Université de Paris-Sud, centre d'Orsay

PROBABILITES ET STATISTIQUES

Cours de licence de Mathématiques
Version 2008

Y. Le Jan
S. Lemaire

Table des Matières

Partie 1

Modélisations des phénomènes aléatoires

(théorie élémentaire des probabilités)

Chapitre 1

Notions fondamentales

1.1 Enjeu et formalisation

Comme souvent en mathématique, la théorie probabiliste part d'une intuition concrète, au départ un peu vague, celle du hasard, ou de l'incertitude. La démarche consiste, comme toujours, à dégager quelques "évidences" fondamentales puis, pas à pas, en suivant une logique rigoureuse, à construire un appareil capable de réduire cette intuition à un calcul. L'enjeu est bien sûr très important puisqu'il s'agit par exemple de développer des méthodes permettant de chiffrer des risques, donc de les comparer, et de décider de la conduite à tenir, en fonction de leur chiffrage et de celui des coûts.

En tant que théorie mathématique, le Calcul des Probabilités ne fait appel, dans ses commencements, qu'à des notions assez simples de théorie des ensembles et de dénombrement. Cependant, l'application pratique de cette théorie simple requiert d'emblée la plus grande attention. Le Calcul des Probabilités vise à évaluer a priori, les chances (ou les risques) dans une expérience (souvent appelée *épreuve*) réalisée dans des conditions déterminées mais dont l'issue comporte un élément d'incertitude ou de hasard, dépendant de facteurs non contrôlés.

Le lancer d'un dé, de deux dés, le tirage au hasard d'un individu dans une population, d'un groupe d'individus distincts, l'observation de la température demain au réveil ou du prix du pétrole dans un mois sont des expériences aléatoires.

Mathématiquement, une telle épreuve est représentée par le *tirage* d'un élément ω dans un ensemble Ω représentant toutes les issues possibles, que nous supposerons fini ou dénombrable[1]. Le développement théorique du calcul des probabilités conduit ensuite à introduire des espaces continus (voir chapitre 7).

Chaque point ω de Ω est affecté d'une *probabilité* $P(\omega)$ qui représente la chance qu'il a d'être tiré (c'est-à-dire la chance qu'a l'épreuve d'avoir l'issue (le résultat) représentée par ω). En langage familier, dire que $P(\omega) = c/b$ où c et b sont des entiers, signifie que ω a c chances sur b d'être tiré. On doit donc supposer que *les $P(\omega)$ sont des nombres compris entre* 0 *et* 1*, dont la somme est égale à* 1. Si Ω est dénombrable, rappelons que $\sum_{\omega\in\Omega} P(\omega) = \sup_{E\in\mathcal{P}_f(\Omega)} \sum_{\omega\in E} P(\omega)$ où $\mathcal{P}_f(\Omega)$ désigne l'ensemble des parties finies de Ω. On vérifie que pour toute bijection σ de $\mathbb{N}$ sur Ω, $\sum_{\omega\in\Omega} P(\omega)$ est la somme de la série de terme général $P(\sigma(n))$. Si Ω est fini et si tous les points ont la même chance d'être tirés, (on dit alors que toutes les issues sont *équiprobables*),

[1] dans un premier temps seulement, bien qu'en pratique ce soit toujours le cas

leur probabilité sera $\frac{1}{|\Omega|}$[2]. L'ensemble Ω, muni de la probabilité P, est appelé *espace probabilisé*. Le *support de* P est constitué par l'ensemble des points ω de Ω tels que $P(\omega)$ soit strictement positif.

Un *événement* susceptible de se produire à l'issue du tirage est représenté par une partie de Ω formée de tous les points représentatifs des issues dans lesquels il se produit. Sa probabilité (la chance qu'il a de se produire) est la somme des probabilités des points qui la constituent.

Ainsi, l'événement qui consiste à obtenir un nombre pair lors du lancer d'un dé sera représenté par le sous-ensemble $\{2, 4, 6\}$ de $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si le dé n'est pas pipé, sa probabilité sera $3 \times \frac{1}{6} = \frac{1}{2}$.

Si l'expérience consiste à tirer un individu ω dans une population Ω, l'événement "ω a les yeux noirs" est représenté par l'ensemble de tous les individus aux yeux noirs. Si le tirage est équitable, la probabilité de cet événement est donnée par la proportion d'individus aux yeux noirs. De façon générale, le recensement d'une population permet de calculer la probabilité de tirer un individu présentant une caractéristique quelconque, pourvu que le tirage soit équitable et que la caractéristique en question soit prise en compte lors du recensement.

1.2 Principes fondamentaux

Le premier principe fondamental du calcul des probabilités, du point de vue de la pratique, est qu'un événement de faible probabilité ne se produit *presque* jamais. La théorie, si elle est pertinente, doit ainsi permettre une prise de risque raisonnable. Bien sûr, s'il peut paraître raisonnable de risquer gros dans une expérience où les chances de succès sont de 99,9% , il ne l'est pas de le faire très souvent car on verra que la probabilité de s'en tirer sans dommage après mille essais n'est plus que de 31% (si le succès d'une expérience n'influence pas celui d'une autre).

Le deuxième principe fondamental concerne la réalisation de plusieurs épreuves, que nous supposerons représentées par des tirages dans des ensembles $\Omega_1, \Omega_2, \ldots, \Omega_n$. Soient P_1, P_2,...,P_n les probabilités associées à chacun des tirages. La réalisation de n épreuves consécutives, qui est aussi une épreuve, est représentée par le tirage d'un multiplet $(\omega_1, ..., \omega_n)$ dans l'ensemble produit $\Omega_1 \times \ldots \times \Omega_n$. Le principe est que *lorsque ces épreuves sont sans influence mutuelle* (on dit qu'elles sont indépendantes), la probabilité de tirer $(\omega_1, \omega_2, ..., \omega_n)$ est donnée par le produit des probabilités $P_1(\omega_1)P_2(\omega_2)...P_n(\omega_n)$.

Ce principe est conforme à l'intuition : on considère bien avoir une chance sur 12 de tirer six et pile lors du lancer d'un dé et d'une pièce, ou avoir une chance sur 2^n de gagner n fois de suite à pile ou face (quels que soient d'ailleurs les n paris effectués). Remarquons enfin que l'on a bien défini ainsi une probabilité sur l'ensemble produit $\Omega_1 \times \Omega_2 \times ... \times \Omega_n$ puisque

$$\sum_{(\omega_1,\omega_2,...,\omega_n)\in\Omega_1\times...\times\Omega_n} P_1(\omega_1)P_2(\omega_2)...P_n(\omega_n) = 1.$$

La probabilité ainsi définie est appelée la *probabilité produit*. Elle est notée

$$P_1 \otimes P_2 \otimes ... \otimes P_n = \bigotimes_{i=1}^{n} P_i.$$

[2] $|A|$ désigne le nombre d'éléments de l'ensemble A.

1.3 Probabilités et fréquences

En pratique, les probabilités $P(\omega)$ sont inconnues ou données par des hypothèses qu'on ne peut jamais vérifier parfaitement. La vraisemblance de telles hypothèses ne peut être vérifiée que par une *expérience statistique*. Il s'agit de recommencer n fois l'épreuve, *dans des conditions identiques*, et de telle manière que les facteurs aléatoires présents lors des différentes épreuves soient *indépendantes* (c'est-à-dire sans influence mutuelle). Mathématiquement si nous notons (Λ, Q) l'épreuve en question, il s'agit de tirer un n-uplet $\omega = (\omega_1, \omega_2, ..., \omega_n)$ dans l'ensemble $\Omega = \Lambda^n$ formé de tous les n-uplets d'éléments de Λ. Les n-uplets peuvent aussi être vus comme des mots de longueur n écrits avec des lettres choisies dans Λ.

Le deuxième principe fondateur donne que *dans ces conditions*, la probabilité de tirer un n-uplet $\omega = (\omega_1, ..., \omega_i, ..., \omega_n)$ est donnée par

$$P(\omega) = Q^{\otimes n}(\omega_1, ..., \omega_i, ..., \omega_n) = Q(\omega_1)...Q(\omega_i)...Q(\omega_n).$$

On peut alors en déduire par le calcul que si le nombre de tirages est grand, la *fréquence* $\frac{1}{n}|\{i \leq n, \ \text{tels que } \omega_i = \omega\}|$ avec laquelle ω a été tiré a de fortes chances d'être proche de $P(\omega)$. De ce fait, la fréquence

$$\frac{1}{n}|\{i \leq n, \ \text{tels que } \omega_i \in A\}|$$

avec laquelle un événement A sera réalisé aura de fortes chances d'être proche de sa probabilité. Plus précisément, on montrera dans la section 3.2, le résultat suivant (loi faible des grands nombres) :

$$\forall\, \epsilon > 0, \ \lim_{n \to +\infty} P\left(\left|\frac{1}{n}|\{i \leq n, \ \text{tels que } \omega_i \in A\}| - P(A)\right| > \epsilon\right) = 0.$$

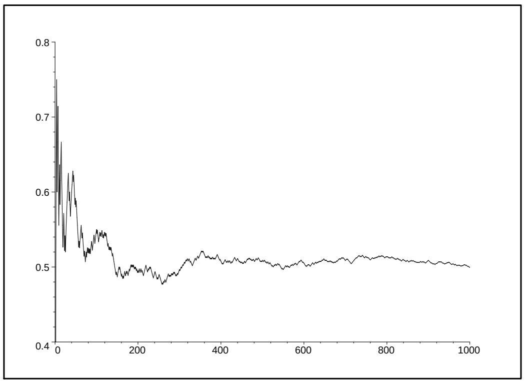


FIG. 1.1 : On a lancé 1000 fois une pièce. La courbe donne en fonction de n, la fréquence des lancers où la pièce est tombée sur “pile” au cours des n premiers lancers

N.B. Revenons sur l'exemple du tirage équitable d'un individu dans une population. Le sondage le plus simple sera constitué par n tirages équitables indépendants successifs d'un individu dans une population. Il s'agit d'un tirage avec remise : après chaque tirage, l'individu tiré doit être remis dans la population. Dans le cas contraire, il est clair que les tirages ne sont plus indépendants. Notons cependant que dans le cas d'une population de grande taille, comparativement à n, il y a de grandes chances pour que le tirage donne n individus différents, et qu'un tirage sans remise a en pratique les mêmes propriétés. En effectuant un sondage sur un assez grand nombre n d'individus, c'est-à-dire en effectuant n tirages sans influence mutuelle, la fréquence avec laquelle on aura tiré des individus aux yeux noirs aura de forte chances d'être proche de leur proportion dans l'ensemble de la population. Cependant, tant qu'il s'agira d'un sondage et non d'un recensement, il demeurera toujours un risque non nul d'erreur importante, et ce risque, parfaitement calculable, sera bien sûr d'autant plus grand que la taille de l'échantillon sera faible.

En fait, la validité des hypothèses sur lesquelles se fonde une expérience statistique est difficile à établir a priori pour un protocole expérimental donné. Leur justification repose surtout sur leur efficacité a posteriori. L'hypothèse d'indépendance des tirages est limitée à certains types d'expérience, dans lesquels les facteurs incontrôlés qui sont à l'origine des aléas fluctuent suffisamment entre deux épreuves consécutives. Par exemple, on sait, par l'intuition et par une longue pratique que les lancers de dés (aléa en latin) animés d'un mouvement chaotique par l'agitation d'un gobelet ou les observations (à des intervalles de temps suffisamment longs) de la hauteur des vagues ou des cours des marchés peuvent être considérés comme des expériences statistiques. Les ordinateurs et les calculatrices fournissent des nombres au hasard (touche random) et des tirages successifs peuvent être considérés comme une expérience statistique[3]. Mais l'usage abusif des statistiques conduit à des erreurs : les fréquences observées dans le passé pour certains phénomènes (les inondations par exemple) n'auront pas grand intérêt pour la gestion des risques si les conditions ont changé (asphaltage, déboisement) ou si une nouvelle ère climatique se dessine.

▷ *Exercice 1.* Décrire quelques précautions nécessaires à la réalisation d'un tirage au sort d'une carte dans un jeu de 52 cartes. Que signifie mathématiquement que le tirage est effectué au hasard ? Comment pourrait-on chercher à le vérifier ?

▷ *Exercice 2.* Comment se pose le problème pour le tirage d'un individu dans une classe, dans une grande ville ? Dans quelle mesure peut-on le ramener au tirage d'un nombre au hasard. De quelle façon ce tirage peut-il être effectué à l'aide d'un jeu de pile ou face ?

[3]Cette affirmation n'est valable que dans certaines limites, qui ne sont pas franchies dans la pratique courante.

Chapitre 2

Les événements

2.1 Relations logiques et représentation ensembliste

Rappelons qu'un événement A est représenté par l'ensemble des points représentatifs des issues où il se produit, qui est une partie de Ω. On emploie généralement la même notation pour un événement et son ensemble représentatif.

De ce fait, l'occurrence simultanée de deux événements A et B sera représentée par l'ensemble des points représentatifs des issues où A et B se produisent, c'est-à-dire par l'*intersection* des ensembles représentatifs de A et de B. Deux événements incompatibles sont représentés par des ensembles disjoints.

De même, la *réunion* des ensembles représentatifs de A et de B représente l'occurrence d'au moins un des événements A ou B. Ceci reste valable dans le cas de plusieurs événements A_1, A_2,..., A_m ou même dans le cas d'une suite infinie d'événements.

La non occurrence de A (qui constitue un événement au même titre que A) est représentée par le *complémentaire* de l'ensemble représentatif de A dans Ω, noté $\Omega \setminus A$ ou A^c s'il n'y a pas d'ambiguité.

La famille $\mathcal{A} = \mathcal{P}(\Omega)$ des événements est stable par ces trois opérations (par convention, l'ensemble vide $\emptyset$ représente l'événement impossible et a pour probabilité 0). On dit qu'elle constitue une σ-algèbre.

événements	représentation ensembliste
événement certain	Ω
événement impossible	$\emptyset$
l'événement A est réalisé	$\omega \in A$
l'événement A n'est pas réalisé	$\omega \in A^c = \Omega \setminus A$
les événements A et B sont réalisés	$\omega \in A \cap B$
A et B sont des événements incompatibles	$A \cap B = \emptyset$
l'événement A ou l'événement B est réalisé	$\omega \in A \cup B$
si l'événement A a lieu alors l'événement B a lieu	$A \subset B$

Tableau 2.1 : Quelques événements associés à une expérience aléatoire dont le résultat est ω et l'ensemble des résultats possibles Ω

► *Exemple 3.* Soit $n \in \mathbb{N}^*$. On tire à pile ou face n fois.

a) Décrire l'ensemble des résultats possibles comme un produit d'ensembles.

b) Ecrire de façon ensembliste l'événement $F =$"pile n'a pas été obtenu lors des 2 premiers lancers".

c) Soit $i \in \{1, \ldots, n\}$. Décrire les éléments de l'événement $E_i =$ "le résultat du i-ième lancer est pile".

d) Ecrire à l'aide des événements E_i l'événement F.

e) Ecrire à l'aide des événements E_i l'événement $G =$" la pièce est tombée au moins une fois sur pile".

Solution. a) Le résultat de chaque lancer est soit "pile", soit "face". Si on note x_i le résultat du i-ème lancer, le résultat de l'expérience "lancer n fois une pièce" peut être noté $(x_1, \ldots, x_n)$. Donc, l'ensemble des résultats possibles, notée Ω, est $\Omega = \{(x_1, \ldots, x_n),\ \forall i\ x_i \in \{pile, face\}\}$ c'est-à-dire l'ensemble produit $\{pile, face\}^n$.
Désignons par $(x_1, \ldots, x_n) \in \Omega$ le résultat des n lancers.

b) L'événement F est réalisé si et seulement si $x_1 = face$ et $x_2 = face$. Donc $F = \{face\} \times \{face\} \times \{pile, face\}^{n-2}$.

c) L'événement E_i est réalisé si et seulement si $x_i = pile$. Donc, $E_i = \{pile, face\}^{i-1} \times \{pile\} \times \{pile, face\}^{n-i}$.

d) F est réalisé si et seulement si E_1 et E_2 ne sont pas réalisés. Donc, $F = (\Omega \setminus E_1) \cap (\Omega \setminus E_2) = (E_1 \cup E_2)^c$.

e) L'événement G est réalisé si et seulement si au moins un des événements E_i est réalisé. Donc, $G = \cup_{i=1}^n E_i$.

On pourrait donner une représentation plus concise en remplaçant pile par 1 et face par 0.

▷ *Exercice 4.* On considère une expérience aléatoire dont les résultats possibles sont représentés par un ensemble Ω. On s'intéresse à trois événements A, B, C associés à cette expérience.
Exprimer en fonction de A, B, C et des opérations ensemblistes les événements ci-dessous et décrire leurs événements complémentaires :

a) A seul se produit.

b) A et C se produisent, mais non B.

c) les trois événements se produisent.

d) l'un au moins des événements se produit.

e) deux événements au moins se produisent.

f) un événement au plus se produit.

g) aucun des trois événements ne se produit.

h) exactement deux événements se produisent.

i) au plus deux événements se produisent.

▷ *Exercice 5.* Soit $A_1, \ldots, A_m$ et B des sous-ensembles de Ω. Démontrer que $(\cup_{i=1}^m A_i) \cap B = \cup_{i=1}^m (A_i \cap B)$ et que $\Omega \setminus (\cap_{i=1}^m A_i) = \cup_{i=1}^m (\Omega \setminus A_i)$.
Formuler ceci en termes d'événements. Donner un exemple concret.

▷ *Exercice 6.* Au poker, représenter par un diagramme les relations logiques entre les événements suivants, pour le tirage d'une main de cinq cartes : tirer une paire, tirer un brelan, tirer une suite (cinq cartes consécutives), tirer une couleur.

2.2 Probabilité d'un événement

En définissant la probabilité $P(A)$ d'un événement A comme la somme des probabilités $P(x)$ sur l'ensemble des éléments x de A, on obtient une application P de $\mathcal{A} = \mathcal{P}(\Omega)$ dans $[0,1]$ qui vérifie les propriétés suivantes :

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

(ii) Si $A_1, A_2, \ldots, A_m, \ldots$, est une suite d'événements deux à deux incompatibles, alors

$$P(\overset{+\infty}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{+\infty} P(A_i).$$

On peut citer deux relations utiles qui se déduisent de (i) et (ii).

(iii) $P(\Omega \setminus A) = 1 - P(A)$.

(iv) Si A et B sont deux événements alors $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Démonstration de (iii) et de (iv) : Ω est la réunion de deux ensembles disjoints A et $\Omega \setminus A$. Donc, $P(\Omega) = P(A) + P(\Omega \setminus A)$ d'après (ii). On obtient (iii) en utilisant que $P(\Omega) = 1$.
Pour obtenir (iv), décomposons l'ensemble $A \cup B$ en deux ensembles disjoints A et $B \setminus A = B \cap (\Omega \setminus A)$. On obtient

$$P(A \cup B) = P(A) + P(B \setminus A). \tag{2.1}$$

Pour obtenir $P(B \setminus A)$ on procède de la même façon que pour (iii) : B est la réunion des ensembles disjoints $B \cap A$ et $B \setminus A$. Donc

$$P(B) = P(B \cap A) + P(B \setminus A). \tag{2.2}$$

On conclut en utilisant l'égalité (2.2) dans l'égalité (2.1). □

▷ *Exercice 7.* Généraliser (iv) au cas de trois événements A, B, C.

▷ *Exercice 8.* Démontrer (ii).

► *Exemple 9.* On lance deux dés à six faces. Décrire l'ensemble Ω des résultats possibles et la probabilité P associée à cette expérience. Donner la probabilité d'obtenir

1. un double,
2. au plus un nombre pair,
3. exactement un nombre pair,
4. deux nombres qui se suivent.

Solution. On représente le résultat d'un lancer des deux dés comme un couple de deux entiers compris entre 1 et 6 (on peut considérer par exemple qu'on lance les dés l'un après l'autre et que l'on note dans l'ordre les résultats des deux lancers). L'ensemble des résultats possibles est $\Omega = \{1, \ldots, 6\}^2$. Chaque élément de Ω a autant de chance d'être le résultat d'un lancer. On munit donc Ω de l'équiprobabilité $P : P(\{\omega\}) = \frac{1}{36}$ pour tout $\omega \in \Omega$. La probabilité d'un événement A est alors $P(A) = \frac{|A|}{|\Omega|}$.
Soit A l'événement "obtenir un double": $A = \{(i,i),\ i \in \{1, \ldots, 6\}\}$. A a 6 éléments donc $P(A) = \frac{1}{6}$.
Soit B l'événement "obtenir au plus un nombre pair". B est le complémentaire de l'événement C "obtenir deux nombres pairs" : $C = \{2,4,6\}^2$ a 9 éléments et donc a pour probabilité $\frac{1}{4}$. Donc, B a pour probabilité $P(B) = 1 - P(C) = \frac{3}{4}$.

Soit D l'événement "obtenir exactement un nombre pair": $D = (\{1,3,5\} \times \{2,4,6\}) \cup (\{2,4,6\} \times \{1,3,5\})$. Donc, $P(C) = \frac{18}{36} = \frac{1}{2}$.
Soit E l'événement "obtenir deux nombres qui se suivent" : E est la réunion disjointe des ensembles $\{(i, i+1),\ i \in \{1, \ldots, 5\}\}$ et $\{(i+1, i),\ i \in \{i \in \{1, \ldots, 5\}\}$. Donc, $P(E) = \frac{10}{36} = \frac{5}{18}$.

▷ *Exercice 10.* On effectue deux expériences aléatoires successivement. L'ensemble des résultats possibles de la première épreuve est noté Ω_1 et l'ensemble des résultats possibles de la deuxième Ω_2. On suppose que ces ensembles sont au plus dénombrables. On note $P_1(x)$ la probabilité que x soit le résultat de la première expérience et $P_2(x)$ la probabilité que x soit le résultat de la seconde expérience. Le résultat de l'ensemble des deux expériences peut être noté comme un couple (x_1, x_2) où x_i désigne le résultat de la i-ème expérience ($i \in \{1,2\}$). Le produit cartésien

$$\Omega_1 \times \Omega_2 = \{(x_1, x_2), x_1 \in \Omega_1,\ x_2 \in \Omega_2\}$$

contient l'ensemble des résultats possibles pour les deux expériences.

1. Montrer que l'on définit bien une probabilité P sur $\Omega_1 \times \Omega_2$ en posant, pour tout $(x_1, x_2) \in \Omega_1 \times \Omega_2$, $P((x_1, x_2)) = P_1(x_1)P_2(x_2)$.
   On appelle P la probabilité produit de P_1 et de P_2. Quand paraît-il légitime d'utiliser P pour décrire l'ensemble des deux expériences ?
2. Soit $A_i \subset \Omega_i$ pour $i \in \{1,2\}$. On note C_i l'événement de $\Omega_1 \times \Omega_2$ suivant : " A_i est réalisé au cours de la i-ème expérience" ($i \in \{1,2\}$). Décrire les événements C_1, C_2 et $C_1 \cap C_2$ comme des sous-ensembles de $\Omega_1 \times \Omega_2$ et déterminer une relation entre leurs probabilités.
3. Généraliser le résultat de la question 1 à une suite de n expériences aléatoires. Comment ceci s'applique-t-il à la description d'une expérience statistique ?
4. Montrer que si chaque P_i est une équiprobabilité, il en est de même de P (autrement dit que dans une épreuve constituée de plusieurs tirages équitables et indépendants, toutes les issues ont la même probabilité).

▷ *Exercice 11.* Soit $n \in \mathbb{N}^*$ et $q \in [0,1]$. A chaque élément $\omega = (\omega_1, \ldots, \omega_n)$ de l'ensemble $\Omega = \{0,1\}^n$, on associe la probabilité $P(\{\omega\}) = q^{\alpha(\omega)}(1-q)^{n-\alpha(\omega)}$ où $\alpha(\omega) = \sum_{i=1}^n \omega_i$. Montrer que P définit bien une probabilité sur Ω. Décrire une expérience statistique représentée par (Ω, P).

▷ *Exercice 12.* Soit Ω un ensemble au plus dénombrable et P une probabilité sur cet ensemble. Soit f une application définie sur Ω et $\tilde{\Omega} = f(\Omega)$. Pour $x \in \tilde{\Omega}$, on pose $Q(x) = P(f^{-1}(x))$[1].
Montrer que Q est une probabilité sur $\tilde{\Omega}$. On appelle Q la probabilité image de P par f et elle est notée $f(P)$. Pour $A \in \mathcal{P}(\tilde{\Omega})$, on a $Q(A) = P(f^{-1}(A))$
Application: Prendre pour Ω la population mondiale et pour f la nationalité. Que représente Q ?

▷ *Exercice 13.*

1. Soit p et n deux entiers strictement positifs. On répartit au hasard p jetons numérotés de 1 à p sur un tableau constitué de n cases numérotés de 1 à n. Chaque jeton est placé sur une case et chaque case peut recevoir plusieurs jetons.
   (a) Décrire l'ensemble Ω des répartitions possibles des p jetons et la probabilité qu'une répartition donnée se réalise au cours d'une telle expérience aléatoire.
   (b) Soit $i \in \{1, \ldots, n\}$. Déterminer la probabilité que la i-ème case reste vide.
   (c) Déterminer la probabilité qu'au moins une case du tableau reste vide.
2. n individus sont montés à des instants différents dans un train composé de n wagons. Calculer la probabilité pour qu'ils ne se soient pas rencontrés lorsque tous les wagons ont la même fréquentation. Montrer qu'elle est plus faible dans le cas contraire.

▷ *Exercice 14.* On considère un système de r particules pouvant être dans un des n niveaux d'énergie $e_1, \ldots, e_n$. On définit l'espace des états du système comme l'ensemble des configurations distinguables. On fait l'hypothèse que chaque configuration distinguable est équiprobable. On considère les trois cas suivants :

[1] $f^{-1}(A)$ est l'image réciproque de l'ensemble A par f : $f^{-1}(A) = \{\omega \in \Omega, f(\omega) \in A\}$

(i) *Statistique de Maxwell-Boltzmann.* Les r particules sont localisées et donc distinguables.

(ii) *Statistique de Bose-Einstein.* Les r particules sont indistinguables.

(iii) *Statistique de Fermi-Dirac.* Les r particules sont indistinguables et il y a au plus une particule par niveau d'énergie (on suppose que $n \geq r$).

a) Pour chacun des modèles (i), (ii), (iii), décrire l'espace des états du système et la probabilité qu'un état donné se réalise.

b) L'état macroscopique du système est déterminé par les nombres r_i de particules dans l'état d'énergie e_i. On modélise l'espace des états macroscopiques par

$$\Omega_{ma} = \{(r_1, \ldots, r_n)\, \text{t.q.}\, r_1 + \cdots + r_n = r\}.$$

Munir Ω_{ma} de la probabilité naturelle dans chacun des cas (i), (ii), (iii). Que peut-on dire dans les cas (ii) et (iii) ?

c) On note $p_{r,n}(k)$ la probabilité qu'un niveau d'énergie donné contienne exactement k particules. Calculer $p_{r,n}(k)$ dans chacun des trois cas (i), (ii) et (iii).

2.3 Probabilité conditionnelle

Supposons qu'une expérience aléatoire ait été réalisée mais que l'on ignore son issue ; on sait seulement que l'événement B (supposé de probabilité non nulle) a eu lieu. On doit définir une nouvelle probabilité P_B qui tienne compte de cette information. D'une part, sachant que l'événement B est réalisé, on posera $P_B(\{x\}) = 0$ si $x \in \Omega \setminus B$. D'autre part, il n'y a pas de raison de changer le rapport entre les probabilités de deux élements de B ce qui amène à poser $P_B(\omega) = cP(\omega)$ pour tout $\omega \in B$, c étant une constante indépendante de ω. Pour que P_B soit une probabilité, c doit valoir $1/P(B)$. La probabilité P_B est maintenant entièrement définie. $P_B(\omega)$ est le plus souvent noté $P(\omega \mid B)$. On a obtenu la formule suivante :

$$P(\omega \mid B) = \frac{P(\omega)}{P(B)} \mathbb{1}_B(\omega)\ \forall \omega \in \Omega.$$

Notons également que d'après ce qui a été annoncé dans la première section, dans une expérience statistique assez longue, la fréquence relative $\frac{|\{i \text{ tels que } x_i = \omega\}|}{|\{i \text{ tels que } x_i \in B\}|}$ a de grande chance d'être proche de $P(\omega \mid B)$.

On vérifie que la probabilité d'un événement A est alors donnée par l'égalité suivante :

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

Elle est appelée la *probabilité conditionnelle de A sachant B*. Dans le cadre d'une expérience statistique, si n est grand, le rapport des fréquences $\frac{|\{i \text{ tels que } x_i \in A \cap B\}|}{|\{i \text{ tels que } x_i \in B\}|}$ a de fortes chances d'être proche de la probabilité conditionnelle $\frac{P(A \cap B)}{P(B)}$.

Cette probabilité conditionnelle représente la chance que l'on a d'avoir tiré un cas où A et B sont réalisés lorsque l'on sait déjà que B est réalisé. Lorsque tous les cas ont la même probabilité, elle représente bien la proportion $\frac{|A \cap B|}{|B|}$. Si l'expérience consiste à tirer au hasard un individu dans une population et si A et B sont formés d'individus présentant une caractéristique donnée, la probabilité conditionnelle $\frac{P(A \cap B)}{P(B)}$ est supérieure à $P(A)$ lorsque ces caractères sont fréquemment associés, ce qui peut faire penser à une relation de cause à effet dans un sens ou dans l'autre ou à une origine commune.

▷ *Exercice 15.* Imaginez des exemples susceptibles d'illustrer la phrase précédente.

► *Exemple 16.* Une personne lance deux dés et nous dit qu'elle a obtenu au moins un nombre pair. Quelle est la probabilité que les deux nombres obtenus soient pairs ?

Solution. On reprend les notations de l'exemple 9. Soit G l'événement "obtenir au moins un nombre pair". On cherche la probabilité conditionnelle de l'événement C sachant G : $P(C|G) = \frac{P(C\cap G)}{P(G)}$. Comme $C \subset G$, $C\cap G = C$ et on a vu que $P(C) = \frac{1}{4}$. G est le complémentaire de l'événement "obtenir deux nombres pairs". Donc $P(G) = 1 - \frac{9}{36} = \frac{3}{4}$. Donc, $P(C|G) = \frac{1}{3}$.

▷ *Exercice 17.* Soit A et B deux événements. On suppose que $0 < P(B) < 1$.
Montrer que les trois inégalités : $P(A|B) \leq P(A)$, $P(B|A) \leq P(B)$ et $P(A|B^c) \geq P(A)$ sont équivalentes. Que signifie cette propriété ? Montrer que le fait qu'un fumeur augmente ses chances de développer un cancer du poumon donne lieu à des formulations de ce type. Donner d'autres exemples.

L'utilisation des probabilités conditionnelles est un moyen de calculer la probabilité d'une intersection d'événements puisque $P(A \cap B) = P(A|B)P(B)$ et donc de décomposer le calcul de la probabilité d'un événement A en introduisant un événement B dont on connait la probabilité. En effet, on a :

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

Ces deux relations se généralisent de la façon suivante :

Proposition 1 *1. Soit $A_1, \ldots, A_m$ m événements, la probabilité de leur intersection est :*

$$P(A_1 \cap \ldots \cap A_m) = P(A_1)P(A_2|A_1)\ldots P(A_m|A_1 \cap \ldots \cap A_{m-1})$$

2. Soit $B_1, \ldots, B_m$ une partition de l'ensemble Ω et A un événement. Alors,

$$P(A) = \sum_{i=1}^{m} P(A|B_i)P(B_i).$$

► *Exemple 18.* On a décelé dans une certaine population une probabilité de 0,01 pour qu'un enfant de moins de trois mois soit atteint par une maladie M. La probabilité qu'un enfant qui n'est pas atteint par M ait une réaction négative à un test T est de 0,9. S'il est atteint par M, la probabilité qu'il ait une réaction positive à T est de 0,95.
Quelle est la probabilité qu'un enfant de moins de trois mois pris au hasard ait une réaction positive au test ? Quelle est la probabilité qu'un enfant de moins de trois mois pris au hasard et ayant une réaction positive soit atteint par M ?

Solution. On note A l'événement "l'enfant a la maladie M", A^c l'événement "l'enfant n'a pas la maladie M", B l'événement "l'enfant a une réaction positive au test", B^c l'événement "l'enfant a une réaction négative au test". D'après l'énoncé, $P(A) = 0,01$, $P(A^c) = 0,99$, $P(B^c|A^c) = 0,9$ et $P(B|A) = 0,95$. Donc,

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) = P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= P(B|A)P(A) + (1 - P(B^c|A^c))P(A^c) = 0,0095 + 0,099 = 0,1085. \end{aligned}$$

La probabilité qu'un enfant de moins de trois mois pris au hasard et ayant une réaction positive soit atteint par M est donnée par $P(A|B) = \frac{P(B\cap A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{95}{1085} \simeq 0.088$. Un tel test serait d'une utilité discutable.

▷ *Exercice 19.* Les ampoules de la marque X sont fabriquées dans deux usines, A et B. 20% des ampoules de l'usine A et 5% de l'usine B sont défectueuses. Chaque semaine l'usine A produit $2n$ ampoules et l'usine B n ampoules.

1. Décrire un espace de probabilité associé à l'expérience aléatoire "tirer une ampoule au hasard dans la production d'une semaine".
2. Quelle est la probabilité que l'ampoule tirée ne soit pas défectueuse ?
3. Si l'ampoule tirée est défectueuse, quelle est la probabilité qu'elle vienne de l'usine A ?

▷ *Exercice 20.* On cherche une lettre qui peut se trouver dans trois piles de papiers différentes. N'ayant pas d'information, on supposera qu'elle a autant de chance de se trouver dans une des trois piles. Notons a_i ($i \in \{1,2,3\}$) la probabilité de ne pas trouver la lettre dans la i-ème pile, après un examen rapide de cette pile, alors qu'elle y ait. Supposons que l'on ait rapidement examiné la première pile et rien trouvé. Quelle est la probabilité qu'elle se trouve en fait dans la j-ème pile ($j \in \{1,2,3\}$) ?

2.4 Indépendance entre deux événements

On dit que deux événements A et B sont *indépendants* si le fait de savoir que B est réalisé ne change pas la probabilité de A, autrement dit si $P(A \mid B) = P(A)$. Comme $P(A|B) = \frac{P(A\cap B)}{P(B)}$, ceci équivaut à supposer que

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

Si l'épreuve consiste à tirer un individu au hasard dans un population et si A et B correspondent à des caractéristiques présentes chez certains individus, l'indépendance indique l'absence de relation causale ou de facteur causal commun.

Lorqu'une épreuve est constituée de plusieurs tirages indépendants, deux événements dépendants de tirages différents sont toujours indépendants (voir exercice 10) La notion d'indépendance n'est pas toujours intuitive comme le montre l'exemple suivant :

▶ *Exemple 21.* Soit n un entier supérieur ou égal à deux. On suppose que toutes les repartitions des sexes des enfants d'une famille à n enfants sont équiprobables. Les deux événements :

- M_n : "la famille de n enfants a des enfants des deux sexes"
- F_n : "la famille de n enfants a au plus une fille"

sont indépendants si et seulement si $n = 3$.

▷ *Exercice 22.* Au poker le tirage de cinq cartes consécutives (quinte) et le tirage de cinq cartes de même couleur sont-ils des événements indépendants ?

On a les propriétés suivantes :

Propriétés.

- *Si un événement A est indépendant d'un événement B, alors il est indépendant du complémentaire de B.*
- *Si un événement A est indépendant de n événements $B_1, B_2, ..., B_n$ deux à deux disjoints, il est indépendant de leur réunion.*

On dira de même que deux événements A et B sont *indépendants conditionnellement à un événement* C si $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$.

N.B. Deux événements A et B peuvent être indépendants et ne pas être indépendants conditionnellement à un événement C. Considérer par exemple les événements suivants associés au lancer de deux dés :

- A : “le résultat du premier dé est impair”
- B : “le résultat du second dé est pair”
- C : “les résultats des deux dés sont de même parité”

$P(A \cap B) = \frac{1}{4} = P(A)P(B)$, $P(A \cap B|C) = 0$ alors que $P(A|C) = P(B|C) = \frac{1}{2}$.

De même, deux événements peuvent ne pas être indépendants mais être indépendants conditionnellement à un autre événement. Par exemple, considérons deux pièces A_1 et A_2 et notons p_i la probabilité que la pièce A_i tombe sur face au cours d'un lancer. On choisit au hasard une des deux pièces et on lance la pièce choisie deux fois. Notons F_i l'événement “on a obtenu face au i-ème lancer” pour $i = 1$ ou 2. Les événements F_1 et F_2 sont indépendants conditionnellement à l'événement “la pièce lancée est A_1”. Par contre, si $p_1 \neq p_2$ alors les événements F_1 et F_2 ne sont pas indépendants.

$\triangleright$ *Exercice 23.* Pour chacune des assertions suivantes donner soit une preuve, soit un contre-exemple.

1. si A et B sont deux événements indépendants et incompatibles alors l'un des deux événements au moins est de probabilité nulle.
2. si l'un des événements A ou B est de probabilité nulle alors A et B sont indépendants et incompatibles.
3. Soit B un événement de probabilité $0 < P(B) < 1$. Un événement A est indépendant de B si et seulement si $P(A|B^c) = P(A|B)$.
4. Si un événement E est indépendant d'un événement F et d'un événement G, alors il est indépendant de $F \cup G$.

Chapitre 3

Variables aléatoires

Dans une épreuve, une *variable aléatoire* (en abrégé v.a.) X est une quantité dont la valeur, a priori incertaine, est déterminée à l'issue du tirage. Elle est donc représentée comme une application X définie sur l'ensemble Ω des résultats possibles de l'épreuve. Par exemple, lors du tirage d'un individu dans une population, la taille, le poids ou le revenu annuel de l'individu tiré sont des variables aléatoires. Lors d'une course hippique, le tiercé gagnant est une variable aléatoire. L'indice boursier du lendemain, la durée de vie d'un homme sont des variables aléatoires, et la loi de cette dernière intéresse les assureurs.

3.1 Loi d'une variable aléatoire

L'ensemble des valeurs possibles de X est l'ensemble $X(\Omega)$. Comme la valeur prise par une variable aléatoire X est fonction du résultat de l'épreuve, on peut affecter à chaque valeur possible $x \in X(\Omega)$ une probabilité, celle de l'événement " X prend la valeur x". Cet événement se note en abrégé "$X = x$" et est représenté par le sous-ensemble de Ω défini par

$$X^{-1}(\{x\}) = \{\omega \in \Omega,\ X(\omega) = x\}.$$

Les nombres $P(X = x)$ pour $x \in X(\Omega)$ sont positifs et leur somme est égale à 1 puisque les événements $\{X = x\}$ pour $x \in X(\Omega)$ forment une partition de l'ensemble Ω. Donc, la famille des nombres $\{P(X = x),\ x \in X(\Omega)\}$ définit une probabilité sur $X(\Omega)$, notée $X(P)$ ou P_X et qui est appelée la *loi* de X (voir exercice 12).

▶ *Exemple 24.* L'indicatrice d'un événement A est une variable aléatoire qui peut prendre deux valeurs 0 ou 1 : $\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in \Omega \setminus A \end{cases}$.
Sa loi est donnée par les relations $P(\mathbb{1}_A = 1) = p,\ P(\mathbb{1}_A = 0) = 1 - p$. Cette loi portée par $\{0, 1\}$ est appelée la *loi de Bernoulli de paramètre* p et notée en abrégé $\mathcal{B}(p)$.

▶ *Exemple 25.* Reprenons l'exemple 9 du lancer de deux dés. On note S la v.a qui donne la somme des deux nombres obtenus avec les dés : pour tout $(i, j) \in \Omega = \{1, \ldots, 6\}^2$, $S((i, j)) = i + j$. Déterminons sa loi. Le tableau suivant donne, en fonction du résultat des deux dés, la valeur de S :

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

L'ensemble des valeurs possibles pour S est $S(\Omega) = \{2, \ldots, 12\}$. Il suffit donc de déterminer les probabilités $P(S = k)$ pour $k \in \{2, \ldots, 12\}$. Comme $P(\omega) = \frac{1}{36}$ pour tout $\omega \in \Omega$, il suffit de calculer le nombre de couples $(i, j) \in \Omega$ tels que $i + j = k$ en utilisant le tableau ou en remarquant que

$$\{(i, j) \in \Omega,\ \text{tels que } i + j = k\} = \{(i, k - i),\ i \in \{\max(k - 6, 1), \ldots, \min(6, k - 1)\}\}.$$

On obtient : $P(S = k) = \frac{k-1}{36}$ si $k \in \{1, \ldots, 7\}$ et $P(S = k) = \frac{13-k}{36}$ si $k \in \{8, \ldots, 12\}$.

▷ *Exercice 26.* Déterminer la loi de la variable aléatoire donnant le maximum des deux chiffres obtenus en lançant deux dés.

▷ *Exercice 27.* Soit X une variable aléatoire, définie sur un ensemble au plus dénombrable Ω muni de la probabilité P et qui prend ses valeurs dans un ensemble au plus dénombrable U. On note par P_X la loi de X : $P_X(u) = P(X = u)$ pour tout $u \in U$.
Soit f une application définie de U et à valeur dans un ensemble V. Vérifier que $f \circ X$ est une variable aléatoire que l'on note $f(X)$ définie sur Ω et montrer que sa loi est définie par les nombres

$$P(f(X) = v) = \sum_{z \in f^{-1}(\{v\})} P_X(z) \text{ pour tout } v \in V.$$

Lorsque la variable aléatoire X prend ses valeurs dans $\mathbb{R}$ ou $\mathbb{C}$, on parlera de *variables aléatoires numériques*. Les variables aléatoires qui prennent leurs valeurs dans $\mathbb{R}^d$ (ou $\mathbb{C}^d$) sont souvent appelées des *vecteurs aléatoires*.
Les composantes $X_1, \ldots, X_d$ d'un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ (ou $\mathbb{C}^d$) sont des variables aléatoires numériques (pour $i \in \{1, \ldots, d\}$, X_i est l'image de X par la projection $\pi_i : (x_1, \ldots, x_d) \mapsto x_i$). On appelle les lois des variables aléatoires X_i pour $i \in \{1, \ldots, d\}$ *les lois marginales de* X.
Inversement, si $X_1, \ldots, X_d$ sont des variables aléatoires numériques définies sur le même espace Ω, c'est-à-dire relatives à la même épreuve, on appelle *la loi conjointe des v.a.* $X_1, \ldots, X_d$ la loi du vecteur aléatoire $X = (X_1, \ldots, X_d)$.
La donnée de la loi conjointe permet de déterminer facilement les lois marginales. Par exemple, exprimons la loi d'une v.a. X à l'aide de la loi conjointe du couple (X, Y). Comme $\{X = x\} = \{(X, Y) \in \{x\} \times Y(\Omega)\}$ pour tout $x \in X(\Omega)$, on a :

$$P_X(x) = \sum_{y \in Y(\Omega)} P_{(X,Y)}(x, y) \quad \text{pour tout } x \in X(\Omega).$$

► *Exemple 28.* Soient A et B deux événements. La loi conjointe de leurs indicatrices $X = (\mathbb{1}_A, \mathbb{1}_B)$ est la loi sur $\{0, 1\} \times \{0, 1\}$ décrite par le tableau ci-dessous :

	0	1
0	$1 - P(A) - P(B) + P(A \cap B)$	$P(A) - P(A \cap B)$
1	$P(B) - P(A \cap B)$	$P(A \cap B)$

Ce résultat se généralise aisément au cas de n événements.

N.B. Deux variables aléatoires identiques ont même loi, mais la réciproque est fausse : par exemple, si toutes les issues sont équiprobables, pour toute variable aléatoire X et pour toute permutation σ de Ω, X et $X \circ \sigma$ ont même loi.

▷ *Exercice 29.* Soit $a \in [0,1]$. Le tableau ci-dessous donne la loi d'un couple de v.a. (X,Y) à valeurs dans $\{1,2,3\}^2$.

$x \setminus y$	1	2	3
1	0	a/3	(1-a)/3
2	(1-a)/3	0	a/3
3	a/3	(1-a)/3	0

Déterminer les lois des marginales. Dépendent-elles de a ?
Calculer la probabilité que X et Y soient égales.

▷ *Exercice 30.* Soit Ω un ensemble fini muni d'une probabilité P et Z une variable aléatoire à valeurs dans $\mathbb{N}^2$ définie sur Ω. Pour $k = (k_1, k_2) \in \{-1,0,1\} \times \{-1,1\}$, la probabilité $P(Z = k)$ est donnée par le tableau suivant

$k_2 \diagdown k_1$	-1	0	1
-1	1/12	1/3	1/12
1	1/6	1/6	1/6

a) Le tableau ci-dessus détermine-t-il la loi de Z ?

b) Soit π_1 et π_2 les applications coordonnées définies par $\pi_1(x,y) = x$ et $\pi_2(x,y) = y$. Donner la loi des variables $X = \pi_1 \circ Z$ et $Y = \pi_2 \circ Z$.

c) Déterminer la loi des variables $X + Y$ et XY.

La loi d'une variable aléatoire peut être représentée graphiquement par un diagramme en bâtons. Par exemple, le diagramme en bâtons représentant la loi de la variable aléatoire S définie dans l'exemple 25 est donnée sur la figure 3.1.

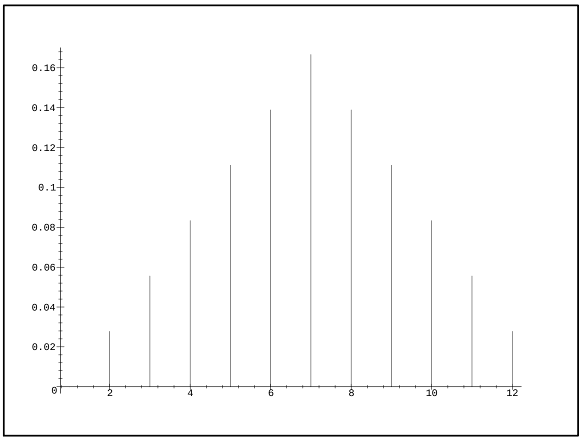


FIG. 3.1 : Loi de la somme des nombres obtenus en lançant deux dés

Les valeurs les plus probables d'une variable aléatoire sont appelées *les modes* de la loi de cette variable aléatoire.

▶ *Exemple 31.* La variable aléatoire S n'a qu'un mode qui est 7.

▷ *Exercice 32.* Dans une pile de n ($n \geq 2$) feuilles dactylographiées, se trouvent les deux lettres que l'on doit envoyer. On enlève une par une les feuilles du paquet jusqu'à ce que l'une des lettres à envoyer se trouve sur le dessus du paquet. On note X_1 la variable aléatoire donnant le nombre de feuilles enlevées. On recommence l'opération jusqu'à trouver la deuxième lettre et on note X_2 la variable aléatoire donnant le nombre supplémentaire de feuilles qu'il a fallu retirer du paquet avant que la deuxième lettre soit sur le dessus du paquet. Sans information supplémentaire, on peut supposer que toutes les positions possibles pour les deux lettres sont équiprobables.

1. Décrire l'ensemble Ω des résultats possibles pour cette expérience aléatoire et la probabilité P mise sur Ω.
2. Déterminer la loi du couple (X_1, X_2) puis la loi de X_1 et de X_2.
3. Calculer la probabilité de l'événement "$X_1 = X_2$".
4. On note $Z = X_1 + X_2 + 2$. Que représente la variable aléatoire Z ? Déterminer sa loi.
5. Quelles sont les modes de ces trois variables aléatoires ? Commenter les résultats.

3.2 Loi Binomiale et loi faible des grands nombres

Soit A un événement possible d'une épreuve décrite par l'ensemble Λ muni de la probabilité Q et $p = Q(A)$ sa probabilité. Au cours de l'expérience statistique consistant à répéter cette épreuve n fois de façon indépendante (représentée par l'ensemble $\Omega = \Lambda^n$ muni de la probabilité produit $P = Q^{\otimes n}$), on compte le nombre de fois où l'événement A a été réalisé. Cela définit une variable aléatoire notée S_n :

$$\begin{array}{rccc} S_n : & \Omega & \to & \{0, \dots, n\} \\ & (\omega_1, \dots, \omega_n) & \mapsto & |\{i,\ \omega_i \in A\}| \end{array}$$

En utilisant les v.a. indicatrices, on a pour tout $(\omega_1, \dots, \omega_n) \in \Omega$:

$$S_n(\omega_1, \dots, \omega_n) = \sum_{i=1}^{n} \mathbb{1}_A(\omega_i)$$

ou encore en désignant par $A^{(i)}$ l'événement de Ω "A est réalisé à la i-ème épreuve", c'est-à-dire $A^{(i)} = \Lambda^{i-1} \times A \times \Lambda^{n-i}$:

$$S_n = \sum_{i=1}^{n} \mathbb{1}_{A^{(i)}}.$$

Proposition 2 *Pour tout n-uplet $(A_1, A_2, ..., A_n)$ d'événements indépendants de même probabilité p définis sur le même espace probabilisé (Ω, P), notons S_n la v.a. $\sum_{i=1}^{n} \mathbb{1}_{A_i}$.*
Pour tout entier k compris entre 0 et n,

$$P(S_n = k) = C_n^k q^k (1-q)^{n-k}$$

On dit que S_n suit la loi binomiale *de paramètres n et p (notée en abrégé $\mathcal{B}(n, p)$)*[1].

