications statistiques

« There are three kinds of lies : lies, damned lies, and statistics. » Attribué à Benjamin Disraeli par Mark Twain dans son Autobiographie

1 Estimation statistique

Dans la démarche statistique, on cherche à traiter des données, en général une série d'observations x_1, x_2, ..., x_n. On suppose qu'elles correspondent à des réalisations de variables aléatoires X_1, X_2, ..., X_n, et on cherche une distribution théorique du vecteur $(X_k)_{1\leqslant k\leqslant n}$ qui reflète correctement les propriétés des observations $(x_k)_{1\leqslant k\leqslant n}$.

Concrètement, les valeurs x_k sont donc connues. Par exemple, x_k représente la durée de vie du moteur de la voiture numéro k que l'on a choisi d'observer. Si le fabriquant est un grand constructeur, on ne peut pas recenser la durée de vie de toutes les voitures fabriquées donc on ne considère qu'un échantillon de taille raisonnable. Le constructeur aimerait à partir de ces valeurs améliorer la fabrication des moteurs. Il serait alors utile de connaitre la loi sous-jacente à la durée de vie d'un moteur. Cette loi est inconnue, mais à partir de l'échantillon $(x_k)_k$, on peut cependant estimer certaines valeurs, comme par exemple la durée de vie moyenne d'un moteur ; on parle alors de problèmes d'estimation et d'intervalle de confiance. Se pose parfois ensuite la question de la validité de l'estimation ; on parle alors de problème de test.

Dans toute la suite, on se restreint au cas le plus simple : **on suppose que les variables aléatoires $(X_k)_{1\leqslant k\leqslant n}$ sont indépendantes et de même loi et que cette loi appartient à une collection fixée a priori, que l'on notera $\{P_\theta\,;\,\theta\in\Theta\}$.**

Un premier exemple : les tirages avec remise (TAR). On considére une pièce de monnaie dont on ignore si elle est truquée ou non, et on veut connaitre la probabilite p de tomber sur « pile ». On lance n fois la pièce et on note X_i le résultat du lancer numéro i, avec $X_i = 1$ si on obtient « pile » (ce qui se produit avec probabilité p) et $X_i = 0$ si on obtient « face » (ce qui se produit avec probabilité $1-p$). Les variables aléatoires X_i sont indépendantes et de même loi de Bernoulli $b(p)$ donc, sur n lancers, le nombre total de « piles » $S_n = X_1 + \cdots + X_n$ suit la loi binomiale $B(n,p)$.

Un deuxième exemple : les tirages sans remise (TSR). On considère une population de taille N divisée en deux classes selon leur caractère, par exemple les personnes de rhésus positif et celles de rhésus négatif. Il y a n_1 individus dans la classe 1 et $N - n_1$ individus dans la classe 2 mais les quantités n_1 et N sont grandes et inconnues. Pour estimer n_1 ou N ou les deux, on considère une sous-population de taille raisonnable n. On compte alors le nombre de personnes de chaque classe dans cette sous-population. Si X_k désigne la classe de l'individu obtenu au tirage numéro k et si la sous-population est choisie de façon uniforme, on voit que les variables aléatoires X_k sont de même loi mais non indépendantes. En effet, $P(X_1 = 1) = P(X_2 = 1) = p_1$ avec $p_1 := n_1/N$ mais

$$P(X_1 = X_2 = 1) = \frac{n_1(n_1-1)}{N(N-1)} \neq p_1^2.$$

On considèrera donc plutôt directement le nombre $0 \leqslant X \leqslant n$ d'individus de la classe 1 parmi les n individus interrogés et on sait que la loi de X est hypergéométrique $H(n_1, n, N)$.

1.1 Estimation ponctuelle

On considère un échantillon $(X_k)_{1\leqslant k\leqslant n}$ tel que chaque X_k est à valeurs dans un ensemble E (par exemple, $E = \mathbb{R}$) et de même loi P_θ qui dépend d'un paramètre inconnu $\theta \in \Theta$. Le but est d'estimer la valeur de θ à partir des valeurs des X_k.

Dans l'exemple TAR, le paramètre inconnu est $\theta = p$, dans l'exemple TSR, le paramètre inconnu est $\theta = (n_1, N)$.

Définition 1.1. *Un estimateur de θ est une variable aléatoire $\widehat{\theta}_n$ telle qu'il existe une fonction $F_n : E^n \to \Theta$ avec $\widehat{\theta}_n = F_n(X_1, X_2, \ldots, X_n)$; c'est donc une fonction de l' échantillon.*

Attention : un estimateur de θ ne doit pas dépendre de θ, mais seulement des observations $(X_k)_{1\leqslant k\leqslant n}$.

Le problème est maintenant de choisir la fonction F_n de façon à estimer correctement θ.

Dans l'exemple TAR, $\widehat{\theta}_n = 10$ et $\widehat{\theta}_n = X_1X_2$ sont des estimateurs de $\theta = p$, tous deux un peu stupides (dire pourquoi). On préfèrera utiliser l'estimateur naturel de

p fourni par la fréquence empirique des succès $\overline{X}_n = S_n/n$, qui vaut le nombre de « piles » divisé par le nombre de lancers.

Pour tout θ dans Θ, on note E_θ l'espérance par rapport à la loi P_θ.