Démonstration : Notons $A_{(I)}$ l'événement $(\cap_{i \in I} A_i) \cap (\cap_{j \in \{1,\dots,n\} \setminus I} (\Omega \setminus A_j))$. Les événements A_i pour $i \in \{1, \dots, n\}$ étant indépendants et de probabilité p, $P(A_{(I)}) = p^k (1-p)^{n-k}$. L'événement $\{S_n = k\}$ est la réunion disjointe des événements $A_{(I)}$ où I varie parmi les $C_n^k = \frac{n!}{k!(n-k)!}$ parties de $\{1, ..., n\}$ formées de k éléments. Donc, $P(S_n = k) = C_n^k p^k (1-p)^{n-k}$. □

La fréquence avec laquelle A est réalisé lors de cette expérience statistique est donnée par le rapport $\frac{S_n}{n}$. Nous pouvons maintenant justifier le résultat annoncé dans la section 1.3 :

Proposition 3 *(Loi faible des grands nombres) Pour tout $\epsilon > 0$ et $n \in \mathbb{N}^*$,*

$$P(|\frac{S_n}{n} - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

Cette probabilité tend vers zéro lorsque n tend vers $+\infty$.

[1] $\mathcal{B}(1, p)$ est la loi de Bernoulli $\mathcal{B}(p)$

Preuve :

$$P(|\frac{S_n}{n} - p| \geq \epsilon) = \sum_{\omega\in\Omega,\ |S_n(\omega)-np|\geq n\epsilon} P(\omega) \leq \sum_{\omega\in\Omega} \frac{(S_n(\omega)-np)^2}{n^2\epsilon^2} P(\omega)$$

Et

$$\sum_{\omega\in\Omega}(S_n(\omega)-np)^2 P(\omega) = \sum_{k=0}^{n}(k-np)^2 P(S_n = k) = \sum_{k=0}^{n}(k-np)^2 C_n^k p^k (1-p)^{n-k}.$$

En utilisant que $iC_n^i = nC_{n-1}^{i-1}$ pour tout $i \in \{1,\dots,n\}$ et donc que $i(i-1)C_n^i = n(n-1)C_{n-2}^{i-2}$ pour tout $i \in \{2,\dots,n\}$ puis en appliquant la formule du binôme, on montre que cette somme est égale à $np(1-p)$. □

▷ *Exercice 33.* Vérifier que

$$\sum_{k=0}^{n}(k-np)^2 C_n^k p^k (1-p)^{n-k} = np(1-p).$$

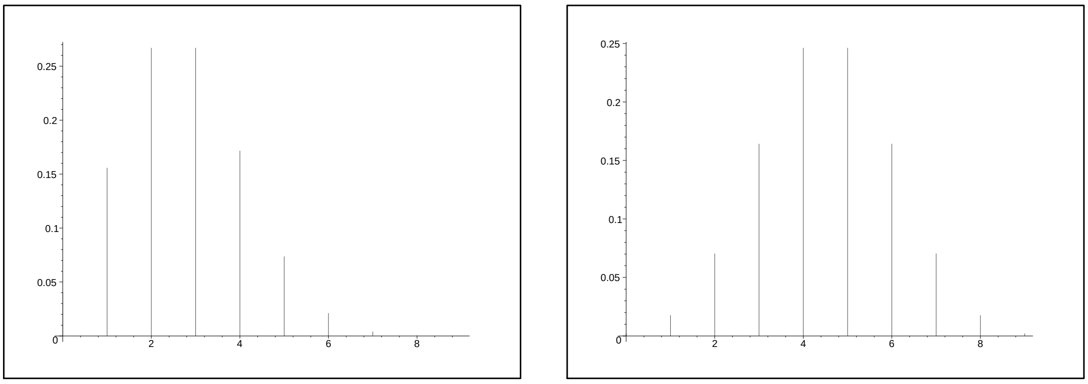


FIG. 3.2 : Diagrammes des lois $\mathcal{B}(9, 0.3)$ et $\mathcal{B}(9, 0.5)$

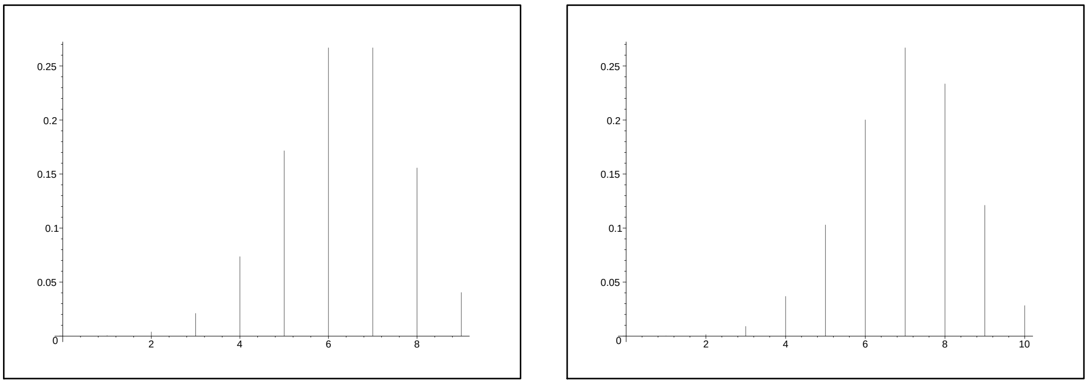


FIG. 3.3 : Diagrammes des lois $\mathcal{B}(9, 0.7)$ et $\mathcal{B}(10, 0.7)$

▶ *Exemple 34.* On lance un dé n fois et on note X_n la variable aléatoire qui donne le nombre de fois où le dé est tombé sur 6. Alors X_n suit la loi binomiale $\mathcal{B}(n, \frac{1}{6})$.

▷ *Exercice 35.* Quelle est la loi du nombre de garçons dans une famille de n enfants ? (Préciser les hypothèses que vous avez faites)

▷ *Exercice 36.* Sachant qu'une population de grande taille contient 2% de porteurs d'un virus, calculez les probabilités que sur un échantillon de 100 individus tirés au hasard, il y ait 0, 1, 2, 3, 4, au moins 5 porteurs du virus.

▷ *Exercice 37.* Tracer les diagrammes de lois binomiales $\mathcal{B}(n, p)$ pour différentes valeurs de n et de p afin de déterminer :

- à quelles conditions sur n et p, le diagramme en bâtons de la loi binomiale $\mathcal{B}(n, p)$ est symétrique,
- quelle est l'expression des modes d'une loi binomiale $\mathcal{B}(n, p)$ en fonction de la valeur des paramètres n et p.

Démontrer ensuite les conjectures que vous avez formulées.

▷ *Exercice 38.* On lance 20 fois une pièce supposée bien équilibrée. On désigne par X la fréquence du nombre de fois où pile a été obtenu au cours de 20 lancers.

1. Représenter la loi de X.
2. Avec quelle probabilité X est-elle strictement au dessus de 0.5 ?
3. Avec quelle probabilité X est-elle comprise entre 0.4 et 0.6 ?
4. Déterminer le plus petit réel $a > 0$ telle que la probabilité que X soit dans l'intervalle $[0.5 - a, 0.5 + a]$ soit supérieure à 95%.
5. On lance la pièce 20 fois. Elle tombe 5 fois sur pile et 15 fois sur face. D'après vous la pièce est-elle bien équilibrée (on justifiera sa réponse en utilisant la question 4) ?
6. On ne suppose plus maintenant que la pièce est bien équilibrée. On note q la probabilité que le dé tombe sur pile au cours d'un lancer, Ω l'ensemble des résultats possibles pour l'expérience consistant à lancer cette pièce 20 fois, P_q la probabilité sur Ω associée à cette expérience et S le nombre de fois où pile a été obtenu au cours des 20 lancers. A l'aide d'une observation s de S, on voudrait estimer la valeur de la probabilité q. Une méthode naturelle consiste à donner à q la valeur (si elle est unique) pour laquelle $P_q(S = s)$ est maximale (méthode du maximum de vraisemblance). Déterminer la valeur de $q(s)$ donnée à q par cette méthode.
7. Lorsque $s = 5$, tracer la fonction $q \mapsto P_q(S = s)$ et donner la valeur de $q(s)$.

3.3 Approximation poissonienne et loi de Poisson

L'expression de la loi binomiale en k peut aussi s'écrire

$$\frac{(np)^k}{k!}(1-p)^{n-k}\prod_{j=0}^{k-1}(1-\frac{j}{n}).$$

Si, pour k fixé, on fait tendre n vers l'infini, $\prod_{j=0}^{k-1}(1-\frac{j}{n})$ tend vers 1. Si de plus on fait tendre simultanément p vers 0 de manière à ce que np tende vers une limite finie λ, $(1-p)^{n-k}$ converge vers $e^{-\lambda}$, et donc la loi binomiale en k converge vers $q_k = e^{-\lambda}\frac{\lambda^k}{k!}$. La famille $(q_k)_{k\in\mathbb{N}}$ définit une probabilité sur $\mathbb{N}$ qui est appelée la *loi de Poisson de paramètre* λ et que l'on note $\mathcal{P}(\lambda)$. Elle fournit une bonne approximation de la loi binomiale lorsque n est grand et p petit, en particulier pour évaluer la loi du nombre d'événements rares apparaissant sur un grand nombre d'épreuves indépendantes (par exemple les accidents d'avion).

▷ *Exercice 39.* Comparer les résultats fournis par l'approximation poissonienne avec les résultats exacts dans l'exercice 36.

▷ *Exercice 40.* Dans une fabrique de galettes, quelle proportion de raisins faut-il mélanger à la pâte pour que 98% des galettes contiennent au moins un raisin ? Imaginer une variante militaire de ce problème.

▷ *Exercice 41.* Déterminer l'expression des modes d'une loi de Poisson en fonction du paramètre de cette loi.

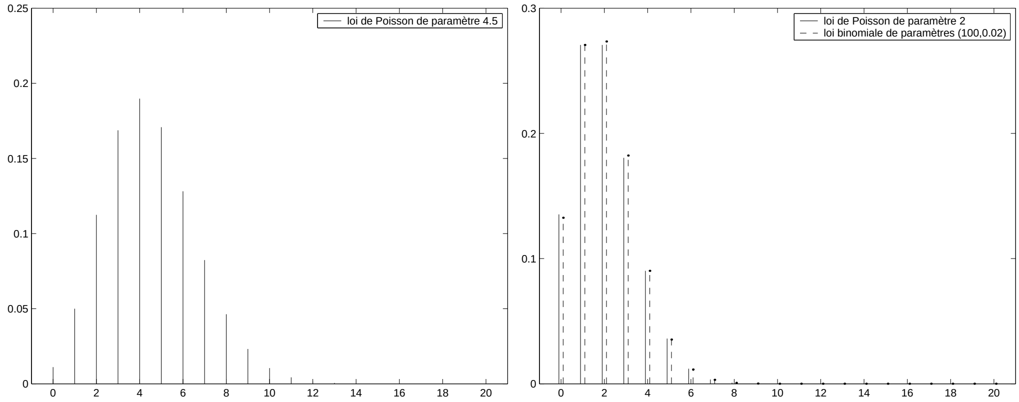


FIG. 3.4 : Figure de gauche : loi de Poisson $\mathcal{P}(4.5)$. Figure de droite : comparaison entre la loi de Poisson $\mathcal{P}(2)$ et la loi binomiale $\mathcal{B}(100, 0.02)$

3.4 Convergence des lois de probabilité

Etant données une suite de probabilités $(\mu_n)_n$ sur un ensemble dénombrable E et une probabilité μ, on dit que μ_n *converge vers* μ si $\mu_n(x)$ converge vers $\mu(x)$ pour tout $x \in E$.

▶ *Exemple 42.* Soit $(p_n)_n$ une suite de réels telle que pour tout $n \in \mathbb{N}$, $p_n \in [0,1]$ et $(np_n)_n$ converge vers un réel a. On vient de voir que la suite de lois $(\mathcal{B}(n, p_n))_n$ converge vers la loi $\mathcal{P}(a)$.

Notons que les lois μ_n et μ peuvent avoir des supports différents. Par exemple, le support de la loi $\mathcal{B}(n,p)$ est $\{0, \dots, n\}$ si $p \in]0,1[$ et le support de la loi $\mathcal{P}(a)$ est $\mathbb{N}$ si $a > 0$.

▷ *Exercice 43.* Montrer que le support de la loi de μ est inclus dans la limite supérieure (dans le sens ensembliste) des supports des lois μ_n.

▷ *Exercice 44.* Montrer par un contre-exemple, qu'une suite de probabilités $(\mu_n)_n$ telle que $(\mu_n(x))_n$ converge pour tout $x \in E$, ne converge pas toujours vers une probabilité μ.
Montrer que c'est le cas si et seulement si pour tout $\epsilon > 0$, il existe une partie finie F de E telle que $\mu_n(F) \geq 1 - \epsilon$ pour tout n.

3.5 Fonctions de répartition.

La fonction de répartition d'une v.a. X à valeurs dans un ensemble $E \subset \mathbb{R}$ au plus dénombrable est la fonction, notée F_X, qui à tout réel t associe la probabilité que X prenne une valeur inférieure ou égale à t :

$$F_X : \begin{array}{ccc} \mathbb{R} & \to & [0,1] \\ t & \mapsto & P(X \leq t) \end{array}$$

C'est donc une fonction croissante qui a une limite égale à 0 en $-\infty$ et 1 en $+\infty$, et qui fait des sauts en chaque valeur $e \in E$ dont la hauteur est $P(X = e)$. Remarquons que la connaissance de la fonction de répartition d'une v.a. est équivalente à la connaissance de sa loi (voir exercice 46).

La probabilité que X prenne une valeur dans un intervalle $]a, b]$ c'est-à-dire la probabilité de l'événement "$X \in]a, b]$" peut se calculer facilement en utilisant la fonction de répartition de X : $P(a < X \leq b) = F_X(b) - F_X(a)$. De façon générale, pour un ensemble $B \subset X(\Omega)$, l'événement "X est à valeurs dans B" se note "$X \in B$" et peut se calculer en faisant la somme des probabilités des événements "$X = x$" pour tout $x \in B$.

► *Exemple 45.* Le graphe de la fonction de répartition de la variable S définie dans l'exemple 25 est donnée à la figure 3.5

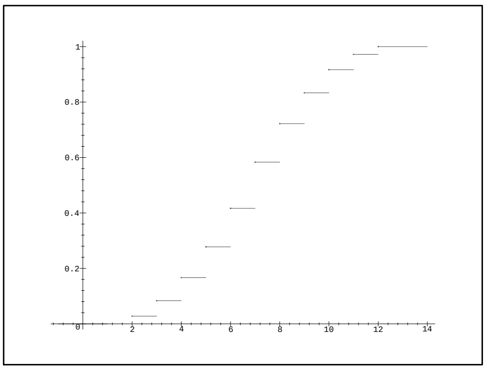


FIG. 3.5 : Fonction de répartition de la variable aléatoire S

La probabilité que S soit impair est

$$\begin{aligned} P(S \in \{3,5,7,9,11\}) &= P(S=3)+P(S=5)+P(S=7)+P(S=9)+P(S=11) \\ &= \frac{1}{2}. \end{aligned}$$

▷ *Exercice 46.* Soit X une variable aléatoire à valeurs réelles prenant un nombre au plus dénombrable de valeurs et $t \in \mathbb{R}$. Montrer que

$$P(X < t) = \lim_{n \to +\infty} P(X \leq t - \frac{1}{n}).$$

En déduire une expression de $P(X = t)$ en fonction uniquement de la fonction de répartition de X.

▷ *Exercice 47.* Un sac contient N jetons numérotés de 1 à N. On en tire au hasard $n \leq N$ en remettant chaque jeton tiré avant de tirer le suivant.

1. Décrire l'ensemble Ω des tirages possibles et la probabilité P associée.
2. Soit $k \in \{1, \ldots, N\}$. On note A_k l'événement "toutes les boules tirées ont un numéro inférieur ou égal à k". Décrire cet événement et déterminer sa probabilité.
3. Soit X la variable aléatoire qui à un tirage de n jetons associe le plus grand numéro des jetons tirés. Quelle est sa fonction de répartition ? En déduire la loi de X.
4. Représenter graphiquement la loi de X à l'aide d'un diagramme en bâtons lorsque $N = 30$, $n = 1$ et $n = 5$.
5. Déterminer les modes de la loi de X en fonction de n et N.
6. Soit $p \in]0,1[$. Déterminer le nombre minimal $n(p)$ de tirages que l'on doit faire pour que la probabilité d'obtenir au moins un jeton de numéro strictement supérieur à $N/2$ soit supérieure à p. Tracer la fonction $p \mapsto n(p)$.

3.6 Quelques lois classiques

3.6.1 Loi uniforme

On dit qu'une v.a. suit une loi uniforme si chacune des valeurs qu'elle est susceptible de prendre a la même probabilité d'apparaître. On note Ent la fonction partie entière définie sur $\mathbb{R}$ par $\mathrm{Ent}(t) = \sup(n \in \mathbb{Z}, n \leq t)$. Pour tout entier k, la fonction de répartition de la loi uniforme sur $\{0, 1, ..., k\}$ est nulle sur $]-\infty, 0[$, égale à 1 sur $]k, +\infty[$, et à $\frac{1+\mathrm{Ent}}{k+1}$ sur $[0, k]$.

► *Exemple 48.* La variable aléatoire qui donne le nombre obtenu en lançant un dé suit la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$.
La variable aléatoire définie dans l'exemple 28 a une loi uniforme sur $\{0, 1\}^2$ si et seulement si A et B sont deux événements indépendants de probabilité $\frac{1}{2}$.

▷ *Exercice 49.* On pourrait considérer que le jour de la semaine ou la date de la naissance d'un individu tiré au hasard dans la population suivent des lois uniformes. Trouver des facteurs qui laisse penser que ce n'est pas tout à fait vrai.

▷ *Exercice 50.* On lance une pièce 2 fois et on note X la variable aléatoire donnant le nombre de fois où la pièce est tombée sur "face". Peut-on construire une pièce asymétrique de sorte que la loi de X soit uniforme sur $\{0, 1, 2\}$?

3.6.2 Loi hypergéométrique

Considérons une population constituée de N individus que l'on peut séparer en deux classes ; n_1 individus de type I et n_2 individus de type II. Parmi cette population, on choisit au hasard un sous-ensemble ω de k individus et on note $X(\omega)$ le nombre d'individus de type I dans ω. La loi de la variable aléatoire X est appelée la loi *hypergéométrique de paramètres* (N, n_1, k) (notée en abrégé $\mathcal{H}(N, n_1, k)$).

Proposition 4 *Pour tout j compris entre 0 et N*[2], $P(X = j) = \frac{\binom{n_1}{j}\binom{n_2}{k-j}}{\binom{N}{k}}$

Preuve : Notons $\mathcal{I}_1$ et $\mathcal{I}_2$ respectivement l'ensemble des individus de type I et II de la population et Ω l'ensemble des parties à k éléments de la population $\mathcal{I}_1 \cup \mathcal{I}_2$. Chaque élément de Ω a autant de chance d'être tiré. On munit donc Ω de la probabilité uniforme P : $P(\{\omega\}) = \frac{1}{\binom{N}{k}}$ pour tout $\omega \in \Omega$. La variable aléatoire X est l'application définie sur Ω par : $X(\omega) =| \omega \cap \mathcal{I}_1 |$ pour tout $\omega \in \Omega$. Comme il y a n_1 individus de type I, pour tout $\omega \in \Omega$, $X(\omega) \leq n_1$. D'autre part, $k - X(\omega)$ est le nombre d'individus de type II présents dans ω. Donc, $k - X(\omega) \leq n_2$. Par conséquent, X est à valeurs dans l'ensemble

$$\{j \in \mathbb{N},\ 0 \leq j \leq k \text{ et } k - n_2 \leq j \leq n_1\} = \{j \in \mathbb{N},\ \max(0, k - n_2) \leq j \leq \min(k, n_1)\}.$$

Soit $j \in \{\max(0, k - n_2), \ldots, \min(k, n_1)\}$. $\omega \in \{X = j\}$ si et seulement si c'est un sous-ensemble constitué de j individus de type I et de $k - j$ individus de type II. Comme il y a $\binom{n_1}{j}$ façons de choisir j individus de type I et $\binom{n_2}{k-j}$ façons de choisir $k - j$ individus de type II, $|\{X = j\}| = \binom{n_1}{j}\binom{n_2}{k-j}$. Donc,

$$P(X = j) = \frac{\binom{n_1}{j}\binom{n_2}{k-j}}{\binom{N}{k}}.$$

□

► *Exemple 51.* Dans une classe de vingt filles et quinze garçons, on choisit au hasard cinq personnes différentes. La loi du nombre de garçons choisis est la loi hypergéométrique $\mathcal{H}(35, 15, 5)$. Celle du nombre de filles sélectionnées est la loi $\mathcal{H}(35, 20, 5)$.
Si par contre, on choisit cinq fois de suite une personne au hasard parmi les élèves de la classe, le nombre de garçons choisis suit la loi binomiale $\mathcal{B}(5, \frac{15}{35})$ et le nombre de filles sélectionnées suit la loi $\mathcal{B}(5, \frac{20}{35})$.

[2] Avec la convention, $\binom{n}{k} = 0$ si $k \notin \{0, \ldots, n\}$

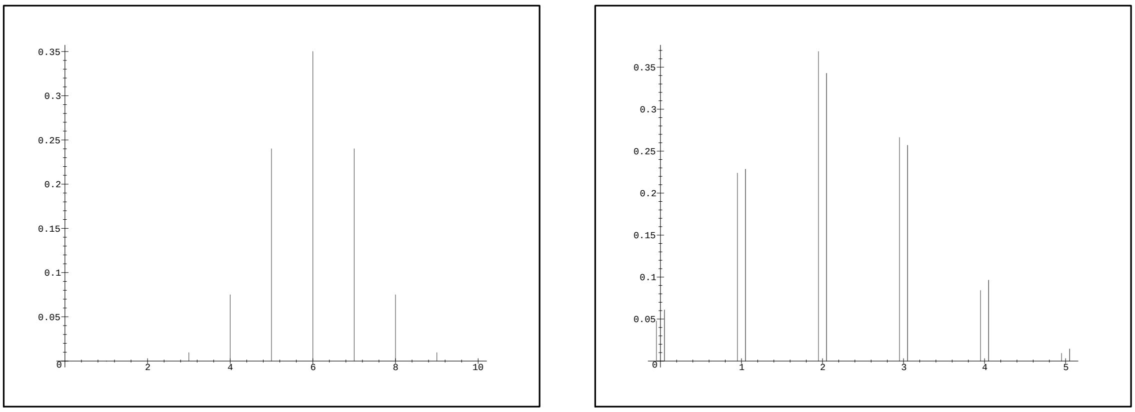


FIG. 3.6 : Figure de gauche : loi Hypergéométrique $\mathcal{H}(20, 12, 10)$. Figure de droite : comparaison de la loi hypergéométrique $\mathcal{H}(35, 15, 5)$ (bâtons de gauche) avec la loi binomiale $\mathcal{B}(5, \frac{15}{35})$ (bâtons de droite)

▷ *Exercice 52.* Le poker : en supposant que toutes les mains de cinq cartes sont tirées avec la même probabilité (dans un jeu de 52 cartes), calculer la probabilité de tirer un carré d'as, une paire de rois. Calculer la probabilité d'avoir un brelan de rois sachant qu'on en a au moins une paire.

▷ *Exercice 53.* Montrer que lorsque N tend vers l'infini et $\frac{n_1}{N}$ tend vers p, la loi hypergéométrique de paramètres (N, n_1, k) converge vers la loi binomiale $\mathcal{B}(n, p)$. Ceci est à rapprocher de la remarque de la page 5.

▷ *Exercice 54.* Pour obtenir des informations sur le nombre N de grenouilles vivant dans un étang, on en capture en premier lieu un nombre m, on les marque puis on les relache. Après un temps suffisamment long pour que les grenouilles relachées se soient dispersées, on capture de nouveau n grenouilles de l'étang. On supposera que le nombre de grenouilles entre les deux captures n'a pas changé et qu'à chaque capture les grenouilles de l'étang ont toute la même probabilité d'être attrapées.

1. On note X la variable aléatoire donnant le nombre de grenouilles marquées attrapées lors de la deuxième capture. Déterminer la loi de X (On décrira l'ensemble Ω_N des résultats possibles pour cette expérience et la probabilité P_N mise sur Ω_N).
2. Lors de la deuxième capture, on a attrapé i grenouilles marquées. En utilisant uniquement cette observation, on voudrait estimer le nombre N. Une méthode appelée “estimation par le maximum de vraisemblance” consiste à prendre pour N la valeur (si elle existe) pour laquelle $P_N(X = i)$ est maximale. Pour $i \geq 1$, déterminer N de cette manière en fonction de i, m et n. Interpréter le résultat. Qu'obtient-on lorsque $i = m = n$?
3. Pour $m = 20$, $n = 10$ et $i = 2$, déterminer la valeur de N donnée par la méthode d'estimation par le maximum de vraisemblance. Tracer le diagramme en bâtons de la loi de X pour cette valeur de N et pour d'autres valeurs de N. Tracer la fonction $x \mapsto P_x(X = i)$.

3.6.3 Loi multinomiale

Dans le modèle conduisant à la loi binomiale, on ne s'intéressait qu'à la réalisation ou à la non réalisation d'un événement A. Considérons maintenant une *partition* de Λ en d événements disjoints $C_1, ..., C_d$ de probabilités respectivement $p_1, \ldots, p_d$. Ces classes peuvent éventuellement être réduites à un point de Λ. Par exemple, dans le cas du tirage d'une carte, on peut partitionner Λ en quatre événements $\{$carreaux$\}$, $\{$coeurs$\}$, $\{$trèfles$\}$ et $\{$piques$\}$. Dans le cas du tirage d'un individu dans une population, on pourra considérer n'importe quelle classification selon le groupe sanguin, la profession ...

Pour tout j compris entre 1 et d, notons N_j le nombre de fois où l'événement C_j a été réalisé lors de l'expérience statistique décrite par les n tirages : $N_j = \sum_{i=1}^n \mathbb{1}_{C_j^{(i)}}$. La fréquence avec laquelle C_j est réalisé est donnée par le rapport $\frac{N_j}{n}$. On peut calculer la loi du vecteur aléatoire $\overrightarrow{N} = (N_1, ..., N_d)$ à l'aide du vecteur $\overrightarrow{p} = (p_1, ..., p_d)$. On a :

Proposition 5 *Pour tout d-uplet* $\overrightarrow{k} = (k_1, ..., k_d) \in \mathbb{N}^d$ *tel que* $\sum_{j=1}^d k_j = n$,

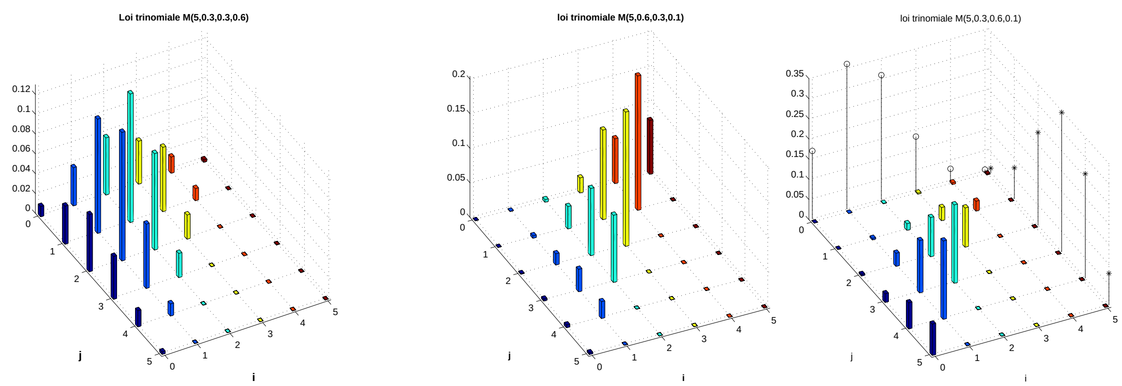
$$P(\overrightarrow{N} = \overrightarrow{k}) = n! \prod_{j=1}^{d} \frac{p_j^{k_j}}{k_j!}.$$

On dit que $\overrightarrow{N}$ *suit la* loi multinomiale de paramètres n et $\overrightarrow{p}$ *(notée en abrégé,* $\mathcal{M}(n, \overrightarrow{p})$*).*

Preuve : L'évènement $\{\overrightarrow{N} = \overrightarrow{k}\}$ peut s'écrire comme réunion disjointe des événements de la forme $A_{(I)} = \cap_{j=1}^d (\cap_{i \in I_j} \{C_j^{(i)}\})$ où I varie parmi les $\frac{n!}{k_1!...k_d!}$ partitions de $\{1, ..., n\}$ en d classes I_j formées de k_j éléments. Pour une telle partition I, $P(A_{(I)}) = p_1^{k_1} ... p_d^{k_d}$. □

N.B. Pour $j \in \{1, \ldots, d\}$, N_j suit la loi binomiale $\mathcal{B}(n, p_j)$.

► *Exemple 55.* Sur les figures ci-dessous, la loi d'un couple de variables aléatoires (X, Y) est représentée par un diagramme en bâtons : la hauteur du bâton positionnée au point de coordonnées (i, j) est la valeur de $P(X = i,\ Y = j)$: sur la figure de gauche $(X, Y, 5 - X - Y)$ suit la loi trinomiale $\mathcal{M}(5, (0.3, 0.3, 0.6))$. Sur la figure du centre, $(X, Y, 5 - X - Y)$ suit la loi trinomiale $\mathcal{M}(5, (0.6, 0.3, 0.1))$. Sur la figure de droite, $(X, Y, 5 - X - Y)$ suit la loi trinomiale $\mathcal{M}(5, (0.3, 0.6, 0.1))$. On a représenté en plus sur cette figure les lois des marginales (c'est-à-dire la loi de X (bâtons surmontés d'un rond) et de la loi de Y (bâtons surmontés d'une étoile).



▷ *Exercice 56.* Quelle est la probabilité pour que le lancer de quatre dés donne deux six et deux trois ?

▷ *Exercice 57.* Soit $\overrightarrow{N}_n = (N_{n,1}, \ldots, N_{n,d})$ un vecteur aléatoire de loi multinomiale $\mathcal{M}(n, (p_{n,1}, \ldots, p_{n,d}))$.

1. Quelle est la loi de $N_{n,1} + N_{n,2}$?
2. De façon générale, soit J un sous-ensemble de $\{1, \ldots, d\}$. Quelle est la loi de $\sum_{j \in J} N_{n,j}$?
3. Supposons que, pour tout $i \in \{1, \ldots, d\}$, $np_{n,i}$ tend vers une constante $\lambda_i > 0$ lorsque n tend vers $+\infty$. Que peut-on dire sur la loi de $\overrightarrow{N}_n$ lorsque n tend vers $+\infty$?

▷ *Exercice 58.* Pour étudier une population de taille N, on a défini r classes d'individus $C_1, \ldots, C_r$ de sorte que chaque individu appartienne à une et une seule de ces classes. Posons $n_i = \mid C_i \mid$ pour $i \in \{1, \ldots, r\}$. On choisit au hasard k individus différents parmi cette population et on désigne par $X_{N,i}$ le nombre d'individus sélectionnés qui appartiennent à la classe C_i.

1. Déterminer la loi de $(X_{N,1},\dots,X_{N,r})$.
2. Déterminer la loi des marginales $X_{N,i}$.
3. On fait tendre N vers $+\infty$ de sorte que les proportions d'individus appartenant à chaque classe C_i tendent vers un réel strictement positif noté p_i. Que peut-on dire de la loi de $X_{N,1}$ lorsque N tend vers $+\infty$? Même question pour la loi de $(X_{N,1},\dots,X_{N,r})$.
4. Au poker, calculer la probabilité de tirer une double paire.

3.6.4 Loi géométrique

Considérons à nouveau le cas de n tirages indépendants $\omega_1,\dots,\omega_n$ dans Λ et d'un événement A de Λ. Soit T le plus petit entier $i \le n$ tel que $\omega_i \in A$. Lorsque $\omega_i \notin A$ pour tout $i \le n$, on convient de poser $T = n+1$. T est donc une v.a. à valeurs dans $\{1, 2, ..., n+1\}$, définie sur l'ensemble $\Omega = \Lambda^n$ représentant tous les tirages possibles. Sa loi est donnée par les formules suivantes :

$$P(T=i) = p(1-p)^{i-1} \text{ si } 1 \le i \le n$$
$$P(T=n+1) = (1-p)^n$$

En effet, pour $i \in \{1\dots,n\}$, $\{T=i\} = (\Omega \setminus A^{(1)}) \cap \dots \cap (\Omega \setminus A^{(i-1)}) \cap A^{(i)}$ et $\{T = n+1\} = (\Omega \setminus A^{(1)}) \cap \dots \cap (\Omega \setminus A^{(n)})$. Les probabilités de ces événements s'obtiennent en utilisant que les événements $A^{(j)}$ sont indépendants et de probabilité p pour tout $j \in \{1, 2, ..., n\}$.
On appellera la loi de T, la *loi géométrique tronquée à n*.

Lorsque n tend vers l'infini, la loi de T converge de façon évidente vers la suite $(p(1-p)^{i-1})_{i\in\mathbb{N}^*}$ qui définit une loi sur $\mathbb{N}^*$ qu'on appelle la *loi géométrique* de paramètre p sur $\mathbb{N}^*$ et que l'on note $\mathcal{G}_{\mathbb{N}^*}(p)$.

▷ *Exercice 59.* Calculer les fonctions de répartition de la loi géométrique et de la loi géométrique tronquée.

▷ *Exercice 60.* On lance une pièce de monnaie N fois. Soit $n \le N$, on s'intéresse au rang du n-ième pile si pile apparaît au moins n fois.

1. Décrire l'ensemble Ω des résultats possibles et la probabilité P mise sur Ω.
2. Décrire l'événement A_N : “ pile apparaît moins de n fois (c'est-à-dire $n-1$ fois au plus) ” et calculer sa probabilité que l'on notera p_N.
3. Tracer $N \mapsto p_N$ pour $n = 3$ et $N \in \{3,\dots,50\}$. Quel est le comportement asymptotique de la suite $(p_N)_N$? Démontrer-le.
4. On note X_N la variable aléatoire égale au rang du n-ième pile sur l'événement A_N^c, et égale à $N+1$ sur l'événement A_N. Déterminer la loi de X_N.
5. Représenter la loi de X_N pour les valeurs de $n = 1$ et $N = 10$, de $n = 3$ et $N = 10$ puis de $n = 3$ et $N = 20$.
6. Que peut-on dire de la loi de X_N lorsque N tend vers $+\infty$?
7. On suppose que $N = 100$. Tracer en fonction de n, sur l'intervalle $[1, 50]$, la probabilité r_n d'avoir obtenu au moins n fois pile au cours des $2n$ premiers tirages.

Chapitre 4

Indépendance des variables aléatoires

4.1 Cas de deux variables aléatoires

On dit que deux variables aléatoires X et Y sont *indépendantes* si et seulement si, pour tout couple (x, y) de valeurs possibles de (X, Y),

$$P(X = x \text{ et } Y = y) = P(X = x)P(Y = y)$$

ce qui signifie que les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants.

▶ *Exemple 61.* Il est clair que dans l'exemple du lancer de deux dés, les chiffres donnés par les deux dés sont des v.a. indépendantes.
Par contre, les variables aléatoires X et Y définies dans l'exercice 30 ne sont pas indépendantes puisque $P((X, Y) = (-1, -1)) \neq P(X = -1)P(Y = -1)$.

Comme conséquences immédiates de la définition d'indépendance, on a les résultats suivants :

Proposition 6 *Soit (X, Y) deux variables aléatoires à valeurs dans les ensembles finis ou dénombrables E et F respectivement. Alors,*

(i) X et Y sont indépendantes si et seulement si la loi de (X, Y) est le produit de ses lois marginales : $P_{(X,Y)} = P_X \otimes P_Y$.

(ii) X et Y sont indépendantes si et seulement pour tout $I \subset E$ et $J \subset F$,

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

(iii) Soit f et g deux applications définies sur les ensembles E et F respectivement. Si X et Y sont indépendantes alors $f(X)$ et $g(Y)$ sont aussi indépendantes.

Preuve : (*ii*) Soit $I \subset E$ et $J \subset F$. On a

$$P(X \in I \text{ et } Y \in J) = \sum_{i \in I} \sum_{j \in J} P(X = i \text{ et } Y = j).$$

Comme X et Y sont indépendants, pour tout $(x, y) \in E \times F$,

$$P(X = x \text{ et } Y = y) = P(X = x)P(Y = y).$$

Donc,

$$
\begin{aligned}
P(X \in I \; et \; Y \in J) &= \sum_{i\in I}\sum_{j\in J} P(X=i)P(Y=j). \\
&= (\sum_{i\in I} P(X=i))(\sum_{j\in J} P(Y=j))
\end{aligned}
$$

Mais,

$$
P(X \in I) = \sum_{i\in I} P(X=i) \qquad \text{et} \qquad P(Y \in J) = \sum_{j\in J} P(Y=j).
$$

Donc, $P(X \in I \; et \; Y \in J) = P(X \in I)P(Y \in J)$. L'implication réciproque est immédiate.
(iii) Soit $u \in f(E)$ et $v \in g(F)$. On doit montrer que

$$
P(f(X) = u \; et \; g(Y) = v) = P(f(X) = u)P(g(Y) = v).
$$

Désignons par I_u et J_v respectivement l'ensemble des antécédents de u et v par f et g.

$$
P(f(X) = u \; et \; g(Y) = v) = \sum_{x\in I_u}\sum_{y\in J_v} P(X = x \; et \; Y = y).
$$

Comme X et Y sont indépendants, pour tout $(x,y) \in E \times F$,

$$
\begin{aligned}
P(f(X) = u \; et \; g(Y) = v) &= \sum_{x\in I_u}\sum_{y\in J_v} P(X=x)P(Y=y) \\
&=(\sum_{x\in I_u} P(X=x))(\sum_{y\in J_v} P(Y=y)) \\
&= P(f(X)=u)P(g(Y)=v).
\end{aligned}
$$

□

▷ *Exercice 62.* Une variable aléatoire X peut-elle être indépendante d'elle-même ?

▷ *Exercice 63.* On lance deux dés. On note X et Y les variables aléatoires donnant respectivement les plus grand et plus petit nombres obtenus. Les variables aléatoires X et Y sont-elles indépendantes ?

▷ *Exercice 64.* On effectue 2 tirages successifs sans remise dans une urne contenant n jetons numérotés de 1 à n. On note X_1 le numéro du premier jeton tiré et X_2 le numéro du 2ème jeton tiré.

1. Quelle est la loi de X_1 ?
2. Quelle est la loi du vecteur (X_1, X_2) ?
3. En déduire la loi de X_2 ?
4. Les variables X_1 et X_2 sont-elles indépendantes ?
5. Déterminer la loi de $X_1 + X_2$.

▷ *Exercice 65.* Soit X et Y deux variables aléatoires indépendantes à valeurs dans $E = \{0, \ldots, n\}$ et $F = \{0, \ldots, m\}$ respectivement. On pose $p_i = P(X = i)$ pour $i \in E$ et $q_i = P(Y = j)$ pour $j \in F$. Ecrire en fonction des lois $p = (p_i)$ et $q = (q_j)$:

1. la loi de (X, Y),
2. la loi de $X + Y$.

▷ *Exercice 66.* Montrer que si X et Y sont deux v.a. indépendantes, suivant respectivement les lois $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$, alors $X + Y$ suit la loi $\mathcal{B}(n + m, p)$.

▷ *Exercice 67.* Montrer que si X et Y sont deux v.a. indépendantes, suivant respectivement les lois $\mathcal{P}(a)$ et $\mathcal{P}(b)$, alors $X+Y$ suit la loi $\mathcal{P}(a+b)$.

▷ *Exercice 68.*

1. Montrer que si X_1 et X_2 suivent des lois géométriques de paramètres r_1 et r_2, sur $[0,N]$, alors la v.a. $Z = \inf(X_1, X_2)$[1] suit une loi géométrique de paramètres $r_1 r_2$, sur $[0,N]$.
2. Si maintenant X_1 et X_2 sont des variables aléatoires de lois géométriques de paramètres r_1 et r_2 sur $\mathbb{N}^*$, quelle est la loi de $Z = \inf(X_1, X_2)$?

4.2 Cas de n variables aléatoires

Plus généralement, on dit que *m variables aléatoires $X_1, ..., X_m$ sont indépendantes* si et seulement si, pour tout m-uplet $(x_1, ..., x_m)$ de valeurs possibles de $(X_1, ..., X_m)$,

$$P((X_1, ..., X_m) = (x_1, ..., x_m)) = \prod_{j=1}^{m} P(X_j = x_j)$$

On voit ainsi que des v.a. sont indépendantes si et seulement si leur loi conjointe coïncide avec le produit de ses lois marginales.

Exemple fondamental Lors d'une suite d'épreuves indépendantes, en particulier lors d'une expérience statistique, des variables aléatoires concernant des tirages différents sont indépendantes. Plus précisément, si X_i est une variable aléatoire rendant compte du i-ième tirage (elle est définie sur Ω_i), alors on définit des variables aléatoires indépendantes $\tilde{X}_1, \ldots, \tilde{X}_n$ sur $\Omega_1 \times ... \times \Omega_n$ en posant pour tout $j \in \{1, \ldots, n\}$, et $\underline{\omega} = (\omega_1, \ldots, \omega_n) \in \Omega_1 \times ... \times \Omega_n$, $\tilde{X}_j(\omega_1, \ldots, \omega_n) = X_j(\omega_j)$. Lors d'une expérience statistique, les variables aléatoires $\tilde{X}_i$ ont en plus la même loi. On dit alors qu'elles forment un *échantillon de taille n* de la loi de X_1 (appelé aussi *n-échantillon de X_1*).

N.B. Si les variables aléatoires $X_1, \ldots, X_m$ sont indépendantes alors pour tout $i \neq j$, les variables aléatoires X_i et X_j sont indépendantes. Mais la réciproque est fausse comme le montre l'exemple suivant : Soit X et Y deux variables aléatoires indépendantes prenant les valeurs -1 et 1 avec probabilité $1/2$ chacune. Posons $Z = XY$. Alors, les variables aléatoires X, Y et Z sont deux à deux indépendantes, mais la famille de variables X, Y et Z n'est pas indépendante.

▷ *Exercice 69.* Généraliser les exercices 66 à 68 aux cas de n variables aléatoires indépendantes.

▷ *Exercice 70.* n personnes se répartissent de façon aléatoire dans les r hôtels, $H_1, \ldots, H_r$, de la ville. On désigne par X_i le nombre de personnes qui vont dans le hôtel H_i pour $i \in \{1, \ldots, r\}$.

1. Soit $i \in \{1, \ldots, r\}$. Quelle est la loi de X_i ?
2. Déterminer la loi du vecteur $(X_1, \ldots, X_r)$.
3. Les variables X_i sont-elles deux à deux indépendantes ?

Indépendance d'une famille d'événements n évènements $A_1, ..., A_n$ sont dit *indépendants* si les variables aléatoires $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ sont indépendantes.

N.B.

[1] On note $\inf(a, b)$ le plus petit des deux nombres a et b.

1. Dans le cas $n = 2$, on retrouve avec cette définition que A et B sont indépendants si et seulement si $P(A \cap B) = P(A)P(B)$.

2. Dans le cas $n = 3$, on trouve (exercice) que trois événements A_1, A_2 et A_3 sont indépendants si et seulement si

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

et si de plus ils sont deux à deux indépendants :

$$P(A_1 \cap A_2) = P(A_1)P(A_2),\ P(A_2 \cap A_3) = P(A_2)P(A_3) \text{ et } P(A_3 \cap A_1) = P(A_3)P(A_1).$$

3. Plus généralement, on montre (exercice) que m événements $A_1, A_2, \ldots, A_m$ sont dits *indépendants* si et seulement si la probabilité de l'intersection d'un nombre quelconque d'entre eux est le produit de leurs probabilités.

► *Exemple 71.* Lors du lancer de deux dés, les trois événements {le premier dé donne 1 ou 2}, {le deuxième dé donne 1 ou 2}, {la somme des dés est impaire} sont indépendants.

► *Exemple 72.* Lors du lancer de deux dés, les trois événements {le premier dé donne un nombre impair}, {le deuxième dé donne un nombre pair}, {les résultats des deux dés sont de même parité} ne sont pas indépendants. Par contre, ces événements sont indépendants deux à deux.

▷ *Exercice 73.* Soit $X_1, \ldots, X_n$ n variables aléatoires indépendantes.
Montrer que les événements $\{X_j = x_j\}$ sont indépendants par récurrence sur m en sommant sur toutes les valeurs possibles de x_m.

▷ *Exercice 74.* Montrer que si les événements $A_1, A_2, ..., A_m$ sont indépendants, alors les $m-1$ événements $A_1 \cap A_2,\ A_3, \ldots, A_m$ sont également indépendants.
Plus généralement, montrer qu'on obtient un système de p événements $B_1, B_2, ..., B_p$ indépendants à partir d'une partition de $\{1, 2, ..., m\}$ en p parties, $I_1, I_2, ..., I_p$, en posant $B_k = \cap_{i \in I_k} A_i$ pour tout $k \in \{1, \ldots, p\}$.

▷ *Exercice 75.* On suppose que la probabilité pour qu'un nouveau-né soit un garçon est de $0,55$. Quelle est la probabilité que sur cinq nouveaux-nés d'une clinique, il y ait exactement deux garçons ? (On précisera les hypothèses d'indépendance utilisées)

▷ *Exercice 76.* Pour détecter une cible, on utilise n systèmes de détection, $S_1, \ldots, S_n$ travaillant indépendamment. Chaque système a une probabilité p de détecter la cible. Déterminer la probabilité que l'alarme soit déclenchée dans les cas suivants :

- l'alarme est déclenchée si les n systèmes ont détecté la cible.
- l'alarme est déclenchée si au moins un des systèmes a détecté la cible.
- l'alarme est déclenchée si au moins deux systèmes ont détecté la cible.

▷ *Exercice 77.* Soit $X_1, \ldots, X_n$ un n-échantillon de variables aléatoires de loi de Bernoulli de paramètre p. Quelle est la loi de leur somme $S = X_1 + \cdots + X_n$?

4.3 Indépendance conditionnelle

On dira que deux v.a. X et Y sont indépendantes conditionnellement à un événement A si et seulement si, pour tout couple (x, y) de valeurs possibles de (X, Y),

$$P(\{X = x\} \cap \{Y = y\} \mid A) = P(\{X = x\} \mid A)P(\{Y = y\} \mid A),$$

ce qui signifie que les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants conditionnellement à A. On dira que X et Y sont indépendantes conditionnellement à une v.a. Z si et seulement si elles sont indépendantes conditionnellement à tout événement de la forme $\{Z = z\}$. Ces notions se généralisent de façon évidente au cas de plusieurs variables.

▶ *Exemple 78.* On dispose de six pièces. On note p_i la probabilité que la i-ème pièce tombe sur "pile" pour $i \in \{1, \ldots, 6\}$. On lance un dé puis n fois la pièce dont le numéro a été obtenu avec le dé. Notons Z le nombre obtenu avec le dé et X_i l'indicatrice de l'événement "pile a été obtenu au i-ème lancer" pour $i \in \{1, \ldots, n\}$. Alors les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes conditionnement à Z. Par contre, les variables aléatoires $X_1, \ldots, X_n$ ne sont indépendantes que lorsque $p_1 = \cdots = p_6$.

▷ *Exercice 79.* Montrer que les variables aléatoires X_1 et X_2 de l'exemple 78 sont idépendantes si et seulement si $p_1 = \cdots = p_6$. (*Indications :* On pourra considérer l'application g_r définie sur $\mathbb{R}^r$ par $g_r(x) = \sum_{k=1}^r x_i^2 - \frac{1}{r}(\sum_{k=1}^r x_i)^2$ pour tout $x = (x_1, \ldots, x_r) \in \mathbb{R}^r$ et montrer par récurrence sur r que $g_k \geq 0$ et que $g_k(x) = 0$ seulement si tous les x_i sont identiques)

4.4 Loi conditionnelle

Etant donné un évènement A, la *loi conditionnelle de X sachant A* est donnée par la famille des probabilités conditionnelles $P(\{X = x\}|A)$.

▶ *Exemple 80.* On lance un dé deux fois et on note X la variable aléatoire donnant le nombre obtenu au premier lancer. Déterminons la loi de X conditionnellement à l'événement "la somme des nombres obtenus au cours des deux lancers est 4".
L'ensemble des résultats possibles pour cette expérience aléatoire est $\Omega = \{1, \ldots, 6\}^2$. On munit Ω de la probabilité uniforme notée P. Notons A cet événement. D'après l'exemple 25, $P(A) = \frac{3}{36}$. X est une variable aléatoire à valeur dans $\{1, \ldots, 6\}$. La loi de X conditionnellement à l'événement A est donnée par les probabilités $P(X = i|A) = \frac{P(\{X=i\}\cap A)}{P(A)}$ pour tout $i \in \{1, \ldots, 6\}$. Pour $i \geq 4$, $P(X = i|A) = 0$ et pour $i \in \{1, 2, 3\}$, $P(X = i|A) = \frac{P(\{(i,4-i)\})}{P(A)} = \frac{1}{3}$. Donc, la loi de X conditionnellement à l'événement A est la loi uniforme sur $\{1, 2, 3\}$.