Définition 1.2. *L'estimateur $\widehat{\theta}_n$ est un estimateur consistant de θ pour la valeur θ si $\widehat{\theta}_n \longrightarrow \theta$ presque sûrement par rapport à P_θ, quand n tend vers l'infini.*

La consistance est la propriété la plus importante d'un estimateur. Ainsi, pour de grands échantillons, l'approximation de θ par $\widehat{\theta}_n$ est correcte. Par exemple, la loi forte des grands nombres affirme que, dans l'exemple TAR, $\overline{X}_n$ est un estimateur consistant de p.

Définition 1.3. *Le biais d'un estimateur $\widehat{\theta}_n$ de θ est*

$$B(\widehat{\theta}_n, \theta) := E_\theta(\widehat{\theta}_n) - \theta.$$

Le biais $B(\widehat{\theta}_n, \theta)$ est donc un nombre, qui dépend de θ et de la fonction F_n définissant l'estimateur $\widehat{\theta}_n$.

L'estimateur $\widehat{\theta}_n$ est dit sans biais si $B(\widehat{\theta}_n, \theta) = 0$ pour toute valeur de θ dans Θ, sinon l'estimateur est dit biaisé.

La valeur moyenne d'un estimateur sans biais est notre inconnue θ. Dans l'exemple TAR, $\overline{X}_n$ est un estimateur sans biais de p.

On veut maintenant comparer différents estimateurs de θ.

Définition 1.4. *Le risque quadratique de l'estimateur $\widehat{\theta}_n$ par rapport à θ est défini par*

$$R(\widehat{\theta}_n, \theta) := E_\theta\left[(\widehat{\theta}_n - \theta)^2\right].$$

Si le risque est faible, l'estimateur est souvent proche de l'inconnue θ, donc on souhaite avoir un risque le plus faible possible.

Définition 1.5. *L'estimateur $\widehat{\theta}_n$ est (quadratiquement) meilleur que l'estimateur $\widetilde{\theta}_n$ pour la valeur θ dans Θ, si $R(\widehat{\theta}_n, \theta) \leqslant R(\widetilde{\theta}_n, \theta)$.*

L'estimateur $\widehat{\theta}_n$ est (quadratiquement) uniformément meilleur que l'estimateur $\widetilde{\theta}_n$ si $\widehat{\theta}_n$ est (quadratiquement) meilleur que $\widetilde{\theta}_n$ pour tout θ dans Θ.

Attention : on ne peut pas toujours comparer uniformément deux estimateurs.

On remarque que

$$R(\widehat{\theta}_n, \theta) = \mathrm{var}_\theta(\widehat{\theta}_n) + B(\widehat{\theta}_n, \theta)^2.$$

Par conséquent, si l'estimateur est sans biais, $R(\widehat{\theta}_n,\theta) = \mathrm{var}_\theta(\widehat{\theta}_n)$. C'est aussi la raison pour laquelle on préfère souvent les estimateurs sans biais.

Attention : on pourrait donc penser utiliser $\widehat{\theta}_n - B(\widehat{\theta}_n,\theta)$ comme estimateur au lieu de $\widehat{\theta}_n$ puisque le risque quadratique est plus petit :

$$R(\widehat{\theta}_n - B(\widehat{\theta}_n,\theta),\theta) = \mathrm{var}_\theta(\widehat{\theta}_n) \leqslant \mathrm{var}_\theta(\widehat{\theta}_n) + B(\widehat{\theta}_n,\theta)^2 = R(\widehat{\theta}_n,\theta).$$

Mais $\widehat{\theta}_n - B(\widehat{\theta}_n,\theta)$ n'est pas en général un estimateur ! En effet, le terme de biais $B(\widehat{\theta}_n,\theta)$ dépend de $\widehat{\theta}_n$, ce qui est parfait puisque $\widehat{\theta}_n$ ne dépend que de l'échantillon observé, mais aussi de θ que l'on ne connaît pas.

1.2 Estimateurs des moments

Un principe général pour estimer θ dans deux situations particulières.

Premier cas : si $\theta = E[\varphi(X_1)]$, alors $\widehat{\theta}_n := \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\varphi(X_k)$ est un estimateur sans biais et consistant.

Deuxième cas : si $\theta = \psi(E[X_1])$ avec ψ continue, alors $\widehat{\theta}_n := \psi(\overline{X}_n)$ est un estimateur, biaisé en général, mais consistant.

1.2.1 Moyenne empirique

Si $\theta = E(X_1)$, on pourra utiliser la moyenne empirique, définie par

$$\overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k.$$

La moyenne empirique est un estimateur sans biais de la moyenne $E(X)$, consistant grâce à la loi des grands nombres, de risque

$$R(\overline{X}_n,\theta) = \frac{1}{n}\mathrm{var}_\theta(X_1).$$

Pour un échantillon de loi de Bernoulli $b(p)$, la loi de $n\overline{X}_n$ est binomiale $B(n,p)$. Pour un échantillon de loi normale $\mathcal{N}(\theta,v)$, la loi de $\overline{X}_n$ est normale $\mathcal{N}(\theta,v/n)$.

1.2.2 Variance et covariance empiriques

En appliquant ce qui précède aux variables aléatoires $(X_k^2)_k$, on voit qu'un estimateur consistant de $E(X_1^2)$ vaut

$$\frac{1}{n}\sum_{k=1}^{n} X_k^2.$$

Comme $\mathrm{var}(X) = E(X^2) - E(X)^2$, on obtient un estimateur consistant de $\mathrm{var}(X_1^2)$ en posant

$$\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X}_n^2 = \frac{1}{n}\sum_{i=k}^{n} \left(X_k - \overline{X}_n\right)^2.$$

Mais $E(\widehat{\sigma}_n^2) = v\,\frac{n-1}{n} \neq v$ donc $\widehat{\sigma}_n^2$ est biaisé.

Proposition 1.6. *Pour tout échantillon de taille $n \geqslant 2$, indépendant et de même loi de carré intégrable et de variance $v = \mathrm{var}(X_1)$,*

$$V_n := \frac{1}{n-1}\sum_{k=1}^{n} \left(X_k - \overline{X}_n\right)^2$$

est un estimateur sans biais et consistant de v.

Pour la covariance, on considère un échantillon $(X_k, Y_k)_{1\leqslant k\leqslant n}$ indépendant et de même loi de carré intégrable dont on veut estimer la covariance

$$C = \mathrm{cov}(X_1, Y_1) = E(X_1Y_1) - E(X_1)E(Y_1).$$

Alors $\widehat{C}_n$ est un estimateur sans biais et consistant de C, si on pose

$$\widehat{C}_n := \frac{1}{n-1}\sum_{k=1}^{n} X_kY_k - \overline{X}_n\overline{Y}_n.$$

Exercice : Où doit-on mettre les parenthèses dans $\widehat{C}_n$ pour que cet estimateur soit effectivement sans biais ?

2 Estimation par intervalle de confiance

Dans la section précédente on proposait une valeur unique $\widehat{\theta}_n$ pour estimer θ. On veut maintenant proposer un ensemble $I_n \subset \Theta_n$ aussi petit que possible, tel que θ appartienne souvent à I_n.

Comme précédemment, on ne dispose pour construire I_n que des observations $(X_k)_{1\leqslant k\leqslant n}$, que l'on suppose indépendantes et de même loi P_θ pour une certaine valeur inconnue du paramètre $\theta \in \Theta$.

Définition 2.1. *Un intervalle de confiance au niveau a est un intervalle aléatoire I_n qui ne dépend que de l'échantillon $X_1,\dots,X_n$, mais pas de θ, et tel que, pour tout θ dans Θ,*

$$P_\theta(\theta \in I_n) \geqslant a.$$

Le nombre $1-a$ représente le taux d'erreur maximal que l'on accepte en prédisant que I_n contient θ.

Une façon de construire des intervalles de confiance consiste à considérer un estimateur $\widehat{\theta}_n$ raisonnable de θ et à trouver sa loi sous chaque P_θ.

Si $P_\theta(\widehat{\theta}_n \in [\theta - s_n, \theta + t_n]) \geqslant a$ pour tout θ dans Θ, alors $I_n = \left[\widehat{\theta}_n - t_n, \widehat{\theta}_n + s_n\right]$ est un intervalle de confiance pour θ au niveau a.

Enfin le meilleur intervalle de confiance est celui dont la longueur est la plus petite.

Définition 2.2. *Soit $(I_n)_{n\geqslant 1}$ une suite d'intervalles de confiance, donc chaque I_n ne dépend que de l'échantillon $X_1,\dots,X_n$. Le niveau de confiance asymptotique de la suite $(I_n)_n$ vaut a si, pour tout θ dans Θ,*

$$\lim_{n\to+\infty} P_\theta(\theta \in I_n) = a.$$

2.1 Estimation de la moyenne avec une variance connue

Soit $(X_k)_{1\leqslant k\leqslant n}$ un échantillon indépendant et de même loi P_θ de carré intégrable pour une certaine valeur θ dans Θ. On pose

$$m := E_\theta(X_1), \qquad v := \mathrm{var}_\theta(X_1).$$

La moyenne empirique $\overline{X}_n$ est un estimateur sans biais et consistant de m. De plus, d'après le théorème central limite, pour tout x positif,

$$P_\theta\left(\sqrt{n}\,|\overline{X}_n - m| \leqslant x\sqrt{v}\right) \underset{n\to+\infty}{\longrightarrow} P\left(|Z| \leqslant x\right),$$

où Z désigne une variable aléatoire de loi gaussienne centrée réduite $\mathcal{N}(0,1)$. En utilisant la table de la loi normale, on choisit x_a tel que $P\left(|Z| \leqslant x_a\right) = a$. Par conséquent,

$$I_n := \left[\overline{X}_n - x_a\sqrt{v/n}, \overline{X}_n + x_a\sqrt{v/n}\right]$$

est un intervalle de confiance asymptotique de m au niveau a.

Par exemple, pour $a = 95\%$, $x_a = 1,96$ convient.

2.2 Estimation de la moyenne avec une variance inconnue

Par exemple, soit $(X_k)_k$ un échantillon indépendant et de même loi exponentielle $\mathcal{E}(1/\theta)$. Donc $m = \theta$ et $v = \theta^2$ et on cherche à estimer θ.