Soit Y une autre variable aléatoire définie sur l'ensemble Ω. Pour tout $y \in Y(\Omega)$ tel que $P(Y = y) > 0$, on peut ainsi définir la loi de X conditionnellement à l'événement $\{Y = y\}$; elle est définie par la famille de probabilités $P(X = x|Y = y)$ pour $x \in X(\Omega)$. L'ensemble de ces lois conditionnelles (appelées lois conditionnelles de X sachant Y) décrivent l'influence de Y sur la valeur de X. La donnée de la loi de Y et des lois conditionnelles de X sachant Y détermine évidemment la loi conjointe de X et de Y (donc la loi de X) et réciproquement. En effet, on a la formule

$$P_{(X,Y)}(x, y) = \begin{cases} P(X = x|Y = y)P_Y(y) & \text{si } P_Y(y) > 0 \\ 0 & \text{si } P_Y(y) = 0 \end{cases}$$

On obtient un autre critère pour l'indépendance de deux variables aléatoires :

Proposition 7 *Deux variables aléatoires X et Y sont indépendantes si et seulement si pour tout y dans le support de la loi de Y, la loi de X conditionnellement à $Y = y$ coïncide avec la loi de X.*

▶ *Exemple 81.* Considérons deux populations E_1 et E_2 contenant des proportions p_1 et p_2 d'individus de grande taille avec $p_1 > p_2$. On effectue l'expérience aléatoire suivante. On choisit au hasard l'une des deux populations de sorte que E_1 soit choisie avec probabilité $q \in]0, 1[$. Puis, on tire au hasard et avec remise deux individus dans la population choisie. Les tailles des deux individus tirés sont des variables aléatoires indépendantes conditionnellement au choix de la population. En général, elles ne sont pas indépendantes.

▷ *Exercice 82.* (Etude de l'exemple 81).

1. Modéliser l'expérience décrite dans l'exemple 81.
2. Calculer la probabilité de tirer un individu de grande taille au premier tirage.
3. Calculer la loi du nombre d'individus de grande taille tirés.
4. Calculer la probabilité de tirer un deuxième individu de grande taille si le premier individu tiré l'était.
5. Calculer la probabilité d'avoir choisi la population E_1 sachant que les deux individus tirés sont de grande taille.

► *Exemple 83.* Reprenons l'exemple 80 et calculons les lois conditionnelles de X relatives à la variable aléatoire S qui donne la somme des deux nombres obtenus au cours des lancers. S est une variable aléatoire à valeurs dans $\{2,\dots,12\}$, sa loi a été calculée à l'exemple 25.
Si $k \in \{2,\dots,7\}$ alors pour tout $i \in \{1,\dots,k-1\}$,

$$P(X=i|S=k) = \frac{P(\{(i,k-i)\})}{P(S=k)} = \frac{1}{k-1}.$$

Si $k \in \{8,\dots,12\}$ alors pour tout $i \in \{k-6,\dots,6\}$,

$$P(X=i|S=k) = \frac{P(\{(i,k-i)\})}{P(S=k)} = \frac{1}{13-k}.$$

Donc la loi de X conditionnellement à l'événement "$S=k$" est la loi uniforme sur $\{1,\dots,k-1\}$ pour tout $k \in \{2,\dots,7\}$ et est la loi uniforme sur $\{k-6,\dots,6\}$ pour tout $k \in \{8,\dots,12\}$.

▷ *Exercice 84.* Soit $\alpha \in [0,1]$. Soit (X,Y) un couple de variables aléatoires à valeurs dans $\{0,1,2\}\times\{0,1\}$. Le tableau suivant donne la probabilité $P\{(X,Y)=(x,y)\}$, pour tout $(x,y) \in \{0,1,2\}\times\{0,1\}$,

$y \setminus x$	0	1	2
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$	0

Déterminer la loi des marginales X et Y, puis les lois conditionnelles de X sachant Y.

▷ *Exercice 85.* Soit X et Y des variables aléatoires à valeurs dans des ensembles au plus dénombrables $E=\{x_i\}_{i\in I}$ et $F=\{y_j\}_{j\in J}$ respectivement. On note

$$p_{ij} = P(\{X=x_i\}\cap\{Y=y_j\}), \qquad p_{i\bullet} = \sum_{l\in J} p_{il} \qquad \text{et} \qquad p_{\bullet j} = \sum_{k\in I} p_{kj}$$

pour tout $(i,j) \in I\times J$.
En utilisant les notations introduites, écrire les lois conditionnelles de X sachant Y et de Y sachant X.

▷ *Exercice 86.* Soit X et Y deux variables aléatoires indépendantes de loi $\mathcal{B}(n,p)$ et $\mathcal{B}(m,p)$ respectivement.
Déterminer les lois conditionnelles de X sachant $X+Y$.

▷ *Exercice 87.* Soit X et Y deux variables aléatoires indépendantes de loi $\mathcal{P}(a)$ et $\mathcal{P}(b)$ respectivement. Déterminer les lois conditionnelles de X sachant $X+Y$.

▷ *Exercice 88.* Soit $n \in \mathbb{N}^*$, $p=(p_1,p_2,p_3)$ une loi de probabilité et $X=(X_1,X_2,X_3)$ un vecteur aléatoire de loi multinomiale $\mathcal{M}(n,p)$.
Quelle est la loi de X_1 ?
Déterminer les lois conditionnelles de (X_2,X_3) sachant X_1.

Chapitre 5

Premières notions de Statistique

5.1 Tests d'hypothèses

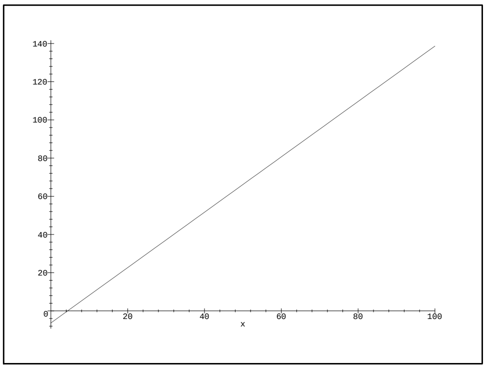
A la base de la méthode statistique se trouve la notion de *test d'hypothèses*. Ces hypothèses concernent la loi d'une variable aléatoire, et on cherche à les établir à partir de l'observation des valeurs prises par cette variable dans une suite finie de tirages indépendants. Considérons pour commencer la méthode du sondage : on cherche à vérifier ou invalider des hypothèses sur la proportion p d'éléments d'un certain type τ dans une population Ω de grande taille, à partir du tirage au hasard avec remise de n éléments parmi lesquels on compte le nombre S_n d'éléments du type τ. S_n est ici appelée la *variable de test* car c'est à partir de sa valeur observée que l'on effectue le test.

Supposons par exemple que l'on sache a priori que p peut prendre deux valeurs, par exemple 2% ou 8% : ce peut être le cas si on connait ces proportions dans deux populations mais qu'on ignore si l'échantillon a été prélevé dans l'une ou l'autre. On appelle Ω la population dans laquelle l'échantillon a été prélevé et on veut préciser si $p = 2\%$ (hypothèse notée H_0) ou $p = 8\%$ (hypothèse notée H_1), à partir de cet échantillon de taille n prélevé dans Ω par une suite de n tirages équitables et indépendants. Remarquons d'emblée qu'aucune certitude absolue ne peut être obtenue par un tel procédé. Même si p vaut 8%, on peut jouer de malchance et ne tirer aucun élément du type τ. Il faudra donc toujours accepter un certain *risque d'erreur.*

On utilisera la notation P_H pour indiquer une probabilité calculée en supposant vraie l'hypothèse H. Il est naturel de se demander si une valeur k de S_n a plus ou moins de chances d'apparaître selon que H_1 ou H_0 est vérifiée. On est ainsi amené à considérer le *rapport des vraisemblances* :

$$\frac{P_{H_1}(S_n = k)}{P_{H_0}(S_n = k)}.$$

Cette expression sera notée $L(k)$. Prenons pour fixer les idées $n = 100$, $L(x) = (\frac{8}{2})^x(\frac{92}{98})^{100-x}$. Son logarithme est une fonction linéaire dont on peut tracer le graphe :



Notons que L est supérieur à 1 pour $k \geq 5$ et inférieur à 1 pour $k < 5$. On serait ainsi conduit à opter pour H_1 lorsque la valeur observée de S_n est supérieure ou égale à 5 et pour H_0 dans le cas contraire. Mais ce choix ne prend pas en compte les risques d'erreur.

Par exemple, $P_{H_1}(S_n \leq 4) = \sum_{k=0}^{4} \binom{100}{k} (0.08)^k (0.92)^{100-k} = 9.0337 \times 10^{-2}$ ce qui signifie que cette règle de choix conduit à 9% d'erreurs lorsque H_1 est vérifiée.

De même, $P_{H_0}(S_n > 4) = 1 - \sum_{k=0}^{4} \binom{100}{k} (0.02)^k (0.98)^{100-k} = 0.05083$, soit 5.1% de risque d'erreur si H_0 est vérifiée. On prendra moins de risque en s'abstenant de conclure lorsque celui-ci est trop élevé. Par exemple, en excluant les valeurs 4 et 5, où d'ailleurs le rapport des vraisemblances est proche de 1, on a :

$$P_{H_1}(S_n \leq 3) = \sum_{k=0}^{3} \binom{100}{k} (0.08)^k (0.92)^{100-k} = 3.6706 \times 10^{-2} \text{ et}$$

$$P_{H_0}(S_n > 5) = 1 - \sum_{k=0}^{5} \binom{100}{k} (0.02)^k (0.98)^{100-k} = 1.5484 \times 10^{-2}.$$

N.B. Si l'on remplace H_1 par $p \geq 8\%$ ou H_0 par $p \leq 2\%$, la procédure sera la même, mais fournira des majorants pour les risques d'erreurs : si l'on observe la valeur 3, on optera pour H_0 avec un risque d'erreur inférieur à 3.7%.

▷ *Exercice 89.* Montrer comment on peut effectuer ce test avec un risque d'erreur de l'ordre de 2%, avec un échantillon de taille 200.

▷ *Exercice 90.* On s'intéresse à la probabilité p d'observer un certain phénotype sur un individu issu d'un croisement. Selon que le phénotype est déterminé par un gène ou par 2 gènes situés sur des chromosomes différents, la probabilité p d'observer ce phénotype sera $p = \frac{3}{4}$ ou $p = \frac{9}{16}$. On réalise 50 croisements indépendants, et on note X le nombre d'individus présentant le phénotype.
A l'aide de l'observation de X, construire un test pour préciser si $p = \frac{3}{4}$ ou $p = \frac{9}{16}$.

5.2 Niveau et puissance des tests.

De façon générale, un test statistique consiste en la donnée de deux hypothèses H_0 et H_1, d'une variable aléatoire observable T appelée *variable de test* dont on connait la loi sous H_0 et d'un *niveau* $\alpha \in [0, 1]$. Le niveau α représente le risque d'erreur (dit aussi *erreur de premier type*) correspondant à un rejet erroné de H_0. Par exemple, si on fixe le niveau du test à 5%, cela

signifie qu'il y a une chance sur 20 que l'on rejette l'hypothèse H_0 alors que celle-ci était la bonne hypothèse.
La variable de test T est choisie de sorte que ses valeurs les plus fréquemment observées soient bien différentes sous l'hypothèse H_1 et sous l'hypothèse H_0.

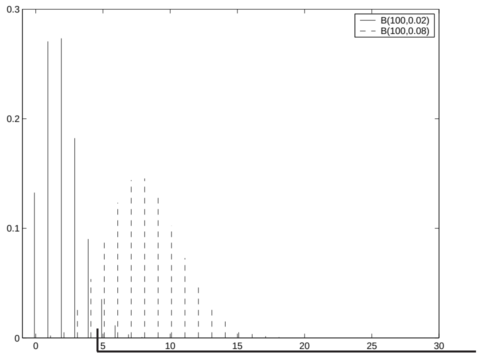
Les deux hypothèses n'apparaissent pas a priori sur un pied d'égalité. L'une notée H_0, sert d'*hypothèse de référence*, H_1 étant alors une *hypothèse alternative*.

► *Exemple 91.* Si l'on s'inquiète des effets du voisinage d'un site industriel sur les maladies respiratoires, l'hypothèse H_0 sera que la probabilité d'être affecté par une maladie respiratoire en résidant dans un voisinage donné de ce site, est la même que dans le reste du pays (c'est-à-dire égale à la fréquence de ces maladies dans l'ensemble de la population). L'hypothèse H_1 sera que la probabilité d'être affecté par une maladie respiratoire en résidant dans un voisinage donné de ce site est supérieure à la fréquence de ces maladies dans l'ensemble de la population.

Si T est la variable de test et si elle a tendance à être plus grande sous l'hypothèse alternative H_1 que sous l'hypothèse de référence H_0, le test de niveau α consiste à rejeter H_0 lorsque la valeur observée t_{obs} de la variable de test T est "exceptionnellement grande"; plus précisément on rejette H_0 lorsque la valeur observée t_{obs} de T est telle que si H_0 est vraie la probabilité que T prenne une valeur aussi grande ou plus grande est inférieure à α : $P_{H_0}(T \geq t_{obs}) \leq \alpha$. Sinon on ne remet pas en cause la validité de H_0. Dans un tel test, l'observation t_{obs} conduit au rejet de H_0 si $t_{obs} \geq a$" où a est la plus petite valeur de $T(\Omega)$ vérifiant $P_{H_0}(T \geq a) \leq \alpha$. L'ensemble $\{T \geq a\}$ est appelée *la zone de rejet de* H_0.
De même, si T a tendance à être plus petite sous l'hypothèse alternative H_1 que sous l'hypothèse de référence H_0, le test de niveau α consiste à rejeter H_0 lorsque la valeur observée t_{obs} de la variable de test T est "exceptionnellement petite" ; plus précisément lorsque la valeur observée t_{obs} est telle que $P_{H_0}(T \leq t_{obs}) \leq \alpha$. Dans un tel test, l'observation t_{obs} conduit au rejet de H_0 si $t_{obs} \leq b$" où b est la plus grande valeur de $T(\Omega)$ vérifiant $P_{H_0}(T \leq b) \leq \alpha$. La zone de rejet de H_0 est donc dans ce cas $\{T \leq b\}$.

Lorsque l'hypothèse alternative H_1 est précise et que l'on peut calculer la loi de T sous H_1, on peut alors déterminer le risque d'erreur (dit de 2ème type) qui correspond à un choix erroné de H_0. Le pourcentage complémentaire est appelé la *puissance* du test : la puissance du test est donc la probabilité de rejeter H_0 avec raison. Par exemple si $T \geq a$ est la zone de rejet de H_0, alors la puissance du test sera $P_{H_1}(T \geq a)$. Remarquons que dans l'exemple du sondage, à moins de de pouvoir augmenter à volonté la taille de l'échantillon, on ne peut pas fixer en même temps l'erreur de 1er type et l'erreur de 2ème type : si l'on réduit le niveau du test, l'erreur de 2ème type augmente nécessairement et donc la puissance du test diminue.

Dans l'exemple du sondage traité ci-dessus, le test de niveau $\alpha = 6\%$, rejette H_0 dès que la valeur observée de S_n est supérieure ou égale à 5. La puissance du test est $P_{H_1}(S_n \geq 5) = 1 - P_{H_1}(S_n \leq 4) = 91\%$. Le test de niveau $\alpha = 2\%$, rejette H_0 dès que la valeur observée de S_n est supérieure ou égale à 6. La puissance du test n'est plus que $P_{H_1}(S_n \geq 6) = 1 - P_{H_1}(S_n \leq 5) = 82\%$.



zone de rejet de H_0 dans le test de niveau 6% de $H_0 : p = 2\%$ contre $H_1 : p = 8\%$

On peut se trouver dans une situation où les hypothèses H_0 et H_1 sont plus complexes par exemple, si H_0 dit que $p \leq 2\%$ et que H_1 dit seulement que p doit être strictement supérieur à 2% : ce sera le cas si l'on cherche à vérifier sur un échantillon que tel procédé de fabrication ne produit pas plus de 2% de pièces défectueuses. Dans cet exemple, pour chaque valeur $p_0 \leq 2\%$, on peut calculer la probabilité de rejeter l'hypothèse H_0 alors que c'est p_0 la bonne valeur du paramètre. L'erreur de premier type est alors définie comme le maximum de ces probabilités : précisement, si R est la zone de rejet de H_0 alors l'erreur de premier type est $\max_{p_0 \leq 2\%} P_{p_0}(R)$ (on utilisera la notation P_{p_0} pour indiquer que la probabilité est calculée avec la valeur p_0 du paramètre). Dans l'exemple du sondage, on peut montrer (voir exercice 94) que

$$\max_{p \leq 2\%} P_p(S_n \geq 5) = P_{p=2\%}(S_n \geq 5)$$

et donc le test de niveau $\alpha = 5\%$ rejettera H_0 si la valeur observée de S_n est supérieure ou égale à 5 et le test de niveau $\alpha = 2\%$ rejettera H_0 si la valeur observée de S_n est supérieure ou égale à 6.

A chaque valeur $p_0 > 2\%$, on peut associer la probabilité de ne pas se tromper en acceptant l'hypothèse H_0, c'est-à-dire $P_{p_0}(R)$. Cela définit une fonction sur $]2\%, 100\%]$ que l'on appelle *la courbe de puissance du test*. Un exemple de courbe de puissance est présentée à la figure 5.1.

N.B. Si l'observation n'est pas dans la zone de rejet de l'hypothèse H_0 pour le test de H_0 : $p \leq 2\%$ contre l'hypothèse $H_1 : p > 2\%$, de niveau $\alpha = 5\%$, il est imprudent de conclure que $p \leq 2\%$. En effet, si le test est effectué en tirant un échantillon de taille 100, la probabilité d'avoir une observation qui ne soit pas dans la zone de rejet de H_0 alors que $p = 5\%$ est $P_{5\%}(S_{100} \leq 4) = 43,6\%$.

► *Exemple 92.* On effectue un sondage avant un référendum : 52% des 1000 personnes interrogées par tirage au sort dans une population ont répondu qu'ils voteraient "oui" au référendum. On pourra effectuer un test afin de savoir si la proportion de personnes interrogées, favorables au "oui" est suffisamment importante pour pouvoir annoncer la victoire probable du "oui" au référendum. Si p est la proportion (inconnue) de personnes décidées à voter pour le "oui" dans la population, on fera un test de l'hypothèse $H_0 : p \leq 0.5$ contre l'hypothèse H_1: $p > 0.5$.

Considérons maintenant un autre exemple. Supposons que les éléments de Ω soient numérotés de 1 à N, N étant inconu. L'hypothèse de référence est $H_0 : N = N_0$ (par exemple $N_0 = 1000$)

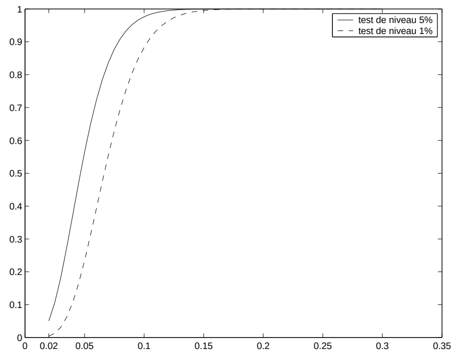


FIG. 5.1 : Courbes de puissances des tests de l'hypothèse $H_0 : p \leq 0.02$ contre l'hypothèse $H_1 : p > 0.02$ de niveau 5% et 1%, dont la variable de test est S_{100} ; la zone de rejet de H_0 est $\{S_{100} \geq 5\}$ pour le test de niveau 5% et $\{S_{100} \geq 7\}$ pour le test de niveau 1%.

et l'hypothèse alternative $H_1 : N < N_0$. A partir d'une suite $\omega = (\omega_1, \dots, \omega_n)$ de n (par exemple 100) tirages indépendants effectués au hasard dans Ω, on note $M_n(\omega)$, le plus grand numéro tiré : $M_n(\omega) = \max\{\omega_i,\ i \in \{1, \dots, n\}\}$. Remarquons que les variables aléatoires X_i définies sur Ω par $X_i(\omega) = \omega_i$, sont des variables aléatoires indépendantes de loi uniforme sur $\Omega = \{1, \dots, N\}$. La loi de M_n sous H_0 est déterminée par sa fonction de répartition : pour tout $k \in \{1, \dots, N_0\}$,

$$P_{H_0}(M_n \leq k) = P_{H_0}(\bigcap_{i=1}^{n}\{X_i \leq k\}) = \prod_{i=1}^{n} P_{H_0}(X_i \leq k) = (\frac{k}{N_0})^n.$$

Sous H_1, M_n a tendance à prendre des valeurs plus petites que sous H_0 car pour tout $k \in \{1, \dots, N\}$, $P_{H_1}(M_n \leq k) = (\frac{k}{N})^n > (\frac{k}{N_0})^n$. Le test de niveau α consistera donc à rejeter H_0 lorsque le plus grand des numéros observés lors des n tirages, noté m_{obs} est tel que $P_{H_0}(M_n \leq m_{obs}) \leq \alpha$, c'est-à-dire $(\frac{m_{obs}}{N_0})^n \leq \alpha$. Autrement dit, on rejettera H_0 si $m_{obs} \leq N_0\alpha^{\frac{1}{n}}$.
Pour $N_0 = 1000$ et $n = 100$, on voit que pour le test de niveau 5%, H_0 ne sera invalidée que si le plus grand numéro tiré est inférieur ou égal à 970. Pour le test de niveau 1%, H_0 sera invalidée si le plus grand numéro tiré est inférieur ou égal à 954.

$\triangleright$ *Exercice 93.* Pendant un jeu de pile ou face, un observateur a compté que la pièce de monnaie était tombée 7 fois sur "face" en 10 lancers. On veut faire un test de l'hypothèse H_0 : la pièce est bien équilibrée, contre l'hypothèse H_1 : la probabilité p que la pièce tombe sur "face" est strictement supérieure à 1/2.

1. Soit Z la variable aléatoire qui à une série de 10 lancers associe le nombre de fois où la pièce est tombée sur "face". Quelle est la loi de Z sous H_0 ? sous H_1 ?
2. Déterminer la zone de rejet de H_0 pour avoir un test de niveau 6 %. Quel est le niveau exact du test ?
3. Quelle est la conclusion de l'observateur avec ce test ?
4. Quelle est la probabilité de conclure avec ce test que la pièce est bien équilibrée alors qu'en réalité $p = 0.6$? alors qu'en réalité $p = 0.7$?

$\triangleright$ *Exercice 94.* On note S_n une variable aléatoire de loi binomiale $\mathcal{B}(n, q)$.

1. Montrer que pour tout $k \in \{0, \dots, n\}$, l'application $q \mapsto P_q(S_n \geq k)$ est croissante sur $[0, 1]$.

Dans l'exemple du sondage, on effectue le test de l'hypothèse $p \leq 0.02$ contre $p > 0.02$ au niveau α.

2. Montrer que la zone de rejet du test est $S_n \geq a$ où a est le plus petit entier tel que $P_{p=0.02}(S_n \geq a) \leq \alpha$.

3. Montrer que la courbe de puissance du test est croissante.
4. La figure 5.1 représente la courbe de puissance du test lorsque $n = 100$ et $\alpha = 0.05$. Utiliser cette courbe pour déterminer approximativement la probabilité, avec ce test, de rejeter l'hypothèse H_0 alors qu'en réalité $p = 4\%$ puis $p = 10\%$.
5. Construire la zone de rejet du test précédent lorsque $n = 200$ et reprendre la question abordée en 4 dans ce cas.

5.3 Tests bilatères

Le test précédent ($H_0 : p \leq 2\%$ contre $H_1 : p > 2\%$) et sa variante ($H_0 : p = 2\%$ contre $H_1 : p > 2\%$) sont appelés des *tests unilatères.* Dans un test bilatère, l'hypothèse H_0 fixe la loi de la variable de test et H_1 est la négation de H_0. Dans l'exemple du sondage, on cherche à vérifier une information sur la valeur du pourcentage p d'éléments d'un certain type dans une population. Si la valeur avancée pour p est p_0, l'hypothèse de référence est $H_0 : p = p_0$ et l'hypothèse alternative pour un test bilatère est $H_1 : p \neq p_0$.
Un test bilatère peut être décrit comme la conjonction de deux tests unilatères ; la somme de leurs niveaux est le niveau du test bilatère.
Dans l'exemple du test $H_0 : p = p_0$ contre $H_1 : p \neq p_0$, pour le sondage, on construit un test bilatère symétrique en rejetant alors H_0 si la valeur observée de S_n notée s_{obs}, apparaît exceptionnellement grande ou exceptionnellement faible c'est-à-dire si

$$P_{H_0}(S_n \geq s_{obs}) \leq \alpha/2 \quad \text{ou} \quad P_{H_0}(S_n \leq s_{obs}) \leq \alpha/2.$$

La zone de rejet de ce test est $\{S_n \leq a \text{ ou } S_n \geq b\}$ où a est le plus grand nombre appartenant à $S_n(\Omega)$ vérifiant $P_{H_0}(S_n \leq a) \leq \frac{\alpha}{2}$ et b est le plus petit nombre appartenant à $S_n(\Omega)$ vérifiant $P_{H_0}(S_n \geq b) \leq \frac{\alpha}{2}$. Le niveau de ce test est bien inférieur ou égal à α puisqu'un rejet erroné de H_0 peut provenir d'une observation ou trop grande, ou trop faible et que les probabilités d'erreurs s'ajoutent (les événements $\{S_n \leq a\}$ et $\{S_n \geq b\}$ sont incompatibles dès que $a < b$).
Par exemple, si l'hypothèse H_0 est " $p = 4\%$" et si $n = 100$, on rejettera H_0 avec un risque d'erreur inférieur à 5% si la valeur observée de S_n est supérieure ou égale à 9 ou égale à 0. En effet,

$$\begin{aligned}
P_{4\%}(S_{100} \leq 1) &= \sum_{k=0}^{1} \binom{100}{k} (0.04)^k (0.96)^{100-k} = 8.7163 \times 10^{-2}, \\
P_{4\%}(S_{100} \leq 0) &= (0.96)^{100} = 0.01687, \\
P_{4\%}(S_{100} \leq 7) &= \sum_{k=0}^{7} \binom{100}{k} (0.04)^k (0.96)^{100-k} = 0.95249 \\
P_{4\%}(S_{100} \leq 8) &= \sum_{k=0}^{8} \binom{100}{k} (0.04)^k (0.96)^{100-k} = 0.98101.
\end{aligned}$$

La courbe de puissance de ce test est représentée à la figure 5.2.

Dans le deuxième exemple (tirage des numéros), on pourrait effectuer le test bilatère de l'hypothèse $H_0 : N = N_0$ contre l'hypothèse $H_1 : N \neq N_0$ en rejetant l'hypothèse H_0 si la valeur observée m_{obs} de M_n est strictement supérieure à N_0 ou si $P_{H_0}(M_n \leq m_{obs}) \leq \alpha$. Ce test est la superposition de deux tests unilatères définis à l'aide de M_n, l'un de niveau 0 et l'autre de niveau α.

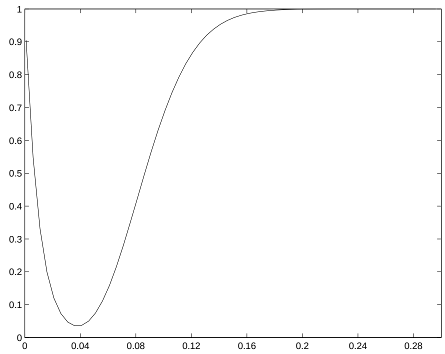


FIG. 5.2 : Courbe de puissance du test bilatère $H_0 : p = 0.04$ contre $H_1 : p \neq 0.04$ de niveau 5%, dont la zone de rejet est $\{S_{100} = 0$ ou $S_{100} \geq 9\}$.

5.4 Domaines de confiance

On considère un paramètre inconnu θ, dont dépend la loi d'une variable de test T ainsi qu'une famille de tests (en général bilatères) définis à l'aide de T. Dans l'exemple du sondage, le paramètre inconnu est la proportion p. Dans l'exemple du tirage de numéros, c'est N l'inconnu. Les tests bilatères de l'hypothèse $H_0 : \theta = \theta_0$ contre l'hypothèse $H_1 : \theta \neq \theta_0$, θ_0 parcourant l'ensemble des valeurs possibles pour le paramètre θ, permettent alors de définir à partir d'un tirage t_{obs} de la variable de test T, un domaine de confiance pour ce paramètre pour un niveau de confiance γ donné : ce domaine est constitué de toutes les valeurs θ_0 du paramètre pour lesquelles l'hypothèse $H_0 : \theta = \theta_0$ est acceptée par le test bilatère de niveau $\alpha = 100\% - \gamma$. On n'a donc pas la certitude que θ se trouve dans ce domaine. Ce domaine est obtenu en excluant toutes les valeurs que l'on peut rejeter avec un risque inférieur à α.
Le domaine aléatoire D_T ainsi défini est tel que pour tout θ, $P_\theta(\theta \in D_T) \geq 100\% - \alpha = \gamma$. En effet, la probabilité que l'hypothèse $\theta = \theta_0$ soit rejetée (c'est-à-dire que $\theta \notin D_T$) par le test $H_0 : \theta = \theta_0$ contre $\theta \neq \theta_0$ de niveau α est par définition inférieure ou égale à α. On peut noter $D_{t_{obs}}$ le domaine de confiance obtenu après observation de T. Notons que cette procédure fournit, en moyenne dans plus de $\gamma\%$ des cas, un domaine qui contient la vraie valeur de θ (voir figure 5.4). Par contre, il est évidemment faux de dire que "D_{obs} a $\gamma\%$ de chance de contenir le paramètre θ", puisque ni D_{obs}, ni θ ne sont aléatoires.
Dans l'exemple du sondage, si s_{obs} est la valeur observée de la variable S_n, le domaine de confiance, constitué de toutes les valeurs p_0 telles que l'hypothèse $p = p_0$ soit acceptée par le test bilatère symétrique, sont celles qui vérifient les inégalités :

$$P_{p_0}(S_n \leq s_{obs}) > \frac{\alpha}{2} \text{ et } P_{p_0}(S_n \geq s_{obs}) > \frac{\alpha}{2}$$

Le domaine de confiance est donc fonction de la valeur observée s_{obs}. Il est clair que le domaine de confiance est d'autant plus étendu que le niveau de confiance requis est grand.
Si par exemple le niveau de confiance requis est de 95%, si $n = 100$ et si la valeur observée de S_n est 3, on cherche les valeurs de p_0 pour lesquelles

$$P_{p_0}(S_{100} \leq 3) > 2.5\% \text{ et } P_{p_0}(S_{100} \geq 3) > 2.5\%$$

c'est-à-dire $f(p_0) > 0.025$ et $g(p_0) > 0.025$, en posant

$$f(x) = \sum_{k=0}^{3} \binom{100}{k} (x)^k (1-x)^{100-k} \text{ et } g(x) = 1 - \sum_{k=0}^{2} \binom{100}{k} (x)^k (1-x)^{100-k}.$$

Les graphes de f et de g sont tracés sur la figure 5.3.

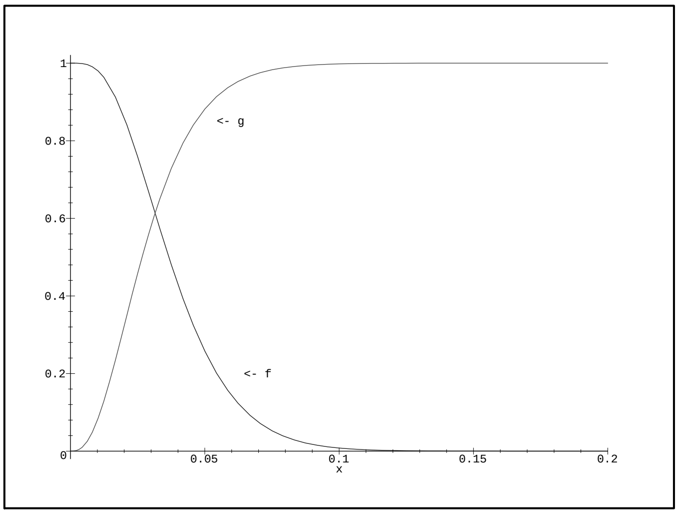


FIG. 5.3 : Graphes des fonctions f et g.

On a $f(0.085) = 2.533 \times 10^{-2}$, $f(0.086) = 2.349 \times 10^{-2}$ $g(0.007) = 0.0336$ et $g(0.006) = 0.0227$. On voit que $]0.6\%,\ 8.6\%[$ est un intervalle de confiance pour la proportion p de niveau de confiance légèrement supérieur à 95% et que $[0.7\%,\ 8.5\%]$ est un intervalle de confiance de niveau légèrement inférieur à 95%.

▷ *Exercice 95.* Vérifier que si la valeur observée de S_n est 4 alors $]1.1\%,\ 10\%[$ est un intervalle de confiance de niveau légèrement supérieur à 95% pour p.

La figure 5.4 montre 50 intervalles de confiance de niveau de confiance 95% obtenus en prélevant 50 fois un échantillon de taille 100 dans une population comportant 8% d'éléments de type τ. Sur les 50 intervalles de confiance, seulement 2 ne contiennent pas le paramètre $p = 8\%$.

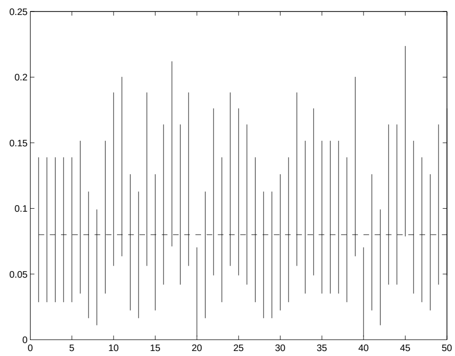


FIG. 5.4 : Intervalles de confiance de niveau de confiance 95% pour la proportion p d'éléments de type τ dans une population. Ces intervalles ont été obtenus en prélevant 50 fois un échantillon de taille 100 dans une population comportant 8% d'éléments de type τ.

N.B. Pour obtenir rapidement un intervalle de confiance pour la proportion p, on peut utiliser l'inégalité suivante établie à la proposition 3

$$P_p(|\frac{S_n}{n} - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2} \text{ pour tout } \epsilon > 0 \text{ et } n \in \mathbb{N}^*.$$

Pour $\alpha \in]0,1[$ et l'observation s_{obs} de S_n, $]\frac{s_{obs}}{n} - \frac{1}{2\sqrt{n\alpha}};\ \frac{s_{obs}}{n} + \frac{1}{2\sqrt{n\alpha}}[$ est un intervalle de confiance de niveau de confiance supérieur ou égal à $1-\alpha$. Pour $n = 100$, $\alpha = 5\%$ et $s_{obs} = 3$, l'intervalle obtenu pour la proportion p est $[0,\ 25.4\%[$, on voit que cette méthode donne un intervalle de confiance beaucoup plus grand qu'avec la méthode précédente.

Dans l'exemple du tirage des numéros, si m_{obs} est le plus grand numéro observé lors des n tirages, le domaine de confiance défini par le test bilatère asymétrique construit précédemment, est constitué de toutes les valeurs N_0 telles que

$$N_0 \geq m_{obs} \text{ et } P_{N_0}(M_n \leq m_{obs}) > \alpha.$$

C'est l'intervalle $[m_{obs},\ m_{obs}\alpha^{-\frac{1}{n}}[$. Par exemple pour $m_{obs} = 970$, l'intervalle de confiance de niveau 95% est $[970,\ 999]$.

▷ *Exercice 96.* Des analyses de dépistage d'un anomalie génétique rare, effectuée sur un échantillon de 1000 individus choisis au hasard dans une population de grande taille ont amené à découvrir un individu porteur de cette anomalie. On note p la proportion de porteurs de l'anomalie dans la population.

1. Tester l'hypothèse $p = 2\%$ contre l'hypothèse $p = 1\%$ au niveau 5%.
2. Donner un domaine de confiance pour p de niveau de confiance 95%.

Indication : on utilisera l'approximation poissonienne.

▷ *Exercice 97.* Des analyses effectuées sur 10000 individus tirés dans une population de grande taille ont montré que 4 d'entre eux sont porteurs d'un virus. Donner des intervalles de confiance de niveau 95% et 99% pour la proportion de la population porteuse du virus. (0n utilisera l'approximation poissonienne).

▷ *Exercice 98.* Dans l'exemple 92, donner un intervalle de confiance pour la proportion p de personnes favorables au “oui” dans la population, de niveau de confiance supérieur à 95%.
Combien de personnes doit-on tirer au sort pour pouvoir donner un encadrement de p à 1% près, de niveau de confiance supérieur à 95% ?

▷ *Exercice 99.* (Suite de l'exercice 94) Montrer que le domaine de confiance pour p défini dans l'exemple du sondage (c'est-à-dire à l'aide de l'observation de la variable S_n) est un intervalle.

Chapitre 6

Espérance et variance d'une variable aléatoire

Dans ce chapitre, Ω désignera un ensemble au plus dénombrable muni d'une probabilité P telle que $P(\omega) > 0$ pour tout $\omega \in \Omega$.

6.1 Espérance

Soit X une variable aléatoire à valeurs positives définie sur (Ω, P). Son *espérance* (on dit aussi sa *moyenne*, surtout lorsque tous les cas sont équiprobables) est par définition la somme sur toutes les issues possibles ω de ses valeurs $X(\omega)$ pondérées par les probabilités $P(\omega)$. Elle est notée $\mathrm{E}(X)$:

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

Lorsque Ω est infini, $\mathrm{E}(X)$ peut être infinie.

▶ *Exemple 100.* Désignons par $\Omega = \{\omega_1, \ldots, \omega_n\}$ l'ensemble des individus d'un groupe et par $X(\omega_i)$ le score, un entier entre 0 et 100, obtenu par la i-ème personne du groupe à une compétition. Munissons Ω de l'équiprobabilité P. Alors $E(X) = \frac{1}{n} \sum_{i=1}^{n} X(\omega_i)$ est la moyenne des scores du groupe.
Remarquons que si on note $x_1 < \ldots < x_r$ les différents scores obtenus dans le groupe, $\sum_{i=1}^{n} X(\omega_i)$ peut aussi s'écrire $\sum_{i=1}^{r} \alpha_i x_i$ avec α_i le nombre de personnes dans le groupe ayant obtenues le score x_i. Comme $\frac{\alpha_i}{n} = P(X = x_i)$, on obtient une autre expression de $E(X)$: $E(X) = \sum_{i=1}^{r} x_i P(X = x_i)$.

Le calcul de l'espérance peut être effectué à l'aide de la loi de X :

Propriétés. *Soit X une variable aléatoire à valeurs positive définie sur (Ω, P).*

- $E(X) = \sum_{x \in X(\Omega)} x P(X = x)$,
- $E(X) = \int_0^{+\infty} P(X \geq t) dt$.

Preuve : Pour établir la première égalité, il suffit de remarquer que la famille des ensembles $(\{X = x\}, x \in X(\Omega))$ définit une partition de Ω. Donc,

$$\mathrm{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{x \in X(\Omega)} \left(\sum_{\omega \in \{X = x\}} X(\omega) P(\omega) \right).$$

Pour tout $x \in X(\Omega)$,

$$\sum_{\omega \in \{X=x\}} X(\omega)P(\omega) = x \sum_{\omega \in \{X=x\}} P(\omega) = xP(X=x).$$

Pour établir la deuxième expression, décomposons $P(X \geq t)$ en $\sum_{x \in X(\Omega)} \mathbb{1}_{\{x \geq t\}} P(X=x)$ dans l'expression de $\int_0^{+\infty} P(X \geq t)dt$ et intervertissons l'intégrale et la somme (cela est possible car tous les termes sont positifs). On obtient

$$\int_0^{+\infty} P(X \geq t)dt = \sum_{x \in X(\Omega)} \int_0^{+\infty} \mathbb{1}_{\{x \geq t\}} dt P(X=x) = \sum_{x \in X(\Omega)} xP(X=x) = E(X).$$

□

▶ *Exemple 101.*

1. Si X est la fonction indicatrice d'un événement $A \subset \Omega$ alors $\mathrm{E}(X) = P(A)$.
2. Si X suit la loi de Poisson $\mathcal{P}(\lambda)$ alors X est intégrable et $\mathrm{E}(X) = \lambda$.

On dit qu'une variable aléatoire à valeurs réelles X définie sur (Ω, P) est intégrable si $\mathrm{E}(|X|)$ est finie. On vérifie immédiatement que :

- Pour tout $X \in \mathcal{L}^1(\Omega, P)$, $E(X) = E(X^+) - E(X^-)$ avec les notations suivantes $x^+ = \max(0, x)$ et $x^{-1} = -\min(0, x)$.
- si Y est une autre variable aléatoire numérique définie sur (Ω, P) alors

$$\mathrm{E}(|X+Y|) \leq \mathrm{E}(|X|) + \mathrm{E}(|Y|),$$

- pour tout $\alpha \in \mathbb{R}$, $\mathrm{E}(|\alpha X|) \leq |\alpha| \mathrm{E}(|X|)$.

L'ensemble $\mathcal{L}^1(\Omega, P)$ des variables aléatoires réelles intégrables est donc un espace vectoriel.

Pour toute variable aléatoire réelle intégrable X, on définit aussi son espérance $\mathrm{E}(X)$ par

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

(la famille $\{X(\omega)P(\omega)\}_{\omega \in \Omega}$ est une famille sommable, donc la somme est bien définie indépendamment de l'ordre des termes).

Propriétés.

1. *Pour tout* $X \in \mathcal{L}^1(\Omega, P)$, $|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$.
2. *L'espérance est linéaire : pour tout* $X, Y \in \mathcal{L}^1(\Omega, P)$ *et* $a \in \mathbb{R}$,

$$\mathrm{E}(X + aY) = \mathrm{E}(X) + a\mathrm{E}(Y).$$

3. *L'espérance est une fonction croissante : si* X *et* Y *sont des variables aléatoires dans* $\mathcal{L}^1(\Omega, P)$ *telles que* $X \leq Y$ *alors* $\mathrm{E}(X) \leq \mathrm{E}(Y)$.

▷ *Exercice 102.* Que peut-on dire d'une variable aléatoire positive définie sur (Ω, P) dont l'espérance est nulle ?

▷ *Exercice 103.* Soit X une variable aléatoire numérique définie sur (Ω, P) et soit a et b deux réels tels que $a \leq X(\Omega) \leq b$. Montrer que X est intégrable et que $a \leq \mathrm{E}(X) \leq b$. A quelle condition a-t-on $E(X) = a$?

▷ *Exercice 104.* On définit l'application $||\cdot||_1$ sur $\mathcal{L}^1(\Omega, P)$ par $||X||_1 = \mathrm{E}(|X|)$ pour tout $X \in \mathcal{L}^1(\Omega, P)$. Vérifier que $||\cdot||_1$ définit une norme sur $\mathcal{L}^1(\Omega, P)$ et montrer que $\mathcal{L}^1(\Omega, P)$ muni de cette norme est un espace de Banach (c'est-à-dire toute suite de Cauchy dans $(\mathcal{L}^1(\Omega, P), ||\cdot||_1)$ converge).

Comme dans le cas d'une variable aléatoire à valeurs positives, on peut exprimer l'expérance de X à l'aide de la loi de X.

Proposition 8 *Pour une variable aléatoire Z quelconque définie sur (Ω, P) et pour toute fonction numérique ϕ définie sur $Z(\Omega)$, on a :*

$$\mathrm{E}(|\phi(Z)|) = \sum_{z \in Z(\Omega)} |\phi(z)| P_Z(z) \tag{6.1}$$

et donc, $\phi(Z)$ est intégrable si et seulement si $\phi \in \mathcal{L}^1(Z(\Omega), P_Z)$. Dans ce cas,

$$\mathrm{E}(\phi(Z)) = \sum_{z \in Z(\Omega)} \phi(z) P_Z(z) \tag{6.2}$$

En particulier, si $X \in \mathcal{L}^1(\Omega, P)$ alors $\mathrm{E}(X) = \sum_{x \in X(\Omega)} x P_X(x)$.

Preuve : La famille des ensembles $(\{Z = z\}, z \in Z(\Omega))$ définit une partition de Ω. Donc,

$$\mathrm{E}(|\phi(Z)|) = \sum_{\omega \in \Omega} |\phi(Z(\omega))| P(\omega) = \sum_{z \in Z(\Omega)} \left(\sum_{\omega \in \{Z = z\}} |\phi(Z(\omega))| P(\omega) \right).$$

Pour tout $z \in Z(\Omega)$,

$$\sum_{\omega \in \{Z = z\}} |\phi(Z(\omega))| P(\omega) = |\phi(z)| \sum_{\omega \in \{Z = z\}} P(\omega) = |\phi(z)| P(Z = z).$$

Donc, $\mathrm{E}(|\phi(Z)|) = \sum_{z \in Z(\Omega)} |\phi(z)| P_Z(z)$.
Lorsque $\phi(Z)$ est intégrable, $\{\phi(Z(\omega)) P(\omega)\}_{\omega \in \Omega}$ est une famille sommable. Donc, les égalités précédentes sont encore vraies sans les valeurs absolues, ce qui permet d'établir (6.2). □

N.B.

1. On voit que l'intégrabilité et l'espérance d'une variable aléatoire ne dépendent que de la loi de cette variable aléatoire.

2. La question de l'intégrabilité ne se pose pas lorsque $X(\Omega)$ est un ensemble fini.

3. Pour $k \in \mathbb{N}$, lorsque X^k est intégrable, $\mathrm{E}(X^k)$ est appelé le moment d'ordre k de X.

4. On définit l'espérance d'un vecteur aléatoire $X = (X_1, \ldots, X_d)$ comme le vecteur $\mathrm{E}(X) = (\mathrm{E}(X_1), \ldots, \mathrm{E}(X_d))$.

► *Exemple 105.*

1. Si X est une variable aléatoire de loi uniforme sur l'ensemble $\{x_1, \ldots, x_n\}$, son espérance est $\mathrm{E}(X) = \frac{1}{n}\sum_{i=1}^n x_i$.
2. Si X est la fonction indicatrice d'un événement $A \subset \Omega$ alors $\mathrm{E}(X) = P(A)$.
3. Si X suit la loi de Poisson $\mathcal{P}(\lambda)$ alors X est intégrable et $\mathrm{E}(X) = \lambda$.
4. Si X suit la loi binomiale $\mathcal{B}(n,p)$ alors $\mathrm{E}(X) = np$ (ce résultat s'obtient sans calcul en remarquant qu'une somme de n variables aléatoires de loi de Bernoulli $\mathcal{B}(p)$ indépendantes suit la loi $\mathcal{B}(n,p)$).

▷ *Exercice 106.* Appliquer la proposition 8 pour montrer que dans l'exercice 25, $E(S) = 7$.

▷ *Exercice 107.* Soit $n \in \mathbb{N}^*$. Montrer que si X^n est intégrable alors pour tout $k \in \{1, \ldots, n\}$, X^k est intégrable.

▷ *Exercice 108.* Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Montrer que l'on a

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} P(X \geq k)\,.$$

Application : déterminer l'espérance de la loi $\mathcal{G}_{\mathbb{N}^*}(p)$.

▷ *Exercice 109.* Soit $A_i,\ i \in \{1, \ldots, n\}$ une famille finie d'événements. Etablir, à partir de la formule du crible pour les fonctions indicatrices

$$\mathbb{1}_{\overset{n}{\underset{i=1}{\cup}} A_i} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{1}_{A_{i_1} \cap \ldots \cap A_{i_k}}\,,$$

l'identité suivante :

$$P(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Simplifier la formule lorsque pour tout $k \in \{1, \ldots, n\}$ et $1 \leq i_1 < \ldots < i_k \leq n$,

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_1 \cap \ldots \cap A_k).$$

Application : soit p et n deux entiers strictement positifs. On répartit au hasard p jetons numérotés de 1 à p sur un tableau constitué de n cases numérotés de 1 à n. Chaque jeton est placé sur une case et chaque case peut recevoir plusieurs jetons.

1. Décrire l'ensemble Ω des répartitions possibles des p jetons et la probabilité qu'une répartition donnée se réalise au cours d'une telle expérience aléatoire.
2. Soit $i \in \{1, \ldots, n\}$. Déterminer la probabilité que la i-ème case reste vide.
3. Déterminer la probabilité qu'au moins une case du tableau reste vide.
4. En moyenne, combien de cases restent vides ?

6.2 Variance

On note $\mathcal{L}^2(\Omega, P)$, l'espace des variables aléatoires réelles de carré intégrable c'est-à-dire telles que $\mathrm{E}(X^2)$ soit fini. Les inégalités :

$$(X+Y)^2 \leq 2(X^2 + Y^2) \text{ et } 2|XY| \leq X^2 + Y^2$$

et la linéarité de l'espérance montrent que $\mathcal{L}^2(\Omega, P)$ est un espace vectoriel et que le produit de deux variables aléatoires de carré intégrable et une variable aléatoire intégrable. Les constantes étant de carré intégrable, ce dernier point montre que les variables aléatoires de carré intégrable sont intégrables : $\mathcal{L}^2(\Omega, P) \subset \mathcal{L}^1(\Omega, P)$.

▷ *Exercice 110.* Montrer que l'application $(X,Y) \mapsto \mathrm{E}(XY)$ définie sur $\mathcal{L}^2(\Omega,P)$ est un produit scalaire. La norme associée est notée $||\cdot||_2$. Montrer que $(\mathcal{L}^2(\Omega,P),||\cdot||_2)$ est un espace de Hilbert. Enoncer l'inégalité de Cauchy-Schwarz dans $(\mathcal{L}^2(\Omega,P),||\cdot||_2)$. En déduire que pour tout $X \in \mathcal{L}^2(\Omega,P)$, $\mathrm{E}(|X|) \leq \sqrt{E(X^2)}$.

On définit la *variance* d'une v.a. réelle $X \in \mathcal{L}^2(\Omega,P)$ par la formule :

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) = \sum_{x \in X(\Omega)} (x - \mathrm{E}(X))^2 P(X = x)$$

ou de façon équivalente par la linéarité de l'espérance,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$$

La variance mesure l'amplitude moyenne des fluctuations de la variable autour de sa moyenne. Remarquons que l'espérance de X est la constante qui approche le mieux la variable aléatoire X au sens suivant :

Propriété. *La fonction* $a \mapsto \mathrm{E}((X-a)^2)$ *atteint son minimum en* $\mathrm{E}(X)$*. De plus, pour tout* $a \in \mathbb{R}$, $\mathrm{E}((X-a)^2) = \mathrm{Var}(X) + (\mathrm{E}(X)-a)^2$.