D'après le théorème de la limite centrale, en gardant les notations de la section précédente, pour n assez grand,

$$P_\theta\left(\sqrt{n}|\overline{X}_n - \theta| \leqslant x_a\theta\right) \approx a.$$

donc un intervalle de confiance au niveau asymptotique a, défini pour tout $n > x_a^2$, est

$$I_n := \left[\frac{\overline{X}_n}{1 + x_a/\sqrt{n}}, \frac{\overline{X}_n}{1 - x_a/\sqrt{n}}\right].$$

De manière générale, lorsque l'on ne peut pas faire autrement, on estime v par V_n et on utilise le fait que $\left(\overline{X}_n - m\right)\sqrt{n/V_n}$ suit asymptotiquement une loi gaussienne centrée réduite $\mathcal{N}(0, 1)$.

3 Tests

3.1 Principe général

On s'intéresse à la répartition des enfants nouveaux-nés en garçons et filles. On dispose des résultats d'un sondage, selon lequel sur 429440 naissances, on a dénombré 221023 filles. On se demande si cette répartition entre filles et garçons est compatible avec l'hypothèse d'équiprobabilité de naissance des garçons et des filles.

On dispose donc d'un ensemble Θ de paramètres, d'une valeur particulière θ_0, et d'un échantillon $(X_k)_{1\leqslant k\leqslant n}$ de loi P_θ pour un paramètre θ inconnu. On veut pouvoir résoudre l'alternative « H_0 contre H_1 », avec

$$H_0 : \theta = \theta_0; \qquad H_1 : \theta \neq \theta_0.$$

L'hypothèse H_0 est appelée l'hypothèse nulle, et H_1 l'hypothèse alternative.

Dans la situation ci-dessus l'échantillon $(X_k)_{1\leqslant k\leqslant n}$ est indépendant et suit la loi de Bernoulli $b(\theta)$, et $\theta_0 = 50\%$.

Pour tester H_0 contre H_1, on définit une zone de rejet R ne dépendant que de $(X_k)_{1\leqslant k\leqslant n}$ et on adopte la stratégie suivante :

Si R est réalisée, on rejette l'hypothèse H_0 et on accepte $H_1 : \theta \neq \theta_0$.
Si R n'est pas réalisée, on accepte l'hypothèse $H_0 : \theta = \theta_0$.

Définition 3.1. *Le niveau de risque de première espèce de R dans le test de H_0 contre H_1 est $\alpha := \sup\{P_\theta(R)\,;|\theta \in H_0\}$.*

Le niveau de risque de première espèce mesure donc le risque de rejeter l'hypothèse H_0 alors qu'elle est réalisée. En général, on fixe un niveau de risque de l'ordre de $\alpha = 5\%$ et on calcule une zone de rejet R adaptée.

Quand $H_0 : \theta = \theta_0$, on obtient $\alpha := P_{\theta_0}(R)$.

On peut utiliser des intervalles de confiance pour calculer des zones de rejet. Supposons par exemple que I est un intervalle de confiance pour θ au niveau de confiance $1 - \alpha$, c'est-à-dire que, pour tout θ dans Θ,

$$P_\theta(\theta \in I) \geqslant 1 - \alpha.$$

Alors $R = \{\theta_0 \notin I\}$ est une zone de rejet pour le test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau de risque α.

L'autre erreur possible associée à une zone de rejet R donnée consiste à accepter H_0 alors que H_0 est fausse.

Définition 3.2. *Le niveau de risque de seconde espèce de R dans le test de H_0 contre H_1 est $\beta := \max\{P_\theta(R)\,;|\theta \in H_1\}$.*

La première erreur α étant en général fixée, une bonne zone de rejet minimise l'erreur de seconde espèce β.

3.2 Estimation du paramètre d'une loi binomiale

L'échantillon $(X_n)_{n\geqslant 1}$ suit la loi de Bernoulli de paramètre p et le paramètre p est inconnu. On fixe p_0 et veut tester $H_0 : p > p_0$ contre $H_1 : p \leqslant p_0$.

Un test de H_0 contre H_1 au niveau de risque α est associé à une zone de rejet R telle que $P_p(R) \leqslant \alpha$ pour tout $p > p_0$. On rappelle que $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=k}^{n} X_k$ et on cherche R de la forme $R = \{\overline{X}_n < x\}$, alors

$$\alpha = \sup\{P_p(R) \,|\, p > p_0\} = P_{p_0}(R).$$

Remarque 3.3 (À omettre en première lecture)**.** *Pour évaluer $P_{p_0}(R)$, on peut procéder comme suit. Pour tout $x < p_0$ et tout s dans $]0,1]$, l'inégalité de Crámer donne*

$$P_{p_0}(\overline{X}_n < x) \leqslant s^{nx} E_{p_0}(s^{-X_1})^n,$$

Comme $E_{p_0}(s^{-X_1}) = p_0 s^{-1} + 1 - p_0$, on peut calculer le minimum en x du membre de droite, ce qui donne après calculs,

$$P_{p_0}(\overline{X}_n < x) \leqslant \left(\frac{1-p_0}{1-x}\right)^{n(1-x)} \left(\frac{p_0}{x}\right)^{nx}.$$

On vérifie que le membre de droite est une fonction croissante de x sur l'intervalle $[0, p_0]$ et vaut 1 en $x = p_0$, donc est bien inférieur à 1 pour tout $x < p_0$ Il reste à choisir x de sorte que la valeur du membre de droite soit au plus α, ce qui est possible pour tout α supérieur à la valeur en $x = 0$, soit $\alpha \geqslant (1-p_0)^n$.

Rappelons que $\overline{X}_n$ est un estimateur consistant et sans biais de p.

3.3 Estimation de la moyenne d'une loi normale

L'échantillon $(X_n)_{n\geqslant 1}$ suit la loi normale $\mathcal{N}(m, v)$ avec v connue et m inconnue.

Rappelons que $\overline{X}_n$ est un estimateur consistant et sans biais de m et que $\overline{X}_n = m + Z_n\sqrt{v/n}$ où Z_n est asymptotiquement gaussienne centrée réduite.

Première situation On veut tester $H_0 : m > m_0$ contre $H_1 : m \leqslant m_0$.

Il s'agit d'un test unilatère. On cherche une zone de rejet $R := \{\overline{X}_n < x\}$ avec $x < m_0$. Alors,

$$\alpha = \sup\{P_m(R) \,|\, m > m_0\} = P_{m_0}\left(Z < -(m_0 - x)\sqrt{n/v}\right),$$

et il reste à consulter la table de la loi gaussienne pour choisir x. Par exemple, $\alpha = 5\%$ donne

$$x = m_0 - 1,96\sqrt{v/n}.$$

Deuxième situation On veut tester $H_0 : m \in [m_1, m_2]$ contre $H_1 : m \notin [m_1, m_2]$.

Il s'agit d'un test bilatère. On cherche une zone de rejet $R := \{\overline{X}_n \notin [x_1, x_2]\}$ avec $x_1 < m_1$ et $x_2 > m_2$. Si on suppose que $x_1 = m_1 - c$ et $x_2 = m_2 + c$ avec c positif, on montre qu'il s'agit de choisir c tel que

$$\alpha = P(Z \geqslant c\sqrt{n/v}) + P(Z \geqslant (c + m_2 - m_1)\sqrt{n/v}).$$

Les cas limites sont $m_2 = +\infty$ pour lequel on retrouve le test unilatère et $m_2 = m_1$ qui donne le test de $H_0 : m = m_1$ contre $H_1 : m \neq m_1$.

4 Exercices

1. Pièces On lance une pièce équilibrée et on souhaite obtenir une proportion de « piles » entre 49% et 51% avec une probabilité au moins égale à 96%. Déterminer le nombre de jets nécessaire en utilisant l'approximation par une loi normale.

2. Repas Un restaurateur peut servir 75 repas, uniquement sur réservation. En pratique, 20% des clients ayant réservé ne viennent pas. Le restaurateur souhaite pouvoir servir tous les clients qui se présentent avec une probabilité supérieure ou égale à 90%. Déterminer le nombre maximal de réservations que le restaurateur peut accepter.

3. Défauts Une entreprise reçoit un lot important de pièces fabriquées en série. L'entreprise n'accepte la livraison que si la proportion p de pièces défectueuses est inférieure à 5%. Dans un échantillon de 200 pièces, on observe que 15 pièces sont défectueuses.
Décrire la conclusion d'un test de $H_0 : p \leqslant 5\%$ contre $H_1 : p > 5\%$ au niveau 1% relatif à une région de rejet $R = \{X \geqslant x\}$, où X désigne le nombre de pièces défectueuses dans l'échantillon.

4. Médicament L'écart type de la teneur d'un composant dans un médicament est de 3 milligrammes. Un nouveau procédé de fabrication vise à diminuer cet écart type. Dans un échantillon de 10 unités fabriquées par le nouveau procédé, on obtient en milligrammes :

$$725, 722, 727, 718, 723, 731, 719, 724, 726, 726.$$

On supppose l'échantillon de loi $\mathcal{N}(m, v)$.
1) On suppose que m est connue et vaut 724. Donner un intervalle de confiance de niveau 95% pour la variance.
2) Dans cette question m est inconnue, tester au niveau 5% si le but recherché est atteint.

5. Électricité Dans une fabrique de compteurs électriques, on vérifie le réglage des compteurs sur un échantillon de 10 compteurs. Lors d'une mesure de 100 unités, les compteurs de l'échantillon enregistrent :

$$983, 1002, 998, 996, 1002, 983, 994, 991, 1005, 986.$$

On supppose l'échantillon de loi $\mathcal{N}(m, \sigma^2)$.

Donner un intervalle de confiance au niveau 95% pour la moyenne.

Tester l'hypothèse $m = 1000$ contre $m \neq 1000$ au niveau 5%.

6. Confiance Échantillon de loi de densité $h_\theta(x) = (2x/\theta^2)\mathbf{1}_{0\leqslant x\leqslant\theta}$. On suppose que n est grand et $0 < \theta \leqslant 2$. Donner un intervalle de confiance pour θ au niveau de confiance 95% basé sur $\overline{X}_n$.

7. Risque Échantillon de loi de densité $f_\theta(x) = (2\theta x - \theta + 1)\mathbf{1}_{0\leqslant x\leqslant 1}$, où $-1 \leqslant \theta \leqslant 1$ est un paramètre que l'on se propose d'estimer.