Preuve : En introduisant $E(X)$ et en utilisant la linéarité de l'espérance, on obtient :

$$\begin{aligned} \mathrm{E}((X-a)^2) &= \mathrm{E}(((X-\mathrm{E}(X)) + \mathrm{E}(X) - a)^2) \\ &= \mathrm{E}((X-\mathrm{E}(X))^2) + (\mathrm{E}(X)-a)^2 + 2(\mathrm{E}(X)-a)\mathrm{E}(X-\mathrm{E}(X)) \\ &= \mathrm{Var}(X) + (\mathrm{E}(X)-a)^2. \end{aligned}$$

□

▷ *Exercice 111.* Vérifier à partir de la linéarité de l'espérance, l'égalité des deux expressions de la variance.

On voit sur la première expression de la variance que :

Propriétés.

- *la variance de X est positive ou nulle,*
- *la variance de X est nulle si et seulement si X est constante.*
- *si* $a \in \mathbb{R}$, *alors* $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.

▶ *Exemple 112.*

1. si X est de loi uniforme sur l'ensemble $\{x_1,\ldots,x_n\}$ alors $\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2$.
2. si X est la fonction indicatrice d'un événement A, alors $\mathrm{Var}(X) = P(A)(1-P(A))$.
3. Si X suit la loi de Poisson $\mathcal{P}(\lambda)$ alors X est de carré intégrable et $\mathrm{Var}(X) = \lambda$.

L'*écart-type* de la v.a. X, notée $\sigma(X)$, est par définition la racine carrée de la variance. Il s'exprime dans les mêmes unités que la variable aléatoire considérée.

Une v.a. d'espérance nulle est dite *centrée*. Une v.a. de variance 1 est dite *réduite*.

▷ *Exercice 113.* Soit Z une variable aléatoire de variance non nulle, vérifier que $\widetilde{Z} = \frac{Z-\mathrm{E}(Z)}{\sigma(Z)}$ est une v.a. centrée réduite.

▷ *Exercice 114.* Soit X une variable aléatoire réelle définie sur (Ω, P). On dit que m est une médiane pour X si $P(X < m) \leq \frac{1}{2} \leq P(X \leq m)$.

1. Soit $x_1, \ldots, x_n$ n réels distincts. Déterminer les médianes de la loi uniforme sur l'ensemble $\{x_1, \ldots, x_n\}$.

Supposons que $X \in \mathcal{L}^1(\Omega, P)$.

1. Montrer que si m est une médiane pour X alors pour tout $a \in \mathbb{R}$, $\mathrm{E}(|X - a|) \geq \mathrm{E}(|X - m|)$.
2. Montrer que si m_1 et m_2 sont deux médianes pour X alors $\mathrm{E}(|X - m_1|) = \mathrm{E}(|X - m_2|)$.
3. Supposons que $X \in \mathcal{L}^2(\Omega, P)$. Montrer que $\mathrm{E}(|X - m|) \leq \sigma(X)$.

6.3 Inégalités de Markov et de Bienaymé-Tchebychev

On a vu que la variance mesure l'amplitude moyenne des fluctuations de la variable autour de sa moyenne. L'inégalité de Bienaymé-Tchebychev va permettre de préciser cette propriété.

Proposition 9 *Si X est une variable aléatoire positive, alors pour tout $\varepsilon > 0$,*

$$P(X \geq \varepsilon) \leq \frac{\mathrm{E}(X)}{\varepsilon} \qquad \textit{(inégalité de Markov).}$$

En particulier, si Y est une variable aléatoire de carré intégrable, alors pour tout $\varepsilon > 0$,

$$P(|Y - \mathrm{E}(Y)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2} \qquad \textit{(inégalité de Bienaymé-Tchebychev).}$$

Preuve : De façon évidente, pour toute variable aléatoire X positive, $\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}} \leq X$. En prenant l'espérance de chaque membre de l'inégalité, on obtient $\varepsilon P(X \geq \varepsilon) \leq \mathrm{E}(X)$.
En appliquant cette inégalité à la variable aléatoire $X = (Y - \mathrm{E}(Y))^2$, on trouve l'inégalité de Bienaymé-Tchebychev. □

En utilisant l'inégalité de Bienaymé-Tchebychev, on peut déduire que si X est une variable aléatoire de carré intégrable et non constante alors plus de 88% des valeurs observées de X se trouvent dans l'intervalle $[\mathrm{E}(X) - 3\sigma(X), \mathrm{E}(X) + 3\sigma(X)]$. Plus généralement, pour tout $a > 0$,

$$P(|X - \mathrm{E}(X)| \geq a\sigma(X)) \leq \frac{1}{a^2}.$$

▷ *Exercice 115.* On suppose que le nombre de pièces sortant d'une usine donnée en une journée est une variable aléatoire d'espérance 50.

1. Peut-on estimer la probabilité que la production de demain dépasse 75 pièces ?
2. Que peut-on dire de plus sur cette probabilité si on sait que l'écart-type de la production quotidienne est de 5 pièces ?

▷ *Exercice 116.* Pour étudier les particules émises par une substance radioactive, on dispose d'un détecteur. On note X la variable aléatoire représentant le nombre de particules qui atteignent le détecteur pendant un intervalle de temps Δt. Le nombre maximal de particules que le détecteur peut compter pendant un intervalle de temps Δt est de 10^3.

1. On suppose que X a pour espérance 10^2. Donner une majoration de la probabilité que X dépasse 10^3.

2. On suppose de plus que X a une variance égale à 10^2. Donner une majoration plus fine de la probabilité que X dépasse 10^3.
3. On suppose maintenant que X suit une loi de Poisson de paramètre m. Soit $a > 1$.
   (a) Montrer que pour tout $\theta > 0$, $P(X > am) \leq E(e^{\theta(X-am)})$.
   (b) En déduire que $P(X > am) \leq e^{-m(1-a+a\ln(a))}$.
   (c) Donner un majorant de la probabilité que X dépasse 10^3 si X suit une loi de Poisson d'espérance 10^2.

6.4 Covariance

La variance est une forme quadratique sur l'espace vectoriel des variables aléatoires réelles de carré intégrable. La *covariance* de deux variables aléatoires X et Y de carré intégrable, notée $\mathrm{Cov}(X,Y)$ est la forme bilinéaire symétrique associée, c'est-à-dire telle que $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$. Elle est donnée par les formules :

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

et on a

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) \pm 2\mathrm{Cov}(X,Y) + \mathrm{Var}(Y) \tag{6.3}$$

▷ *Exercice 117.* Vérifier à partir de la linéarité de l'espérance, l'égalité des deux expressions de la covariance, ainsi que la formule (8.1).

▷ *Exercice 118.* Soit A et B deux événements de probabilité non nulle. On note $X = \mathbb{1}_A$ et $Y = \mathbb{1}_B$. Montrer que $\mathrm{Cov}(X,Y) > 0$ si et seulement si $P(A|B) > P(A)$.

▷ *Exercice 119.* Soit $X_1, \ldots, X_n$ des variables aléatoires de carré intégrable. On pose $S_n = X_1 + \cdots + X_n$. Montrer que

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{Cov}(X_i, X_j)$$

N.B. On définit la *matrice de covariance* d'un vecteur aléatoire de carré intégrable $\overrightarrow{Z} = (Z_1, ..., Z_d)$, comme la matrice $d \times d$ dont le coefficient (i,j) est $\mathrm{Cov}(Z_i, Z_j)$.

▷ *Exercice 120.* Dans l'exercice 109, déterminer la variance de la variable aléatoire donnant le nombre de cases vides du tableau après la répartition des jetons.

▷ *Exercice 121.* Montrer que la matrice de covariance d'un vecteur aléatoire $\overrightarrow{Z} = (Z_1, ..., Z_d)$ est une matrice symétrique positive.
Soit A une matrice $d \times d$. Déterminer la matrice de covariance du vecteur aléatoire $\overrightarrow{Y} = A\overrightarrow{Z}$ en fonction de celle de $\overrightarrow{Z}$.

Proposition 10 *La covariance de deux variables aléatoires indépendantes de carré intégrable est nulle et la variance de leur somme est égale à la somme de leurs variances.*

Preuve : Considérons deux v.a. indépendantes X et Y de carré intégrable. La première assertion signifie que $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$. Le premier membre s'écrit, en appliquant la proposition (8) au vecteur aléatoire (X,Y) et à la fonction $\phi(x,y) = xy$,

$$\sum_{x\in X(\Omega), y\in Y(\Omega)} xyP((X,Y) = (x,y)).$$

Le deuxième membre s'écrit

$$(\sum_{x\in X(\Omega)} xP(X=x))(\sum_{y\in Y(\Omega)} yP(Y=y)).$$

On conclut par la définition de l'indépendance des variables aléatoires.
La deuxième assertion procède directement de la première par l'identité (8.1). □

N.B. Le fait que deux variables aient une covariance nulle n'entraîne pas qu'elles sont indépendantes comme le montre l'exemple suivant :

▶ *Exemple 122.* Soit X et Y deux variables aléatoires indépendantes de loi de Bernoulli de paramètre $1/2$. Les variables aléatoires $X+Y$ et $|X-Y|$ ont une covariance nulle et elles ne sont pas indépendantes.

Corollaire 11 *Si* $(X^{(1)},...,X^{(n)})$ *est un échantillon de la variable aléatoire* X *de carré intégrable et si l'on pose* $S_n = X^{(1)} + ... + X^{(n)}$*, alors* $\mathrm{E}(S_n) = n\mathrm{E}(X)$ *et* $\mathrm{Var}(S_n) = n\mathrm{Var}(X)$.

Preuve : Il suffit d'observer que S_n est la somme de n variables aléatoires indépendantes et d'appliquer la linéarité de l'espérance ainsi que la proposition précédente. □

▶ *Exemple 123.* Soit (Λ, Q) un espace de probabilité. Soit $C_1,\dots,C_d$ une partition de Λ en événements de probabilité $p_i = Q(C_i)$ pour $i\in\{1,\dots,d\}$. La matrice de covariance du vecteur $X = (\mathbb{1}_{C_1},\dots,\mathbb{1}_{C_d})$ est la matrice $K = (K_{i,j})_{(i,j)\in\{1,\dots,d\}^2}$ définie par $K_{i,j} = \begin{cases} p_i(1-p_i) & \text{si } i=j \\ -p_ip_j & \text{si } i\neq j \end{cases}$
D'après la proposition 21 et la section 3.6.3, il en résulte que la matrice de covariance d'un vecteur aléatoire de loi multinomiale $\mathcal{M}(n,(p_1,\dots,p_d))$ est nK.
En particulier, l'espérance et la variance d'une variable aléatoire de loi binomiale $\mathcal{B}(n,p)$ sont respectivement np et $np(1-p)$.

L'inégalité de Schwarz donne une majoration de la covariance de deux variables aléatoires :

Propriétés. *Soit* X *et* Y *deux v.a. réelles de carré intégrable définies sur* (Ω, P).
On a $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$ *et l'égalité est réalisée si et seulement si il existe* $\lambda\in\mathbb{R}$ *et* $c\in\mathbb{R}$ *tels que* $Y = \lambda X + c$.

Si X et Y sont des variables aléatoires de variances non nulles, on définit le *coefficient de corrélation de* X *et de* Y comme étant le rapport

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

C'est un réel dont la valeur absolue est majorée par 1 avec égalité si et seulement si $\exists\lambda\in\mathbb{R}^*$ et $c\in\mathbb{R}$ tels que $Y=\lambda X + c$ (dans ce cas, le signe de $\mathrm{Corr}(X,Y)$ est donné par celui de λ).

▶ *Exemple 124.* Soit X et U deux variables aléatoires indépendantes centrées d'écart-type σ non nul. Notons Y la variable aléatoire définie par $Y = aX + \sqrt{1-a^2}U$. Alors, le vecteur $Z = (X,Y)$ est centré, de matrice de covariance $\Gamma = \begin{pmatrix} \sigma^2 & a\sigma^2 \\ a\sigma^2 & \sigma^2 \end{pmatrix}$. En particulier, la corrélation entre X et Y est donnée par a.
Sur la figure 8.1 sont représentées, pour différentes valeurs de a, 100 réalisations du vecteur aléatoire Z lorsque X et U suivent la loi uniforme sur $\{-10,-9,\dots,10\}$.

▷ *Exercice 125.* Pour une élection, une population de n individus a eu à choisir entre voter pour le candidat A ou le candidat B. On note m le nombre de personnes qui a voté pour A. On interroge au hasard k individus différents dans cette population ($1\leq k\leq N$).

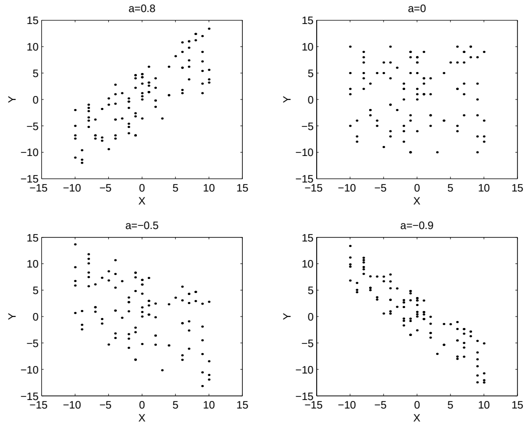


FIG. 6.1 : Nuages de points constitués chacun de 100 réalisations du couple (X,Y) lorsque $Y = aX + \sqrt{1-a^2}U$, X et U étant des variables aléatoires indépendantes de loi uniforme sur $\{-10,-9,\dots,9,10\}$.

1. On désigne par $a_1,\dots,a_m$ les m personnes qui ont voté pour A. Pour $i \in \{1,\dots,m\}$, on note X_i l'indicatrice de l'événement "la personne a_i est interrogée". Quelle est la loi de X_i ?
2. Déterminer l'espérance ainsi que la matrice de covariance de $(X_1,\dots,X_m)$.
3. On pose $S = \sum_{i=1}^m X_i$. Quelle est la loi de S ?
4. Déterminer le nombre moyen de personnes votant pour A sur k personnes tirées au hasard.
5. Déterminer la variance de S.

6.5 Régression linéaire et espérance conditionnelle

6.5.1 Espérance conditionnelle

Etant données deux variables aléatoires X (quelconque) et Y (réelle) de carré intégrable définies sur (Ω, P), on peut chercher (par exemple lorsque X peut être observée plus facilement que Y), à approcher Y par une fonction de X (que nous noterons ϕ) le mieux possible "aux sens des moindres carrés", c'est-à-dire de manière à ce que l'erreur quadratique $\mathrm{E}((Y-\phi(X))^2)$ soit minimale.

Si l'on connait la loi conjointe de X et de Y, on peut déterminer ϕ à l'aide des lois conditionnelles.

Proposition 12 *Posons* $\phi(x) = \mathrm{E}(Y \mid X = x) = \sum_{y \in Y(\Omega)} yP(Y = y \mid X = x)$ *pour tout* $x \in X(\Omega)$ *et notons* $\mathcal{F}_X = \{f(X),\ f \in \mathcal{L}^2(X(\Omega), P_X)\}$.
Alors, $\phi(X) \in \mathcal{L}^2(\Omega, P)$. *C'est la projection orthogonale de* Y *sur le sous-espace vectoriel fermé* $\mathcal{F}_X$ *de* $(\mathcal{L}^2(\Omega,P), ||\cdot||_2)$.
En particulier,

$$\inf_{Z\in\mathcal{F}_X} \mathrm{E}((Y-Z)^2) = \mathrm{E}((Y-\phi(X))^2) = \mathrm{E}(Y^2) - \mathrm{E}(\phi(X)^2)$$

Preuve : Comme $Y \in \mathcal{L}^2(\Omega,P)$, Y est intégrable et donc $\mathrm{E}(|Y||X=x)$ est finie. Par conséquent, $\phi(x)$ est bien définie et on a $\phi(x) = \mathrm{E}(Y 1\!\!1_{\{X=x\}})(P(X=x))^{-1}$. D'après l'inégalité de Cauchy-

Schwarz, $(\mathrm{E}(Y\mathbb{1}_{\{X=x\}}))^2 \leq \mathrm{E}(Y^2\mathbb{1}_{\{X=x\}})P(X=x)$. Donc,

$$\mathrm{E}(\phi(X)^2) = \sum_{x\in X(\Omega)} \phi(x)^2 P(X=x) \leq \sum_{x\in X(\Omega)} \mathrm{E}(Y^2\mathbb{1}_{X=x}) = \mathrm{E}(Y^2) < +\infty.$$

La projection orthogonale $\pi(Y)$ de Y sur $\mathcal{F}_X$ pour le produit scalaire $\langle\cdot,\cdot\rangle\ :\ (X,Y)\mapsto \mathrm{E}(XY)$ est le vecteur de $\mathcal{F}_X$ qui vérifie $\langle Y-\pi(Y), f(X)\rangle = 0$ pour tout $f\in\mathcal{L}^2(X(\Omega),P_X)$. Calculons donc $\mathrm{E}(Yf(X))$ pour $f\in\mathcal{L}^2(X(\Omega),P_X)$:

$$\begin{aligned}
\mathrm{E}(Yf(X)) &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} yf(x)P(X=x \text{ et } Y=y)\\
&= \sum_{x\in X(\Omega)}\left(\sum_{y\in Y(\Omega)} yP(Y=y|X=x)\right) f(x)P(X=x)\\
&= \mathrm{E}(\phi(X)f(X))
\end{aligned}$$

Donc, $\pi(Y) = \phi(X)$. La dernière relation est la traduction d'une propriété des projections orthogonales :

$$\text{pour tout } Z\in\mathcal{F}_X,\ ||Y-Z||_2^2 = ||Y-\pi(Y)||_2^2 + ||\pi(Y)-Z||_2^2 \text{ (théorème de Pythagore)}$$

qui s'obtient en remarquant que $\langle Y-\pi(Y), \pi(Y)-Z\rangle = 0$. Donc, $||Y-Z||_2^2 \geq ||Y-\pi(Y)||_2^2$ avec égalité si et seulement si $Z=\pi(Y)$.
Enfin, $||Y-\pi(Y)||_2^2 = ||Y||_2^2 - ||\pi(Y)||_2^2$ en prenant $Z=0$ dans le théorème de Pythagore. □

La variable aléatoire $\phi(X)$ est notée $\mathrm{E}(Y|X)$ et est appelée *l'espérance conditionnelle de Y sachant X*.

N.B. Lorsque Y est simplement intégrable, on peut encore définir l'espérance conditionnelle de Y sachant X comme étant la variable aléatoire $\phi(X)$ où ϕ est l'application définie par $\phi(x) = \mathrm{E}(Y|X=x)$ pour tout $x\in X(\Omega)$. Mais, $\mathrm{E}(Y|X)$ ne peut plus s'interpréter comme une projection orthogonale.

▷ *Exercice 126.* Soit X et Y des variables aléatoires indépendantes de loi respectivement $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Déterminer $\mathrm{E}(X|X+Y)$.

▷ *Exercice 127.* Soit (X,Y,Z) une variable aléatoire de loi trinomiale $\mathcal{M}(n,(p_1,p_2,p_3))$. Déterminer $\mathrm{E}(Y|X)$.

▷ *Exercice 128.* Soit X et Y des variables aléatoires réelles définies sur (Ω,P) de carré intégrable.

1. Montrer que $\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}(Y)$.
2. Montrer que si X et Y sont indépendants alors $\mathrm{E}(Y|X) = \mathrm{E}(Y)$.
3. Montrer que si $\mathrm{E}(Y|X)$ est la variable aléatoire constante égale à $\mathrm{E}(Y)$ alors $\mathrm{Cov}(X,Y)=0$.
4. Soit U et V deux variables aléatoires indépendantes de loi $\mathcal{B}(p)$ avec $p\in]0,1[$. On pose $X=U-V$ et $Y=U+V$. Déterminer la covariance entre X et Y, puis $\mathrm{E}(Y|X)$.
5. Montrer que pour tout $g\in\mathcal{F}_X$, $\mathrm{E}(Yg(X)|X) = g(X)\mathrm{E}(Y|X)$.

6.5.2 Regression linéaire

Soit (X,Y) un couple de variables aléatoires réelles de carré intégrable, dont on suppose seulement connues les espérances et la matrice des covariances. On peut alors trouver la meilleure approximation (au sens de l'erreur quadratique[1]) de Y par une fonction affine de X.

Proposition 13 *Supposons la v.a. X non constante. La meilleur approximation (au sens des moindres carrés) de Y par une fonction affine de X est donnée par la variable aléatoire* $\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X-\mathrm{E}(X))+\mathrm{E}(Y)$*, l'erreur quadratique étant égale à* $\mathrm{Var}(Y)(1-\mathrm{Corr}(X,Y)^2)$.

Autrement dit, l'application $\psi\ :(a,b)\mapsto E((Y-a-bX)^2)$ définie sur $\mathbb{R}^2$ admet un unique minimum, le point de coordonnées $\hat{a}=\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$ et $\hat{b}=\mathrm{E}(Y)-\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\mathrm{E}(X)$. La valeur minimale de ψ est $\psi(\hat{a},\hat{b})=\mathrm{Var}(Y)(1-\mathrm{Corr}(X,Y)^2)$.

Preuve : Notons V_X le sous-espace vectoriel de $\mathcal{L}^2(\Omega,P)$ engendré par 1 et X. La meilleur approximation de Y par une fonction affine de X est donnée par la projection orthogonale $\pi(Y)$ de Y sur V_X. V_X est un espace de dimension 2 qui admet pour base orthonormale $(1,\frac{X-\mathrm{E}(X)}{\sigma(X)})$. Donc, $\pi(Y)=\langle Y,1\rangle 1+\langle Y,X-\mathrm{E}(X)\rangle\frac{(X-\mathrm{E}(X))}{\mathrm{Var}(X)}=\mathrm{E}(Y)+\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X-\mathrm{E}(X))$. □

► *Exemple 129.* Dans l'exemple 193, $\mathrm{E}(Y|X)=aX$ et l'erreur quadratique est $\mathrm{E}((Y-aX)^2)=(1-a^2)\sigma^2$. Sur les nuages de points représentant 100 réalisations du couple (X,Y) présentés figure 8.1, on a superposé la droite $y=ax$, appelée droite de regression de Y par rapport à X (figure 8.2).

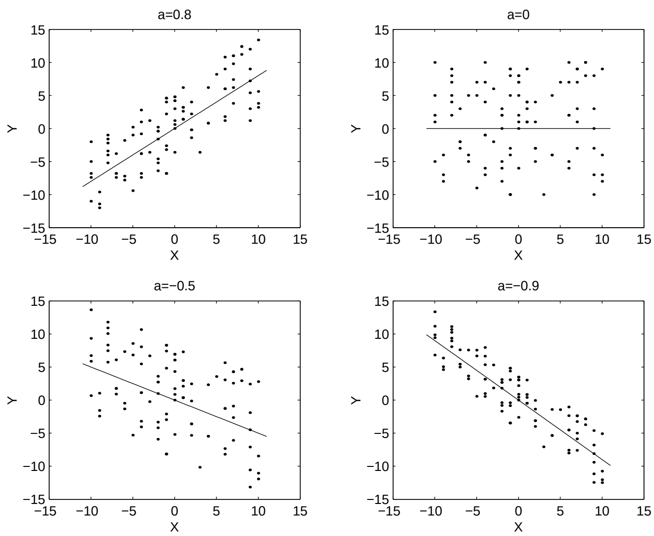


FIG. 6.2 : Nuages de points constitués chacun de 100 réalisations du couple (X,Y) lorsque $Y=aX+\sqrt{1-a^2}U$, X et U sont des variables aléatoires indépendantes de loi uniforme sur $\{-10,-9,\ldots,9,10\}$.

▷ *Exercice 130.* Généraliser la proposition 23 au cas où X est un vecteur aléatoire dont la matrice de covariance K_X est inversible : montrer que la meilleur façon d'approcher Y par une fonction affine des composantes du vecteur X est de prendre la variable aléatoire

$$\sum(K_X^{-1})_{ij}\mathrm{Cov}(X_j,Y)(X_i-\mathrm{E}(X_i))+\mathrm{E}(Y).$$

▷ *Exercice 131.* Montrer que l'erreur quadratique est toujours plus grande avec la méthode de régression linéaire qu'en utilisant l'espérance conditionnelle.

[1]On dit aussi au sens des moindres carrés.

Partie 2

Théorie des variables aléatoires

Une compréhension approfondie des chapitres suivants nécéssitera la connaissance de l'intégrale de Lebesgue sur $\mathbb{R}$. La définition et les propriétés de l'intégrale de Lebesgue sont rappelées en appendice. Cependant une connaissance élémentaire du calcul intégrale peut suffire si on admet que les propriétés démontrées dans le cas dénombrables s'étendent aux espaces continus en remplaçant les sommes par des intégrales. On pourra alors passer rapidement sur les propriétés relatives aux σ-algèbres. Il est néanmoins utile de parcourir l'ensemble des chapitres pour en maîtriser la terminologie.

Chapitre 7

Espaces de probabilité généraux

7.1 Densités de probabilité

Nous abordons dans cette section des résultats asymptotiques dont la compréhension suppose une connaissance élémentaire du calcul intégral pour les fonctions d'une ou de plusieurs variables réelles. Certaines variables aléatoires réelles sont susceptibles de prendre un très grand nombre de valeurs, chacune de ces valeurs ayant une probabilité infime d'être prise (c'est par exemple le cas pour le résultat d'un sondage d'intentions de vote avant un référendum, effectué sur un échantillon de 1000 personnes, comme le montrent les représentations graphiques de la figure 7.1). La fonction de répartition F_X d'une telle variable aléatoire est très proche d'une fonction continue et souvent même d'une fonction dérivable, définie sur $\mathbb{R}$ et croissant de 0 à 1. La dérivée d'une telle fonction est appelée *densité de probabilité*.

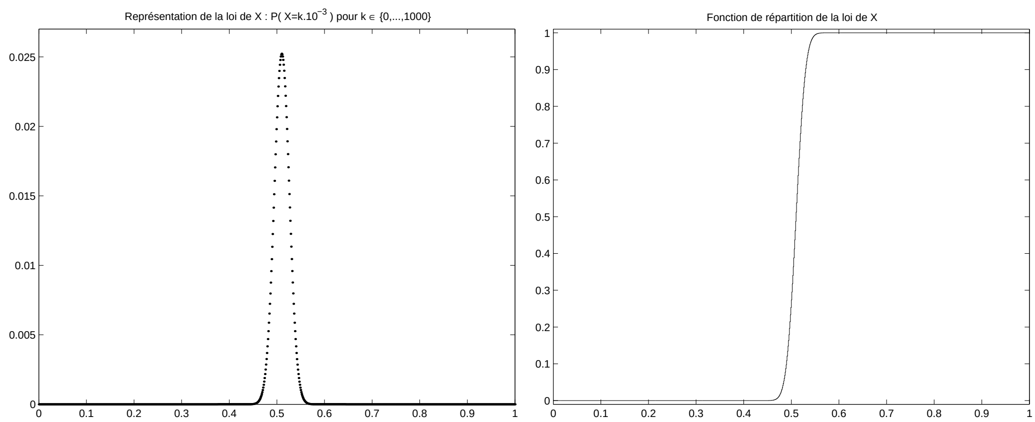


FIG. 7.1 : Loi et fonction de répartition empirique de la proportion X de personnes votant en faveur du "oui" à un référendum sur 1000 personnes interrogées au hasard dans une population où 51% des personnes votent pour le "oui".

Plus précisément, on a la définition suivante :

Définition. Une fonction ϕ définie sur $\mathbb{R}$, positive, borélienne (par exemple, continue par morceaux) est appelée *densité de probabilité* si et seulement si son intégrale $\int_{-\infty}^{+\infty} \phi(x)dx$ est égale à 1.

On dit qu'une densité est portée par l'ensemble Σ des points où elle ne s'annule pas.

N.B. Nous utilisons dans ce chapitre l'intégrale de Lebesgue des fonctions boréliennes sur $\mathbb{R}^n$ qui étend l'intégrale de Riemann des fonctions continues par morceaux. Les propriétés de cette intégrales, que l'on pourra au besoin admettre, sont rappelées en appendice. La notion de fonction borélienne est aussi rappelée en appendice. Disons seulement ici que la classe des fonctions boréliennes est la plus petite classe de fonctions contenant les fonctions continues et stable par toutes les limites monotones. En pratique, les densités usuelles sont continues par morceaux et l'usage que nous faisons de l'intégrale de Lebesgue vise essentiellement à simplifier les démonstrations notamment par l'utilisation des théorèmes de convergence.

7.1.1 Densités usuelles et changement de variable

Densité uniforme sur l'intervalle $[a, b]$. Cette densité est égale à $\frac{1}{b-a}$ sur $\Sigma = [a, b]$ et nulle ailleurs.

Densité exponentielle de paramètre λ. Cette densité est égale à $\lambda \exp(-\lambda x)$ sur $\Sigma = [0, +\infty[$ et nulle ailleurs.

Changement de variable. Si u est une fonction strictement monotone à dérivée continue, il est facile de voir que la fonction $\phi_u(x) = \dfrac{\phi(u^{-1}(x))}{u'(u^{-1}(x))}$ est une densité. On a pour tout $t \in \mathbb{R}$, $\int_{-\infty}^t \phi_u(x)dx = \int_{-\infty}^{u^{-1}(t)} \phi(x)dx$. Donc, si $F_X(t) \simeq \displaystyle\int_{-\infty}^t \phi(x)dx$, alors $F_{u(X)}(t) \simeq \displaystyle\int_{-\infty}^t \phi_u(x)dx$. Dans le cas où u est strictement monotone par morceaux, on doit prendre $\phi_u(x) = \displaystyle\sum_{y, u(y)=x} \frac{\phi(y)}{|u'(y)|}$.

Densité gaussienne. Posons $\gamma_{0,1}(x) = \dfrac{1}{\sqrt{2\pi}} \exp(-\dfrac{x^2}{2})$. La fonction $\gamma_{0,1}$ est une densité de probabilité[1] (appelée densité gaussienne) portée par $\mathbb{R}$. En effectuant un changement de variable linéaire, on montre que pour tout $m \in \mathbb{R}$ et $\sigma > 0$, la fonction γ_{m,σ^2} définie sur $\mathbb{R}$ par : $\gamma_{m,\sigma^2}(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp(-\dfrac{(x-m)^2}{2\sigma^2})$ pour tout $x \in \mathbb{R}$ est encore une densité de probabilité.

▷ *Exercice 132.* Effectuer les calculs justifiant les résultats précédents.

▷ *Exercice 133.* Si $\phi = \gamma_{0,1}$ et si $u(x) = x^2$, montrer que $\phi_u(x) = \frac{1}{\sqrt{2\pi x}} \exp(-\frac{x}{2}) \mathbb{1}_{x>0}$. Cette densité est notée χ_1^2.

7.1.2 Convergence des fonctions de répartition, premiers exemples

La loi d'une variable aléatoire réelle est approximativement donnée par la densité ϕ si pour tout intervalle $[a, b]$, $P(X \in [a, b]) \simeq \int_a^b \phi(x)dx$, ce qui équivaut à supposer que $F_X(t) \simeq \int_{-\infty}^t \phi(x)dx$

[1] Pour montrer que $\int_{-\infty}^{+\infty} \exp(-\frac{x^2}{2})dx = \sqrt{2\pi}$, on calcule le carré de l'intégrale qui s'écrit, d'aprés le théorème de Fubini, $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp(-\frac{x^2+y^2}{2})dxdy$. Cette intégrale double est facile à calculer en passant en coordonnées polaires.

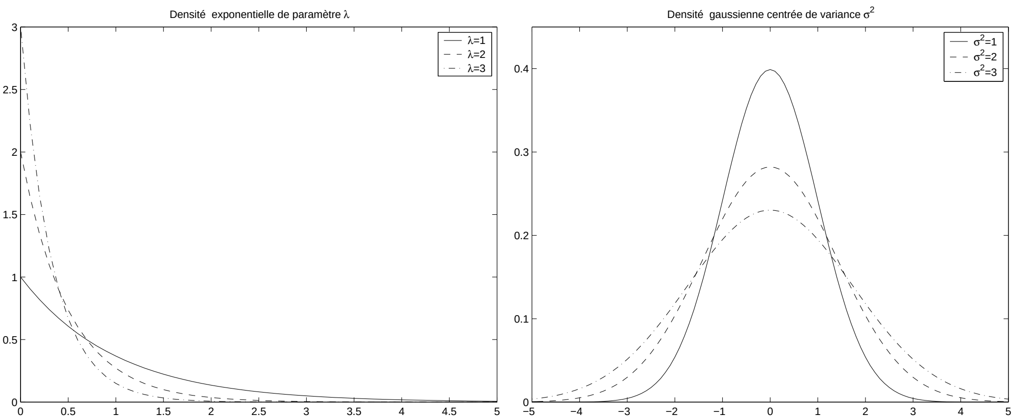


FIG. 7.2 : Représentation graphique de quelques densités exponentielles (sur la figure de gauche) et gaussiennes (sur la figure de droite)

pour tout $t \in \mathbb{R}$. Dans la pratique, les événements de probabilité infime sont le plus souvent considérés comme négligeables. De la même façon, la probabilité d'un événement n'a besoin d'être connue qu'approximativement. Les résultats qui suivent permettent d'approcher une fonction de répartition $F_X(t)$ difficile, voir impossible à calculer exactement par la primitive $F_\phi(t) = \int_{-\infty}^t \phi(x)dx$ d'une densité ϕ convenablement choisie. De façon rigoureuse, on pose la définition suivante :

Définition. *Une suite de lois $(\pi_n)_n$ converge vers une densité ϕ* si la suite de leurs fonctions de répartitions $(F_{\pi_n})_n$ convergent vers F_ϕ en tout point.

En pratique, lorsque n est grand, on remplace la loi π_n par ϕ en controlant si possible par un calcul ou une expérience numérique, la qualité de l'approximation. Donnons maintenant quelques résultats précis d'approximation.

Lois uniformes. Pour tout entier n, considérons la loi uniforme sur l'ensemble $I_n = \{\frac{i}{n}, 1 \leq i \leq n\}$. Sa fonction de répartition est la fonction F_n définie par :

$$F_n(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{1}{n} Ent(nt) & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t \geq 1 \end{cases}$$

Lorsque n tend vers l'infini, on voit très facilement que F_n converge vers la fonction F définie par $F(t) = \min(t, 1)\mathbb{1}_{t>0}$ pour tout $t \in \mathbb{R}$. Cette fonction est la primitive de la densité uniforme sur $[0, 1]$. Donc, la loi uniforme sur I_n converge vers la densité uniforme sur $[0, 1]$. De plus, l'erreur que l'on commet en approximant $F_n(t)$ par $F(t)$ est majorée par $\frac{1}{n}$.

Lois exponentielles et lois géométriques. La fonction de répartition de la loi géométrique de paramètre p sur $\mathbb{N}^*$ est la fonction $t \mapsto 1 - (1-p)^{Ent(t)}$. On peut transporter cette loi en une loi π_n sur $\{\frac{i}{n}, i \in \mathbb{N}^*\}$ dont la fonction de répartition sera $t \mapsto 1 - (1-p)^{Ent(nt)}$. Si X_n est une variable aléatoire de loi géométrique $\mathcal{G}_{\mathbb{N}^*}(\frac{\lambda}{n})$, π_n est la loi de $\frac{X_n}{n}$. Intuitivement, $\frac{X_n}{n}$ représente l'instant du premier gain lors d'un jeu de pile ou face, où les lancers sont effectués à des intervalles de temps $\frac{1}{n}$ et où on a $\frac{\lambda}{n}$ chances de gagner à chaque lancer. En prenant $p(n) = \frac{\lambda}{n}$, pour n assez grand, on voit que la fonction de répartition converge vers $t \mapsto 1 - \exp(-\lambda t)$. L'erreur est aussi de l'ordre de $\frac{1}{n}$ (le vérifier !). La convergence montre que la décroissance de la probabilité de gain à une partie est compensée par l'accélération du jeu. On obtiendrait le même résultat à partir d'une loi géométrique tronquée en $k(n)$ pourvu que $\frac{k(n)}{n}$ tende vers l'infini avec n.

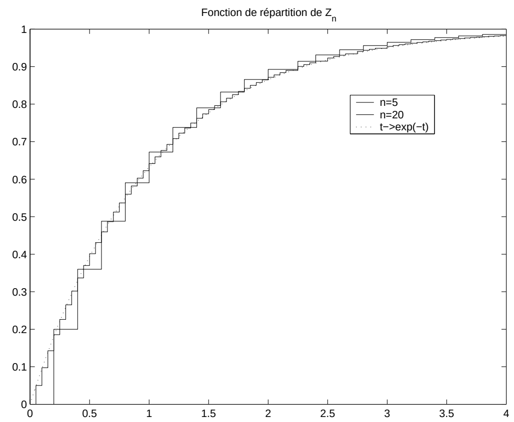


FIG. 7.3 : Illustration de la convergence de la loi de Z_n lorsque nZ_n suit la loi $\mathcal{G}_{\mathbb{N}^*}(\frac{1}{n})$.

N.B.

1. Si les lois d'une suite de variables aléatoires $(X_n)_n$ convergent vers une densité ϕ et si u est une fonction strictement monotone par morceaux à dérivée continue, il est facile de voir que la suite des lois des variables aléatoires $(u(X_n))_n$ convergent vers la densité ϕ_u.

2. Lorsque l'on considère les lois d'une suite de variables, ces variables n'ont pas à être définies sur le même espace probabilisé. Dans la plupart des exemples que nous étudions, elles sont définies sur des espaces Ω^n représentant des suites de tirages indépendants de plus en plus longues.

▷ *Exercice 134.* Une urne contient à l'instant 1, une boule blanche et une boule noire. A chaque instant n, on modifie la composition de l'urne de la façon suivante : on retire au hasard une boule de l'urne et on la remplace par deux boules de la même couleur que celle retirée. On note B_n le nombre de boules blanches dans l'urne avant la n-ième modification et X_n la proportion de boules blanches au même moment.

1. Montrer, par récurrence sur n, que X_n suit la loi uniforme sur l'ensemble
$$V_n = \{\frac{k}{n+1},\ k = 1, \dots, n\}.$$
2. Montrer que la loi de X_n converge vers une densité f que l'on déterminera.
3. Considérer le cas où l'urne contient initialement deux boules blanches et une boule noire. Montrer qu'alors pour tout $n \geq 2$ et pour tout $i \in \{2, \dots, n+1\}$, $P(X_n = \frac{i}{n+2}) = \frac{2(i-1)}{(n+1)(n+2)}$. En déduire que la loi de $(X_n)_n$ converge vers une densité g différente de f que l'on expicitera.
Application : On s'intéresse à la propagation d'une information sur un fait divers dans une population. On modélise la propagation de la façon suivante : à l'instant 0, 2 personnes sont en possession d'une information I_1 sur ce fait et une personne est en possession d'une information I_2 qui diffère de I_1. A chaque instant $k \in \mathbb{N}$, un individu qui n'est pas au courant de ce fait divers, rencontre une des personnes possédant soit l'information I_1, soit l'information I_2, celle-ci

lui divulgue l'information en sa possession. Suivant la personne rencontrée, l'individu sera alors lui-même en possession de l'information I_1 ou de l'information I_2. On supposera que les rencontres se font au hasard et indépendamment les unes des autres. Quelle est la probabilité qu'il y ait deux fois plus de personnes en possession de l'information I_1 que de personnes en possession de l'information I_2 au bout d'un très long temps ?

▷ *Exercice 135.* On considère le polygone $\mathcal{P}_n$ régulier à n côtés centré en l'origine et inscrit dans le cercle unité. On tire au hasard un sommet du polygone et on note X_n la distance qui sépare ce sommet de l'axe des abscisses. Montrer que la loi de X_n converge vers une densité f que l'on déterminera.

7.1.3 Densités à plusieurs variables

On définit de la même façon une *densité sur* $\mathbb{R}^d$ comme une fonction positive ϕ définie sur $\mathbb{R}^d$, borélienne (par exemple, continue par morceaux), dont l'intégrale sur $\mathbb{R}^d$,

$$\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \phi(x_1, ..., x_d) dx_1 \dots dx_d$$

est égale à 1.

Notons que si $\phi_1, ..., \phi_d$ sont des densités sur $\mathbb{R}$, l'application $(x_1, \dots, x_d) \mapsto \phi_1(x_1) \cdots \phi_d(x_d)$ est une densité sur $\mathbb{R}^d$ appelée *densité produit*.

Densités uniformes sur un ouvert de $\mathbb{R}^d$. Si D est un ouvert borné de $\mathbb{R}^d$ alors la fonction $\frac{\mathbf{1}_D}{vol(D)}$ est une densité de probabilité sur $\mathbb{R}^d$ appelée "*densité uniforme sur D*" ($vol(D)$ désigne le volume de D).

Densité gaussienne sur $\mathbb{R}^d$. Pour tout $(\sigma_1, ..., \sigma_d) \in (\mathbb{R}_+^*)^d$ et $(m_1, \dots, m_d) \in \mathbb{R}^d$,

$$(x_1, \dots, x_d) \mapsto \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{i=1}^d \sigma_i} \exp(-\sum_{i=1}^d \frac{(x_i - m_i)^2}{2\sigma_i^2})$$

est une densité sur $\mathbb{R}^d$ comme produit de d densités gaussiennes sur $\mathbb{R}$. Si C est la matrice diagonale dont les coefficients diagonaux sont $\sigma_1^2, \dots, \sigma_d^2$, cette densité s'écrit :

$$\gamma_{\overrightarrow{m},C}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det(C)}} \exp\Big(-\frac{1}{2} \sum_{i,j=1}^d (C^{-1})_{ij} (x_i - m_i)(x_j - m_j)\Big) \ \forall x \in \mathbb{R}^d. \tag{7.1}$$

Plus généralement, on montre par un changement de variable orthogonal, que pour tout vecteur $\overrightarrow{m}$ de $\mathbb{R}^d$, et toute matrice positive non dégénérée C de rang d, la fonction $\gamma_{\overrightarrow{m},C}$ définie par la formule (7.1) est une densité de probabilité sur $\mathbb{R}^d$ (appelée également *densité gaussienne*). Cette densité est une densité produit si et seulement si C est diagonale.

Convergence des lois d'une suite de vecteurs aléatoires. La loi d'un vecteur aléatoire $\overrightarrow{X} = (X_1, ..., X_d)$ est approximativement donnée par la densité ϕ si pour toute famille d'intervalles $[a_1, b_1], \dots, [a_d, b_d]$,

$$P(X_i \in [a_i, b_i] \text{ pour tout } 1 \leq i \leq d) \simeq \int_{a_1}^{b_1} \dots \int_{a_d}^{b_d} \phi(x_1, \dots, x_d) dx_1 \dots dx_d.$$

De façon mathématiquement plus précise, on pose la définition suivante :

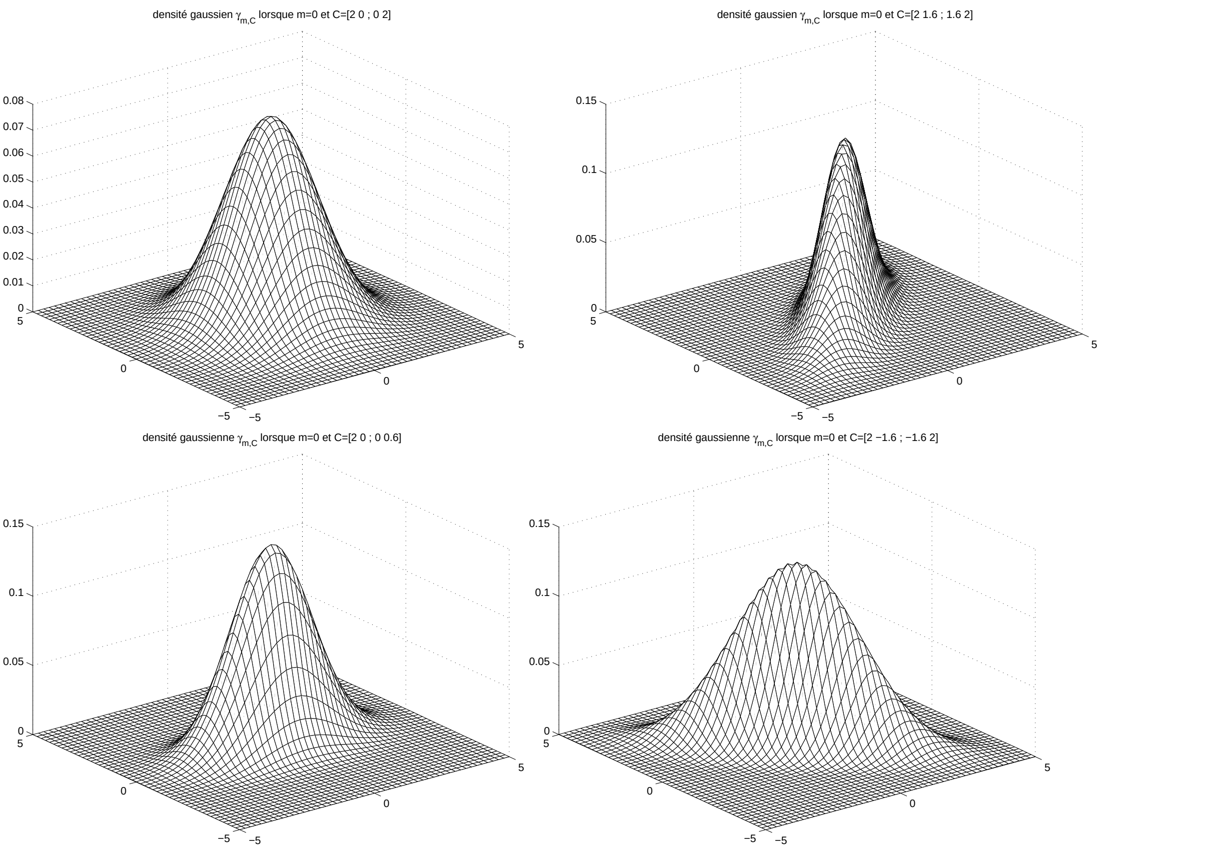


FIG. 7.4 : Représentation graphique de densités gaussiennes $\gamma_{m,C}$ sur $\mathbb{R}^2$

Définition. *Les lois d'une suite de vecteurs aléatoires* $\overrightarrow{X}^{(n)}$ *de* $\mathbb{R}^d$ *convergent lorsque* n *tend vers* $+\infty$ *vers la densité* ϕ si pour toute famille d'intervalles $]a_1, b_1], \ldots,]a_d, b_d]$,

$$P(\overrightarrow{X}^{(n)} \in]a_1, b_1] \times \ldots \times]a_d, b_d]) \underset{n \to +\infty}{\longrightarrow} \int_{a_1}^{b_1} \ldots \int_{a_d}^{b_d} \phi(x_1, \ldots, x_d) dx_1 \ldots dx_d.$$

Ceci équivaut à supposer que la fonction de répartition de $\overrightarrow{X}^{(n)}$ au point $(t_1, \ldots, t_d)$ définie par $F_{\overrightarrow{X}^{(n)}}(t_1, \ldots, t_d) = P(X_i^{(n)} \leq t_i \text{ pour tout } 1 \leq i \leq d)$ converge lorsque n tend vers l'infini vers $\displaystyle\int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_d} \phi(x_1, \ldots, x_d) dx_1 \ldots dx_d$ pour tout $(t_1, \ldots, t_d) \in \mathbb{R}^d$.
Remarquons que si les variables aléatoires $X_1^{(n)}, \ldots, X_d^{(n)}$ sont indépendantes pour tout n et si leurs lois convergent vers des densités $\phi_1, \ldots, \phi_d$ respectivement, alors la loi du vecteur $\overrightarrow{X}^{(n)} = (X_1^{(n)}, \ldots, X_d^{(n)})$ converge vers la densité produit $(x_1, \ldots, x_d) \mapsto \phi_1(x_1) \ldots \phi_d(x_d)$.

Plus généralement, si D est un ouvert à frontière régulière et si la loi d'un vecteur aléatoire $\overrightarrow{X}$ est approximativement donnée par la densité ϕ,

$$P(\overrightarrow{X} \in D) \simeq \int_D \phi(x_1, \ldots, x_d) dx_1 \ldots dx_d.$$

De façon mathématiquement plus précise, on peut dire que si $(\overrightarrow{X_n})_n$ est une suite de vecteurs aléatoires dont les lois convergent vers la densité ϕ, les probabilités $P(\overrightarrow{X_n} \in D)$ convergent vers $\displaystyle\int_D \phi(x_1, ..., x_d) dx_1 ... dx_d$.

Pour $d = 2$, ce résultat est facile si D est un rectangle. Il peut être démontré en approchant l'ouvert par une suite de réunions de rectangles. En dimension quelconque, on remplacera les rectangles par les pavés, c'est-à-dire les produits d'intervalles.

Changement de variable. Si u est une transformation continûment différentiable et inversible définie sur $\mathbb{R}^d$, il résulte de la formule de changement de variable dans les intégrales multiples que la fonction ϕ_u définie par

$$\phi_u(\overrightarrow{x}) = \frac{\phi(u^{-1}(\overrightarrow{x}))}{\det(Du)(u^{-1}(\overrightarrow{x}))} \text{ pour tout } \overrightarrow{x} \in \mathbb{R}^d$$

est une densité. Si les lois d'une suite de vecteurs aléatoires $\overrightarrow{X}_n$ convergent vers ϕ, celles des vecteurs $u(\overrightarrow{X}_n)$ convergent vers ϕ_u.

▷ *Exercice 136.* Si $(X_n, Y_n)_n$ est une suite de couples de v.a. indépendantes de lois uniformes sur $\{1, 2, ..., n\}$, montrer que la loi de $\frac{|X_n - Y_n|}{n}$ converge vers une densité que l'on calculera.
Application : un dépanneur opère sur une section d'autoroute de 112 km, quelle est la probabilité pour que deux dépannages consécutifs aient lieu à plus de 70 km de distance ?

7.2 σ-algèbres et probabilités

Dans la section précédente, nous avons vu que les densités de probabilité pouvaient, en passant à la limite, remplacer les lois de variables aléatoires susceptibles de prendre un très grand nombre de valeurs, chacune ayant une probabilité infime d'apparaître. On peut en fait les considérer comme des lois de probabilité de variables aléatoires susceptibles de prendre un continuum de valeurs. Cela n'a peut-être pas beaucoup de justification empirique (les mesures étant toujours données avec un nombre fini de chiffres) mais en pratique, cela facilite de façon très appréciable la description des lois et les calculs les concernant. Il existe une difficulté théorique due au fait qu'une variable aléatoire prenant un continuum de valeurs ne peut être définie sur un espace dénombrable. Pour surmonter cette difficulté, la théorie doit être étendue à un ensemble Ω quelconque (par exemple, $\mathbb{R}$, $\mathbb{R}^d$, $\{0, 1\}^{\mathbb{N}}$ etc). Il est alors nécessaire, pour définir des probabilités naturelles sur ces ensembles, de restreindre quelque peu la définition des événements.

Définition. On appelle *σ-algèbre (ou tribu)* sur Ω une famille $\mathcal{A}$ de parties de Ω contenant Ω et stable par les opérations ensemblistes suivantes :

(i) passage au complémentaire : si $A \in \mathcal{A}$ alors le complémentaire de A dans Ω appartient à $\mathcal{A}$.

(ii) réunion dénombrable : si $(A_i)_{i\in\mathbb{N}^*}$ est une suite d'éléments de $\mathcal{A}$ alors la réunion de ces ensembles $\cup_{i\in\mathbb{N}^*} A_i \in \mathcal{A}$

Ces deux propriétés (i) et (ii) entrainent que $\mathcal{A}$ est stable par intersection dénombrable : si $(A_i)_{i\in\mathbb{N}^*}$ est une suite d'éléments de $\mathcal{A}$ alors $\cap_{i\in\mathbb{N}^*} A_i$ appartient à $\mathcal{A}$.
La famille $\mathcal{A}$ choisie est souvent appelée la *σ-algèbre des événements de* Ω.

Exemples

1. $\mathcal{P}(\Omega)$ est une σ-algèbre sur Ω.

2. Si $(U_i)_{i\in I}$ est une partition au plus dénombrable de Ω alors l'ensemble constitué de l'ensemble vide et des réunions finies ou dénombrables d'ensembles de la partition :

$$\mathcal{A} = \{\emptyset, \underset{j\in J}{\cup} U_j,\ J \subset I\}$$

est une σ-algèbre sur Ω.

En effet, $\Omega = \cup_{i\in I} U_i$, donc Ω appartient à $\mathcal{A}$. Une réunion au plus dénombrable d'éléments de $\mathcal{A}$ est encore une réunion au plus dénombrable d'ensembles de la partition $(U_i)_{i\in I}$; elle appartient donc à $\mathcal{A}$. Le complémentaire d'un ensemble U_j de la partition $(U_i)_{i\in I}$ est égal à $\cup_{i\neq j} U_i$; c'est donc un élément de $\mathcal{A}$. Plus généralement, pour tout $J \subset I$, le complémentaire de $\cup_{j\in J} U_j$ est $\cup_{i\in I\setminus J} U_i$; c'est donc aussi un élément de $\mathcal{A}$.

3. Une intersection quelconque de σ-algèbres sur Ω est encore une σ-algèbre.

4. Sur $\mathbb{R}$ la plus petite σ-algèbre contenant les intervalles ouverts $]a,b[$ est appelée la σ-algèbre borélienne sur $\mathbb{R}$ et est notée $\mathcal{B}(\mathbb{R})$. Notons que l'expression “la plus petite σ-algèbre” est justifiée par le fait qu'une intersection quelconque de σ-algèbres est encore une σ-algèbre. Elle contient tous les ouverts, tous les fermés, tous les intervalles.

5. Plus généralement, on définit la σ-algèbre borélienne sur $\mathbb{R}^d$ comme la plus petite σ-algèbre contenant les produits d'intervalles ouverts $]a_1,b_1[\times\cdots\times]a_d,b_d[$ (appelés pavés ouverts). On la note $\mathcal{B}(\mathbb{R}^d)$. Elle contient tous les ouverts, tous les fermés, tous les pavés.

▷ *Exercice 137.* Montrer qu'une intersection quelconque de σ-algèbres sur Ω est encore une σ-algèbre.

▷ *Exercice 138.* Soit $\mathcal{T}$ la plus petite σ-algèbre contenant tous les intervalles de la forme $]-\infty,a]$, $a\in\mathbb{R}$. Montrer que $\mathcal{T}$ coïncide avec la plus petite σ-algèbre contenant tous les intervalles ouverts $]a,b[$, $a<b$ (notée $\mathcal{B}(\mathbb{R})$.
NB : le résulat reste vrai, par exemple si on remplace les intervalles de la forme $]-\infty,a]$, $a\in\mathbb{R}$ par les intervalles de la forme $]-\infty,a[$, $a\in\mathbb{R}$, ou bien par les intervalles de la forme $]a,b]$, $a<b$.

Solution. Soit $a\in\mathbb{R}$. Comme $]-\infty,a] = \cup_n]n, a+\frac{1}{n}[$, $\mathcal{B}(\mathbb{R})$ contient $]-\infty,a]$, $a\in\mathbb{R}$. Donc $\mathcal{T}$ est inclus dans $\mathcal{B}(\mathbb{R})$. Pour tout $a<b$, $]a,b[=]a,+\infty[\cap]-\infty,b[$. Comme $]a,+\infty[$ est le complémentaire de $[a,+\infty[$ et comme $]-\infty,b[=\cap_n]-\infty,b+\infty 1n]$, $\mathcal{T}$ contient tous les intervalles finis ouverts. Donc, $\mathcal{T}$ contient $\mathcal{B}(\mathbb{R})$.

7.2.1 Σ-algèbre engendrée

Etant donné une classe $\mathcal{C}$ de parties de Ω, on note $\sigma(\mathcal{C})$ la plus petite σ-algèbre contenant $\mathcal{C}$, appelée *σ-algèbre engendrée par $\mathcal{C}$*. Le résultat technique fondamental sur les σ-algèbres engendrées est le théorème suivant :

Théorème 14 *Soit $\mathcal{C}$ un ensemble de parties de Ω, stable par intersection finie et $\mathcal{L}$ un ensemble de parties de Ω, contenant Ω, stable par passage au complémentaire et par réunion dénombrable de parties disjointes. Si $\mathcal{L}$ contient $\mathcal{C}$, il contient aussi $\sigma(\mathcal{C})$.*

Preuve : L'idée de la preuve est de construire une σ-algèbre $\mathcal{T}$ contenant $\mathcal{C}$ et incluse dans $\mathcal{L}$. On aura alors $\sigma(\mathcal{C}) \subset \mathcal{T} \subset \mathcal{L}$.
Remarquons qu'une σ-algèbre a en plus des propriétés de $\mathcal{L}$ la propriété d'être stable par intersection finie. Remarquons d'autre part que si $\mathcal{L}_1$ et $\mathcal{L}_2$ sont deux parties de Ω ayant les mêmes propriétés que $\mathcal{L}$ alors $\mathcal{L}_1 \cap \mathcal{L}_2$ a les mêmes propriétés que $\mathcal{L}$.
Désignons par $\mathcal{L}_0$ l'intersection des ensembles de parties Ω contenant $\mathcal{C}$ et qui ont les mêmes propriétés que $\mathcal{L}$. On va montrer que $\mathcal{L}_0$ est stable par intersection fini et donc est une σ-algèbre.
Pour $A \subset \Omega$, posons $\mathcal{L}_A = \{B,\ A \cap B \subset \mathcal{L}_0\}$.

1. Comme $\mathcal{C}$ est stable par intersection finie, $\mathcal{C} \subset \mathcal{L}_A$ pour tout $A \in \mathcal{C}$.

2. Soit $A \in \mathcal{C}$, montrons que $\mathcal{L}_A$ a les mêmes propriétés que $\mathcal{L}$. Il est facile de voir que $\mathcal{L}_A$ contient Ω et est stable par réunion disjointe. Il reste à montrer que $\mathcal{L}_A$ est stable par passage au complémentaire : soit $B\mathcal{L}_A$. Comme $A \cap B$ et A sont dans $\mathcal{L}_0$, $(A \cap B) \cup A^c \in \mathcal{L}_0$ et donc son complémentaire qui coïncide avec $A \cap B^c$ est dans $\mathcal{L}_0$, ce qui montre que B^c appartient à $\mathcal{L}_A$.

3. Par construction de $\mathcal{L}_0$, on en déduit que pour tout $A \in \mathcal{C}$, $\mathcal{L}_0 \subset \mathcal{L}_A$. Cela signifie que pour tout $A \in \mathcal{C}$, et tout $B \in \mathcal{L}_0$, $A \cap B \in \mathcal{L}_0$. Donc, pour tout $B \in \mathcal{L}_0$, $\mathcal{C} \subset \mathcal{L}_B$. On montre comme à l'étape 2, que pour tout $B \in \mathcal{L}_0$, $\mathcal{L}_B$ a les mêmes propriétés que $\mathcal{L}$. Donc, pour tout $B \in \mathcal{L}_0$, $\mathcal{L}_0 \subset \mathcal{L}_B$, ce qui montre que $\mathcal{L}_0$ est stable par intersection fini.

□

Espace produit Etant donnés deux espaces $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$, l'ensemble $\Omega_1 \times \Omega_2$ peut être muni de la σ-algèbre produit $\mathcal{A}_1 \otimes \mathcal{A}_2$ qui est définie comme la plus petite σ-algèbre contenant tous les "pavés mesurables", c'est-à-dire tous les ensembles de la forme $A_1 \times A_2$ avec $A_1 \in \Omega_1$ et $A_2 \in \Omega_2$.
Cette définition se généralise naturellement au cas d'un produit d'un nombre fini d'espaces.
En particulier, la σ-algèbre borélienne sur $\mathbb{R}^d$ est la σ-algèbre produit de d σ-algèbres boréliennes sur $\mathbb{R}$: $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$ (d fois).

Lemme 15 (Lemme de section) *Soit $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Pour tout $\omega_1 \in \Omega_1$, la section $A^2_{\omega_1} = \{\omega_2 \in \Omega_2,\ (\omega_1, \omega_2) \in A\}$ est un événement de $\mathcal{A}_2$.*
De même, pour $\omega_2 \in \Omega_2$, la section $A^1_{\omega_2} = \{\omega_1 \in \Omega_1,\ (\omega_1, \omega_2) \in A\}$ est un événement de $\mathcal{A}_1$.

Preuve : Pour démontrer la premier assertion, on applique le théorème 14 en prenant pour $\mathcal{C}$ la classe des pavés mesurables et pour $\mathcal{L}$ la classe des parties $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ telles $A^2_{\omega_1} \in \mathcal{A}_2$. □

7.2.2 Mesures de probabilité

Une probabilité sur $(\Omega, \mathcal{A})$ est alors définie comme dans le cas d'un espace dénombrable :

Définition. Soit Ω un ensemble muni d'une σ-algèbre $\mathcal{A}$. Une *probabilité sur* $(\Omega, \mathcal{A})$ est une application $P\ : \mathcal{A} \to [0, 1]$ telle que :

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

(ii) P est σ-additive : si $(A_i)_{i\in\mathbb{N}^*}$ est une suite d'éléments de $\mathcal{A}$ deux à deux incompatibles, alors

$$P(\overset{+\infty}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{+\infty} P(A_i).$$

Le triplet $(\Omega, \mathcal{A}, P)$ est appelé *espace de probabilité.*

Mesure de Dirac. Soit E un ensemble quelconque muni d'une σ-algèbre $\mathcal{E}$ et a un élément de E. En posant, pour tout $A \in \mathcal{E}$,

$$\delta_a(A) = \begin{cases} 1 & \text{si A contient a} \\ 0 & \text{si A ne contient pas a} \end{cases}$$

on définit une probabilité sur $(E, \mathcal{E})$; cette probabilité est appelée la *mesure de Dirac en a.* En effet, $\delta_a(E) = 1$ puisque $z \in E$. Soit $(A_i)_{i\in\mathbb{N}^*}$ une suite d'éléments de $\mathcal{E}$ deux à deux incompatibles. Si $e \in \cup_{i\in\mathbb{N}^*} A_i$, il existe un unique $i \in \mathbb{N}^*$ tel que $a \in A_i$ et donc $\sum_{i\in\mathbb{N}^*} \delta_a(A_i) = 1$. De même, si $e \notin \cup_{i\in\mathbb{N}^*} A_i$ alors pour tout $i \in \mathbb{N}^*$, $a \notin A_i$ et donc $\sum_{i\in\mathbb{N}^*} \delta_a(A_i) = 0$. Cela montre la σ-additivité de δ_a.

Mesure de probabilité discrète. Soit E un ensemble quelconque muni d'une σ-algèbre $\mathcal{E}$. Si $(a_n)_{n\in\mathbb{N}^*}$ est une suite de points de E et si $(p_n)_{n\in\mathbb{N}^*}$ est une suite de réels positifs telle que $\sum_{n=1}^{+\infty} p_n = 1$, alors la somme des mesures de Dirac en les points a_n pondérée par les poids p_n, $\sum_{n=1}^{\infty} p_n\delta_{a_n}$, est une mesure de probabilité sur $(E, \mathcal{E})$.

Probabilité à densité. Si ϕ est une densité de probabilité sur $\mathbb{R}^d$, l'application μ définie sur les boréliens de $\mathbb{R}^d$ par

$$\mu(A) = \int_B \phi(x_1, \ldots, x_d)dx_1 \ldots dx_d.$$

est une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

▷ *Exercice 139.* Démontrer ce résultat en utilisant les propriétés de la mesure de Lebesgue données en appendice.

μ est alors souvent notée $\phi(x_1, \ldots, x_d)dx_1 \ldots dx_d$ et est appelée *la probabilité de densité* ϕ. Elle est déterminée par ses valeurs sur les pavés ouverts d'après le théorème général suivant :

Théorème 16 (Théorème d'unicité) *Soit Ω un ensemble et $\mathcal{C}$ un sous-ensemble de $\mathcal{P}(\Omega)$ tel que $A, B \in \mathcal{C}$ implique $A \cap B \in \mathcal{C}$. Soit $\mathcal{A}$ la plus petite σ-algèbre contenant $\mathcal{C}$. Si μ_1 et μ_2 sont deux probabilités définies sur $(\Omega, \mathcal{A})$ qui coïncident sur $\mathcal{C}$, alors elles sont égales.*

Preuve : C'est une conséquence directe du théorème 14: on prendra

$$\mathcal{L} = \{A \in \sigma(\mathcal{C}), \mu_1(A) = \mu_2(A)\}.$$

□

▶ *Exemple 140.*

1. Si D est un ouvert borné de $\mathbb{R}^d$ alors la probabilité de densité uniforme sur D est la probabilité μ définie par : pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, $\mu(B) = \int_B \frac{\mathbf{1}_D(x)}{vol(D)} dx = \frac{vol(B\cap D)}{vol(D)}$. Elle est aussi appelée *loi uniforme sur* D et est notée $\mathcal{U}_D$.
2. La probabilité de densité gaussienne $\gamma_{m,C}$ sur $\mathbb{R}^d$ est aussi appelée *loi normale* et est notée $\mathcal{N}(m, C)$. La fonction de répartition de la loi normale $\mathcal{N}(0, 1)$ est tabulée.

Probabilité conditionnelle Comme dans le cas d'un espace dénombrable, si A est un événement de l'espace probabilisé $(\Omega, \mathcal{A}, P)$, tel que $P(A) > 0$, on définit une nouvelle probabilité Q sur Ω muni de la σ-algèbre $\mathcal{A}$ en posant $Q(B) = \frac{P(A\cap B)}{P(A)}$ pour tout $B \in \mathcal{A}$. Elle est notée $P(\cdot|A)$.

▶ *Exemple 141.* Si P est la probabilité de densité uniforme sur un ouvert borné D de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et si A est un ouvert inclus dans D de volume non nul, alors $P(\cdot|A)$ est la probabilité de densité uniforme sur A. En effet, par définition pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, $P(B) = \frac{vol(B\cap D)}{vol(D)}$. Donc, pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, $P(B|A) = \frac{P(B\cap A)}{P(A)} = \frac{vol(B\cap A)}{vol(A)}$.

Mesure de probabilité produit Etant donnés deux espaces probabilisés $(\Omega_1, \mathcal{A}_1, P_1)$ et $(\Omega_2, \mathcal{A}_2, P_2)$, l'espace $\Omega_1 \times \Omega_2$ peut être muni de la σ-algèbre produit $\mathcal{A}_1 \otimes \mathcal{A}_2$. On montrera au chapitre suivant qu'il existe une probabilité Q sur $\mathcal{A}_1 \otimes \mathcal{A}_2$ vérifiant $Q(A_1 \times A_2) = P_1(A_1)P_2(A_2)$ pour tout $A_1 \in \Omega_1$ et $A_2 \in \Omega_2$. Cette probabilité est unique d'après le théorème d'unicité appliquée à la classe des pavés. Elle est appelée la *probabilité produit* et est notée $P_1 \otimes P_2$.
Cette définition se généralise naturellement au cas d'un produit d'un nombre fini d'espaces probabilisés : soit $(\Omega_i, \mathcal{A}_i, P_i)$ pour $i \in \{1, \dots, n\}$ n espaces probabilisés. Sur l'espace produit $(\Omega_1 \times \cdots \times \Omega_n, \otimes_{i=1}^n \mathcal{A}_i)$, il existe une unique probabilité Q telle que

$$Q(A_1 \times \cdots \times A_n) = \prod_{i=1}^n P(A_i) \text{ pour tout } A_1 \times \cdots \times A_n \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_n.$$

La définition de probabilité produit étend de façon évidente celle donnée dans le cas d'espaces dénombrables.

▶ *Exemple 142.* Si $\mu_1, \dots, \mu_d$ sont des probabilités sur $\mathbb{R}$ de densités $\phi_1, \dots, \phi_d$ respectivement, alors la probabilité produit $\mu = \bigotimes_{i=1}^{r} \mu_i = \mu_1 \otimes \cdots \otimes \mu_d$ a elle-même une densité définie par :

$$(x_1, \dots, x_r) \mapsto \phi(x_1) \dots \phi_r(x_r).$$

En effet, pour tout pavé ouvert $D =]a_1, b_1[\times \cdots \times]a_r, b_r[$,

$$\mu(D) = \prod_{i=1}^r \mu(]a_i, b_i[) = \prod_{i=1}^r \int_{a_i}^{b_i} \phi_i(x_i)dx_i = \int_D \phi(x_1, \dots, x_r)dx_1 \dots dx_r.$$

Fonction de répartition. Si μ est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, le théorème d'unicité 16 montre qu'elle est caractérisée par sa fonction de répartition $F_\mu \ : t \mapsto \mu(]-\infty, t])$ (on l'applique à l'ensemble $\mathcal{C}$ des intervalles de la forme $]-\infty,\ t]$). F_μ est une fonction croissante, continue à droite et ayant pour limites 0 en $-\infty$ et 1 en $+\infty$.
On peut montrer par ailleurs (voir exercice 143) que toute fonction F définie sur $\mathbb{R}$, croissante, continue à droite et ayant pour limites 0 en $-\infty$ et 1 en $+\infty$ est la fonction de répartition d'une probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Elle vérifie[2] alors $\mu([a, b]) = F(b) - F(a^-)$ pour tout $a \leq b$.

[2] $F(a^-)$ est la limite de $F(a - h)$ lorsque h décroit vers 0

▷ *Exercice 143.*) Soit μ une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Démontrer que sa fonction de répartition F_μ est croissante, continue à droite et a pour limites 0 en $-\infty$ et 1 en $+\infty$.
Indication : pour montrer que F_μ est continue à droite en un point t, on pourra introduire la suite croissante d'ensembles $(]-\infty, t - \frac{1}{n}])_n$.
En déduire que pour tout $a \in \mathbb{R}$, $\mu(\{a\}) = F(a) - F(a^-)$.

▷ *Exercice 144.* Mélange de probabilités.

1. Montrer que si P_1 et P_2 sont des probabilités sur un espace $(\Omega, \mathcal{A})$ et si $\alpha \in [0,1]$ alors $\alpha P_1 + (1-\alpha)P_2$ est aussi une probabilité sur $(\Omega, \mathcal{A})$. Par exemple, $\alpha\delta_1 + (1-\alpha)\delta_0$ est la loi de Bernoulli $\mathcal{B}(\alpha)$.
2. Dans le cas où P_1 et P_2 sont des probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ayant pour densités respectivement ϕ_1 et ϕ_2, déterminer la densité de $\alpha P_1 + (1-\alpha)P_2$.

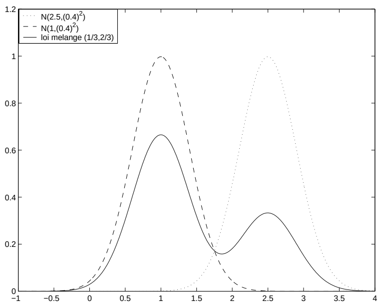


FIG. 7.5 : Représentation en trait continu de la densité de la loi $\frac{1}{3}P_1 + \frac{2}{3}P_2$ lorsque P_1 et P_2 sont respectivement les lois $\mathcal{N}(2.5, (0.4)^2)$ et $\mathcal{N}(1, (0.4)^2)$.

3. Plus généralement, montrer que si $(P_i)_{i \in I}$ est une famille au plus dénombrable de probabilités sur $(\Omega, \mathcal{A})$ et si $(\alpha_i)_{i\in I}$ est une famille de réels positifs telle que $\sum_{i\in I} \alpha_i = 1$ alors $\sum_{i \in I} \alpha_i P_i$ est encore une probabilité sur $(\Omega, \mathcal{A})$. Par exemple, si $\lambda > 0$, $\sum_{k\in\mathbb{N}} e^{-\lambda}\frac{\lambda^k}{k!}\delta_k$ est la loi de Poisson de paramètre λ.

▷ *Exercice 145.*

1. Montrer que la fonction de répartition de la mesure de Dirac en un point $a \in \mathbb{R}$ est $\mathbb{1}_{[a,+\infty[}$.
2. On pose $\mu = \frac{1}{3}\delta_0 + \frac{1}{3}\delta_1 + \frac{1}{3}\mathbb{1}_{[0,1]}(x)dx$. Vérifier que μ est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et calculer sa fonction de répartition.

▷ *Exercice 146.* On considère l'expérience aléatoire suivante : lancer une fléchette sur une cible circulaire de rayon 10 cm. On note p la probabilité que la fléchette n'atteigne pas la cible ou ne s'y fixe pas et on suppose que la probabilité que la fléchette se fixe dans une région A de la cible est proportionnelle à l'aire de A.

1. Construction de l'espace de probabilité $(\Omega, \mathcal{A}, P)$ associé à cette expérience aléatoire : on prend le centre de la cible comme l'origine du plan $\mathcal{P}$ contenant la cible et on définit la cible comme $D = \{(x,y) \in \mathbb{R}^2,\ x^2 + y^2 \leq 100\}$. Le point de rencontre de la fléchette avec le plan $\mathcal{P}$ est déterminé par ses coordonnées dans $\mathcal{P}$, on peut donc poser $\Omega = \mathbb{R}^2$ l'ensemble des résultats possibles et $\mathcal{A} = \{A \cap D, D^c \cup (A \cap D),\ \text{tel que } A \in \mathcal{B}(\mathbb{R}^2)\}$.
Montrer que $\mathcal{A}$ est une σ-algèbre sur Ω et décrire la probabilité P sur $(\Omega, \mathcal{A})$ qui satisfait aux hypothèses de l'énoncé.
2. Soit $t \in [0, 10]$. Calculer la probabilité qu'une fléchette lancée sur la cible se fixe à moins de t cm du centre de la cible puis la probabilité qu'elle se fixe à moins de t cm du bord de la cible.

7.3 Variables aléatoires et lois

Soit Ω un espace muni d'une σ-algèbre $\mathcal{A}$. Une variable aléatoire réelle X est définie comme une application de Ω dans $\mathbb{R}$, telle que pour tout intervalle ouvert $]a, b[$ de $\mathbb{R}$, l'ensemble $X^{-1}(]a,\ b[) =$

$\{\omega \in \Omega,\ X(\omega) \in]a, b[\}$ soit un événement c'est-à-dire un élément de $\mathcal{A}$. On vérifie aisément (voir exercice 152) que si X est une variable aléatoire réelle, alors pour tout borélien B de $\mathbb{R}$, $X^{-1}(B)$ appartient à $\mathcal{A}$.
D'autre part, sur l'espace $\mathbb{R}^d$ muni de la σ-algèbre borélienne, les variables aléatoires sont aussi appelées fonctions boréliennes.

▶ *Exemple 147.* Les fonctions continues par morceaux sont des fonctions boréliennes. L'indicatrice d'un borélien est une fonction borélienne.

▶ *Exemple 148.* Si X est une fonction définie $(\Omega, \mathcal{A})$ à valeurs dans un ensemble fini ou dénombrable $\mathcal{E} \subset \mathbb{R}$, alors X est une variable aléatoire si et seulement si $X^{-1}(\{e\}) \in \mathcal{A}$ pour tout $e \in \mathcal{E}$.

Propriétés.

(a) Toute limite monotone de variables aléatoires est une variable aléatoire.

(b) Toute variable aléatoire bornée inférieurement est limite croissante d'une suite de variables aléatoires simples (c'est-à-dire ne prenant qu'un nombre fini de valeurs).

(c) Les variables aléatoires forment un espace vectoriel,

(d) L'image d'une variable aléatoire par une fonction borélienne est encore une variable aléatoire .

Preuve :

(a) Si (X_n) croit vers X, alors pour tout $a \in \mathcal{A}$, $X^{-1}(]a, +\infty[) = \cup_n X_n^{-1}(]a, +\infty[) \in \mathcal{A}$ et $X^{-1}(] - \infty, a]) = \cap_n X_n^{-1}(] - \infty, a]) \in \mathcal{A}$. Donc, pour tout $a < b$, $X^{-1}(]a, b]) = X^{-1}(]a, +\infty[) \cap X^{-1}(] - \infty, b]) \in \mathcal{A}$. Comme $X^{-1}(]a, b[) = \cup_m X^{-1}(]a, b - \frac{1}{m}])$, $X^{-1}(]a, b[) \in \mathcal{A}$.

(b) Soit X une variable aléatoire positive. Pour $n \in \mathbb{N}^*$, on pose

$$X_n = \sum_{k=0}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\}} + n\mathbb{1}_{X \geq n}.$$

La suite $(X_n)_n$ est une suite croissante de variables aléatoires telle sur l'événement $\{X < n\}$ $|X_n - X| \leq \frac{1}{n}$. On en déduit que $(X_n)_n$ converge vers X.

(c) Ce résultat vrai de façon évidente pour les variables aléatoires simples, s'étend au cas général, en utilisant (b).

(d) Soit X une variable aléatoire définie sur $(\Omega, \mathcal{A})$ et f une fonction borélienne. Pour tout $B \in \mathcal{B}(\mathbb{R})$, $\{f(X) \in B\} = \{X \in f^{-1}(B)\}$. Comme f est une fonction borélienne, $f^{-1}(B)$ est un borélien et comme X est une variable aléatoire $\{X \in f^{-1}(B)\} \in \mathcal{A}$. Donc $f(X)$ est une variable aléatoire sur $(\Omega, \mathcal{A})$.

□

▷ *Exercice 149.* Soit $(X_n)_n$ est une suite de variables aléatoires définie sur $(\Omega, \mathcal{A})$ telle que pour tout $\omega \in \Omega$, $(X_n(\omega))_n$ converge vers un réel $X(\omega)$.
Montrer que X est une variable aléatoire

Solution. $X = \limsup X_n = \lim_{\to+\infty} \sup_{k>n} X_k$. Pour tout $t \in \mathbb{R}$, $\{\sup_{k>n} X_k \leq t\} = \cap_{k>n}\{X_k \leq t\} \in \mathcal{A}$. Donc d'après la propriété (b), X est une variable aléatoire comme limite croissante de variables aléatoires .

7.3.1 Loi d'une variable aléatoire

Proposition 17 *Si P est une probabilité sur Ω muni de la σ-algèbre $\mathcal{A}$, l'application P_X qui, à tout borélien B de $\mathbb{R}$, associe $P_X(B) = P(X^{-1}(B))$ est une probabilité définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; elle est appelée* la loi de *X.*

Preuve : La preuve est une simple vérification des hypothèses définissant une probabilité à partir des propriétés de l'image réciproque d'un ensemble par une application. □

La fonction de répartition de X est définie comme la fonction de répartition de la loi de X : c'est donc la fonction définie sur $\mathbb{R}$ par $t \mapsto P(X \leq t)$.
Rappelons que d'après le téorème d'unicité 16, la loi de X est entièrement caractérisée par la donnée $P_X(]a, b[)$ pour tout $a < b$ ou par la donnée de sa fonction de répartition.
Si X est à valeurs dans un ensemble fini ou dénombrable $\mathcal{E} \subset \mathcal{A}$, alors la loi de X est entièrement déterminée par la donnée de $P(X = e)$ pour tout $e \in \mathcal{E}$ (en effet, pour tout $B \in \mathcal{A}$, $P_X(B) = \sum_{e\in B\mathcal{E}} P(X = e)$).

► *Exemple 150.* On considère l'espace de probabilité $(\Omega, \mathcal{A}, P)$ avec $\Omega = \mathbb{R}^2$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^2)$ et P la probabilité uniforme sur le rectangle $I = [0, 1] \times [0, 1]$. Alors, l'application $X\ : (\omega_1, \omega_2) \mapsto \omega_1$ est une variable aléatoire réelle de loi uniforme sur $[0, 1]$.
En effet, pour tout intervalle ouvert $]a, b[$ de $\mathbb{R}$, $\{X \in]a, b[\} =]a, b[\times\mathbb{R} \in \mathcal{B}(\mathbb{R}^2)$ et

$$P_X(]a, b[) = P(]a, b[\times\mathbb{R}) = Aire((]a, b[\times\mathbb{R}) \cap I) = longueur(]a, b[\cap[0, 1]).$$

Deux variables aléatoires sont dites *égales presque sûrement (p.s.)*, si elles ne différent que sur un ensemble de probabilité nulle.

N.B. Deux variables aléatoires qui sont égales presque sûrement, ont la même loi. En effet, si X et Y sont deux variables aléatoires définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ alors pour tout $A \in \mathcal{A}$, $P(X \in A) = P(X \in A \text{ et } Y \in A) + P(X \in A \text{ et } Y \notin A)$. Donc, $|P(X \in A) - P(Y \in A)| \leq 2P(X \neq Y)$.

► *Exemple 151.* Soit μ est une probabilité ayant une densité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $a < b$ deux réels. Alors les fonctions X et Y définies sur l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ par $X = \mathbb{1}_{]a,b[}$ et $Y = \mathbb{1}_{[a,b]}$ sont deux variables aléatoires qui sont égales presque sûrement.

▷ *Exercice 152.* Soit X une variable aléatoire réelle définie sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$. Montrer que $\mathcal{T} = \{C \in \mathcal{B}(\mathbb{R}),\ X^{-1}(C) \in \mathcal{A}\}$ est une σ-algèbre. En déduire que $\mathcal{T} = \mathcal{B}(\mathbb{R})$.

Solution. Comme $X^{-1}(\mathbb{R}) = \Omega$, $\mathbb{R} \in \mathcal{T}$. Soit $A \in \mathcal{T}$, $\mathbb{R} \setminus A$ est un borélien de $\mathbb{R}$ et $X^{-1}(\mathbb{R} \setminus A) = \Omega \setminus X^{-1}(A)$. Comme $X^{-1}(A) \in \mathcal{A}$ et comme $\mathcal{A}$ est stable par passage au complémentaire, $\mathbb{R} \setminus A \in \mathcal{T}$. Soit $(A_i)_i$ une famille au plus dénombrable d'éléments de $\mathcal{T}$, alors $A = \cup_{i\in I} A_i$ est encore un borélien et $X^{-1}(A) = \cup_{i\in I} X^{-1}(A_i)$. Comme $X^{-1}(A_i) \in \mathcal{A}$ pour tout $i \in I$ et comme $\mathcal{A}$ est une σ-algèbre, $X^{-1}(A) \in \mathcal{A}$. Donc, $A \in \mathcal{T}$. On a montré que $\mathcal{T}$ est une σ-algèbre. X étant une variable aléatoire , $\mathcal{T}$ contient tous les intervalles ouverts de $\mathbb{R}$. Donc $\mathcal{T}$ contient la σ-algèbre borélienne de $\mathbb{R}$.

▷ *Exercice 153.* (suite de l'exercice 146)
Le score du lanceur est comptabilisé de la façon suivante : si la fléchette s'est fixée sur la cible, son score correspond à 20 moins la distance exprimée en cm entre le centre de la cible et la position de la fléchette sur la cible, si la fléchette n'est pas sur la cible, son score est 0.
On note X le score obtenu à la suite du lancer d'un fléchette.

1. Exprimer X comme une fonction définie sur Ω et montrer que X est bien une variable aléatoire c'est-à-dire que pour $t \in \mathbb{R}$, $\{X \leq t\}$ est un élément de $\mathcal{A}$.
2. Déterminer la fonction de répartition de X.

7.3.2 Vecteurs aléatoires

Plus généralement, un vecteur aléatoire d-dimensionnel $\vec{X}$ (ou une variable aléatoire à valeurs dans $\mathbb{R}^d$) est définie comme un d-uplet de variables aléatoires réelles. On vérifie que pour tout borélien B de $\mathbb{R}^d$, $\vec{X}^{-1}(B)$ est un événement (voir exercice 156).
Les fonctions continues par morceaux sont des fonctions boréliennes. L'indicatrice d'un borélien est une fonction borélienne.

Image d'un vecteur aléatoire par une application. Si X est une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$ et si g est une fonction borélienne de $\mathbb{R}^d$ dans $\mathbb{R}^r$ alors $Y = g(X)$ est encore une variable aléatoire définie sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$. La loi de Y est une probabilité sur $(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r))$. Elle est reliée à la loi de X de la façon suivante : pour tout $A \in \mathcal{B}(\mathbb{R}^r)$, $P_Y(A) = P_X(g^{-1}(A))$ puisque

$$P_Y(A) = P(\{\omega \in \Omega,\ g(X(\omega)) \in A\}) = P(X \in g^{-1}(A)).$$

Lois marginales Si $X = (X_1, \dots, X_r)$ est un vecteur aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ de loi μ sur $\mathbb{R}^r$ alors pour tout $i \in \{1, \dots, r\}$, la loi de la marginale X_i est la probabilité P_{X_i} sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par :

$$P_{X_i}(A) = \mu(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{r-i}) \text{ pour tout } A \in \mathcal{B}(\mathbb{R}).$$

En particulier, si la loi de X a pour densité ϕ alors pour tout $t \in \mathbb{R}$,

$$\begin{aligned} P(X_i \leq t) &= P(X \in \mathbb{R}^{i-1} \times]-\infty, t] \times \mathbb{R}^{r-i}) \\ &= \int_{-\infty}^{t} \left(\int_{\mathbb{R}^{r-1}} \phi(x_1, \dots, x_r) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_r \right) dx_i. \end{aligned}$$

Donc, la loi de X_i a pour densité

$$u \mapsto \int_{\mathbb{R}^{r-1}} \phi(x_1, \dots, x_{i-1}, u, x_{i+1}, \dots, x_r) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_r.$$

► *Exemple 154.* Supposons que (X_1, X_2) suive la loi uniforme sur le disque unité de $\mathbb{R}^2$: $D = \{(x, y) \in \mathbb{R}^2,\ x^2 + y^2 \leq 1\}$. Alors X_1 et X_2 ont pour densité la fonction f définie sur $\mathbb{R}$ par : $f(x) = \frac{1}{\pi} \int \mathbb{1}_{\{x^2+y^2 \leq 1\}} dy$ pour tout $x \in \mathbb{R}$. Donc, si $|x| > 1$, $f(x) = 0$ et si $|x| \leq 1$, $f(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1 - x^2}$.

► *Exemple 155.*

1. Si X est une variable aléatoire de loi normale $\mathcal{N}(m,\sigma^2)$, alors $Y = \frac{X-m}{\sigma}$ suit la loi $\mathcal{N}(0,1)$.
2. Si U suit la loi uniforme sur $[0,1]$, alors pour tout $k \in \mathbb{N}^*$, $Ent(kU)$ suit la loi uniforme sur $\{0,\ldots,k-1\}$.

▷ *Exercice 156.* Soit $\vec{X}$ une vecteur aléatoire d-dimensionnel défini sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$. Montrer que $\mathcal{T} = \{C \in \mathcal{B}(\mathbb{R}^d),\ X^{-1}(C) \in \mathcal{A}\}$ est une σ-algèbre qui contient tous les pavés ouverts de $\mathbb{R}^d$. En déduire que $\mathcal{T} = \mathcal{B}(\mathbb{R}^d)$.

▷ *Exercice 157.* Soit F une fonction définie sur $\mathbb{R}$, croissante, continue à droite et ayant pour limites 0 en $-\infty$ et 1 en $+\infty$. Soit G son pseudo-inverse continu à droite, $G^s = \inf(\{u,\ F(u) \geq s\})$ pour tout $s \in]0,1[$ (noter que le graphe de G est le symétrique du graphe de F par rapport à la diagonale). Soit U une variable aléatoire de loi uniforme sur $(0,1)$. Montrer que $G(U)$ a pour fonction de répartition F.

▷ *Exercice 158.* Soit X une variable aléatoire de loi exponentielle $\mathcal{E}(a)$.

1. Déterminer la loi de bX lorsque $b \in \mathbb{R}^*$.
2. Déterminer la loi de la partie entière de X.
3. Déterminer la loi de $-\log(U)$ lorsque U est une variable aléatoire de loi uniforme sur $[0,1]$.

▷ *Exercice 159.* On dit qu'une variable aléatoire X suit la loi lognormale $\mathcal{LN}(m,\sigma^2)$ si $\log(X)$ suit la loi $\mathcal{N}(m,\sigma^2)$. Déterminer la densité de la loi $\mathcal{LN}(m,\sigma^2)$.

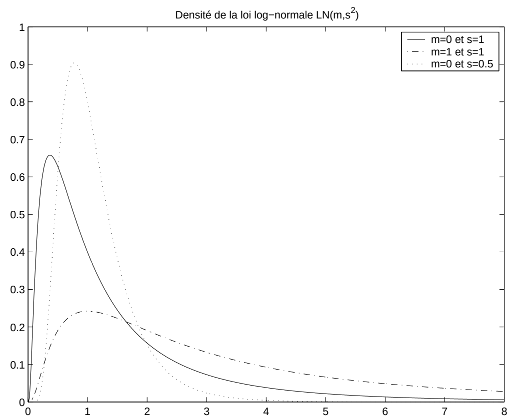


FIG. 7.6 : Représentation des densités de quelques lois lognormales

▷ *Exercice 160.* Soit $(X_1,\ldots,X_d)$ une vecteur gaussien de loi $\mathcal{N}(m,C)$ où C est une matrice positive de rang d.
Déterminer la loi des marginales X_i, $i \in \{1,\ldots,d\}$.

7.3.3 σ-algèbre engendrée par une famille de variables aléatoires

Soit $\{X_i, \imath \in I\}$ une famille (le plus souvent finie ou dénombrable) de variables aléatoires réelles. On note $\sigma(X_i, i \in I)$ la σ-algèbre engendrée par les événements de la forme $X_i^{-1}(A)$, $i \in I$, $A \in \mathcal{B}(\mathbb{R})$.

▶ *Exemple 161.* Soit X une variable aléatoire définie sur $(\Omega,\mathcal{A})$. L'ensemble $\{X^{-1}(B),\ B \in \mathcal{B}(\mathbb{R})\}$ est une σ -algèbre (conséquence directe des propriét'es de l'image réciproque d'un ensemble par une application): c'est donc la σ-algèbre engendrée par X.
En particulier, pour tout $A \in \mathcal{A}$, la tribu engendrée par l'indicatrice de l'ensemble A est l'ensemble $\{\emptyset, A, A^c, \Omega\}$.
Plus généralement, si $X = (X_1,\ldots,X_d)$ est un vecteur aléatoire définie sur $(\Omega,\mathcal{A})$, alors $\{X^{-1}(B),\ B \in \mathcal{B}(\mathbb{R}^d)\} = \sigma(X_1,\ldots,X_d)$.

Preuve : Comme précédemment, on montre facilement, que l'ensemble $\mathcal{F} = \{X^{-1}(B),\ B \in \mathcal{B}(\mathbb{R}^d)\}$ est une σ-algèbre. Elle contient les ensembles $X_i^{-1}(A)$ pour $A \in \mathcal{B}(\mathbb{R})$ puisque $X_i^{-1}(A) = X^{-1}(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{d-i})$. Donc, $\mathcal{F}$ contient $\sigma(X_1, \dots, X_d)$.
Il reste à montrer que si $\mathcal{G}$ est une tribu contenant tous les ensembles de la forme $X_i^{-1}(A)$ pour $i \in \{1, \dots, d\}$ et $A \in \mathcal{B}(\mathbb{R})$, alors $\mathcal{F} \subset \mathcal{G}$. Posons $\mathcal{C} = \{C \subset \mathbb{R}^d,\ X^{-1}(C) \in \mathcal{G}\}$. L'ensemble des pavés mesurables de $\mathbb{R}^d$ est inclus dans $\mathcal{C}$. D'autre part, on vérifie que $\mathcal{C}$ est une tribu. Donc, elle contient $\mathcal{B}(\mathbb{R}^d)$. Cela signifie que $\mathcal{F}$ est incluse dans $\mathcal{G}$ et donc que $\mathcal{F} = \sigma(X_1, \dots, X_d)$. □

Proposition 18 *Soit* $X = (X_1, \dots, X_d)$ *et* $Y = (Y_1, \dots, Y_r)$ *deux vecteurs aléatoires définies sur* $(\Omega, \mathcal{A})$.
Alors $\sigma(Y) \subset \sigma(X)$ *si et seulement si il existe une application borélienne* $f : \mathbb{R}^d \mapsto \mathbb{R}^r$ *telle que* $Y = f(X)$.

Preuve : Si $Y = f(X)$ alors pour tout $B \in \mathcal{B}(\mathbb{R}^r)$, $\{Y \in B\} = \{X \in f^{-1}(B)\}$. Comme f est borélienne, $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^d)$. Donc, $\{Y \in B\} \in \sigma(X_1, \dots, X_d)$.
Pour simplifier les notations, on va démontrer la réciproque dans le cas $d = r = 1$. La preuve se généralise facilement au cas d et r quelconques. Supposons que $\sigma(Y) \subset \sigma(X)$.
Vérifions l'assertion pour les vas simples : si $Y = \sum_{i=1}^n y_i \mathbb{1}_{A_i}$ avec $A_i = \{Y = y_i\}$, alors $A_i \in \sigma(X)$ pour tout $i \in \{1, \dots, n\}$. D'après la description de $\sigma(X)$ donné dans l'exemple 161, pour tout $i \in \{1, \dots, n\}$, il existe $B_i \in \mathcal{B}(\mathbb{R})$ tel que $A_i = X^{-1}(B_i)$. On en déduit que $Y = f(X)$ avec $f = \sum_{i=1}^n y_i \mathbb{1}_{B_i}$ qui est bien une fonction borélienne.
Supposons simplement Y positive. Y peut être approchée par une suite croissantes variables aléatoires simples Y_n telles que $\sigma(Y_n) \in \sigma(X)$. $Y_n = \sum_{k=-n2^n}^{n2^n} \frac{k}{2^n} \mathbb{1}_{\{\frac{k-1}{2^n} \le X < \frac{k}{2^n}\}}$. Donc, pour tout n, il existe une fonction borélienne f_n telle que $Y_n = f_n(X)$. f_n est de la forme $f_n = \sum_{j=1}^{r_n} a_{i,n} \mathbb{1}_{A_{i,n}}$. On pose alors $\tilde{f}_n = \sum_{j \in T_n} a_{i,n} \mathbb{1}_{A_{i,n}}$ avec $T_n = \{i,\ A_{i,n} \cap X(\omega) \neq \emptyset\}$. Alors $(\tilde{f}_n)_n$ est une suite croissante de fonctions boréliennes et $Y_n = \tilde{f}_n(X)$ pour tout n. Posons $\tilde{f} = \sup_n \tilde{f}_n$ et $f = \tilde{f} \mathbb{1}_{\{\tilde{f} < +\infty\}}$ alors f est une fonction borélienne et $Y = f(X)$.
Dans le cas général, on se ramème au cas où Y est positive en considérant les parties positives et négatives Y^+ et Y^-. □

7.4 Indépendance

On peut maintenant donner la définition générale de l'indépendance de variables aléatoires.

Définition. Soit $X_1, \dots, X_n$ n variables aléatoires définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. *Les variables aléatoires* $X_1, \dots, X_n$ *sont dites indépendantes* si leur loi conjointe (c'est-à-dire la loi du vecteur $(X_1, \dots, X_n)$) est égale au produit des lois des variables aléatoires X_i, $i = 1, \dots, n$.

► *Exemple 162.*

1. Si (X_1, X_2) est un vecteur aléatoire de loi uniforme sur un rectangle R de $\mathbb{R}^2$ alors X_1 et X_2 sont indépendantes et suivent des lois uniformes.
2. Si $X = (X_1, X_2)$ est un vecteur aléatoire de loi uniforme sur le disque unité $D(0,1)$ alors X_1 et X_2 ne sont pas indépendantes. Pour le montrer, il suffit de constater que $P((X_1, X_2) \in [\frac{\sqrt{2}}{2}, 1] \times [\frac{\sqrt{2}}{2}, 1]) = 0$, alors que $P(X_1 \in [\frac{\sqrt{2}}{2}, 1])$ et $P(X_2 \in [\frac{\sqrt{2}}{2}, 1])$ sont non nuls.
3. La densité gaussienne $\gamma_{m,C}$ définie au paragraphe 7.1.3 est une loi produit si et seulement si C est une matrice diagonale. Un vecteur aléatoire dont la loi est donnée par une densité gaussienne est appelé un vecteur gaussien.

▷ *Exercice 163.* Soit $N \in \mathbb{N}^*$. On "choisit au hasard un réel" dans l'intervalle $[0, N[$: le réel choisi est modélisé par une variable aléatoire X de loi uniforme sur $[0, N[$. On note Y sa partie entière et Z sa partie fractionnaire : $Z = X - Y$. Montrer que Y et Z sont des variables aléatoires indépendantes.

Solution. On pose $\Omega = [0, N[$. On munit Ω de la tribu borélienne restreinte à $[0, N[$ (c'est-à-dire $\mathcal{B}([0, N[) = \{A \cap [0, N[,\ A \in \mathcal{B}(\mathbb{R})\}$ et de la loi uniforme P. On définit Yet Z sur Ω par $Y : \omega \mapsto [\omega]$ et $Z : \omega \mapsto \omega - [\omega]$. Alors pour tout $k \in \{0, \ldots, N-1\}$ et $0 \leq a < b \leq 1$, $\{Y = k$ et $Z \in [a, b[\} = [a + k,\ b + k[$. C'est donc un événement de probabilité $\frac{b-a}{N}$. D'autre part, $\{Y = k\} = [k, k+1[$ est un événement de probabilité $\frac{1}{N}$ et $\{Z \in [a, b[\} = \cup_{k=0}^{N-1} [a + k,\ b + k[$ est un événement de probabilité $b - a$. Donc $P(\{Y = k$ et $Z \in [a, b[\}) = P(Y = k)P(Z \in [a, b[)$, ce qui montre que Y et Z sont indépendants.

7.4.1 Indépendance des événements et des σ-algèbres

L'indépendance des événements se définit comme dans le cas des espaces dénombrables :

Définition. n évènements $A_1, ..., A_n$ sont dit *indépendants* si les variables aléatoires $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ sont indépendantes.

On retrouve alors, dans ce contexte général, que m événements $A_1, A_2, \ldots, A_m$ sont dits *indépendants* si et seulement si la probabilité de l'intersection d'un nombre quelconque d'entre eux est le produit de leurs probabilités.

Définition. n σ-algèbres $\mathcal{A}_1, ..., \mathcal{A}_n$ sont dit *indépendants* si pour tout $A_1 \in \mathcal{A}_1$,..., $A_n \in \mathcal{A}_n$ les événements $A_1, \ldots, A_n$ sont indépendants.

Notons que $X_1, \ldots, X_n$ sont de variables aléatoires indépendantes si et seulement si les σ-algèbres $\sigma(X_1), \ldots, \sigma(X_n)$ sont indépendantes

▷ *Exercice 164.* Soit $\Omega, \mathcal{A}, P)$ un espace de probabilité et soit $\mathcal{T}_1$ et $\mathcal{T}_2$ deux σ-algèbres incluses dans $\mathcal{A}$. Montrer que si $A \in \mathcal{T}_1 \cap \mathcal{T}_\in$ et si $\mathcal{T}_1$ et $\mathcal{T}_2$ sont indépendantes alors $P(A) = 0$ ou $P(A) = 1$.

Solution. Si $\mathcal{T}_1$ et $\mathcal{T}_2$ sont indépendantes, alors pour tout $A \in \mathcal{T}_1 \cap \mathcal{T}_\in$, $P(A \cap A) = P(A)^2$, c'est-à-dire $P(A) = 0$ ou $P(A) = 1$.

▷ *Exercice 165.* Soit X une variable aléatoire et f une fonction borélienne. Montrer que si X et $f(X)$ sont indépendantes, alors $f(X)$ est presque sûrement constante.