Trouver a et b tels que l'estimateur $T_n = a\overline{X}_n + b$ est sans biais, pour tout θ. Calculer le risque quadratique de T_n.

Calculer la limite de $P_\theta(\sqrt{n}|T_n - \theta| \leqslant x)$ quand $n \to +\infty$, pour x réel positif. En déduire un intervalle de confiance de niveau 99% pour θ.

5 Exercices supplémentaires

1. Un échantillon de 478 électeurs choisis aléatoirement indique que 255 d'entre eux vont voter pour A. Évaluer des intervalles de confiance à 1% et à 5% pour la proportion d'électeurs de A.

2. On effectue un contrôle de fabrication sur des pièces dont une proportion p est défectueuse. On contrôle un lot de 200 pièces et on trouve 20 pièces défectueuses. Donner des intervalles de confiance pour l'estimation de p, au niveau 95% puis au niveau 99%.

3. Des appareils électriques de chauffage ont une moyenne de vie de fonctionnement de 20 000 heures avec un écart-type de 7 000 heures. À l'aide d'un changement de composant, le fabricant affirme que la durée de vie moyenne peut être accrue.

On a testé un échantillon de 127 appareils et on a observé une durée de vie moyenne de 21 000 heures. Préciser si on peut soutenir cette affirmation au risque de 5%, au risque de 1%.

4. Une pièce jetée 660 fois tombe 312 fois sur pile, préciser si on doit penser que cette pièce est bien équilibrée, ou non.

5. Le fabricant d'une nouvelle solution anti-rouille annonce que son produit est efficace à 90%. Dans un échantillon de 500 pièces le résultat est probant pour 420 d'entre elles. Préciser si l'affirmation du fabricant est légitime.

6. Deux machines A et B fabriquent en série la même pièce. Lors d'une expertise de la production, on remarque que la machine A a produit 2700 pièces dont 50 sont défectueuses alors que sur les 1600 pièces produites par la machine B, 35 sont défectueuses. Préciser si on doit en conclure que la machine A est mieux réglée que la machine B, ou non.

Capes de Mathématiques | **Université Joseph Fourier**
Préparation à l'écrit | **Année 2008-2009**

Un problème de Capes blanc

Préambule

Dans tout le sujet, n désigne un entier fixé $\geqslant 2$.

On utilise les notations et observations suivantes.

- $[[n]]$ désigne l'ensemble $\{1, 2, \ldots, n\}$.
- M désigne l'espace vectoriel des matrices carrées d'ordre n à coefficients réels.
- $\mathrm{Sp}(A)$ désigne l'ensemble des valeurs propres (réelles ou complexes) de A dans M.
- C désigne l'espace vectoriel des matrices unicolonnes à n lignes à coefficients réels.
- L désigne l'espace vectoriel des matrices unilignes à n colonnes à coefficients réels.
- I désigne la matrice identité d'ordre n. Les colonnes de I, notées $(E_i)_{i\in[[n]]}$, forment la base canonique de C et ses lignes $(E_i^t)_{i\in[[n]]}$ forment la base canonique de L.
- Pour A dans M, φ_A désigne l'application $\varphi_A : C \to C$, $X \mapsto AX$ qui est linéaire de matrice A par rapport à la base canonique de C.
- Pour A dans M, ψ_A désigne l'application $\psi_M : L \to L$, $X \mapsto XA$ qui est linéaire de matrice A^t par rapport à la base canonique de L.
- U désigne le vecteur colonne $U = (1, \ldots, 1)^t$, c'est-à-dire l'élément de C dont toutes les composantes sont égales à 1.
- Pour toute matrice $A = (a_{i,j})$ (carrée ou rectangulaire), $A \geqslant 0$ désigne le fait que $a_{i,j} \geqslant 0$ pour tous i et j, et $\|A\|$ désigne le nombre $\max |a_{i,j}|$, qui est le maximum de la valeur absolue des coefficients de A.
- Un vecteur ligne $V = (v_1, \ldots, v_n)$ dans L est dit stochastique si $\displaystyle\sum_{i=1}^{n} v_i = 1$ et si $v_i \geqslant 0$ pour tout i dans $[[n]]$.
- Les espaces vectoriels M, C et L sont considérés comme des espaces normés pour la norme $\| \cdot \|$.

- On note $M^+ = \{A \in M \mid A \geqslant 0\}$, $C^+ = \{X \in C \mid X \geqslant 0\}$ et $L^+ = \{X \in L \mid X \geqslant 0\}$.

Nota La partie IV et le début de la partie V sont indépendants de la partie III.

I Matrices stochastiques

Soit S l'ensemble des matrices $A = (a_{i,j})$ de M vérifiant les conditions suivantes :

(1) $A \geqslant 0$.

(2) Pour tout i dans $[[n]]$, $\displaystyle\sum_{j=1}^{n} a_{i,j} = 1$.

On appelle matrices stochastiques les éléments de S.

1. Soit A dans M. On considère les propriétés suivantes :

(1') $\varphi_A(C^+) \subset C^+$. (1") $\psi_A(L^+) \subset L^+$. (2') $\varphi_A(U) = AU = U$.

Montrer les équivalences : (1) si et seulement si (1') si et seulement si (1").

Montrer l'équivalence : (2) si et seulement si (2').