Solution. Posons $Y = f(X)$. Les σ-algèbres $\sigma(Y)$ et $\sigma(X)$ sont indépendantes et $\sigma(Y)$ est inclus dans $\sigma(X)$. Par conséquent, pour tout $t \in \mathbb{R}$, $\{Y \leq t\}$ qui appartient à $\sigma(Y)$ et $\sigma(X)$ est de probabilité 0 ou 1, d'après l'exercice 164. Posons $t^* = \inf\{t \in \mathbb{R}, P(Y \leq t) = 1$. Alors $P(Y = t^*) = P(Y \leq t) - \lim_n P(Y \leq t - \frac{1}{n}) = 1$.

▷ *Exercice 166.* Soit $X = (X_1, X_2)$ un vecteur de loi uniforme sur le disque unité D centré à l'origine. On note R la norme de ce vecteur et Θ l'angle que fait ce vecteur avec l'axe des abscisses.
Montrer que les variables aléatoires R et Θ sont indépendantes et déterminer leurs densités.

▷ *Exercice 167.* (La Recherche n^o33 Avril 73 : la mobilité des foules)
On s'intéresse à la vitesse de déplacement de piétons sur un parcours déterminé. On admet le modèle suivant : la vitesse d'un individu isolé, choisi au hasard, est une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$ avec

- $m = 1.4m/s$ et $\sigma = 0.4m/s$ pour une femme,
- $m = 1.6m/s$ et $\sigma = 0.4m/s$ pour un homme.

La population est composée d'un tiers d'hommes et de deux tiers de femmes. Soit V la vitesse de déplacement d'un individu isolé, choisi au hasard, dont on ignore le sexe.

1. Calculer la probabilité que V soit inférieur à $1.2m/s$.
2. La vitesse d'un individu a été trouvé inférieure à $1.2m/s$. Quelle est la probabilité pour qu'il s'agisse d'une femme ?
3. Déterminer l'expression explicite de la densité de V.

7.4.2 Famille de variables aléatoires indépendantes

Définition. Soit $\{X_i,\ i \in I\}$ une famille de variables aléatoires . On dit que ces variables aléatoires sont indépendantes si pour tout n et pour tout n-uplet $(i_1, \ldots, i_n)$ d'éléments de I, les variables aléatoires $X_{i_1}, \ldots, X_{i_n}$ sont indépendantes.

▷ *Exercice 168.* Soit $\{X_i,\ i \in I\}$ une famille de variables aléatoires.
Montrer que si $I_1, \ldots, I_n$ est une partition de I, les σ-algèbres $\sigma(X_i,\ i \in I_1)$, ...,$\sigma(X_i,\ i \in I_n)$ sont indépendantes.

7.4.3 Convolution de lois

Si X et Y sont des variables aléatoires réelles indépendantes de loi respectivement μ et ν définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ alors $Z = X + Y$ est une variable aléatoire réelle. Sa fonction de répartition F est définie en tout point $t \in \mathbb{R}$ par

$$F(t) = P((X,Y) \in D_t) = \mu * \nu(D_t) \text{ avec } D_t = \{(x,y) \in \mathbb{R}^2,\ x + y \leq t\}.$$

On appelle la loi de Z *la convolée des lois* μ *et* ν et on la note $\mu * \nu$.
Si μ et ν ont des densités notées respectivement ϕ et ψ alors pour tout $t \in \mathbb{R}$,

$$F(t) = \int_{D_t} \phi(x)\psi(y)dxdy.$$

En utilisant le théorème de Fubini-Tonelli et en faisant le changement de variable $y \mapsto x + y$ à x fixé, on obtient :

$$F(t) = \int_{\mathbb{R}} \phi(x) \left(\int_{\mathbb{R}} \psi(z - x) 1\!\!1_{z \leq t} dz \right) dx = \int_{-\infty}^{t} \left(\int_{\mathbb{R}} \phi(x)\psi(z - x)dx \right) dz$$

Donc, $\mu * \nu$ a pour densité $z \mapsto \int_{\mathbb{R}} \phi(x)\psi(z - x)dx = \int_{\mathbb{R}} \phi(z - x)\psi(x)dx$.

▷ *Exercice 169.* (Convolution de densités gaussiennes)
Montrer que si X et Y sont deux variables aléatoires indépendantes de loi gaussienne respectivement $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ alors $X + Y$ est une variable aléatoire de loi gaussienne $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$

Solution. Pour simplifier les calculs, on se ramème au cas où $m_1 = m_2 = 0$ et $\sigma_1^2 + \sigma_2^2 = 1$ en utilisant que si Z suit la loi $\mathcal{N}(0, 1)$ alors pour tout $\sigma > 0$ et $m \in \mathbb{R}$, $\sigma Z + m$ suit la loi $\mathcal{N}(m, \sigma)$. Sous ces hypothèses, on doit montrer que $X + Y$ a pour loi $\mathcal{N}(0, 1)$. Comme X et Y sont indépendantes, $X + Y$ a pour densité la fonction

$$g : z \mapsto \int_{\mathbb{R}} \frac{1}{2\pi\sigma_1\sqrt{1 - \sigma_1^2}} \exp(-\frac{1}{2}(\frac{x^2}{\sigma_1^2} + \frac{(z - x)^2}{1 - \sigma_1^2}))dx.$$

En remarquant que $\frac{x^2}{\sigma_1^2} + \frac{(z-x)^2}{1-\sigma_1^2} = z^2 + \frac{1}{1-\sigma_1^2}(\frac{x}{\sigma_1} - \sigma_1 z)^2$ et en utilisant que

$$x \mapsto \frac{1}{\sqrt{2\pi}\sqrt{1 - \sigma_1^2}} \exp(-\frac{1}{2(1 - \sigma_1^2)}(\frac{x}{\sigma_1} - \sigma_1 z)^2)$$

est la densité de la loi $\mathcal{N}(\frac{\sigma_1^2}{\sqrt{1-\sigma_1^2}} z, \sigma_1^2(1 - \sigma_1^2))$ et donc d'intégrale 1, on obtient que g est bien la densité de la loi $\mathcal{N}(0, 1)$.

▷ *Exercice 170. Application de l'exercice 169 :* sur les paquets de sucre d'une marque M, il est écrit : poids moyen en sucre 1000 g avec un écart-type de 10 g. On note m le poids moyen en sucre des paquets de cette marque (exprimé en grammes). On se propose de tester si m vaut bien 1000 g comme cela est écrit sur les paquets. Pour cela, on dispose d'une balance et d'un paquet de sucre de marque M. En tenant compte du poids de l'emballage, on fera l'hypothèse que le résultat de la pesée de ce paquet est une réalisation d'une variable aléatoire X_1 de loi $\mathcal{N}(m+10,(10)^2)$.

1. Construire un test de niveau 5% de l'hypothèse $H_0 : m = 1000$ contre l'hypothèse $H_1 : m < 1000$ en utilisant la variable aléatoire X_1.
2. Quelle est la probabilité de garder l'hypothèse H_0 avec ce test si le poids moyen en sucre d'un paquet de la marque M est en réalité de 990 g ?
3. La balance indique que le poids du paquet est de 1000 g. Que concluez-vous ?
4. On pèse trois autres paquets de sucre de la même marque avec la même balance ; le résultat de ces trois pesées est 990 g, 1010 g et 1000 g. On interprétera ces valeurs comme des réalisations de trois variables aléatoires X_2, X_3 et X_4 de même loi que X_1 et on supposera que les variables aléatoires X_1, X_2, X_3 et X_4 sont indépendantes.
   Reprendre les questions 1, 2 et 3 en utilisant maintenant la variable aléatoire $Y = \frac{1}{4}\sum_{i=1}^{4} X_i$.
5. Les réponses à la question 4 sont-elles modifiées si on remplace dans le test l'hypothèse H_0 par l'hypothèse $H_{0'} : m \geq 1000$ (on justifiera sa réponse) ?

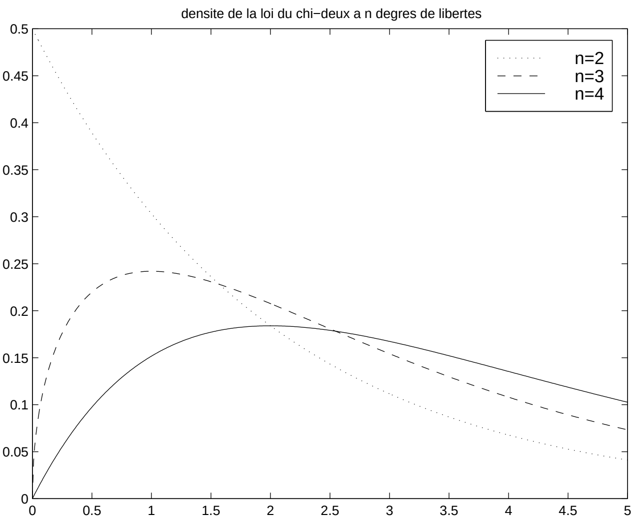
▷ *Exercice 171.* (*Loi Gamma*) On appelle loi Gamma de paramètres $a > 0$ et $b > 0$ (notée $\Gamma(a,b)$) la loi dont la densité est donnée par la fonction :

$$x \mapsto \frac{b^a}{\gamma(a)} x^{a-1} \exp(-bx) \mathbb{1}_{\{x\geq 0\}} \text{ où } \gamma(a) = \int_0^{+\infty} t^{a-1} \exp(-t) \mathbb{1}_{\{t\geq 0\}} dt.$$

Lorsque a est un entier, la loi est aussi connue sous le nom de loi d'Erlang.

1. Vérifier qu'il s'agit bien d'une densité de probabilité.
2. Soit a, b, c des réels strictement positifs. Soit X et Y deux variables aléatoires indépendantes de loi $\Gamma(a,b)$ et $\Gamma(c,b)$. Montrer que cX suit la loi $\Gamma(a,\frac{b}{c})$ et que $X+Y$ a pour loi $\Gamma(a+c,b)$.
3. Le temps (exprimé en minutes) qui s'écoule entre deux appels consécutifs à un standard téléphonique est supposé aléatoire de loi $\mathcal{E}(0.5)$ et indépendant des autres appels téléphoniques. Quelle est la loi du temps qui s'écoule entre le premier et le 10-ème appel ?

Les densités du chi-deux. La densité du *chi-deux à d degrés de liberté, notée* χ^2_d est définie comme le produit de convolution de d densités $\chi^2_1(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} \exp(-\frac{x}{2}) \mathbb{1}_{x>0}$. Par conséquent, si $Y_1, \ldots, Y_d$ est un d-échantillon de variables aléatoires de loi $\mathcal{N}(0,1)$ définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$, alors la loi de $\sum_{i=1}^{d} X_i^2$ a pour densité χ^2_d (Les lois du χ^2 sont tabulées).



Remarquons que d'après ce qui précède et l'exercice 133, si la suite des lois de vecteurs aléatoires de $\mathbb{R}^d$ $\overrightarrow{X_n}$ convergent vers la densité produit γ_{0,I_d}[3], alors les lois de $||\overrightarrow{X_n}||^2$ convergent vers la densité χ^2_d.

▷ *Exercice 172.*

1. Montrer que χ^2_2 est égale à la densité exponentielle de paramètre $\frac{1}{2}$.
2. Plus généralement, montrer que $\chi^2_d(x) = \frac{2^{-\frac{d}{2}}}{\gamma(\frac{d}{2})} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{x>0}$ et vérifier que c'est aussi la densité de la loi $\Gamma(\frac{d}{2}, \frac{1}{2})$.
3. *Application :* Une lampe de poche fonctionne avec une pile de 3V. On suppose que la durée de fonctionnement d'une lampe de poche avec une pile de marque M suit une loi exponentielle de paramètre $\lambda > 0$. Lorsque la lampe ne fonctionne plus, on remplace la pile par une pile de la même marque.
   Si la durée moyenne de fonctionnement de cette lampe avec une telle pile est de $3h$, combien de piles faut-il prévoir pour qu'il y ait moins d'une chance sur 20 que la lampe tombe en panne faute de pile au cours d'une nuit de $9h$?

▷ *Exercice 173.* Une expérience de travaux pratiques consiste à déterminer la teneur en calcium d'une solution. Les enseignants considèrent que le résultat d'un dosage est une variable aléatoire gaussienne, d'espérance 5mg/l, et dont l'écart-type σ dépend de l'habileté de l'expérimentateur, mais ne doit pas excéder 0.05 mg/l pour quelqu'un de normalement entrainé.
Un étudiant effectue une série de n dosages indépendants et note, pour tout $i \in \{1, \dots, n\}$, X_i le résultat du i-ème dosage. Soit $S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - 5)^2$.

1. Quelle est la loi de $Z_n := nS_n^2/\sigma^2$?
2. Soit σ_0 un réel strictement positif supérieur à σ. Exprimer la fonction de répartition de nS_n^2/σ_0^2 en fonction de celle de Z_n. En déduire la relation entre la densité de nS_n^2/σ_0^2 et celle de Z_n.
3. Pour déteminer si l'étudiant est habile ou non, on se propose de faire le test de $H_0 \ : \ \sigma = 0.05$ contre $H_1 \ : \ \sigma > 0.05$ au niveau 5%. Justifier le critère de choix suivant : on rejette H_0 si la valeur observée $s^2_{n,obs}$ de S_n^2 est exceptionnellement grande c'est-à-dire si $P_{0.05}(S_n^2 \geq s^2_{n,obs}) < 0.05$.
4. Que peut-on dire, avec ce test, de l'habilité de l'étudiant si $n = 7$ et si les résultats expérimentaux sont : 5.04, 5.13, 5.07, 5, 5.08, 4.99, 5.11 mg/l ?
5. La réponse à la question précédente serait-elle différente si on effectue plutôt un test de H_0 : $\sigma \leq 0.05$ contre H_1 : $\sigma > 0.05$, de niveau 5% ?

[3] I_d désigne la matrice identité de dimension d.

Chapitre 8

Espérance et variance

8.1 Espérance

Nous allons étendre la notion d'espérance dans le cas général. Dans le cas d'une variable aléatoire X à valeurs positives définie sur un ensemble Ω au plus dénombrable, nous avons vu que $E(X) = \int_0^{+\infty} P(X \geq t)dt$. Cette formule garde un sens pour une variable aléatoire à valeurs positives définie sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ quelconque.

Définition. Soit X une variable aléatoire à valeurs positives. L'espérance de X, notée $E(X)$, est définie par $E(X) = \int_0^{\infty} P(X \geq t)dt$.

N.B.

- pour tout $A \in \mathcal{A}$, $E(\mathbb{1}_A) = P(A)$ puisque $P(\mathbb{1}_A \geq t) = P(A)\mathbb{1}_{0<t\leq 1}$.
- pour une variable aléatoire simple $X = \sum_{i=1}^n x_i \mathbb{1}_{A_i}$, on a $E(X) = \sum_{i=1}^n x_i P(A_i)$.
- $E(X)$ est aussi notée $\int X(\omega)dP(\omega)$ ou simplement $\int XdP$.

► *Exemple 174.*

- Si X est à valeurs dans un ensemble au plus dénombrable $\mathcal{X} \subset \mathbb{R}_+$, alors

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

Preuve : Pour montrer cette égalité, il suffit de remplacer $P(X \geq t)$ par $\sum_{x\in\mathcal{X}} \mathbb{1}_{\{x\geq t\}} P(X = x)$ dans l'expression de $E(X)$ et d'intervertir l'intégrale et la somme (ce qui est justifié puisque tous les termes sont positifs). □

- Si X est une variable aléatoire positive qui a pour densité f, alors

$$E(X) = \int_0^{+\infty} xf(x)dx.$$

Preuve : Par définition, $E(X) = \int_0^{+\infty}(\int_t^{+\infty} f(x)dx)dt$. En intervertissant les deux intégrales grâce au théorème de Fubini-Tonelli, on obtient $E(X) = \int_0^{+\infty} f(x)(\int_0^{\infty} \mathbb{1}_{\{t\leq x\}}dt) = \int_0^{+\infty} xf(x)dx$. □

- Soit X est une variable aléatoire positive dont la loi est une loi mélange $P_X = \sum_{i=1}^{+\infty} \alpha_i\mu_i$ avec μ_i des mesures de probabilité définie sur $\mathbb{R}_+$ et α_i des réels positifs de somme 1. Désignons par Y_i une variable aléatoire de loi μ_i pour tout $i \in \mathbb{N}^*$.
Alors $E(X) = \sum_{i=1}^{+\infty} \alpha_i E(Y_i)$.

Preuve : Cette formule s'obtient en utilisant que pour tout $t \geq 0$, $P(X \geq t) = \sum_{i=1}^{+\infty} \alpha_i \mu_i([t, +\infty[)$.
□

L'espérance vérifie les propriétés suivantes :

Propriétés. *Notons C_+ l'ensemble des variables aléatoires positives définies sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$.*

(i) pour tout $X, Y \in C_+$ telles que $X \leq Y$, $E(X) \leq E(Y)$,

(ii) pour toute suite croissante $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires positives $(E(X_n))_n$ converge vers $E(\sup_n(X_n))$ (convergence croissante).

(iii) pour tout $a \in \mathbb{R}$ et $X, Y \in C_+$, $E(X + aY) = E(X) + aE(Y)$ (linéarité).

(iv) Si $(X_n)_n$ est une suite de variables aléatoires appartenant à C_+ alors $E(\underline{\lim}_n X_n) \leq \underline{\lim}_n E(X_n)$ (Lemme de Fatou).

(v) Si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires positives, alors $E(\sum_{i=0}^{+\infty} X_i) = \sum_{i=0}^{+\infty} E(X_i)$ (σ-additivité)

Preuve :

(i) Si $X \leq Y$ alors pour tout $t \geq 0$, l'événement $\{X \geq t\}$ est inclus dans l'événement $\{Y \geq t\}$, ce qui montre que $E(X) \leq E(Y)$.

(ii) Pour tout $t \geq 0$, $(\{X_n \geq t\})_n$ est une suite croissante dont la réunion est l'événement $\{\sup_n X_n \geq t\}$. On conclut en utilisant le théorème de convergence monotone pour l'intégrale de Lebesgue.

(iii) D'après la remarque 8.1, l'espérance est linéaire pour des variables aléatoires simples. On conclut d'après (ii), sachant que toute variable aléatoire positive est limite d'une suite croissante de variables aléatoires simples.

(iv) $\underline{\lim} X_n$ est la limite de la suite croissante de variables aléatoires $(\inf_{k\geq n} X_k)_n$. D'après (i), pour tout $n \in \mathbb{N}$ et $j \geq n$, $E(X_j) \geq E(\inf_{k\geq n} X_k)$. Donc, pour tout $n \in \mathbb{N}$, $\inf_{j\geq n} E(X_j) \geq E(\inf_{k\geq n} X_k)$. D'après (ii), $(E(\inf_{k\geq n} X_k))_n$ est une suite croissante qui converge vers $E(\sup_n(\inf_{k\geq n} X_k))$. On a donc

$$\underline{\lim} E(X_n) = \lim_{n\to+\infty} \inf_{j\geq n} E(X_j) \geq E(\sup_n(\inf_{k\geq n} X_k)) = E(\underline{\lim} X_n).$$

(v) D'après (ii), $E(\sum_{i=1}^{n} X_i)$ converge vers $E(\sum_{i=1}^{+\infty} X_i)$. Par linéarité de l'espérance,

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) \text{ pour tout } n.$$

Donc, $E(\sum_{i=1}^{n} X_i)$ converge aussi vers $\sum_{i=1}^{+\infty} E(X_i)$ ce qui établit la σ-additivité de l'espérance.

□

N.B. On peut montrer que E est l'unique fonction sur C_+ à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$, telle que pour tout $A \in \mathcal{A}$, $E(\mathbb{1}_A) = P(A)$ et telle que les propriétés (ii) et (iii) soient vérifiées.

N.B. Si X est un vecteur aléatoire et g une fonction borélienne positive,

$$E(g(X)) = \int_{\mathbb{R}^d} g(x) dP_X(x).$$

En effet, pour tout $t \in \mathbb{R}$ $P(g(X) \geq t) = P_X(g \geq t)$ (g étant une variable aléatoire définie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Donc $E(g(X)) = \int_0^\infty P_X(g \geq t) = \int_{\mathbb{R}} g dP_X$.

▷ *Exercice 175.* (*suite de l'exercice 146*) Déterminer l'espérance de X.

▷ *Exercice 176.* Montrer que pour tout $X \in C_+$ et $a > 0$, $P(X \geq a) \leq \frac{E(X)}{a}$.
Indication : donner un majorant de l'espérance de la variable aléatoire $X\mathbb{1}_{X \geq a}$.
En déduire que si $X \in C_+$ et si $E(X) = 0$ alors $P(X = 0) = 1$.

L'ensemble des variables aléatoires réelles X définies sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$ et intégrables c'est-à-dire telles que $E(|X|) < +\infty$ est noté $\mathcal{L}^1(\Omega, \mathcal{A}, P)$. L'espérance se prolonge en une forme linéaire sur $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ en posant $E(X) = E(X^+) - E(X^-)$. Les propriétés suivantes se déduisent de celles de l'espérance d'une variable aléatoire positive :

Propriétés.

1. $|E(X)| \leq E(|X|)$.
2. *L'ensemble* $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ *est un espace vectoriel.*
3. *La fonction définie sur* $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ *par* $||X||_1 = \mathrm{E}(|X|)$ *pour tout* $X \in \mathcal{L}^1(\Omega, P)$ *est une semi-norme :*
   - *pour tout* $a \in \mathbb{R}$, $||aX||_1 = |a| ||X||_1$,
   - *pour tout* $X, Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, $||X + Y||_1 \leq ||X||_1 + ||Y||_1$.
4. *(Théorème de convergence dominée) Soit* X *une variable aléatoire définie sur un espace de probabilité* $(\Omega, \mathcal{A}, P)$ *et* $(X_n)_n$ *une suite de variables aléatoires définies sur* $(\Omega, \mathcal{A}, P)$ *telle que :*
   (i) *pour tout* $\omega \in \Omega$, $(X_n(\omega))_n$ *converge vers* $X(\omega)$.
   (ii) *il existe une variable aléatoire positive telle que* $E(Y) < +\infty$ *et* $|X_n| \leq Y$ *pour tout* $n \in \mathbb{N}$.

   Alors, X *est intégrable et* $(E(|X_n - X|))_n$ *converge vers* 0.

Preuve : La propriété 1, s'obtient en écrivant la valeur absolue d'un réel à l'aide des parties positive et négative de ce nombre. Les propriétés 2 et 3 est une conséquence directe des propriétés de l'espérance des variables aléatoires positives. Le théorème de convergence dominée s'obtient en appliquant le lemme de Fatou aux suites $(|X_n - X|)_n$ et $(2Y - |X_n - X|)_n$. □

N.B. $||\cdot||_1$ n'est pas en général une norme car $||X||_1 = 0$ implique seulement que $P(X = 0) = 1$. En considérant l'espace des classes de variables aléatoires, appartenant à $\mathcal{L}^1(\Omega, \mathcal{A}, P)$, définies par la relation d'équivalence $X \sim Y$ si $P(X = Y) = 1$, on peut montrer qu'on obtient un espace de vectoriel de Banach (c'est-à-dire normé et complet). Cette espace est noté $(L^1(\Omega, \mathcal{A}, P), ||\cdot||_1)$.

Exemples de calculs d'espérances

Cas d'une variable aléatoire discrète Si X est une variable aléatoire discrète c'est-à-dire telle que $X(\Omega)$ soit un ensemble au plus dénombrable et si f est une fonction numérique, alors $Y = f(X)$ est intégrable si et seulement si

$$\sum_{x \in X(\Omega)} \mid f(x) \mid P(X = x) < +\infty$$

et dans ce cas $E(f(X)) = \sum_{x \in X(\Omega)} f(x)P(X = x)$ (on retrouve la situation de l'introduction).

▶ *Exemple 177.* Si X est une variable aléatoire à valeurs dans $\mathbb{N}$ de loi μ alors pour tout $s \in [-1, 1]$, s^X est une variable aléatoire intégrable et $E(s^X) = \sum_{k=0}^{\infty} s^k \mu(k)$.
En particulier, si X est une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$, alors pour tout $s \in [-1, 1]$, $E(s^X) = e^{\lambda(s-1)}$.

Cas d'une variable aléatoire à densité Si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^r$ dont la loi a pour densité ϕ et si g est une fonction borélienne de $\mathbb{R}^r$ dans $\mathbb{R}$, alors $Z = g(X)$ est intégrable si et seulement si

$$\int_{\mathbb{R}^r} |g(x_1, \ldots, x_r)| \phi(x_1, \ldots, x_r) dx_1 \ldots dx_r < +\infty$$

et dans ce cas

$$E(g(X)) = \int_{\mathbb{R}^r} g(x_1, \ldots, x_r) \phi(x_1, \ldots, x_r) dx_1 \ldots dx_r.$$

En particulier, si X est une variable aléatoire réelle intégrable dont la loi a pour densité ϕ, son espérance est la moyenne : $E(X) = \displaystyle\int_{\mathbb{R}} x\phi(x)dx$.

▶ *Exemple 178.* Si X est une variable aléatoire de loi $\mathcal{N}(0, 1)$ alors X est intégrable puisque l'intégrale $\int_{\mathbb{R}} |x| \exp(-\frac{x^2}{2})dx$ est finie et $E(X) = 0$ car la densité de X est paire. Son moment d'ordre 2 est $E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 \exp(-\frac{x^2}{2})dx$. En faisant une intégration par parties, on obtient que $E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-\frac{x^2}{2})dx = 1$.

▶ *Exemple 179.* (*Loi de Cauchy*) Si X est une variable aléatoire de loi uniforme sur $(-\frac{\pi}{2}, \frac{\pi}{2})$, alors $Y = \tan(X)$ est une variable aléatoire de densité $x \mapsto \frac{1}{\pi} \frac{1}{1+x^2}$ et donc n'est pas intégrable.

Cas d'une loi mélange Soit X est une variable aléatoire dont la loi est une loi mélange $P_X = \sum_{i=1}^{+\infty} \alpha_i \mu_i$ avec μ_i des mesures de probabilité définie sur $\mathbb{R}$ et α_i des réels positifs de somme 1. Désignons par Y_i une variable aléatoire de loi μ_i pour tout $i \in \mathbb{N}^*$.
Si pour tout $i \in \mathbb{N}^*$, Y_i est intégrable alors X est intégrable et $E(X) = \sum_{i=1}^{+\infty} \alpha_i E(Y_i)$ (c'est une généralisation directe du cas des lois mélanges sur $\mathbb{R}_+$).

N.B. Si X est un vecteur aléatoire et g une fonction borélienne telle que $E(|g(X)|)$ est finie, alors $E(g(X)) = \int_{\mathbb{R}^d} g(x)dP_X(x)$.

▷ *Exercice 180.* Soit X une variable aléatoire de loi $\mathcal{N}(0, \sigma^2)$, déterminer pour tout $p \in \mathbb{N}^*$ le moment d'ordre p de X c'est-à-dire $E(X^p)$.

8.1.1 Construction de la mesure produit et théorème de Fubini

Théorème 19 *(i) Il existe une unique probabilité $P_1 \otimes P_2$ sur $\mathcal{A}_1 \otimes \mathcal{A}_2$ telle que $P_1 \otimes P_2(A_1 \times A_2) = P_1(A_1)P_2(A_2)$ pour tout $A_1 \in \Omega_1$ et $A_2 \in \Omega_2$.*

(ii) Pour tout $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, les applications $\omega_2 \mapsto P_1(A^1_{\omega_2})$ définie sur Ω_2 et $\omega_1 \mapsto P_2(A^2_{\omega_1})$ définie sur Ω_1 sont des variables aléatoires et on a

$$P_1 \otimes P_2(A) = \int_{\Omega_2} P_1(A^1_{\omega_2})dP_2(\omega_2) = \int_{\Omega_1} P_2(A^2_{\omega_1})dP_1(\omega_1).$$

(rappelons que $A^1_{\omega_2} = \{\omega_1,\ (\omega_1,\omega_2) \in A\}$ et $A^2_{\omega_1} = \{\omega_2,\ (\omega_1,\omega_2) \in A\}$).

Preuve : L'unicité a été vu page 69. Le lemme de section 15 assure que $A^1_{\omega_2} \in \mathcal{A}_1$ et $A^2_{\omega_1} \in \mathcal{A}_2$. Soit $\mathcal{L}$ la classe des événements A pour lesquels l'application $\omega_2 \mapsto P_1(A^1_{\omega_2})$ définie sur Ω_2 est une variable aléatoire . Soit $\mathcal{C}$ la classe des pavés mesurables. Le théorème 14 assure que $\mathcal{L}$ contient $\mathcal{A}_1 \otimes \mathcal{A}_2$. L'intégrale $\int_{\Omega_2} P_1(A^1_{\omega_2})dP_2(\omega_2)$ est donc bien définie et il résulte de la σ-additivité de l'espérance que l'application $A \mapsto \int_{\Omega_2} P_1(A^1_{\omega_2})dP_2(\omega_2)$ définie définie bien une probabilité sur $\mathcal{A}_1 \otimes \mathcal{A}_2$. On peut faire de même en échangeant les rôles de ω_1 et ω_2 et l'unicité assure que les deux expressions coïncident. □

Corollaire 20 *Soit Z une variable aléatoire positive sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.*
Pour tout $\omega_2 \in \Omega_2$, $Z(\cdot,\omega_2)$ est une variable aléatoire définie sur $(\Omega_1,\mathcal{A}_1)$ et $\omega_2 \mapsto \int_{\Omega_1} Z(u,\omega_2)dP_1(u)$ est une variable aléatoire définie sur $(\Omega_2,\mathcal{A}_2)$. De même, pour tout $\omega_1 \in \Omega_1$, $Z(\omega_1,\ \cdot)$ est une variable aléatoire définie sur $(\Omega_2,\mathcal{A}_2)$ et $\omega_1 \mapsto \int_{\Omega_2} Z(\omega_1,u)dP_2(u)$ est une variable aléatoire définie sur $(\Omega_1,\mathcal{A}_1)$.
De plus on a,

$$\int_{\Omega_1\times\Omega_2} Zd(P_1 \otimes P_2) = \int_{\Omega_1} (\int_{\Omega_2} Z(\omega_1,\omega_2)dP_2(\omega_2))dP_1(\omega_1) = \int_{\Omega_2} (\int_{\Omega_1} Z(\omega_1,\omega_2)dP_1(\omega_1))dP_2(\omega_2).$$

Preuve : Le théorème affirme que le résultat est vrai lorque Z est l'indicatrice d'un événement de $\mathcal{A}_1 \otimes \mathcal{A}_2$. Cela permet d'établir le résultat lorsque Z est une variable aléatoire simple. Pour établir le résultat dans le cas général, il suffit alors d'utiliser que Z est la limite d'une suite croissantes de variables aléatoires simples et la propriété de convergence croissante de l'espérance. □

N.B. Le théorème et son corollaire se généralisent au cas de n espaces

N.B. Le théorème de Fubini sur $\mathbb{R}^2$ est un cas particulier de ce corollaire : si $g(x,y)$ est à support compact et si ϕ est une densité de probabilité strictement positive sur $\mathbb{R}$, le théorème précédent appliqué à $dP_1(x) = dP_2(x) = \phi(x)dx$ et à $\frac{g(x,y)}{\phi(x)\phi(y)}$ une démonstration du théorème de Fubini pour l'intégrale de Lebesgue.

8.2 Variance

8.2.1 Variables aléatoires de carré intégrable

On note $\mathcal{L}^2(\Omega,P)$, l'espace des variables aléatoires réelles de carré intégrable c'est-à-dire telles que $\mathrm{E}(X^2)$ soit fini. Les inégalités :

$$(X+Y)^2 \leq 2(X^2+Y^2) \text{ et } 2|XY| \leq X^2+Y^2$$

et la linéarité de l'espérance montrent que $\mathcal{L}^2(\Omega, P)$ est un espace vectoriel et que le produit de deux variables aléatoires de carré intégrable et une variable aléatoire intégrable. Les constantes étant de carré intégrable, ce dernier point montre que les variables aléatoires de carré intégrable sont intégrables : $\mathcal{L}^2(\Omega, P) \subset \mathcal{L}^1(\Omega, P)$.

N.B. $(X, Y) \mapsto \mathrm{E}(XY)$ définie sur $\mathcal{L}^2(\Omega, P)$ est une forme bilinéaire symétrique positive (elle n'est pas en général, définie positive car $E(X^2) = 0$ implique seulement que $P(X = 0) = 1$. D'après l'inégalité de Cauchy-Schwarz, pour toute variable aléatoire $X, Y \in \mathcal{L}^2(\Omega, P)$, $E(XY)^2 \leq E(X^2)E(Y^2)$.

8.2.2 Variance

On définit la *variance* d'une v.a. réelle $X \in \mathcal{L}^2(\Omega, P)$ par la formule :

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2)$$

ou de façon équivalente par la linéarité de l'espérance,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$$

La variance mesure l'amplitude moyenne des fluctuations de la variable autour de sa moyenne. Elle ne dépend que de la loi de X.

Remarquons que l'espérance de X est la constante qui approche le mieux la variable aléatoire X au sens suivant :

Propriété. *La fonction $a \mapsto \mathrm{E}((X - a)^2)$ atteint son minimum en $\mathrm{E}(X)$. De plus, pour tout $a \in \mathbb{R}$, $\mathrm{E}((X - a)^2) = \mathrm{Var}(X) + (\mathrm{E}(X) - a)^2$.*

Preuve : En introduisant $E(X)$ et en utilisant la linéarité de l'espérance, on obtient :

$$\begin{aligned}
\mathrm{E}((X - a)^2) &= \mathrm{E}(((X - \mathrm{E}(X)) + \mathrm{E}(X) - a)^2) \\
&= \mathrm{E}((X - \mathrm{E}(X))^2) + (\mathrm{E}(X) - a)^2 + 2(\mathrm{E}(X) - a)\mathrm{E}(X - \mathrm{E}(X)) \\
&= \mathrm{Var}(X) + (\mathrm{E}(X) - a)^2.
\end{aligned}$$

□

▷ *Exercice 181.* Vérifier à partir de la linéarité de l'espérance, l'égalité des deux expressions de la variance.

On voit sur la première expression de la variance que :

Propriétés.

- *La variance de X est positive ou nulle,*
- *La variance de X est nulle si et seulement si X est constante presque sûrement.*

- *Si* $a \in \mathbb{R}$, *alors* $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.

▶ *Exemple 182.*

1. Si X est de loi uniforme sur l'ensemble $\{x_1, \dots, x_n\}$ alors $\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2$.
2. Si X est la fonction indicatrice d'un événement A, alors $\mathrm{Var}(X) = P(A)(1 - P(A))$.
3. Si X suit la loi de Poisson $\mathcal{P}(\lambda)$ alors X est de carré intégrable et $\mathrm{Var}(X) = \lambda$.
4. La variance de la loi uniforme sur l'intervalle $[a, b]$ est $\frac{(b-a)^2}{12}$.
5. La variance de la loi exponentielle de paramètre λ est $\frac{1}{\lambda^2}$.
6. La loi gaussienne $\mathcal{N}(m, \sigma^2)$ a pour variance σ^2.
   Il suffit pour montrer cette affirmation, de traiter le cas où $m = 0$ et $\sigma^2 = 1$, puisque si Y suit la loi $\mathcal{N}(0, 1)$ alors $\sigma Y + m$ suit la loi $\mathcal{N}(0, 1)$. Les propriétés de la variance impliquent alors que $\mathrm{Var}(\sigma Y + m) = \sigma^2\mathrm{Var}(Y)$.
   Par intégration par parties, on obtient que le moment d'ordre 2 de la loi $\mathcal{N}(0, 1)$ vaut 1.

Définition. L'*écart-type* de la variable aléatoire X, notée $\sigma(X)$, est par définition la racine carrée de la variance.

L'écart-type s'exprime dans les mêmes unités que la variable aléatoire considérée.

Définition. Une variable aléatoire d'espérance nulle est dite *centrée*. Une variable aléatoire de variance 1 est dite *réduite*.

▷ *Exercice 183.* Soit Z une variable aléatoire de variance non nulle, vérifier que $\widetilde{Z} = \frac{Z-\mathrm{E}(Z)}{\sigma(Z)}$ est une variable aléatoire centrée réduite.

▷ *Exercice 184.* Soit X une variable aléatoire réelle définie sur (Ω, P). On dit que m est une médiane pour X si $P(X < m) \leq \frac{1}{2} \leq P(X \leq m)$.

1. Soit $x_1, \dots, x_n$ n réels distincts. Déterminer les médianes de la loi uniforme sur l'ensemble $\{x_1, \dots, x_n\}$.

Supposons que $X \in \mathcal{L}^1(\Omega, P)$.

1. Montrer que si m est une médiane pour X alors pour tout $a \in \mathbb{R}$, $\mathrm{E}(|X - a|) \geq \mathrm{E}(|X - m|)$.
2. Montrer que si m_1 et m_2 sont deux médianes pour X alors $\mathrm{E}(|X - m_1|) = \mathrm{E}(|X - m_2|)$.
3. Supposons que $X \in \mathcal{L}^2(\Omega, P)$. Montrer que $\mathrm{E}(|X - m|) \leq \sigma(X)$.

8.3 Covariance

La variance est une forme quadratique sur l'espace vectoriel des variables aléatoires réelles de carré intégrable. La *covariance* de deux variables aléatoires X et Y de carré intégrable, notée $\mathrm{Cov}(X, Y)$ est la forme bilinéaire symétrique associée, c'est-à-dire telle que $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$. Elle est donnée par les formules :

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

et on a

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) \pm 2\mathrm{Cov}(X, Y) + \mathrm{Var}(Y) \tag{8.1}$$

▷ *Exercice 185.* Vérifier à partir de la linéarité de l'espérance, l'égalité des deux expressions de la covariance, ainsi que la formule (8.1).

▷ *Exercice 186.* Soit A et B deux événements de probabilité non nulle. On note $X = \mathbb{1}_A$ et $Y = \mathbb{1}_B$. Montrer que $\mathrm{Cov}(X, Y) > 0$ si et seulement si $P(A|B) > P(A)$.

▷ *Exercice 187.* Soit $X_1, \dots, X_n$ des variables aléatoires de carré intégrable. On pose $S_n = X_1 + \cdots + X_n$. Montrer que

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{Cov}(X_i, X_j)$$

N.B. On définit la *matrice de covariance* d'un vecteur aléatoire de carré intégrable $\overrightarrow{Z} = (Z_1, ..., Z_d)$, comme la matrice $d \times d$ dont le coefficient (i, j) est $\mathrm{Cov}(Z_i, Z_j)$.

▷ *Exercice 188.* Dans l'exercice 109, déterminer la variance de la variable aléatoire donnant le nombre de cases vides du tableau après la répartition des jetons.

▷ *Exercice 189.* Montrer que la matrice de covariance d'un vecteur aléatoire $\overrightarrow{Z} = (Z_1, ..., Z_d)$ est une matrice symétrique positive.
Soit A une matrice $d \times d$. Déterminer la matrice de covariance du vecteur aléatoire $\overrightarrow{Y} = A\overrightarrow{Z}$ en fonction de celle de $\overrightarrow{Z}$.

Proposition 21 *La covariance de deux variables aléatoires indépendantes de carré intégrable est nulle et la variance de leur somme est égale à la somme de leurs variances.*

Preuve : Considérons deux variables aléatoires indépendantes X et Y de carré intégrable. La première assertion signifie que $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$. Le membre de droite s'écrit $\mathrm{E}(XY) = \int_{\mathbb{R}^2} xy dP_{(X,Y)}(x, y)$. Comme les variables aléatoires X et Y sont indépendantes, la loi de (X, Y) est la loi produit $P_X \otimes P_Y$. On conclut en utilisant le théorème de Fubini,

$$\mathrm{E}(XY) = \int_{\mathbb{R}} x dP_X(x) \int_{\mathbb{R}} y dP_Y(y) = E(X)E(Y).$$

La deuxième assertion procède directement de la première par l'identité (8.1). □

N.B. Le fait que deux variables aient une covariance nulle n'entraîne pas qu'elles sont indépendantes comme le montre l'exemple suivant :

▶ *Exemple 190.* Soit X et Y deux variables aléatoires indépendantes de loi de Bernoulli de paramètre $1/2$. Les variables aléatoires $X+Y$ et $|X-Y|$ ont une covariance nulle et elles ne sont pas indépendantes.

▶ *Exemple 191.* Si (X, Y) est un vecteur de loi uniforme sur sur le disque unité de $\mathbb{R}^2$, $D = \{(x, y) \in \mathbb{R}^2,\ x^2 + y^2 < 1\}$ est $\frac{1}{\pi}\mathbb{1}_D$. Alors, $\mathrm{Cov}(X, Y) = 0$ et X et Y ne sont pas indépendantes.

Corollaire 22 *Si $(X^{(1)}, ..., X^{(n)})$ est un échantillon de la variable aléatoire X de carré intégrable et si l'on pose $S_n = X^{(1)} + ... + X^{(n)}$, alors $\mathrm{E}(S_n) = n\mathrm{E}(X)$ et $\mathrm{Var}(S_n) = n\mathrm{Var}(X)$.*

Preuve : Il suffit d'observer que S_n est la somme de n variables aléatoires indépendantes et d'appliquer la linéarité de l'espérance ainsi que la proposition précédente. □

▶ *Exemple 192.* Soit (Λ, Q) un espace de probabilité. Soit $C_1, \dots, C_d$ une partition de Λ en événements de probabilité $p_i = Q(C_i)$ pour $i \in \{1, \dots, d\}$. La matrice de covariance du vecteur $X = (\mathbb{1}_{C_1}, \dots, \mathbb{1}_{C_d})$ est la matrice $K = (K_{i,j})_{(i,j)\in\{1,\dots,d\}^2}$ définie par $K_{i,j} = \begin{cases} p_i(1-p_i) & \text{si } i = j \\ -p_i p_j & \text{si } i \neq j \end{cases}$
D'après la proposition 21 et la section 3.6.3, il en résulte que la matrice de covariance d'un vecteur aléatoire de loi multinomiale $\mathcal{M}(n, (p_1, \dots, p_d))$ est nK.
En particulier, l'espérance et la variance d'une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ sont respectivement np et $np(1-p)$.

L'inégalité de Schwarz donne une majoration de la covariance de deux variables aléatoires :

Propriétés. *Soit X et Y deux v.a. réelles de carré intégrable définies sur (Ω, P).*
On a $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$ et l'égalité est réalisée si et seulement si il existe $\lambda \in \mathbb{R}$ et $c \in \mathbb{R}$ tels que $Y = \lambda X + c$.

Si X et Y sont des variables aléatoires de variances non nulles, on définit le *coefficient de corrélation de X et de Y* comme étant le rapport

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

C'est un réel dont la valeur absolue est majorée par 1 avec égalité si et seulement si $\exists \lambda \in \mathbb{R}^*$ et $c \in \mathbb{R}$ tels que $Y = \lambda X + c$ (dans ce cas, le signe de $\mathrm{Corr}(X,Y)$ est donné par celui de λ).

► *Exemple 193.* Soit X et U deux variables aléatoires indépendantes centrées d'écart-type σ non nul. Notons Y la variable aléatoire définie par $Y = aX + \sqrt{1-a^2}U$. Alors, le vecteur $Z = (X,Y)$ est centré, de matrice de covariance $\Gamma = \begin{pmatrix} \sigma^2 & a\sigma^2 \\ a\sigma^2 & \sigma^2 \end{pmatrix}$. En particulier, la corrélation entre X et Y est donnée par a.
Sur la figure 8.1 sont représentées, pour différentes valeurs de a, 100 réalisations du vecteur aléatoire Z lorsque X et U suivent la loi uniforme sur $\{-10,-9,\ldots,10\}$.

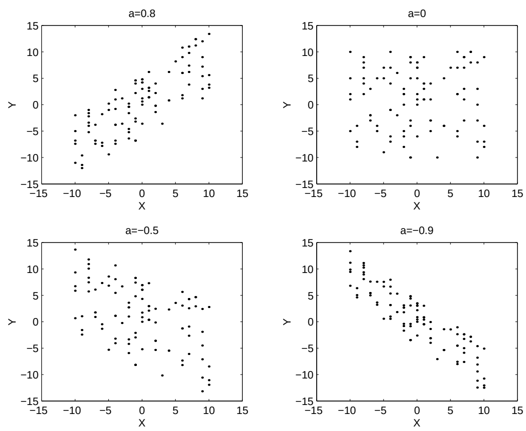


FIG. 8.1 : Nuages de points constitués chacun de 100 réalisations du couple (X,Y) lorsque $Y = aX + \sqrt{1-a^2}U$, X et U étant des variables aléatoires indépendantes de loi uniforme sur $\{-10,-9,\ldots,9,10\}$.

▷ *Exercice 194.* Pour une élection, une population de n individus a eu à choisir entre voter pour le candidat A ou le candidat B. On note m le nombre de personnes qui a voté pour A. On interroge au hasard k individus différents dans cette population ($1 \leq k \leq N$).

1. On désigne par $a_1,\ldots,a_m$ les m personnes qui ont voté pour A. Pour $i \in \{1,\ldots,m\}$, on note X_i l'indicatrice de l'événement "la personne a_i est interrogée". Quelle est la loi de X_i ?
2. Déterminer l'espérance ainsi que la matrice de covariance de $(X_1,\ldots,X_m)$.
3. On pose $S = \sum_{i=1}^m X_i$. Quelle est la loi de S ?
4. Déterminer le nombre moyen de personnes votant pour A sur k personnes tirées au hasard.
5. Déterminer la variance de S.

8.4 Régression linéaire

Etant données deux variables aléatoires X (quelconque) et Y (réelle) de carré intégrable définies sur (Ω, P), on peut chercher (par exemple lorsque X peut être observée plus facilement que Y), à approcher Y par une fonction de X (que nous noterons ϕ) le mieux possible "aux sens des moindres carrés", c'est-à-dire de manière à ce que l'erreur quadratique $\mathrm{E}((Y - \phi(X))^2)$ soit minimale.

Si (X, Y) est un couple de variables aléatoires réelles de carré intégrable, dont on suppose seulement connues les espérances et la matrice des covariances, la proposition suivante montre qu'on peut trouver la meilleure approximation (au sens au sens des moindres carrés) de Y par une fonction affine de X.

Proposition 23 *Supposons la v.a. X non constante. La meilleur approximation (au sens des moindres carrés) de Y par une fonction affine de X est donnée par la variable aléatoire*

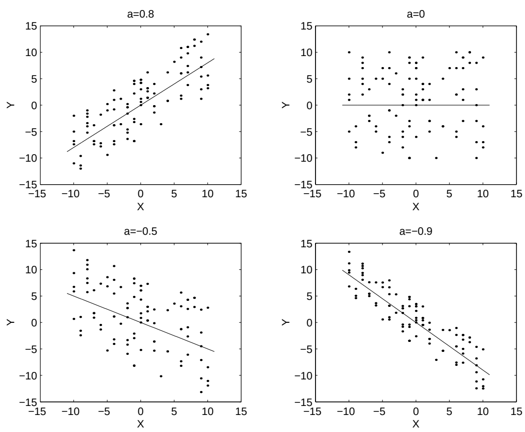
$$\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mathrm{E}(X)) + \mathrm{E}(Y),$$

l'erreur quadratique étant égale à $\mathrm{Var}(Y)(1 - \mathrm{Corr}(X,Y)^2)$.

Autrement dit, l'application $\psi\ : (a,b) \mapsto E((Y - b - aX)^2)$ définie sur $\mathbb{R}^2$ admet un unique minimum, le point de coordonnées $\hat{a} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$ et $\hat{b} = \mathrm{E}(Y) - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\mathrm{E}(X)$. La valeur minimale de ψ est $\psi(\hat{a}, \hat{b}) = \mathrm{Var}(Y)(1 - \mathrm{Corr}(X,Y)^2)$.

Preuve : Notons V_X le sous-espace vectoriel de $\mathcal{L}^2(\Omega, P)$ engendré par 1 et X. La meilleur approximation de Y par une fonction affine de X est donnée par la projection orthogonale $\pi(Y)$ de Y sur V_X. V_X est un espace de dimension 2 qui admet pour base orthonormale $(1, \frac{X-\mathrm{E}(X)}{\sigma(X)})$. Donc, $\pi(Y) = \langle Y, 1\rangle 1 + \langle Y, X - \mathrm{E}(X)\rangle \frac{(X-\mathrm{E}(X))}{\mathrm{Var}(X)} = \mathrm{E}(Y) + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mathrm{E}(X))$. □

► *Exemple 195.* Dans l'exemple 193, aX est la meilleur façon d'approcher Y par une fonction affine de X et l'erreur quadratique est $\mathrm{E}((Y - aX)^2) = (1 - a^2)\sigma^2$.
A la figure 8.2, on a repris les nuages de points représentant 100 réalisations du couple (X, Y) présentés figure 8.1 et on a superposé la droite $y = ax$, appelée droite de regression de Y par rapport à X.



Nuages de points constitués chacun de 100 réalisations du couple (X, Y) lorsque $Y = aX + \sqrt{1 - a^2}U$, X et U sont des variables aléatoires indépendantes de loi uniforme sur $\{-10, -9, \ldots, 9, 10\}$.

FIG. 8.2 :

▷ *Exercice 196.* Soit $\mathcal{E}_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ un ensemble de n points du plan. Notons (X, Y) un couple de variables aléatoires dont la loi est la loi uniforme μ_n sur l'ensemble $\mathcal{E}_n$ c'est-à-dire $\mu_n(\{(x_i, y_i)\}) = \frac{1}{n}$ pour tout $i \in \{1, \ldots, n\}$.

1. Donner l'expression de la droite de régression linéaire de Y en X c'est-à-dire la droite d'équation $y = a_n x + b_n$ où a_n et b_n sont choisis de sorte que $a_n X + b_n$ soit la meilleure façon d'approcher Y par une fonction affine de X au sens des moindres carrés. On donnera une expression de a_n et b_n en fonction des éléments de $\mathcal{E}_n$.
2. Pour $(\alpha, \beta) \in \mathbb{R}^2$, on note $S(\alpha, \beta) = \sum_{i=1}^n (y_i - \alpha x_i - \beta)^2$.
   Donner une interprétation géométrique de $S(\alpha, \beta)$ par rapport à l'ensemble $\mathcal{E}_n$ (on pourra faire un schéma). Puis, déterminer un couple de réels $(\hat{\alpha}, \hat{\beta})$ qui minimise la fonction S.

▷ *Exercice 197.* Généraliser la proposition 23 au cas où X est un vecteur aléatoire dont la matrice de covariance K_X est inversible : montrer que la meilleur façon d'approcher Y par une fonction affine des composantes du vecteur X est de prendre la variable aléatoire

$$\sum (K_X^{-1})_{ij} \mathrm{Cov}(X_j, Y)(X_i - \mathrm{E}(X_i)) + \mathrm{E}(Y).$$

Chapitre 9

Loi faible des grands nombres et applications

9.1 Inégalités de Markov et de Bienaymé-Tchebychev

On a vu que la variance mesure l'amplitude moyenne des fluctuations de la variable autour de sa moyenne. L'inégalité de Bienaymé-Tchebychev va permettre de préciser cette propriété.

Proposition 24 *Si X est une variable aléatoire positive, alors pour tout $\varepsilon > 0$,*

$$P(X \geq \varepsilon) \leq \frac{\mathrm{E}(X)}{\varepsilon} \qquad \textit{(inégalité de Markov).}$$

En particulier, si Y est une variable aléatoire de carré intégrable, alors pour tout $\varepsilon > 0$,

$$P(|Y - \mathrm{E}(Y)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2} \qquad \textit{(inégalité de Bienaymé-Tchebychev).}$$

Preuve : De façon évidente, pour toute variable aléatoire X positive, $\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}} \leq X$. En prenant l'espérance de chaque membre de l'inégalité, on obtient $\varepsilon P(X \geq \varepsilon) \leq \mathrm{E}(X)$.
En appliquant cette inégalité à la variable aléatoire $X = (Y - \mathrm{E}(Y))^2$, on trouve l'inégalité de Bienaymé-Tchebychev. □

En utilisant l'inégalité de Bienaymé-Tchebychev, on peut déduire que si X est une variable aléatoire de carré intégrable et non constante alors plus de 88% des valeurs observées de X se trouvent dans l'intervalle $[\mathrm{E}(X) - 3\sigma(X), \mathrm{E}(X) + 3\sigma(X)]$. Plus généralement, pour tout $a > 0$,

$$P(|X - \mathrm{E}(X)| \geq a\sigma(X)) \leq \frac{1}{a^2}.$$

▷ *Exercice 198.* On suppose que le nombre de pièces sortant d'une usine donnée en une journée est une variable aléatoire d'espérance 50.

1. Peut-on estimer la probabilité que la production de demain dépasse 75 pièces ?
2. Que peut-on dire de plus sur cette probabilité si on sait que l'écart-type de la production quotidienne est de 5 pièces ?

▷ *Exercice 199.* Pour étudier les particules émises par une substance radioactive, on dispose d'un détecteur. On note X la variable aléatoire représentant le nombre de particules qui atteignent le détecteur pendant un intervalle de temps Δt. Le nombre maximal de particules que le détecteur peut compter pendant un intervalle de temps Δt est de 10^3.

1. On suppose que X a pour espérance 10^2. Donner une majoration de la probabilité que X dépasse 10^3.
2. On suppose de plus que X a une variance égale à 10^2. Donner une majoration plus fine de la probabilité que X dépasse 10^3.
3. On suppose maintenant que X suit une loi de Poisson de paramètre m. Soit $a > 1$.
   (a) Montrer que pour tout $\theta > 0$, $P(X > am) \leq E(e^{\theta(X-am)})$.
   (b) En déduire que $P(X > am) \leq e^{-m(1-a+a\ln(a))}$.
   (c) Donner un majorant de la probabilité que X dépasse 10^3 si X suit une loi de Poisson d'espérance 10^2.

9.2 Loi faible des grands nombres

La loi faible des grands nombres démontrée dans la section 3.2 peut maintenant être étendue aux variables aléatoires.

Proposition 25 *Soit* $(X^{(1)}, ..., X^{(n)})$ *un échantillon de la v.a. réelle* X *de carré intégrable. Posons* $S_n = X^{(1)} + ... + X^{(n)}$*. Alors*

$$\text{pour tout } \varepsilon > 0, \quad P(|\frac{S_n}{n} - \mathrm{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{n\varepsilon^2}.$$

En particulier, cette expression tend vers 0 quand n *tend vers l'infini.*

N.B.

1. Comme le choix de ε est arbitraire, ceci montre que la *moyenne empirique* $\frac{S_n}{n}$ a de fortes chances d'être très proche de $\mathrm{E}(X)$ lorsque n est très grand.

2. On peut montrer que le résultat de convergence reste vrai si X est simplement intégrable, mais on ne dispose plus de la majoration.

Preuve de la proposition : Ce résultat fondamental procède immédiatement du corollaire 22 et de l'inégalité de Bienaymé-Tchebychev (proposition 24) appliquée à la v.a. $\frac{S_n}{n}$: pour tout $\varepsilon > 0$, $P(|\frac{S_n}{n} - \mathrm{E}(\frac{S_n}{n})| \geq \varepsilon) \leq \frac{\mathrm{Var}(S_n)}{n^2\varepsilon^2}$ avec $\mathrm{E}(\frac{S_n}{n}) = \mathrm{E}(X)$ et $\mathrm{Var}(S_n) = n\mathrm{Var}(X)$. □

N.B. Si X est une variable aléatoire bornée, on dispose d'une inégalité plus précise que l'inégalité de Bienaymé-Tchébychev lorsque n est assez grand. Elle est appelée l'inégalité de Hoeffding : *soit* X *une variable aléatoire bornée telle que* $a \leq X \leq b$ *et soit* $(X_1, \ldots, X_n)$ *un* n*-échantillon de variables aléatoires de même loi que* X*. Alors, pour tout* $\epsilon > 0$,

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathrm{E}(X)| \geq \epsilon) \leq 2\exp(-2\frac{n\epsilon^2}{(b-a)^2}).$$

Une preuve et une application de cette inégalité à la construction d'intervalles de confiance sont proposées à l'exercice 200. Prenons un exemple : si X est une variable aléatoire de loi de Bernoulli $\mathcal{B}(0.5)$ et $n = 1000$, $P(|\frac{1}{n}\sum_{i=1}^{n} X_i - 0,5| \geq 0,05) \simeq 1,7.10^{-3}$. La majoration par l'inégalité de Bienaymé-Chebychev est $\frac{1}{4n\varepsilon^2} = 0.1$ alors que la majoration par l'inégalité de Hoeffding est $2\exp(-2n\varepsilon^2) \simeq 1,35.10^{-2}$.

▷ *Exercice 200.* Soit X une variable aléatoire bornée telle que $a \leq X \leq b$ et soit $(X_1, \ldots, X_n)$ un n-échantillon de variables aléatoires de même loi que X. L'objectif de l'exercice est d'établir puis d'utiliser l'inégalité de Hoeffding suivante :

$$\text{pour tout } \varepsilon > 0,\ P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathrm{E}(X)| \geq \varepsilon) \leq 2\exp(-2\frac{n\varepsilon^2}{(b-a)^2}). \tag{9.1}$$

1. Soit Y une variable aléatoire telle que $0 \leq Y \leq 1$. Notons μ son espérance. Soit $(Y_1, \ldots, Y_n)$ un n-échantillon de variables aléatoires de même loi que Y et $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
   (a) Montrer que pour tout $\varepsilon > 0$ et $\theta > 0$,
   $$P(\bar{Y}_n - \mu \geq \varepsilon) = P(e^{\theta \sum_{i=1}^{n} Y_i} \geq e^{n\theta(\mu+\varepsilon)}).$$
   (b) En déduire que pour tout $\varepsilon > 0$ et $\theta > 0$, $P(\bar{Y}_n - \mu \geq \varepsilon) \leq e^{n(\lambda(\theta)-\theta(\mu+\varepsilon))}$ avec $\lambda(\theta) = \ln(E(e^{\theta Y}))$.
   Indication : utiliser l'inégalité de Markov.
   (c) Montrer que la fonction $g : \theta \mapsto E(\exp(\theta Y))$ est une fonction C^2 sur $]0, +\infty[$ et que $g'(\theta) = E(Y\exp(\theta Y))$ et $g''(\theta) = E(Y^2\exp(\theta Y)) \leq E(Y\exp(\theta Y))$.
   (d) En calculant les deux premières dérivées de la fonction λ, montrer que pour tout $\theta > 0$, $\lambda(\theta) \leq \theta\mu + \frac{\theta^2}{8}$.
   Indications : utiliser que pour tout $x \in \mathbb{R}$, $x - x^2 \leq \frac{1}{4}$ afin de majorer la dérivée seconde de λ, puis appliquer le développement de Taylor-Maclaurin à la fonction λ.
   (e) En déduire que pour tout $\varepsilon > 0$, $P(\bar{Y}_n - \mu \geq \varepsilon) \leq \exp(-2n\varepsilon^2)$.
   Indication : déterminer le minimum de la fonction $\theta \mapsto \theta\mu + \frac{1}{8}\theta^2 - \theta(\mu+\varepsilon)$ sur $]0, +\infty[$.
   (f) Montrer qu'on a aussi pour tout $\varepsilon > 0$, $P(\bar{Y}_n - \mu \leq -\varepsilon) \leq \exp(-2n\varepsilon^2)$ et donc que pour tout $\varepsilon > 0$,
   $$P(|\bar{Y}_n - \mu| \geq \varepsilon) \leq 2\exp(-2n\varepsilon^2) \tag{9.2}$$
   Indication : appliquer l'inégalité trouvée à la question (e) avec les variables aléatoires $(1-Y_i)$.
2. En déduire l'inégalité de Hoeffding (9.1).
   Indication : appliquer l'inégalité (9.2) avec les variables aléatoires $Y_i = \frac{X_i - a}{b-a}$.
3. *Application :* afin d'estimer la probabilité p qu'un événement A soit réalisé au cours d'une expérience $\mathcal{E}$, on répète cette expérience n fois dans les mêmes conditions. On note S_n le nombre de fois où l'événement A a été réalisé au cours de ces n expériences.
   (a) Soit $\alpha \in]0, 1[$. On pose $\varepsilon_n(\alpha) = \sqrt{\frac{1}{2n}\ln(\frac{2}{\alpha})}$. On note $s_{n,obs}$ la valeur observée de S_n. Montrer que $]\frac{s_{n,obs}}{n} - \varepsilon_n(\alpha), \frac{s_{n,obs}}{n} + \varepsilon_n(\alpha)[$ est un intervalle de confiance pour le paramètre p de niveau de confiance supérieur ou égal à $(1-\alpha)100\%$.
   (b) Si sur 1000 expériences, l'événement A a été réalisé 510 fois, que peut-on dire de p ?
   (c) Comment doit-on choisir n si on veut un intervalle de confiance pour p de longueur inférieure à $0,01$ et de niveau de confiance supérieur ou égal à 90% ?

9.3 Estimation statistique

Si $(X^{(1)}, ..., X^{(n)})$ est un échantillon d'une v.a. X intégrable, on dit que $\frac{S_n}{n}$, qui est souvent noté $\overline{X}_n$ est un *estimateur* de $\mathrm{E}(X)$. Cela signifie qu'une observation de cette v.a. peut être utilisée pour donner une valeur approchée de $\mathrm{E}(X)$. Un estimateur Z_n d'un paramètre θ est toujours défini à partir d'un échantillon de taille n d'une v.a. dont la loi détermine ce paramètre. Il doit être *consistant* ce qui signifie que la probabilité pour que la valeur estimée diffère de θ de plus d'un ε arbitraire fixé doit tendre vers 0 lorsque la taille de l'échantillon tend vers l'infini : Z_n est consistant si pour tout $\varepsilon > 0$, $P(|Z_n - \theta| > \varepsilon)$ tend vers 0 lorsque n tend vers $+\infty$[1].

[1]On dit aussi que Z_n converge en probabilité vers θ lorsque la taille n de l'échantillon tend vers l'infini.

Un estimateur de θ est dit *sans biais* si son espérance est égale à θ. Dans le cas contraire, il est dit *biaisé*, et la différence de son espérance avec θ est appelé le biais de l'estimateur. Enfin, on dit que l'estimateur est *asymptotiquement sans biais* si son biais tend vers 0 lorsque la taille de l'échantillon tend vers $+\infty$.

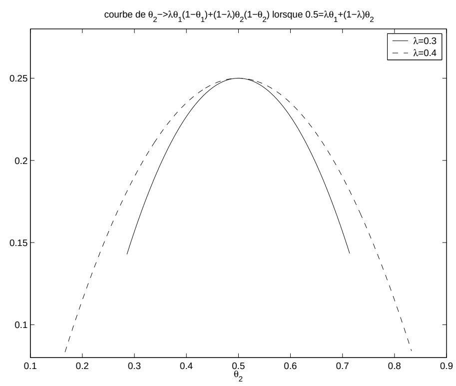
L'erreur quadratique est définie par $\mathrm{E}((Z_n - \theta)^2)$. Remarquons que si Z_n est un estimateur sans biais de θ, l'erreur quadratique est égale à la variance de Z_n.

► *Exemple 201.*

1. Dans l'exemple du sondage décrit dans la section 5.1, la variable aléatoire $\frac{S_n}{n}$ est un estimateur sans biais de la proportion p et l'erreur quadratique est égale à $\frac{p(1-p)}{n}$, elle est donc majorée par $\frac{1}{4n}$.
2. si l'expérience aléatoire consiste à tirer un entier ω au hasard entre 1 et N , et si on effectue n tirages indépendants, un estimateur consistant mais biaisé de N est donné par $M_n = \max(\omega_i,\ i \in \{1, \ldots, n\})$. En effet, pour tout $\epsilon \in]0, 1]$, comme M_n est à valeurs dans $\{1, \ldots, N\}$,

$$P(|M_n - N| \geq \epsilon) \leq P(M_n \leq N - 1) = (\frac{N-1}{N})^n,$$

   majoration qui tend vers 0 lorsque n tend vers $+\infty$, ce qui montre que M_n est un estimateur consistant. C'est un estimateur biaisé puisque $M_n \leq N$ et que $P(M_n < N) > 0$, ce qui entraîne que $\mathrm{E}(M_n) < N$.
3. *Sondages stratifiés :* si une population Ω est divisée en deux groupes dont les proportions λ et $1-\lambda$ sont connues, et si les proportions d'intention de vote pour le "oui" lors d'un référendum ne sont pas les mêmes dans les deux groupes (par exemple, de l'ordre de 20% dans l'un des groupes et de 70% dans l'autre groupe), on obtiendra une meilleure estimation de la proportion d'intention de vote pour le "oui" dans la population totale en sondant $n\lambda$ personnes dans le premier groupe et $n(1-\lambda)$ personnes dans le deuxième groupe qu'en sondant n individus tirés au hasard dans l'ensemble de la population. En effet, notons $S_n^{(1)}$ le nombre de personnes ayant l'intention de voter "oui" sur les λn personnes interrogées au hasard (avec remise) dans le premier groupe, $S_n^{(2)}$ le nombre de personnes ayant l'intention de voter "oui" sur les $(1-\lambda)n$ personnes interrogées au hasard (avec remise) dans le deuxième groupe et S_n le nombre de personnes ayant l'intention de voter "oui" sur n personnes interrogées au hasard dans l'ensemble de la population. Alors, $\mathrm{Var}(S_n^{(1)} + S_n^{(2)}) = n(\lambda\theta_1(1-\theta_1) + (1-\lambda)\theta_2(1-\theta_2))$ où θ_i est la proportion d'intentions de vote pour le "oui" dans le i-ème groupe. La proportion d'intentions de vote pour le "oui" dans la population totale est $\theta = \lambda\theta_1 + (1-\lambda)\theta_2$. On a $\mathrm{Var}(S_n) = n\theta(1-\theta) \geq \mathrm{Var}(S_n^{(1)} + S_n^{(2)})$ avec égalité si et seulement si $\theta_1 = \theta_2$ (car $x \mapsto x(1-x)$ est une fonction strictement concave). L'erreur quadratique pour le sondage avec stratification est d'autant plus faible que l'écart entre θ_1 et θ_2 est grand (voir figure ci-dessous).



▷ *Exercice 202.* Soit $(X^{(1)}, ..., X^{(n)})$ un échantillon d'une v.a. X de carré intégrable. Montrer que $\sigma_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X^{(i)} - \overline{X})^2$ est un estimateur consistant et sans biais de $\mathrm{Var}(X)$.

▷ *Exercice 203.* Montrer que si Z_n est un estimateur de θ asymptotiquement sans biais dont la variance tend vers 0 lorsque n tend vers $+\infty$ alors Z_n est un estimateur consistant de θ.

▷ *Exercice 204.* Soit Y_n et Z_n des estimateurs consistants de respectivement θ et λ. Montrer que $Y_n + Z_n$ et $Y_n Z_n$ sont des estimateurs consistants de respectivement $\theta + \lambda$ et $\theta\lambda$. Montrer que si $\theta \neq 0$, $\frac{1}{Y_n}$ est un estimateur consistant de $\frac{1}{\theta}$.

9.4 Mesure empirique

Pour toute valeur possible x d'une v.a. X, notons N_x le nombre d'indices i pour lesquels $X^{(i)} = x$. On sait, d'après la loi des grands nombres, appliquée à l'indicatrice $\mathbb{1}_{\{X=x\}}$, que lorsque n est assez grand, $\frac{N_x}{n}$ a une grande probabilité d'être très proche de $P(X = x)$. La loi de X est donnée par les nombres $P(X = x)$ et constitue une probabilité sur $X(\Omega)$. De la même façon, pour tout $\omega \in \Omega$, les nombres $\frac{N_x(\omega)}{n}$ constituent une probabilité sur $X(\Omega)$ (puisque $\sum_{x\in X(\Omega)} N_x(\omega) = n$), qu'on appelle la mesure empirique de l'échantillon $(X^{(1)}, ..., X^{(n)})$ en ω et qu'on notera $\mu_n(\omega)$. La mesure empirique est évidemment aléatoire. Des réalisations de la mesure empirique avec n grand peuvent être utilisées pour approcher la loi de X (voir figure 9.1).

Si $X \in \mathcal{L}^1(\Omega, P)$, la moyenne de μ_n[2], $\sum_{x\in X(\Omega)} x\frac{N_x}{n}$, coïncide avec $\overline{X}_n$ appelé *moyenne empirique* de l'échantillon (voir figure 9.2).

Pour une variable aléatoire réelle X, on note $F_n(\omega)$ pour tout $\omega \in \Omega$, la fonction de répartition de la probabilité $\mu_n(\omega)$; c'est la fonction définie sur $\mathbb{R}$ par

$$F_n(\omega)(t) = \mu_n(\omega)(]-\infty, t]) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X^{(i)}(\omega)\leq t} \text{ pour tout } t \in \mathbb{R}.$$

Elle est appelée la fonction de répartition empirique de l'échantillon $(X^{(1)}, \ldots, X^{(n)})$. La fonction de répartition empirique est en tout point t un estimateur sans biais et consistant de la fonction de répartition F de X en t. En fait, on peut montrer que pour n assez grand, $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$ a une très grande probabilité d'être proche de 0. Des réalisations de F_n pour n grands permettent donc d'approcher F (voir figure 9.3).

▷ *Exercice 205.* Montrer que si $X \in \mathcal{L}^2(\Omega, P)$, la variance de μ_n, $\sum_{x\in X(\Omega)}(x - \overline{X}_n)^2\frac{N_x}{n}$ coïncide avec $\frac{n-1}{n}\sigma_n^2$ qui est un estimateur consistant, mais biaisé, de $\mathrm{Var}(X)$.

▷ *Exercice 206.* Montrer que :

1. pour tout $\varepsilon > 0$ et pour tout $x \in X(\Omega)$, $P(\left|\frac{N_x}{n} - P(X = x)\right| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$.
2. si X est une v.a. réelle alors pour tout $\varepsilon > 0$ et pour tout $t \in \mathbb{R}$, $P(|F_n(t) - F(t)| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$.

▷ *Exercice 207.* Soit Y une variable aléatoire réelle définie sur un ensemble Ω muni d'une probabilité P et à valeurs dans un ensemble $\mathcal{E}$ au plus dénombrable. On note μ sa loi. Soit $(Y_1, \ldots, Y_n)$ un n-échantillon de variables aléatoires indépendantes de même loi que Y. Pour tout $\omega \in \Omega$, on note $\mu_n(\omega)$ la mesure empirique de l'échantillon $(Y_1, \ldots, Y_n)$ en ω.

1. On suppose que $\mathcal{E}$ est fini. Montrer que pour tout $\epsilon > 0$, $P(\max_{e\in\mathcal{E}} |\mu_n(\{e\}) - \mu(\{e\})| \geq \epsilon)$ tend vers 0 lorsque n tend vers $+\infty$.

[2] La moyenne d'une probabilité portée par un ensemble de nombres réels est la moyenne de toute v.a. dont elle est la loi (en particulier de la fonction identité $x \to x$ définie sur son support). Il en est de même pour la variance.

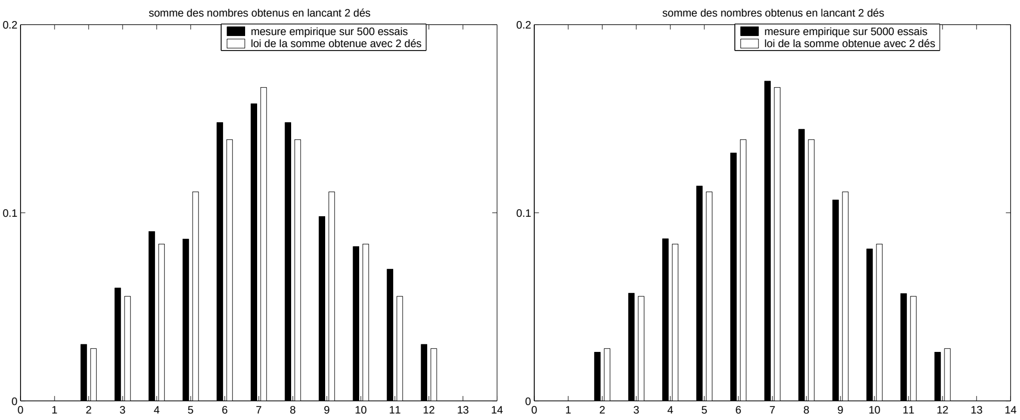


FIG. 9.1 : Comparaison entre la mesure empirique associée aux résultats de 500 puis 5000 lancers de deux dés avec la loi de la somme des nombres obtenus avec deux dés

FIG. 9.2 : Moyenne empirique de la somme des nombres obtenus avec deux dés sur les n premiers lancers des deux dés en fonction de n.

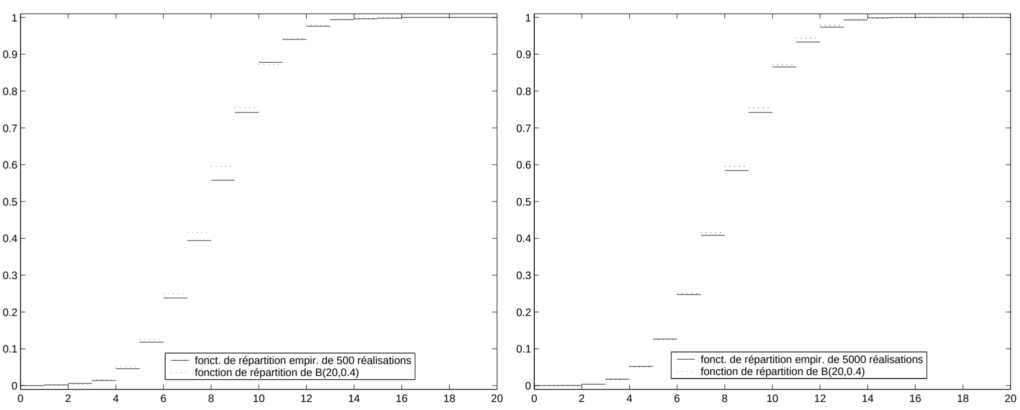


FIG. 9.3 : Comparaison entre la fonction de répartition empirique de 500, puis 5000 réalisations d'une variable aléatoire de loi binomiale $\mathcal{B}(20, 0.4)$ et la fonction de répartition de cette loi.

2. On suppose que $\mathcal{E}$ est dénombrable.

   (a) Montrer que pour tout $\delta > 0$, il existe une partie finie G_δ de $\mathcal{E}$ telle que $\mu(\mathcal{E} \setminus G_\delta) \leq \delta$.

   (b) Montrer que pour tout $\epsilon > 0$, $P(\mu_n(\mathcal{E} \setminus G_{\epsilon/2}) > \epsilon)$ tend vers 0 lorsque n tend vers $+\infty$.

   (c) En déduire que pour tout $\epsilon > 0$, $P(\max_{e\in\mathcal{E}} |\mu_n(\{e\}) - \mu(\{e\})| \geq \epsilon)$ tend vers 0 lorsque n tend vers $+\infty$.

9.5 Le théorème de Shannon

Codage binaire et compression. Remarquons que le codage binaire d'un alphabet à d signes nécessite $[\log_2(d)]+1$ bits ([ ] désignant la partie entière). Pour n grand, le codage d'un message de longueur n écrit avec cet alphabet nécessite a priori $[n\log_2(d)]+1$ bits. Mais dans un type de message donné, les signes n'apparaissent généralement pas avec la même fréquence, loin de là ! Si $\overrightarrow{p}=(p_1,\dots,p_d)$ est le vecteur de leurs fréquences, en supposant, en première approximation, que les signes successifs du message apparaissent de façon indépendante les uns des autres, le théorème de Shannon montre que la plupart des messages peuvent être codés avec un peu plus de $nH(\overrightarrow{p})$ bits où

$$H(\overrightarrow{p})=-\sum_{i=1}^{d}p_i\log_2(p_i)\leq\log_2(d)\text{ avec égalité si et seulement si }p_1=\ldots=p_d=\frac{1}{d}.$$

Comme n est très grand, la différence peut être considérable. Cette observation est à la base de l'idée de compression des données.
$H(\overrightarrow{p})$ est appelée l'entropie de la loi $\overrightarrow{p}$, sa définition s'étend à des lois définies sur un ensemble dénombrable. On définit de même l'entropie d'une variable aléatoire X (définie sur un ensemble Ω au plus dénombrable muni d'une probabilité P qui charge tous les points de Ω) comme étant l'entropie de sa loi P_X et on la note $H(X)$:

$$H(X)=-\sum_{x\in X(\Omega)}P_X(x)\log_2(P_X(x))=-\mathrm{E}(\log_2(P_X(X)))$$

Notons bien que le X en indice de P n'est pas aléatoire ! P_X désigne ici la loi de X.

▷ *Exercice 208.* Montrer que :

1. si X n'est pas constante, $H(X)$ est strictement positive,
2. si X a d valeurs possibles, $H(X)$ est maximale lorsque P_X est uniforme et vaut alors $\log_2(d)$ (faire par exemple une récurrence sur d).

Soit $\overrightarrow{X}=(X^{(i)},\ 1\leq i\leq n)$ un échantillon de taille n formé de n v.a. indépendantes de loi P_X. L'ensemble des résultats possibles de l'expérience statistique ainsi représentée est l'ensemble $X(\Omega)^n$ des suites $\overrightarrow{x}=(x_1,...,x_n)$ de n éléments de $X(\Omega)$. Le théorème suivant montre que la suite obtenue a de grande chance de se trouver dans un ensemble beaucoup plus petit.

Théorème 26 (Shannon) *Fixons $\varepsilon>0$ et posons*

$$A_n^{\varepsilon}=\{\overrightarrow{x}\text{ tels que }|-\log_2(P(\overrightarrow{X}=\overrightarrow{x}))-nH(X)|\leq n\varepsilon\}.$$

Alors $P(A_n^{\varepsilon})$ tend vers 1 quand n tend vers l'infini et A_n^{ε} contient au plus $2^{n(H(X)+\varepsilon)}$ éléments.

Notons que $2^{n(H(X)+\varepsilon)}$ pourra, pour n grand, être très inférieur à $|X(\Omega)^n|$ si P_X n'est pas uniforme.

Preuve du théorème : Remarquons que

$$\log_2(P(\overrightarrow{X} = \overrightarrow{x})) = \sum_{i=1}^{n} \log_2(P(X^{(i)} = x_i)) = \sum_{i=1}^{n} \log_2(P_X(x_i)).$$

La première assertion résulte de la loi des grands nombres appliquées aux v.a. $-\log_2(P_X(X^{(i)}))$ dont l'espérance vaut $H(X)$.
Pour la deuxième assertion, remarquons que si $\overrightarrow{x}$ est un élément de A_n^ε,

$$P(\overrightarrow{X} = \overrightarrow{x}) \geq 2^{-n(H(X)+\varepsilon)}.$$

Donc

$$|A_n^\varepsilon|\, 2^{-n(H(X)+\varepsilon)} \leq P(A_n^\varepsilon) \leq 1.$$

□

Chapitre 10

Le théorème limite central

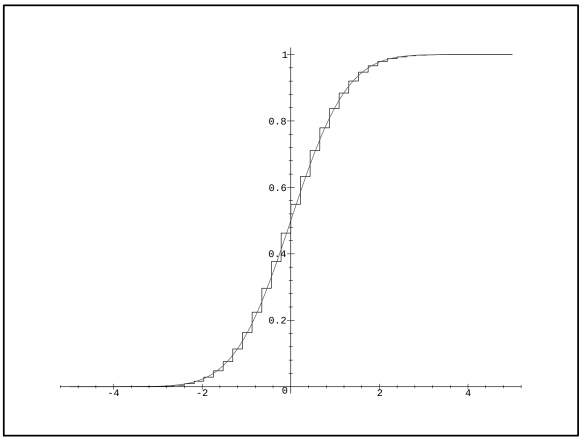
Commençons par un exemple : traçons sur un même graphique la fonction de répartition $F_{\widetilde{X}}$ de la v.a. centrée réduite $\widetilde{X} = \frac{X - \mathrm{E}(X)}{\sigma(X)}$ associée à une v.a. X suivant la loi binomiale $\mathcal{B}(100; 0.3)$ et la fonction de répartition de la loi normale centrée réduite

$$F_{\gamma_{0,1}} \;: x \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)dt.$$

▷ *Exercice 209.* En utilisant un logiciel de calcul formel, on tracera les graphes de

$$F_{\widetilde{X}}(x) = \sum_{k=0}^{Ent(30+x\sqrt{21})} \binom{100}{k} (.3)^k (.7)^{100-k} \text{ et de } F_{\gamma_{0,1}}(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{1}{2}\sqrt{2}x\right).$$

On pourra faire varier les paramètres de la loi binomiale $\mathcal{B}(n,p)$ (dans les limites des possibilités de calcul de la machine, vite dépassées si n est grand).



La proximité de ces deux graphes illustre un fait général qu'exprime le théorème suivant :

Théorème 27 *Considérons un échantillon $(X_1, \ldots, X_n)$ d'une v.a. réelle X de carré intégrable, de moyenne m et de variance σ^2. Lorsque la taille n de l'échantillon tend vers l'infini, la fonction de répartition de $\frac{S_n - nm}{\sqrt{n}}$ converge uniformément vers $F_{\gamma_{0,\sigma^2}}$. De façon équivalente, la fonction de répartition de $\tilde{S}_n = \frac{S_n - nm}{\sigma\sqrt{n}}$ converge uniformément vers $F_{\gamma_{0,1}}$.*

On peut se ramener au cas où $m = 0$ en remplaçant X par $X - m$. Nous allons montrer que la fonction de répartition de $\tilde{S}_n$ converge en tout point vers $F_{\gamma_{0,1}}$. La convergence uniforme s'obtient ensuite en utilisant le lemme de Dini suivant :

Lemme 28 *Soit f_n pour tout $n \in \mathbb{N}$ et f des fonctions définies sur $\mathbb{R}$, à valeurs dans un intervalle borné $[a, b]$, croissantes et continues à droite. Si pour tout $t \in \mathbb{R}$, $(f_n(t))_n$ converge vers $f(t)$ alors $\sup_{t\in\mathbb{R}} |f_n(t) - f(t)|$ tend vers 0 lorsque n tend vers $+\infty$.*

Pour simplifier la preuve de la convergence en tout point de la fonction de répartition de $\tilde{S}_n$, nous supposerons que $\mathrm{E}(|X|^3)$ est fini (ce qui entraine que X est de carré intégrable). Cette hypothèse n'est pas nécessaire.

Preuve du théorème 27 : On doit montrer que pour tout $t \in \mathbb{R}$,

$$P(\frac{S_n}{\sqrt{n}} \leq t) - \int_{-\infty}^{t} \gamma_{0,\sigma^2}(z)dz$$

converge vers 0 lorsque n tend vers $+\infty$. En notant ϕ_t la fonction indicatrice de l'intervalle $]-\infty, t]$, cela revient à montrer que pour tout $t \in \mathbb{R}$, $\mathrm{E}(\phi_t(\frac{S_n}{\sqrt{n}})) - \int_{-\infty}^{t} \phi_t(z)\gamma_{0,\sigma^2}(z)dz$ converge vers 0.

1. Dans une première partie, on va montrer que le résultat est vrai si on remplace ϕ_t par une fonction ϕ de classe C^3 dont la dérivée troisième est bornée.
   Introduisons des variables aléatoires indépendantes $Z_1, \dots, Z_n$, de loi de densité γ_{0,σ^2} et indépendantes de $X_1, \dots, X_n$. Posons $S_0 = 0$, $S_k = \sum_{i=1}^{k} X_i$ et $T_k = Z_{k+1} + \ldots + Z_n$ et $V_k = \phi(\frac{S_k+T_{k+1}}{\sqrt{n}})$ pour $k \in \{0, \dots, n\}$. Il s'agit de montrer qu'il existe une constante C ne dépendant pas de ϕ telle que

   $$|\mathrm{E}(V_n) - \mathrm{E}(V_0)| \leq C||\phi'''||_\infty n^{-\frac{1}{2}}.$$

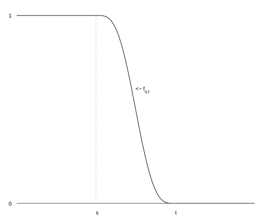
   Décomposons $\mathrm{E}(V_n) - \mathrm{E}(V_0)$ en $\sum_{k=1}^{n} \mathrm{E}(V_k - V_{k-1})$. Soit $k \in \{1, \dots, n\}$. En appliquant la formule de Taylor-Lagrange deux fois au point $S_{k-1} + T_k$, on obtient :

   $$\mathrm{E}(V_k - V_{k-1}) = E(\phi'(\frac{1}{\sqrt{n}}(S_{k-1}+T_k)))\frac{1}{\sqrt{n}}E(X_k - Z_k) + \frac{1}{2}\phi''(\frac{1}{\sqrt{n}}(S_{k-1}+T_k)))\frac{1}{n}E(X_k^2 - Z_k^2) + R_k$$

   avec $|R_k| \leq \frac{1}{6n^{\frac{3}{2}}}||\phi'''||_\infty E((|X_k| + |Z_k|)^3)$. Les deux premiers termes du développement s'annulent puisque X_k et Z_k ont même espérance et même variance.
   En sommant en k, par l'inégalité triangulaire, on obtient la majoration voulue pour $|\mathrm{E}(V_n) - \mathrm{E}(V_0)|$.

2. La convergence de la fonction de répartition de $\frac{S_n}{\sqrt{n}}$ en un point t s'obtient en encadrant l'indicatrice de l'intervalle $]-\infty, t]$ par des fonctions de classe C^3 dont la dérivée troisième est bornée et en appliquant le résultat de la première partie de la preuve.
   Soit $s < t$ deux réels. Considérons une fonction $f_{s,t}$ de classe C^3 telle que :
   $0 \leq f_{s,t} \leq 1$, $f_{s,t} = 1$ sur $]-\infty, -s]$ et $f = 0$ sur $[t, +\infty[$. On peut prendre par exemple $f_{s,t}(x) = f(\frac{x-s}{t-s})$ où f est la fonction définie par

   $$f(x) = \begin{cases} 1 \text{ si } x \in]-\infty, 0], \\ 1 - 140\int_0^x t^3(1-t)^3 dt \\ \quad = 1 + 20x^7 - 70x^6 + 84x^5 - 35x^4 \text{ si } x \in [0, 1] \\ 0 \text{ si } x \in [1, +\infty[\end{cases}$$



En utilisant l'encadrement

$$f_{s,t} \leq 1\!\!1_{]-\infty,t]} \leq f_{t,u} \text{ pour tout } s < t < u,$$

on montre les inégalités suivantes :

$$\forall u > t,\ \overline{\lim} P(\frac{S_n}{\sqrt{n}} \leq t) \leq F_{0,\sigma^2}(u) \text{ et } \forall s < t, \quad F_{0,\sigma^2}(s) \leq \underline{\lim} P(\frac{S_n}{\sqrt{n}} \leq t).$$

On conclut en utilisant la continuité de F_{0,σ^2}.

□

N.B.

1. Le théorème limite central permet d'approximer la fonction de répartition de la moyenne empirique d'un échantillon suffisamment important : si $(X_1, \ldots, X_n)$ est un n-échantillon d'une variable aléatoire réelle de carré intégrable, d'espérance m et de variance σ^2, pour tout $t \in \mathbb{R}$, $\sup_{t\in\mathbb{R}} |P(\bar{X}_n \leq t) - \int_{-\infty}^{\frac{\sqrt{n}}{\sigma}(t-m)} \gamma_{0,1}(u)du|$ tend vers 0 lorsque n tend vers $+\infty$.

2. Au cours de la preuve du théorème limite central, nous avons établi la majoration suivante : si $(X_1, \ldots, X_n)$ est un n-échantillon d'une variable aléatoire réelle X centrée ayant un moment d'ordre 3 fini, $S_n = \sum_{i=1}^n X_i$ et si ϕ est une fonction de classe C^3 dont la dérivée troisième est bornée, alors

$$|E(\phi(\frac{S_n}{\sqrt{n}}) - \int_{-\infty}^{+\infty} \phi(z)\gamma_{0,1}(z)dz| \leq ||\phi'''||_\infty \frac{6}{\sqrt{n}}(2E(X^2)^{\frac{3}{2}} + E(|X^3|)).$$

En appliquant ce résultat pour un échantillon de la variable aléatoire $\frac{X-p}{\sqrt{p(1-p)}}$ lorsque X suit la loi de Bernoulli $\mathcal{B}(p)$, on obtient :

$$|E(\phi(\frac{S_n - np}{\sqrt{n(p(1-p)}}) - \int_{-\infty}^{+\infty} \phi(z)\gamma_{0,1}(z)dz| \leq ||\phi'''||_\infty \frac{3}{\sqrt{n}}(1 + \frac{1}{\sqrt{p(1-p)}}).$$

Remarquons que la majoration diminue lorsque n augmente, mais augmente lorsque p ou $1-p$ décroit vers 0. La figure 10.1 représente

$$\Delta(p,n) = \max_{k\in\{0,\ldots,n\}} |P(S_n \leq k) - \int_{-\infty}^{\frac{k-np}{\sqrt{np(1-p)}}} \gamma_{0,1}(u)du|$$

en fonction de p pour deux valeurs de n ; cela montre que l'erreur commise en approximant la fonction de répartition d'une loi $\mathcal{B}(n,p)$ à l'aide de la fonction de répartition de la densité gaussienne dépend de n et de p. Lorsque p ou $(1-p)$ est petit, de sorte que np ou $n(1-p)$ n'est pas trop grand, on utilise plutôt l'approximation poissonienne.

▷ *Exercice 210.* Ecrire en détail le point 2 de la preuve du théorème limite central.

▷ *Exercice 211.* Montrer que lorsque n est grand, on peut approcher la fonction de répartition de la loi du χ^2 à n degrés de liberté par $t \mapsto \int_{-\infty}^{\frac{t-n}{\sqrt{2n}}} \gamma_{0,1}(u)du$.

▷ *Exercice 212.* Pour étudier les particules émises par une substance radioactive, on dispose d'un détecteur. On note X la variable aléatoire représentant le nombre de particules qui atteignent le détecteur pendant un intervalle de temps Δt. Le nombre maximal de particules que le détecteur peut compter pendant un intervalle de temps Δt est de 140.

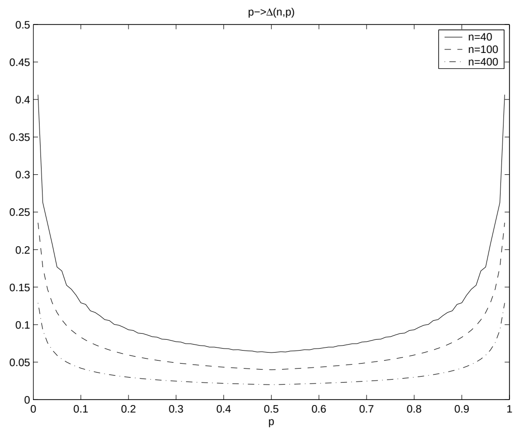


FIG. 10.1 : Représentation de la courbe de $\Delta(p,n) := \max(|P(S_n \leq k) - \int_{-\infty}^{\frac{k-np}{\sqrt{np(1-p)}}} \gamma_{0,1}(u)du|, k \in \{0,\ldots,n\})$ en fonction de p lorsque S_n est une variable aléatoire de loi $\mathcal{B}(n,p)$.

1. On suppose que X a pour espérance 100. Donner une majoration de la probabilité que X dépasse 140.
2. On suppose de plus que X a un écart-type égal à 10. Donner une majoration plus fine de la probabilité que X dépasse 140.
3. On suppose de plus que X suit une loi de Poisson. Donner une approximation de la probabilité que X dépasse 140 en utilisant le théorème limite central.

▷ *Exercice 213.* On veut transporter entre deux villes A et B dans des conditions de confort acceptables 1600 voyageurs se présentant pratiquement en même temps à la gare de A. On met à leur disposition deux trains identiques. On suppose que chaque individu choisit au hasard l'une ou l'autre rame et qu'il n'a pas le temps d'en changer. Combien faut-il prévoir de places assises dans chaque rame si l'on veut que la probabilité que des voyageurs soient obligés de rester debout soit inférieure à 3.10^{-3} ?

10.1 Exemples d'application en statistiques

Reprenons l'exemple du sondage où l'on cherche à estimer la valeur de la proportion p d'éléments d'un certain type τ dans une population Ω de grande taille, à partir d'un tirage au hasard et sans remise de n éléments dans cette population. Désignons par S_n la variable aléatoire qui, à un tirage de n éléments dans cette population associe le nombre d'éléments de type τ. Notons $X_n = \frac{S_n}{n}$ et p_0 un réel de $]0,1[$. Si on souhaite vérifier ou invalider l'hypothèse que la valeur de p est p_0 au niveau α, on effectuera le test $H_0\;: p = p_0$ contre l'hypothèse $H_1\;: p \neq p_0$ au niveau α. Pour ce test, on rejette l'hypothèse H_0 si on observe une valeur x_{obs} de X_n qui est exceptionnellement éloignée de p_0, c'est-à-dire si $P_{H_0}(|X_n - p_0| \geq |x_{obs} - p_0|) \leq \alpha$. Si np_0 et $n(1-p_0)$ sont suffisamment grands, on pourra utiliser l'approximation donnée par le théorème limite central : pour tout réel t, $P(\frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}}|X_n - p_0| \geq t) \simeq 1 - \int_{-t}^{t} \gamma_{0,1}(u)du$. Notons t_α le réel qui vérifie $\int_{-t_\alpha}^{t_\alpha} \gamma_{0,1}(u)du = 1 - \alpha$. On est conduit à rejeter l'hypothèse "$p = p_0$" au niveau approximativement α si $\frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}}|x_{obs} - p_0| \geq t_\alpha$. Une valeur approchée de t_α peut-être obtenue à partir de la table de valeurs numériques de la fonction de répartition de la loi $\mathcal{N}(0,1)$: la densité gaussienne $\gamma_{0,1}$ étant paire, t_α est le réel qui vérifie $F_{\mathcal{N}(0,1)}(t_\alpha) = 1 - \frac{\alpha}{2}$. Par exemple, pour un test de niveau 5%, on rejette l'hypothèse $H_0\;: p = p_0$ si $\frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}}|x_{obs} - p_0| \geq 1.96$.

La région de confiance pour p de niveau de confiance approximativement $1-\alpha$ que l'on obtient à l'aide de cette famille de tests est l'ensemble $I(x_{obs})$ des réels $p_0 \in]0,1[$ pour lesquels on conserve l'hypothèse H_0 pour l'observation x_{obs}. Donc,

$$I(x_{obs}) = \{p_0 \in]0,1[\text{ pour lesquels } |x_{obs} - p_0| < t_\alpha \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}}\}. \tag{10.1}$$

En prenant le carré des deux membres de l'inégalité, on voit que $I(x_{obs}) =]p_-, p_+[$ où $p_- \leq p_+$ sont les racines du polynome en z : $(x_{obs} - z)^2 - \frac{t_\alpha}{\sqrt{n}} z(1-z)$ c'est-à-dire

$$p_\pm = \left(x_{obs} + \frac{t_\alpha^2}{n} \pm \sqrt{x_{obs}\frac{t_\alpha^2}{n} + \frac{t_\alpha^4}{4n^2} - x_{obs}^2 \frac{t_\alpha^2}{n}} \right) / (1 + \frac{t_\alpha^2}{n}).$$

Cet intervalle a été construit en utilisant une approximation déduite du théorème limite central pour un échantillon de variables aléatoires de loi de Bernoulli $\mathcal{B}(p_0)$, il n'est donc pas très utile de résoudre explicitement cette équation. En considérant l'expression (10.1) de $I(x_{obs})$, on peut se contenter de remarquer que $I(x_{obs})$ est inclus dans l'ensemble

$$J(x_{obs}) = \{p_0 \in]0,1[\text{ tels que } |x_{obs} - p_0| < \frac{t_\alpha}{\sqrt{4n}}\} =]x_{obs} - \frac{t_\alpha}{\sqrt{4n}}, x_{obs} + \frac{t_\alpha}{\sqrt{4n}}[.$$

Donc, $J(x_{obs})$ est un intervalle de niveau de confiance supérieur à celui de $I(x_{obs})$.

Pour estimer l'espérance d'un grand échantillon $(X_1, \ldots, X_n)$ d'une variable aléatoire X de carré intégrable de loi quelconque, il n'est pas toujours possible d'utiliser le théorème limite central directement si on ne connait pas l'expression de la variance de X. Dans ce cas, on pourra estimer l'écart-type à l'aide d'un estimateur consistant :

Corollaire 29 *Soit $(X_1, \ldots, X_n)$ un n-échantillon d'une variable aléatoire réelle X de carré intégrable non-constante et T_n un estimateur consistant de l'écart-type de X. Notons $\bar{X}_n$ la moyenne empirique de l'échantillon. Alors, la fonction de répartition de $\frac{\sqrt{n}}{T_n}(\bar{X}_n - m)$ converge uniformément vers la fonction de répartition de la loi $\mathcal{N}(0,1)$.*

Les étapes de la preuve de ce corollaire sont données dans l'exercice 218.

On peut prendre pour T_n, la racine carrée de l'estimateur empirique non biaisé de la variance c'est-à-dire

$$\hat{\sigma}_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}.$$

En appliquant ce corollaire, on obtient, lorsque n est suffisamment grand, l'intervalle de confiance pour m de niveau approximativement $1-\alpha$ suivant :

$$]\bar{X}_n - t_\alpha \frac{\hat{\sigma}_n}{\sqrt{n}}, \ \bar{X}_n + t_\alpha \frac{\hat{\sigma}_n}{\sqrt{n}}[.$$

Dans le cas particulier où $(X_1, \ldots, X_n)$ est un n-échantillon de loi $\mathcal{B}(p)$, la variance de la loi $\mathcal{B}(p)$ étant $p(1-p)$, $(\bar{X}_n(1-\bar{X}_n))^{\frac{1}{2}}$ est un estimateur consistant de l'écart-type de la loi

$\mathcal{B}(p)$. Cela conduit à un autre intervalle de confiance de niveau approximativement $1-\alpha$ pour la proportion p, lorsque n est suffisamment grand :

$$]\bar{X}_n - t_\alpha \left(\frac{\bar{X}_n(1-\bar{X}_n)}{n}\right)^{\frac{1}{2}}, \ \bar{X}_n + t_\alpha \left(\frac{\bar{X}_n(1-\bar{X}_n)}{n}\right)^{\frac{1}{2}}[.$$

La figure 10.2 montre que là encore l'erreur que l'on commet en approximant la fonction de répartition de $\frac{\sqrt{n}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}(\bar{X}_n - p)$ par celle de la loi $\mathcal{N}(0,1)$ dépend de n mais aussi de p.