2. Montrer que pour toutes matrices A et B de S et tout réel t dans $[0, 1]$, la matrice $tA + (1-t)B$ est dans S (c'est-à-dire que S est une partie convexe de M).

3. Montrer que pour toute suite $(A_k)_{k\in\mathbb{N}}$ de matrices de S convergeant vers B dans M, la matrice B est dans S (c'est-à-dire que S est une partie fermée de M). Montrer que la partie S est compacte.

4. Soit $V = (v_1, \ldots, v_n)$ un vecteur ligne stochastique. Montrer que pour toute matrice A de S, le vecteur ligne VA est stochastique.

5. Montrer que S est une partie de M, stable pour la multiplication matricielle.

6.a. Soit A dans S telle que $AA^t = I$. Montrer que chaque ligne et chaque colonne de A a tous ses coefficients nuls sauf un qui vaut 1. Préciser l'action de φ_A sur la base canonique (E_i) de C.

6.b. Que peut-on dire de $G = \{A \in S \mid AA^t = I\}$?

II Éléments propres des matrices stochastiques

Soit $A = (a_{i,j})$ une matrice de S.

1. Montrer que 1 est valeur propre de A.

2. Montrer que $\mathrm{Sp}(A)$ est inclus dans $\{z \in \mathbb{C} \mid |z| \leqslant 1\}$.

[Pour traiter cette question, il pourra être utile de considérer pour λ dans $\mathrm{Sp}(A)$, un vecteur colonne $Y = (y_1, \dots, y_n)^t \neq 0$ tel que $AY = \lambda Y$ et $\mu = \max\{|y_i|, i \in [[n]]\}$.]

3. Soit λ un élément de $\mathrm{Sp}(A)$. Justifier le fait qu'il existe un élément p de $[[n]]$ tel que $|\lambda - a_{p,p}| \leqslant 1 - a_{p,p}$.

4. On suppose ici que $a_{i,i} > 1/2$ pour tout i dans $[[n]]$. Montrer que A est inversible.

5. On suppose ici que $a_{i,i} > 0$ pour tout i dans $[[n]]$.

5.a. Montrer qu'il existe un réel α dans $]0, 1]$ tel que $\mathrm{Sp}(A)$ est inclus dans l'ensemble $D_\alpha = \{z \in \mathbb{C} \mid |z - \alpha| \leqslant 1 - \alpha\}$.

Interpréter géométriquement le résultat précédent et faire la figure correspondante.

5.b. Que peut-on dire du module des valeurs propres de A différentes de 1 ?

6. On suppose ici que $a_{i,j} > 0$ pour tous i et j dans $[[n]]$. Montrer que les valeurs propres (complexes) de A autres que 1, sont de module strictement inférieur à 1, et préciser le rang de la matrice $A - I$.

7. Soit λ dans $\mathrm{Sp}(A)$ avec $|\lambda| = 1$.

7.a. Soit Y un vecteur colonne $Y = (y_1, \dots, y_n)^t \neq 0$ tel que $AY = \lambda Y$. On note

$$\mu = \max\{|y_i| \mid i \in [[n]]\}, \qquad K = \{k \in [[n]] \mid |y_k| = \mu\}.$$

Construire une application f de K dans K telle que pour tout $k \in K$, $y_{f(k)} = \lambda y_k$.

7.b. En déduire qu'il existe un entier p compris entre 1 et n tel que $\lambda^p = 1$.

III Convergence

Soit A dans S. On s'intéresse dans cette partie à la convergence éventuelle de la suite $(A^k)_{k\in\mathbb{N}}$. Pour tout k dans $\mathbb{N}$, on note $a_{i,j}^{(k)}$ le coefficient d'indice (i, j) de A^k pour tous i et j dans $[[n]]$.

1. La suite $(A^k)_{k\in\mathbb{N}}$ peut-elle converger si A possède une valeur propre λ de module 1 et différente de 1 ?

2. On suppose que la suite $(A^k)_{k\in\mathbb{N}}$ converge vers B dans M.

2.a. Montrer que B appartient à S et $B^2 = B$.

2.b. Montrer qu'on a aussi $BA = AB = B$. Ces égalités traduisent des propriétés remarquables des colonnes et des lignes de B, lesquelles ?

2.c. Lorsque $n = 2$, donner toutes les possibilités de matrices B.

3. On suppose ici que A est diagonalisable dans l'espace des matrices à coefficients complexes et qu'elle ne possède pas de valeur propre de module 1 autre que 1. Montrer que la suite $(A^k)_{k\in\mathbb{N}}$ converge.

4. Soit $\varepsilon = \min\{a_{i,j} \mid i, j \in [[n]]\}$. On suppose ici que $\varepsilon > 0$. Pour tout k dans $\mathbb{N}$, et tout j dans $[[n]]$, on note :

$$\alpha_j^{(k)} = \min\{a_{i,j}^{(k)} \mid i \in [[n]]\}, \quad \beta_j^{(k)} = \max\{a_{i,j}^{(k)} \mid i \in [[n]]\}, \quad \delta_j^{(k)} = \beta_j^{(k)} - \alpha_j^{(k)}.$$

4.a. Montrer que pour tout k dans $\mathbb{N}$, et tout j dans $[[n]]$, on a :

$$\alpha_j^{(k)} \leqslant \alpha_j^{(k+1)} \leqslant \beta_j^{(k+1)} \leqslant \beta_j^{(k)}, \qquad \delta_j^{(k+1)} \leqslant (1-2\varepsilon)\delta_j^{(k)}.$$

4.b. En déduire que $(A^k)_{k\in\mathbb{N}}$ converge. Si $B = \lim_{k\to\infty} A^k$, comparer les lignes de B.

5. Dans chacun des cas suivants, préciser, en faisant le minimum possible de calculs, si la suite $(A^k)_{k\in\mathbb{N}}$ converge ou non :

$$(1)\ A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix} \ ; \qquad (2)\ A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \ ;$$

$$(3)\ A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \ ; \qquad (4)\ A = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 1/3 & 1/6 & 1/2 \end{pmatrix}.$$

IV Chaînes de Markov

Soit une suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ définies sur un même espace de probabilité $(\Omega, \mathcal{T}, P)$ à valeurs dans $[[n]]$. On dit que la suite $(X_k)_{k\in\mathbb{N}}$ est une *chaîne de Markov* (à n états) si les propriétés suivantes sont réalisées :

(CM1) Il existe une matrice $T = (p_{i,j})$ dans M, appelée *matrice de transition* telle que pour tout entier $k \geqslant 1$ et tous i et j dans $[[n]]$, on ait

$$P(X_k = j \mid X_{k-1} = i) = p_{i,j}.$$

Autrement dit, l'état X_k dépend de l'état X_{k-1} de manière invariable.

(CM2) Pour tout entier $k \geqslant 2$ et tous i, j, i_{k-2}, …, i_0 dans $[[n]]$, on a

$$P(X_k = j \mid X_{k-1} = i, X_{k-2} = i_{k-2}, \dots, X_0 = i_0) = P(X_k = j \mid X_{k-1} = i).$$

Autrement dit, l'état X_k ne dépend que de l'état X_{k-1}.

On considère une telle chaîne de Markov $(X_k)_{k\in\mathbb{N}}$. On note $T = (p_{i,j})$ dans M sa matrice de transition. Pour tout entier k, on note $P(X_k = i) = q_i^{(k)}$ pour tout i dans $[[n]]$ et Q_k le vecteur ligne $Q_k = (q_1^{(k)}, \ldots, q_n^{(k)})$ (la loi de X_k). La loi de X_0 donnée par Q_0 que l'on notera simplement $Q_0 = (q_1, \ldots, q_n)$ est appelée *la loi initiale* de la chaîne.

1. Montrer que T est dans S (T est stochastique) et que pour tout $k \geqslant 0$, Q_k est un vecteur ligne stochastique, c'est-à-dire $\displaystyle\sum_{i=1}^{n} q_i^{(k)} = 1$.

2. Pour tout entier $k \geqslant 0$, montrer que $Q_{k+1} = Q_k T$. En déduire l'expression de la loi de X_k donnée par Q_k à l'aide de T et de la loi initiale donnée par Q_0.

3. Montrer qu'il existe une loi initiale telle que $Q_0 = Q_0 T$. Une telle loi est dite *stationnaire.*

[On sera amené à montrer que si un vecteur $V = (v_1, \ldots, v_n)$ vérifie $VT = V$ alors $V' = (|v_1|, \ldots, |v_n|)$ vérifie aussi $V'T = V'$.]

4. On suppose que la loi initiale est stationnaire, c'est-à-dire $Q_0 = Q_0 T$. Montrer qu'il existe une matrice $(q_{i,j})$ de S telle que $q_{i,j} = P(X_k = j \mid X_{k+1} = i)$ pour tout k entier.

V Un exemple de chaîne de Markov

Une urne contient initialement deux boules rouges et deux boules bleues. On décide de faire une succession de tirages avec la règle suivante : pour tout $k \geqslant 0$, la boule tirée au tirage k est laissée de côté au tirage $k+1$ et n'est remise dans l'urne que pour effectuer le tirage $k+2$. On considère la suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ où X_k est le nombre de boules rouges dans l'urne après le tirage k. Tous les tirages sont équiprobables sauf le tirage initial de numéro $k = 0$, pour lequel on décide que $P(X_0 = 1) = p$ pour un certain p dans $]0, 1[$ fixé.

0. Déterminer la loi de X_1, sa moyenne $E(X_1)$ et sa variance $\mathrm{var}(X_1)$.

1. Justifier que la suite $(X_k)_{k\in\mathbb{N}}$ est une chaîne de Markov dont on précisera la matrice de transition T et la loi initiale.

2. Montrer qu'il y a une seule loi stationnaire que l'on précisera.

3. En s'aidant des parties I à III, montrer que la suite des lois des X_k converge vers la loi stationnaire.

On note ici $R_k = [X_k = 1]$ l'événement "Au tirage k, une boule rouge a été tirée". On considère la variable aléatoire N égale au numéro d'ordre du premier tirage d'une boule rouge (autrement dit N est égal au plus petit entier k tel que R_k est réalisé).

4. Exprimer pour $k \geqslant 0$ l'événement $[N = k]$ à l'aide de R_i pour $i \leqslant k$.

5. En déduire la loi de N, sa moyenne $E(N)$ et sa variance $\mathrm{var}(N)$.