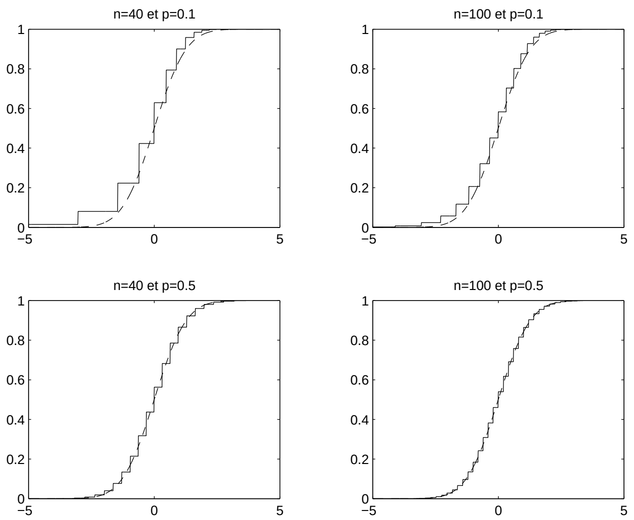


Fig. 10.2 : Comparaison entre la fonction de répartition de la loi $\mathcal{N}(0,1)$ (en pointillés) et celle de $\frac{\sqrt{n}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}(\bar{X}_n - p)$ lorsque $(X_1,\dots,X_n)$ est un n-échantillon de loi $\mathcal{B}(p)$ (en traits continus)

▷ *Exercice 214.* On effectue un sondage sur un échantillon de 10000 personnes à la veille d'un référendum : 4903 d'entre elles s'apprêtent à voter oui, et 5097 à voter non. Quel risque d'erreur court-on en prédisant la victoire du non ?

▷ *Exercice 215.* On lance une pièce 100 fois. Elle tombe 40 fois sur pile. Déterminer un intervalle de confiance pour la probabilité p que la pièce tombe sur "pile" :

1. de niveau supérieur ou égal à 90% en utilisant l'inégalité de Hoeffding (voir exercice 200),
2. de niveau approximativement 90% en utilisant une approximation gaussienne.

Peut-on dire que la pièce est bien équilibrée ?

▷ *Exercice 216.* Reprendre la question 4 de l'exercice 173 lorsque $n = 100$ et lorsque l'observation de S_{100}^2 est 0.003.

▷ *Exercice 217.* Chaque jour, un train subit un retard aléatoire au départ, évalué en minutes, indépendant des retards des autres jours et dont la loi est approximativement $\mathcal{E}(\lambda)$. Sur 400 jours, le retard moyen est de 10 mn. Donner un intervalle de confiance de niveau approximativement 98% pour λ.

▷ *Exercice 218.* Soit F une fonction de répartition continue d'une variable aléatoire réelle, $(U_n)_n$ et $(Z_n)_n$ deux suites de variables aléatoires définies sur un espace de probabilité $(\Omega,\mathcal{A},P)$ telles que :

- $(U_n)_n$ converge en probabilité vers un réel u, c'est-à-dire $\forall\epsilon>0$, $P(|U_n-u|>\epsilon) \underset{n\to+\infty}{\to} 0$,
- pour tout $t\in\mathbb{R}$, $P(Z_n\le t) \underset{n\to+\infty}{\to} F(t)$.

L'objectif de l'exercice est de montrer que la fonction de répartition de U_nZ_n converge en tout point vers la fonction $t\mapsto F(\frac{t}{u})$.

1. Supposons que $u > 0$. On notera F_n la fonction de répartition de Z_n et G_n celle de $U_n Z_n$. Soit $0 < \delta < u$ et $t \in \mathbb{R}$.
   (a) Montrer que $G_n(t) \leq P(Z_n U_n \leq t \text{ et } |U_n - u| \leq \delta) + P(|U_n - u| > \delta)$.
   (b) En déduire que $G_n(t) \leq F_n(\frac{t}{u+\delta})\mathbb{1}_{t<0} + F_n(\frac{t}{u-\delta})\mathbb{1}_{t\geq 0} + P(|U_n - u| > \delta)$.
   (c) Montrer de même que $P(Z_n U_n > t) \leq P(Z_n > \frac{t}{u+\delta})\mathbb{1}_{t\geq 0} + P(Z_n > \frac{t}{u-\delta})\mathbb{1}_{t<0} + P(|U_n - u| > \delta)$.
   (d) En déduire que la suite $(G_n(t))_n$ converge vers $F(\frac{t}{u})$.
2. En procédant de façon analogue, traiter les cas $u < 0$ et $u = 0$.
3. *Application :* utiliser ce résultat pour en déduire le corollaire 29.

N.B. Lorsque F n'est pas continue, en remplaçant l'hypothèse sur $(Z_n)_n$ par : en tout point $t \in \mathbb{R}$ tel que $F(t)$ est continue $(F_n(t))_n$ converge vers $F(t)$, on peut montrer avec ce schéma de preuve, qu'en tout point s tel que $\frac{s}{u}$ est un point de continuité de F, $(G_n(s))_n$ converge vers $F(\frac{s}{u})$.

▷ *Exercice 219.* Soit $(X_1, \ldots, X_n)$ un n-échantillon d'une variable aléatoire réelle X de carrée intégrable. Vérifier que $\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$ est un estimateur consistant de l'écart-type de X.

10.2 Le théorème limite central vectoriel

Le théorème limite central se généralise au cas de vecteurs aléatoires de la façon suivante :

Théorème 30 *Considérons un échantillon $(\overrightarrow{X}^{(1)}, ..., \overrightarrow{X}^{(n)})$ d'un vecteur aléatoire $\overrightarrow{X}$, de moyenne $\overrightarrow{m}$ et de matrice de covariance non dégénérée C. On pose $\overrightarrow{S_n} = \sum_{i=1}^n \overrightarrow{X}^{(i)}$. Lorsque la taille n de l'échantillon tend vers l'infini, la fonction de répartition de $\frac{\overrightarrow{S_n} - n\overrightarrow{m}}{\sqrt{n}}$ converge uniformément vers la fonction de répartition de la loi normale $\mathcal{N}(0, C)$.*

▷ *Exercice 220.*

1. On lance un dé bien équilibré n fois. On note $X_n^{(i)}$ le nombre de fois où le dé est tombé sur i au cours de ces n lancers pour $i \in \{1, \ldots, 6\}$ et $\overrightarrow{Y}_n$ le vecteur $(\frac{1}{n}X_n^{(1)}, \ldots, \frac{1}{n}X_n^{(6)})$.
   (a) Montrer que $(\overrightarrow{Y}_n)_n$ converge en probabilité vers le vecteur $\overrightarrow{l} = (\frac{1}{6}, \ldots, \frac{1}{6})$ c'est-à-dire que pour tout $\epsilon > 0$, $P(||\overrightarrow{Y}_n - \overrightarrow{l}||_\infty > \epsilon)$ tend vers 0.
   (b) Par quelle loi peut-on approcher la loi du vecteur $\sqrt{n}(\overrightarrow{Y}_n - \overrightarrow{l})$ lorsque n est très grand ?
2. Plus généralement, on s'intéresse à une expérience aléatoire dont les résultats possibles sont notés $e_1, \ldots, e_d$. On note p_i la probabilité que le résultat de l'expérience soit e_i pour tout $i \in \{1, \ldots, d\}$ et $\overrightarrow{p} = (p_1, \ldots, p_d)$. On effectue n fois cette expérience. On note $\overrightarrow{X}_n$ le vecteur de dimension d dont la i-ème coordonnée est le nombre moyen d'expériences dont le résultat a été e_i. Montrer que $\overrightarrow{X}_n$ est un estimateur consistant du vecteur $\overrightarrow{p}$, puis donner une approximation de la loi du vecteur $\sqrt{n}(\overrightarrow{X}_n - \overrightarrow{p})$ lorsque n est très grand.

▷ *Exercice 221.* Une molécule formée d'une longue chaîne plane d'atomes identiques est modélisée ainsi : deux atomes consécutifs positionnés en A_i et A_{i+1} sont à une distance ϵ, la direction du vecteur $\overrightarrow{A_iA_{i+1}}$ est une variable aléatoire de loi uniforme sur l'ensemble des directions possibles et indépendante des directions des autres vecteurs $\overrightarrow{A_jA_{j+1}}$.

Quelle est la loi approchée du vecteur $\overrightarrow{A_1A_n}$ où A_1 désigne la position du premier atome de la chaîne et A_n la position du dernier atome de la chaine ? En déduire la loi approchée du carré de la distance entre le premier atome et le dernier atome de la chaine.

Chapitre 11

Lois et test du chi-deux

11.1 Un exemple de test

Considérons une expérience aléatoire $\mathcal{E}$ décrite par un ensemble Λ muni de la probabilité Q. Soit $C_1, \ldots, C_d$ une partition de Λ en événements de probabilité $p_i = Q(C_i)$ pour $i \in \{1, \ldots, d\}$. Pour $i \in \{1, \ldots, d\}$, notons N_j le nombre de fois où l'événement C_j a été réalisé lors de l'expérience statistique consistant à répéter n fois l'expérience aléatoire $\mathcal{E}$: $N_j = \sum_{i=1}^n \mathbb{1}_{C_j^{(i)}}$. On sait que le vecteur aléatoire $\overrightarrow{N}$ suit la loi multinomiale de paramètres n et $p = (p_1, \ldots, p_d)$ où p_i est la probabilité que X prenne la valeur i pour tout $i \in \{1, \ldots, d\}$. Pour tester l'hypothèse suivante sur la loi de X

$$H_0 \ : p = (\pi_1, \ldots, \pi_d)$$

avec π_k des réels positifs tels que $\sum_{k=1}^d \pi_k = 1$, on utilise fréquemment la variable de test

$$Z = \sum_{k=1}^{d} \frac{(N_k - n\pi_k)^2}{n\pi_k}.$$

Si H_0 est vraie, cette variable aléatoire a pour espérance $d-1$ quelle que soit la taille de l'échantillon (voir exemple 192). Par contre, si H_0 est fausse, son espérance est supérieure à $n(\sum_{k=1}^d \frac{(p_k-\pi_k)^2}{\pi_k})$.
Pour un échantillon d'assez grande taille, Z aura tendance à prendre de grandes valeurs, ce qui conduit à rejeter H_0 lorsque la valeur observée z_{obs} de la v.a. de test Z est exceptionnellement grande, c'est-à-dire si $P_{H_0}(Z \geq z_{obs})$ est inférieure au niveau du test.

Considérons par exemple le cas où $d = 3$ et $p_1 = p_2 = p_3 = \frac{1}{3}$ et où nous disposons d'un échantillon de taille $n = 45$. La fonction de répartition de Z en x se calcule en décomposant l'événement $\{Z \leq x\}$ en fonction des valeurs possibles pour $\overrightarrow{N}$:

$$\begin{aligned} F_Z^{H_0}(x) &= \sum_{i=0}^{45}\sum_{j=0}^{i} P(\overrightarrow{N} = (j, i-j, n-i))H(x - \frac{(j-15)^2}{15} - \frac{(i-j-15)^2}{15} - \frac{(30-i)^2}{15}) \\ &= \sum_{i=0}^{45}\left(\sum_{j=0}^{i}\binom{45}{i}\binom{i}{j}(\frac{1}{3})^{45} H(x - \frac{(j-15)^2}{15} - \frac{(i-j-15)^2}{15} - \frac{(30-i)^2}{15})\right) \end{aligned}$$

où $H(x) = \mathbb{1}_{\{x\geq 0\}}$. Son graphe est tracé sur la figure 11.1, en même temps que celui de la fonction $1-\exp(-x/2)$. La proximité des deux courbes sera expliquée dans la section 11.

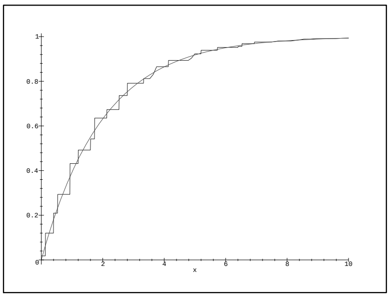


FIG. 11.1 : Représentations graphiques de $F_Z^{H_0}$ et de la fonction $x \mapsto 1-\exp(-\frac{x}{2})$

On voit que la probabilité de dépasser la valeur 6 est voisine de 5% et que celle de dépasser la valeur 8 est voisine de 2%. Cependant ces chiffres ne doivent pas faire illusion : dans les cas où l'hypothèse H_1 est simplement la négation de H_0, un résultat de test positif signifie simplement que l'hypothèse de référence n'est pas infirmée. Si par exemple nous précisions H_1 en prenant: $P_{H_1}(X=1) = P_{H_1}(X=2) = \frac{1}{4}$, et $P_{H_1}(X=3) = \frac{1}{2}$, la fonction de répartition de Z sous l'hypothèse H_1 s'écrirait :

$$F_Z^{H_1}(x) = \sum_{i=0}^{45} (\frac{1}{4})^i (\frac{1}{2})^{45-i} \left(\sum_{j=0}^{i} \binom{45}{i} \binom{i}{j} H(x - \frac{(j-15)^2}{15} - \frac{(i-j-15)^2}{15} - \frac{(30-i)^2}{15}) \right).$$

Son graphe est tracé sur la figure 11.2.

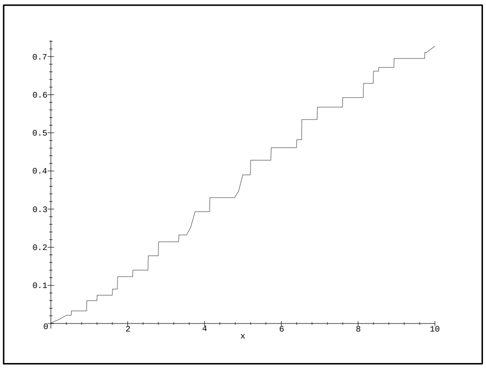


FIG. 11.2 : Graphe de la fonction $F_Z^{H_1}$

On voit que $F_Z^{H_1}(6) = 46\%$. La puissance du test de niveau 5% est donc très faible (54%). Celle du test de niveau 2% est encore moindre. Si l'on cherchait à équilibrer les deux types de risques d'erreur, on prendrait le point d'intersection des courbes de $F_Z^{H_0}$ et de $1-F_Z^{H_1}$ (figure 11.3), ce qui donne des risques d'erreur de 20% si l'on opte pour H_1 dès que z_{obs} dépasse 3.

N.B. Nous avons présenté ce calcul pour montrer les limites des tests de confirmation d'hypothèse. Si nous avions à choisir entre H_0 et H_1, il vaudrait mieux employer comme variable de test $\overrightarrow{N}$

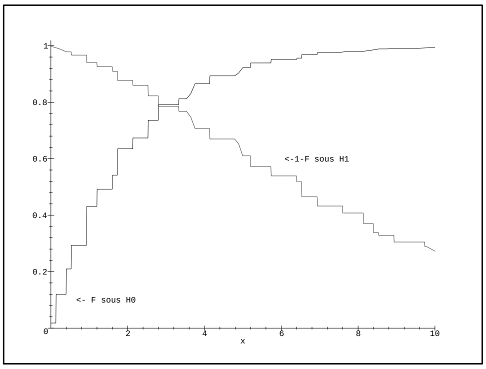


FIG. 11.3 : Graphes des fonctions $F_Z^{H_0}$ et $1 - F_Z^{H_1}$

et étudier le rapport des vraisemblances pour déterminer la zone de rejet de H_0 en suivant la méthode de la section 5.

▷ *Exercice 222.* Montrer qu'en utilisant la variable aléatoire de test $\overrightarrow{N}$ issue de l'observation du résultat de 45 expériences, un critère naturel pour choisir entre les hypothèses $H_0 \;: p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ et $H_1 \;: p = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ est de rejeter H_0 si $N_3 \geq 19$. Montrer que le niveau de ce test est de $18,5\%$ et que sa puissance est de $88,4\%$.
Modifier ce critère afin d'obtenir un test de niveau inférieur ou égal à 5% et calculer la puissance de ce nouveau test.

11.2 Un théorème de convergence

Revenons à la situation étudiée section 11.1. Soit $(X_1, ..., X_n)$ un échantillon de la loi d'une variable aléatoire X portée par un ensemble à d éléments $\{x_1, ..., x_d\}$. Pour tout k compris entre 1 et d, notons $N_n^{(k)}$ le nombre d'indices $i \leq n$ pour lesquels $X_i = x_k$. On sait que le vecteur aléatoire $\overrightarrow{N_n}$ suit la loi multinomiale de paramètres n et $\overrightarrow{p} = (p_1, ..., p_d)$ où $p_k = P(X = x_k)$.

Théorème 31 *Quand la taille n de l'échantillon tend vers l'infini, la fonction de répartition de la variable aléatoire $Z_n = \sum_{k=1}^d \frac{(N_n^{(k)} - np_k)^2}{np_k}$ converge vers la fonction de répartition de la loi du χ^2 à $(d-1)$ degrés de liberté.*

Preuve : Avant de démontrer le résultat dans le cas général, faisons la preuve lorsque $d = 2$. On a $Z_n = \frac{(N_n^{(1)} - np_1)^2}{np_1} + \frac{(N_n^{(2)} - np_2)^2}{np_2}$. En utilisant que $N_n^{(1)} + N_n^{(2)} = n$ et $p_1 + p_2 = 1$, le deuxième rapport s'écrit : $\frac{(N_n^{(1)} - np_1)^2}{n - np_1}$. Donc, $Z = \frac{(N_n^{(1)} - np_1)^2}{n}(\frac{1}{p_1} + \frac{1}{1-p_1}) = \frac{(N_n^{(1)} - np_1)^2}{np_1(1-p_1)}$. D'après le théorème limite central, la fonction de répartition de la loi $\frac{N_n^{(1)} - np_1}{\sqrt{np_1(1-p_1)}}$ converge uniformément vers celle de la loi $\mathcal{N}(0,1)$ lorsque n tend vers $+\infty$. Donc, la fonction de répartition de Z_n converge vers celle de la loi du χ^2 à un degré de liberté.
La preuve dans le cas général fait appel aux fondements de la géométrie euclidienne en dimension d. On pourra sans inconvénient supposer que $d = 3$ pour mieux la visualiser car la preuve est la même en dimension supérieure. La variable Z_n représente la norme euclidienne du vecteur aléatoire $\overrightarrow{V}_n$ de coordonnées $\frac{N_n^{(k)} - np_k}{\sqrt{np_k}}$. De plus, comme $\sum_{k=1}^d N_n^{(k)} = n$ et $\sum_{k=1}^d p_k = 1$, ce vecteur est orthogonal au vecteur constant $\overrightarrow{\nu} = (\sqrt{p_1}, \ldots, \sqrt{p_d})$.

Soit $(\overrightarrow{e_j}, 1 \le j \le d-1)$ une base orthonormée quelconque de l'hyperplan orthogonal à $\overrightarrow{\nu}$. La variable aléatoire Z_n peut s'écrire $\sum_{j=1}^{d-1} \left\langle \overrightarrow{V}_n, \overrightarrow{e_j} \right\rangle^2$. Remarquons que l'échantillon de X fournit (en posant $\overrightarrow{U}_i^{(k)} = \frac{\mathbf{1}_{\{X_i = x_k\}} - p_k}{\sqrt{p_k}}$) un échantillon $(\overrightarrow{U}_1, .., \overrightarrow{U}_n)$ du vecteur $\overrightarrow{U} = (\frac{\mathbf{1}_{\{X=x_1\}} - p_1}{\sqrt{p_1}}, \ldots, \frac{\mathbf{1}_{\{X=x_d\}} - p_d}{\sqrt{p_d}})$ et qu'on a : $\left\langle \overrightarrow{V}_n, \overrightarrow{e_j} \right\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\langle \overrightarrow{U}_i, \overrightarrow{e_j} \right\rangle$. Donc la variable Z_n s'écrit aussi comme la norme euclidienne du vecteur $\frac{1}{\sqrt{n}} \sum_{i=1}^n \overrightarrow{Y}_i$ avec $\overrightarrow{Y}_i$ le vecteur de coordonnées $\left\langle \overrightarrow{U}_i, \overrightarrow{e_j} \right\rangle$. Comme le vecteur aléatoire $(d-1)$-dimensionnel $\overrightarrow{Y}$ de coordonnées $\left\langle \overrightarrow{U}, \overrightarrow{e_j} \right\rangle$ est de moyenne nulle et comme sa matrice de covariance est égale à l'identité (exercice !), d'après le théorème 30, la fonction de répartition du vecteur $\frac{1}{\sqrt{n}} \sum_{i=1}^n \overrightarrow{Y}_i$ converge vers la densité gaussienne de moyenne nulle et de covariance identité sur $\mathbb{R}^{d-1}$. Donc la fonction de répartition de Z_n converge vers celle de la loi du χ^2 à d degrés de liberté (voir la remarque suivant la définition de la loi χ^2_d, page 78). $\square$

11.3 Le test d'ajustement du chi-deux

Revenons au test présenté dans la section 11.1. La variable Z_n est utilisée pour tester l'hypothèse H_0 : "$P_{H_0}(X = x_k) = p_k$ pour tout k", faite sur la loi de X, à partir du tirage d'un échantillon de taille n. Si la valeur observée z_0 est telle que $P_{H_0}(Z_n \ge z_0)$ est inférieur ou égal au niveau du test, qui est fixé a priori, l'hypothèse est infirmée. Lorsque n est relativement petit, les logiciels de calcul formel d'emploi courant permettent d'effectuer des calculs exacts mais lorsque n devient grand, leurs possibilités sont vite dépassées et il faut utiliser l'approximation de la loi de Z_n donnée dans le théorème précédent. Cette approximation est d'ailleurs souvent excellente comme le montre la figure 11.1 où le graphe de la fonction de répartition de Z_{45} suit parfaitement celui de $F_{\chi^2_2}(t) = 1 - \exp(-\frac{t}{2})$ pour $d = 3$ et $p_1 = p_2 = p_3 = \frac{1}{3}$.[1] Le test prend alors le nom de *test d'ajustement du chi-deux* et on peut l'effectuer à l'aide d'un simple stylo, pourvu que l'on dispose de *tables* donnant les valeurs utiles de $F_{\chi^2_{d-1}}$.

▷ *Exercice 223.* A l'issue d'une expérience de 1000 tirages, un générateur de chiffres aléatoires a donné les résultats suivants :

chiffre	1	2	3	4	5	6	7	8	9	0
nombre d'apparitions	87	103	90	110	81	108	85	123	90	123

Tester au niveau 5% l'hypothèse selon laquelle le générateur simule de façon satisfaisante un tirage uniforme sur les entiers $\{0, \ldots, 9\}$.

Solution. Notons X le résultat d'un tirage d'un entier entre 0 et 9 à l'aide de ce générateur et $p = (p_0, \ldots, p_9)$ sa loi. On cherche à tester l'hypothèse H_0 : "p est la loi $\mathcal{U}_{\{0,\ldots,9\}}$" contre l'hypothèse H_1 : "p n'est pas la loi $\mathcal{U}_{\{0,\ldots,9\}}$" au niveau 5%. Notons $N^{(i)}$ le nombre d'apparitions du chiffre i sur 1000 tirages de chiffres à l'aide de ce générateur et $Z = \sum_{i=0}^9 \frac{(N^{(i)} - 100)^2}{100}$.
On rejette l'hypothèse H_0 au niveau 5% si la valeur observée z_{obs} de Z est telle que $P_{H_0}(Z \ge z_{obs}) \le 0.05$. Comme pour tout $i \in \{0, \ldots, 9\}$, $1000p_i = 100$ est suffisamment grand, on peut approximer la fonction de répartition de la loi de Z sous H_0 par celle du χ^2 à 9 degrés de liberté. Donc, on peut approximer $P_{H_0}(Z \ge z_{obs})$ par $1 - F_{\chi^2_9}(z_{obs})$. A partir des valeurs observées pour les variables aléatoires $N^{(i)}$ qui sont données dans le tableau, on obtient $z_{obs} \simeq 21.86$. D'après la table de valeurs numériques de la loi du χ^2, $F_{\chi^2_9}(z_{obs})$ est compris entre 0.99 et 0.995. Donc, au niveau 5%, on rejette l'hypothèse H_0. On rejette encore l'hypothèse H_0 au niveau 1% mais pas au niveau 0.5%.

[1] Pour les besoins habituels de la statistique, on considére que l'approximation est valable dès que n dépasse 100 et que tous les effectifs théoriques moyens np_k dépassent 10.

▷ *Exercice 224.* A la suite de la formulation des lois de Mendel, Bateson a effectué des croisements avec des pois de senteur afin d'étudier la couleur (pourpre ou rouge) et la forme du pollen (allongée ou ronde). Ces croisements ont été fait sur des hybrides qui sont de couleur pourpre et ont un pollen de forme allongé ; en supposant que la couleur et la forme du pollen sont chacun controlés par un gène qui a deux formes différentes (appelées allèles) notées S et s pour la couleur et T, t pour la forme du pollen, le génotype d'un pois hybride pour ces deux caractères est (Ss, Tt). Les majuscules désignent les allèles dominants (ici la couleur pourpre et la forme allongée du pollen). Les résultats des croisements sont donnés dans le tableau ci-dessous :

	couleur pourpre	couleur rouge
pollen allongé	1528	117
pollen rond	106	381

En faisant l'hypothèse que les quatre possibilités de transmission des allèles pour chacun de ces deux gènes sont équiprobables, peut-on affirmer que les gènes contrôlant ces deux caractères sont transmis indépendamment l'un de l'autre ?
Comment s'assurer que l'hypothèse sur la transmission équiprobable des quatre possibilités de transmission des allèles est valide pour chacun de ces deux gènes ?

▷ *Exercice 225.* Le couvert végétal du domaine vital d'un élan d'Amérique se composait de peuplements feuillus ($25,8\%$ de la superficie du domaine vital), de peuplements mixtes (38% de la superficie), de peuplements résineux ($25,8\%$ de la superficie) et d'un marécage ($10,4\%$ de la superficie). Dans ce domaine, l'élan fut localisé à 511 reprises au cours de l'année. Sur 511 localisations, 118 se situaient dans les feuillus, 201 dans les peuplements mixtes, 110 dans les résineux et 82 dans le marécage[2].
L'élan fréquente-t-il indifféremment les quatre types de végétation de son domaine vital ?

11.4 Estimation de paramètres

En pratique, il arrive que l'hypothèse ne donne pas la valeur de certains paramètres de la loi. On est alors conduit à estimer la valeur de ces paramètres à l'aide des données observées. On peut montrer qu'alors le théorème 31 reste valable pourvu que l'*on diminue le nombre de degrés de liberté de la loi du* χ^2 *du nombre de paramètres estimés.* Nous ne démontrerons pas ce résultat mais remarquons tout de même le fait suivant : Si $d = 2$, la loi de X est une loi de Bernoulli ne dépendant que du choix d'un paramètre p_1. Si p_1 est estimé par $\frac{N_n^{(1)}}{n}$ (et donc $p_2 = 1 - p_1$ par $\frac{N_n^{(2)}}{n}$), la v.a. $\widehat{Z}_n$ obtenue en remplaçant dans Z_n les paramètres par leurs estimations est identiquement nulle. Sa loi ne peut en aucun cas être approchée par une densité du χ^2 à un degré de liberté.

Considérons un autre exemple moins simple. On cherche à confirmer, par un test effectué sur n tirages indépendants, une hypothèse d'indépendance faite sur un couple de variables aléatoires $X = (S, T)$. Soient $\{s_1, ..., s_{d_1}\}$ les valeurs possibles de S et $\{t_1, ..., t_{d_2}\}$ les valeurs possibles de T. On doit distinguer différentes situations :

- Si l'on connait a priori les lois de S et T, il s'agit de confirmer l'hypothèse selon laquelle la loi de X est donnée par le produit de ces deux lois. Si n est suffisamment grand, on pourra appliquer le test du chi-deux en utilisant la distribution du χ^2 à $(d_1 d_2 - 1)$ degrés de liberté (c'est la cas dans l'exercice 224). La variable de test s'écrit $Z_n = \sum_{k=1}^{d_1} \sum_{l=1}^{d_2} \frac{(N_n^{(k,l)} - np_k q_l)^2}{np_k q_l}$ où $N_n^{(k,l)}$ désigne le nombre d'indices i pour lesquels $X_i = (s_k, t_l)$, $p_k = P(S = s_k)$ et $q_l = P(T = t_l)$.

[2]Source : B. Scherrer "Biostatistique", éditeur gaetan morin, 1984, page 556

- Si on ne connait pas la loi de T, on estimera chacun des paramètres q_l par la fréquence $\widehat{q_l} = \sum_{k=1}^{d_1} \frac{N_n^{(k,l)}}{n}$ avec laquelle la valeur s_l apparaît dans l'échantillon. La loi de la variable de test $\widehat{Z}_n = \sum_{k=1}^{d_1} \sum_{l=1}^{d_2} \frac{(N_n^{(k,l)} - np_k\widehat{q_l})^2}{np_k\widehat{q_l}}$ sera alors approchée par la densité du χ^2 à $(d_1d_2 - d_2)$ degrés de liberté car $d_2 - 1$ paramètres ont été estimés[3]. On peut grosso modo justifier ce dernier point en observant que $\widehat{Z}_n$ s'écrit $\sum_{l=1}^{d_2} Z_n^{(l)}$ avec

$$Z_n^{(l)} = \sum_{k=1}^{d_1} \frac{(N_n^{(k,l)} - p_k\widehat{q_l})^2}{np_k\widehat{q_l}} = \sum_{k=1}^{d_1} \frac{(N_n^{(k,l)} - p_k\frac{n_l}{n})^2}{p_k n_l},$$

  où $n_l = \sum_k N_n^{(k,l)}$ désigne le nombre de tirages de la valeur t_l. Chaque variable $Z_n^{(l)}$ ne dépend que des valeurs de $S^{(i)}$ tirées lorsque $T^{(i)} = t_l$. Elles dépendent donc des résultats de tirages différents et comme S est supposée indépendante de T, elles sont indépendantes conditionnellement aux n_l. Quand n, et donc, avec une grande probabilité, quand les n_l deviennent grands, leur loi est approximativement donnée par la densité du χ^2 à $d_1 - 1$ degrés de liberté et tend donc à ne plus dépendre des n_l. Les variables aléatoires Z_l sont donc proches de l'indépendance et la loi de leur somme est donc approximativement donnée par une densité du χ^2 à $d_2(d_1 - 1)$ degrés de liberté.

- Si on ne connait ni la loi de T ni celle de S, on estimera les paramètres q_l mais aussi p_k. Il faudra employer la loi du χ^2 à $d_1d_2 - (d_1 + d_2 - 2) - 1 = (d_1 - 1)(d_2 - 1)$ degrés de liberté.

▷ *Exercice 226.* Sur un échantillon de 1000 personnes exposées à un certain facteur de risque, il y avait 645 hommes et 355 femmes. 115 hommes et 95 femmes apparaissent avoir développé un pathologie liée à ce facteur. Tester au niveau 5% l'hypothèse selon laquelle le risque est supérieur chez les femmes.

▷ *Exercice 227.* Mille pièces défectueuses ont été classées selon le type de défectuosité A, B, C ou D et la division où la pièce a été produite.

Peut-on considérer que le type de défectuosité est indépendant de la division de production au niveau 5% ?

	A	B	C	D
1	31	70	40	114
2	50	130	86	130
3	35	129	65	120

▷ *Exercice 228.*

1. On note F_d la fonction de répartition de la loi du χ^2 à d degrés de liberté. Montrer que pour tout $t > 0$, $F_d(t) > F_{d+1}(t)$.
2. On dispose de l'observation $(x_1, \ldots, x_{1000})$ d'un 1000-échantillon d'une variable aléatoire X à valeurs dans $\{0, \ldots, 4\}$. La moyenne $\bar{x} = \frac{1}{1000} \sum_{i=1}^{1000} x_i$ de cette échantillon est 2. Au niveau 5%, on a rejeté l'hypothèse H_0 : "X suit une loi binomiale" à l'aide d'un test du χ^2. La conclusion sera-t-elle la même si on remplace l'hypothèse H_0 par l'hypothèse H_0' : "X suit la loi $\mathcal{B}(4, 0.5)$" ?

[3] Il n'y a que $d_2 - 1$ paramètres estimés du fait que la somme des q_l est égale à 1

Appendix A

Rappels sur la manipulation des ensembles et sur la sommation

1.1 Notations sur les ensembles

A, B désignent des ensembles et $(A_i)_{i\in I}$ une famille d'ensembles indexée par l'ensemble I.

1.1.1 Sous-ensembles

- *"$A \subset B$" (A est un sous-ensemble de B)* :

$$A \subset B \text{ si et seulement si pour tout } x \in A,\ x \in B.$$

  N.B. $\emptyset$ est un sous-ensemble de n'importe quel ensemble.

- "$\mathcal{P}(A)$" : ensemble constitué de tous les sous-ensembles de A.
- "$A = B$" : $A \subset B$ et $B \subset A$.

▶ *Exemple 229.* $\mathcal{P}(\{1,2\})$ a 4 éléments qui sont $\emptyset$, $\{1\}$, $\{2\}$ et $\{1,2\}$.

1.1.2 Produit cartésien

- *"$A \times B$" (produit cartésien de l'ensemble A par l'ensemble B)* :

$$A \times B = \{(x, y), x \in A \text{ et } y \in B\}.$$

- "$\prod_{i=1}^{p} A_i$":$=A_1 \times \cdots \times A_p = \{(x_1, \ldots, x_p), x_i \in A_i \text{ pour tout } i \in \{1, \ldots, p\}\}$

  N.B. $(x_1, \ldots, x_p)$ est appelé un p-uple.

  N.B. Lorsque $A_1 = \ldots = A_p$, $A_1 \times ... \times A_p$ est noté A^p :

$$A^p = \{(x_1, \ldots, x_p),\ x_i \in A \text{ pour tout } i \in \{1, \ldots, p\}\}.$$

1.1.3 Intersection

- *"$A \cap B$" (intersection des ensembles A et B)* : ensemble des éléments qui sont dans A et dans B i.e.
$$A \cap B = \{x,\ x \in A \text{ et } x \in B\}.$$
- *"$x \in \cap_{i\in I} A_i$"* : pour tout $i \in I$, $x \in A_i$.
- *"A et B sont disjoints"* : leur intersection est vide : $A \cap B = \emptyset$.
- *"$\{A_i\}_{i\in I}$ est une famille d'ensembles disjoints"* : pour tout $i, j \in I$ tels que $i \neq j$, $A_i \cap A_j = \emptyset$.

1.1.4 Réunion

- *"$A \cup B$" (union des ensembles A et B)* : ensemble des éléments qui sont dans A ou dans B i.e.
$$A \cup B = \{x,\ x \in A \text{ ou } x \in B\}.$$
- *"$x \in \cup_{i\in I} A_i$"* : il existe au moins un indice $i \in I$ tel que $x \in A_i$.
- *"$\{A_i\}_{i\in I}$ est une partition de l'ensemble B"* : famille d'ensembles disjoints telle que $B = \cup_{i\in I} A_i$

► *Exemple 230.*

- Soit A un ensemble. La famille constituée des singletons de A est une partition de A, puisque $A = \cup_{x\in A}\{x\}$ et que $\{x\} \cap \{y\} = \emptyset$ si $x \neq y$.
- Soit A et B deux ensembles. La famille $\{\{x\} \times B\}_{x\in A}$ définit une partition de l'ensemble $A \times B$.

▷ *Exercice 231.* Pour $r \in \mathbb{N}$, on pose $E_r = \{(p, q) \in \mathbb{N}^2,\ p + q = r\}$. Montrer que $\{E_r\}_{r\in\mathbb{N}}$ définit une partition de $\mathbb{N}^2$.

1.1.5 Différence

- *"$A \setminus B$"* : ensemble des éléments de A qui ne sont pas dans B i.e.
$$A \setminus B = \{x \in A,\ x \notin B\}.$$
- Si B est un sous-ensemble de A alors $A \setminus B$ est appelé *le complémentaire de B dans A*. On le notera aussi en abrégé B^c s'il n'y a pas d'ambiguité.

▷ *Exercice 232.* Soient E, F et G trois sous-ensembles d'un ensemble Ω. Démontrer les relations suivantes :

a) $E \cap F \subset E \subset E \cup F$
b) si $E \subset F$ alors $F^c \subset E^c$
c) $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ et $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$
d) $E \cup F = E \cup (F \cup F \cap E^c)$
e) $F = (F \cap E) \cup (F \cap E^c)$
f) $(E \cap F)^c = E^c \cup F^c$ et $(E \cup F)^c = E^c \cap F^c$

Généraliser ces propriétés à des familles d'événements.

▷ *Exercice 233.* Soit $A_1, \ldots, A_n$ n ensembles. Construire une partition de l'ensemble $\cup_{i=1}^n A_i$ en n ensembles $B_1, \ldots, B_n$ tels que pour tout $1 \leq m \leq n$, $\cup_{i=1}^m B_i = \cup_{i=1}^m A_i$.

1.1.6 Image d'un ensemble par une application

Soit $f : E \to F$ une application.

- *"Image d'un sous-ensemble A de E par f"* : $f(A) = \{f(x), x \in A\}$.
- *"Image réciproque d'un sous-ensemble B de F par f"* : $f^{-1}(B) = \{x \in E,\ f(x) \in B\}$.

N.B. on utilise aussi la notation $\{f \in B\}$ pour $f^{-1}(B)$.

► *Exemple 234.* Soit $f : x \mapsto x^2$ définie sur $\mathbb{R}$. Alors $f^{-1}(\{a\}) = \{-\sqrt{a}, \sqrt{a}\}$ si $a \geq 0$ et $f^{-1}(\{a\}) = \emptyset$ si $a < 0$. L'image réciproque de $[0, 1[$ par f s'écrit $f^{-1}([0, 1[) = \{f < 1\} =]-1, 1[$.

▷ *Exercice 235.* Soit $f : E \to F$ une application.
Soit A et B deux sous-ensembles de E. Montrer les relations suivantes :

- $f(A \cup B) = f(A) \cup f(B)$,
- $f(A \cap B) \subset f(A) \cap f(B)$,
- $A \subset f^{-1}(f(A))$.

Soit A et B des sous-ensembles de F. Montrer les relations suivantes :

- $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$,
- $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$,
- $f(f^{-1}(A)) \subset A$,
- $f^{-1}(F \setminus A) = E \setminus f^{-1}(A)$.

▷ *Exercice 236.* Soit $f : E \to F$ une application entre deux ensembles. Montrer que $\{f^{-1}(\{y\})\}_{y \in F}$ définit une partition de l'ensemble E.

1.1.7 Fonction indicatrice d'un ensemble

Soit Ω un ensemble et A un sous-ensemble de Ω, on désigne par fonction indicatrice de A (relativement à Ω), notée $\mathbb{1}_A$ la fonction définie sur Ω par :

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

▷ *Exercice 237.*

a) Soient A et B deux sous-ensembles de Ω. Vérifier que $\mathbb{1}_A \leq \mathbb{1}_B$ si et seulement si $A \subset B$.

b) Exprimer les fonctions indicatrices du complémentaire de A dans Ω, de $A \cap B$, de $A \times B$ et de $A \cup B$ à l'aide des fonctions indicatrices de A et B.

c) Soit f une fonction définie sur un ensemble E à valeurs dans Ω. La fonction $\mathbb{1}_A \circ f$ est-elle la fonction indicatrice d'un ensemble ?

1.1.8 Suite d'ensembles

Soit $(A_n)_{n \in \mathbb{N}}$ une suite de sous-ensemble d'un ensemble Ω.

- *"(A_n) est une suite croissante d'ensembles"* : pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$.
- *"(A_n) est une suite décroissante d'ensembles"* : pour tout $n \in \mathbb{N}$, $A_{n+1} \subset A_n$.

► *Exemple 238.*

- $(]\frac{1}{n}, +\infty[)_n$ est une suite croissante d'ensembles.

- $(\cup_{k=0}^{n} A_k)_n$ est une suite croissante d'ensembles et $(\cup_{k=n}^{+\infty} A_k)_n$ est une suite décroissante d'ensembles.

▷ *Exercice 239.* Décrire $\cup_{n\in\mathbb{N}^*} A_n$ lorsque $A_n = [\frac{1}{n}, +\infty[$ pour tout $n \in \mathbb{N}^*$.

- "$\underline{\lim} A_n$" *(limite inférieure de la suite (A_n))* : ensemble des éléments de E qui sont dans tous les ensembles A_n sauf éventuellement dans un nombre fini d'entre eux

$$\underline{\lim} A_n = \bigcup_{n\in\mathbb{N}} (\bigcap_{k\geq n} A_k).$$

- "$\overline{\lim} A_n$" *(limite supérieure de la suite (A_n))* : ensemble des éléments de E qui sont dans une infinité d'ensembles de la suite (A_n)

$$\overline{\lim} A_n = \bigcap_{n\in\mathbb{N}} (\bigcup_{k\geq n} A_k).$$

▷ *Exercice 240.* Soit B et C deux ensembles. Pour tout $n \in \mathbb{N}$, on pose $A_{2n} = B$ et $A_{2n+1} = C$. Déterminer $\underline{\lim} A_n$ et $\overline{\lim} A_n$.

▷ *Exercice 241.* Soit $(A_n)_n$ et $(B_n)_n$ deux suites de sous-ensembles d'un ensemble Ω. Montrer que :

- $\underline{\lim} A_n \subset \overline{\lim} A_n$,
- $(\overline{\lim} A_n)^c = \underline{\lim} A_n^c$,
- $(\underline{\lim} A_n)^c = \overline{\lim} A_n^c$,
- $(\overline{\lim} A_n) \cup (\overline{\lim} B_n) = \overline{\lim}(A_n \cup B_n)$,
- $(\underline{\lim} A_n) \cap (\overline{\lim} B_n) \subset \overline{\lim}(A_n \cap B_n) \subset (\overline{\lim} A_n) \cap (\overline{\lim} B_n)$,
- $\mathbb{1}_{\overline{\lim} A_n} = \overline{\lim} \mathbb{1}_{A_n}$.

1.2 Ensembles finis

1.2.1 Cardinal d'un ensemble

Si E est un ensemble fini non vide, le nombre de ses éléments (c'est l'unique entier $p \in \mathbb{N}^*$ tel qu'on puisse trouver une bijection de $\{1, \ldots, p\}$ dans E) est appelé *le cardinal de E*.
Plusieurs notations sont utilisées pour désigner le cardinal d'un ensemble fini E : $\mathrm{Card}(E)$, $|E|$ ou encore $\#E$. Par convention, on pose $\mathrm{Card}(\emptyset) = 0$.
Si E est un ensemble fini de cardinal p, toute bijection $f : \{1, \ldots, p\} \to E$ définit une façon de numéroter les éléments de E. On a : $E = \{f(1), \ldots, f(p)\}$.
Pour dénombrer un ensemble fini E, il suffit de construire une bijection de E sur un autre ensemble F dont on connait le cardinal.

▷ *Exercice 242.* n joueurs de tennis participent à un tournoi par élimination. A chaque tour, le tirage des rencontres est fait au hasard ; si le nombre de joueurs est impair, l'un d'entre eux, choisi au hasard, ne participe pas à ce tour. Les gagnants de chaque tour et éventuellement le joueur qui n'y a pas participé jouent le tour suivant, jusqu'à ce que, finalement, il n'y ait plus qu'un gagnant.
Quel est le nombre de matches joués au total dans le tournoi ?

Propriétés. *Soit A et B deux ensembles finis.*

1. $A \cup B$ *est fini et* $\mathrm{Card}(A \cup B) = \mathrm{Card}(A) + \mathrm{Card}(B) - \mathrm{Card}(A \cap B)$,

2. $A \times B$ *est fini et* $\mathrm{Card}(A \times B) = \mathrm{Card}(A)\mathrm{Card}(B)$.

3. $\mathcal{P}(A)$ *est fini et* $\mathrm{Card}(\mathcal{P}(A)) = 2^{\mathrm{Card}(A)}$.

1.2.2 Somme d'une famille finie de nombres

Soit I un ensemble fini à $p \in \mathbb{N}^*$ éléments et $(x_i)_{i\in I}$ une famille de nombres indexés par l'ensemble I.
Pour additionner ces nombres, on choisit un ordre, c'est-à-dire une bijection $f : \{1, \ldots, p\} \to I$, mais le résultat $\sum_{j=1}^p x_{f(j)} = x_{f(1)} + \ldots + x_{f(p)}$ ne dépend pas de l'ordre dans lequel on a additionné ces nombres. On note la somme obtenue $\sum_{i\in I} x_i$:

$$\textit{si } f : \{1, \ldots, p\} \to I \textit{ est une bijection, alors } \sum_{i\in I} x_i = \sum_{j=1}^{p} x_{f(j)}.$$

Par convention, si $I = \emptyset$ alors $\sum_{i\in I} x_i = 0$.
L'associativité de l'addition permet de sommer par "paquets" :

$$\textit{Si } \{I_j\}_{j=1,\ldots,n} \textit{ est une partition de l'ensemble } I, \textit{ alors } \sum_{i\in I} x_i = \sum_{j=1}^{n} \sum_{i\in I_j} x_i.$$

▷ *Exercice 243.* On dispose d'un tableau qui donne le nombre a_i de jours où i voitures sont passées à un péage donné entre 7h et 8h du matin pendant le mois de septembre 2001 pour i allant de 0 au nombre maximum m de voitures observées. Donner une expression pour le nombre moyen de voitures passées à ce péage, entre 7h et 8h au mois de septembre 2001, en fonction des nombres a_i, $i \in \{0, \ldots, m\}$.

N.B. Souvent l'ensemble des indices I sur lequel on somme est remplacé, dans la notation $\sum_{i\in I} x_i$, par les conditions qui caractérisent les éléments de I. Par exemple, soit $(x_{i,j})_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}$ une famille de nombres indexés par $\mathbb{N}^* \times \mathbb{N}^*$,

- $\sum_{1\le i<j\le n} x_{i,j}$ est égale à $\sum_{l\in L} x_l$ avec $L = \{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*,\ i < j \le n\}$.
- $\sum_{i+j=n} x_{i,j}$ est égale à $\sum_{l\in L} x_l$ avec $L = \{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*,\ i + j = n\} = \{(i, n-i),\ i \in \{1, \ldots, n-1\}\}$.

▷ *Exercice 244.* Soit E et Ω deux ensembles finis tels que $E \subset \Omega$. Que vaut $\sum_{x\in\Omega} \mathbb{1}_E(x)$?

▷ *Exercice 245.* Soit f une application définie sur un ensemble E à valeurs dans un ensemble fini F. Vérifier que $f = \sum_{a\in F} a\mathbb{1}_{f^{-1}(\{a\})}$.

▷ *Exercice 246.* Soit I et J deux ensembles finis et $\{x_{i,j}\}_{(i,j)\in I\times J}$ une famille de nombres indexés par $I \times J$. Montrer que

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

▷ *Exercice 247.* Soit n et m deux entiers naturels tels que $m \le n$.
Soit $(x_{i,j})_{\substack{i=0,\ldots,n\\ j=0,\ldots,n}}$ une famille de nombres réels. Intervertir les doubles sommes suivantes :

$$S_1 = \sum_{i=0}^{n}\sum_{j=i}^{n} x_{i,j}, \quad S_2 = \sum_{i=0}^{n}\sum_{j=0}^{n-i} x_{i,j} \quad \text{et} \quad S_3 = \sum_{i=0}^{m}\sum_{j=i}^{n} x_{i,j}$$

▷ *Exercice 248.*

a) Soient $n \ge 2$ et $b_i,\ i \in \{1, \ldots, n\}$ des réels. Montrer que

$$\prod_{i=1}^{n}(1 + b_i) = 1 + \sum_{k=1,}^{n} \sum_{1\le i_1<\ldots<i_k\le n} b_{i_1} \ldots b_{i_k}.$$

b) Soient $A_i,\ i \in \{1, \dots, n\}$ des ensembles finis. Etablir la formule du crible :

$$\mathbb{1}_{\overset{n}{\underset{i=1}{\cup}} A_i} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{1}_{A_{i_1} \cap \dots \cap A_{i_k}}.$$

c) En déduire l'égalité suivante :

$$\mathrm{Card}(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathrm{Card}(A_{i_1} \cap \dots \cap A_{i_k}). \qquad \text{(A.1)}$$

d) *Application :* Une séquence d'ADN est un enchaînement de nucléotides dont il existe 4 types qui diffèrent par leur base : A (adénine), C (cytosine), G (guanine) et T (thymine).
Si on suppose que tous les agencements de ces 4 types de nucléotides sont possibles, combien y a-t-il de séquences différentes d'ADN constituées de $p \ge 4$ nucléotides ?
Parmi ces séquences, combien contiennent les 4 types de nucléotides ?

1.2.3 Analyse combinatoire

▷ *Exercice 249.* Soit f une application d'un ensemble X dans un ensemble Y. On suppose que :

- f est surjective,
- Y est un ensemble à q éléments,
- pour tout $y \in Y$, l'ensemble $f^{-1}(\{y\}) = \{x \in X, f(x) = y\}$ a p éléments.

Montrer que X a alors pq éléments.

Définition. On appelle *p-arrangement d'un ensemble E*, un p-uple $(x_1, \dots, x_p)$ de p éléments de E tous distincts.
Un p-arrangement d'un ensemble E à p élément est aussi appelé une *permutation de E*.

▷ *Exercice 250.*

a) Si E est un ensemble à n éléments, montrer que le nombre de p-arrangements de E est

$$A_n^p := n(n-1)\dots(n-p+1) = \frac{n!}{(n-p)!}.$$

Indication : on pourra faire une récurrence sur p.

b) Combien y a-t-il de nombres constitués de 4 chiffres distincts ?

N.B. Lorsque n tend vers $+\infty$, $n! \sim (\frac{n}{e})^n \sqrt{2\pi n}$ (*Formule de Stirling*)

Définition. On appelle *p-combinaison d'un ensemble E*, un sous-ensemble de E à p éléments.

▷ *Exercice 251.*

a) Si E est un ensemble à n éléments, montrer que le nombre de p-combinaisons de E est

$$C_n^p := \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}.$$

Indication : on pourra appliquer le résultat de l'exercice 249.

b) Combien peut-on constituer de n-uples différents contenant k fois le symbole 1 et $n-k$ fois le symbole 0 ?
Indication : définir une bijection entre l'ensemble à dénombrer et l'ensemble des k-combinaisons de $\{1, \dots, n\}$.

▷ *Exercice 252.* De combien de façons peut-on répartir n objets distincts en r groupes de tailles respectives $n_1, \dots, n_r$ avec $\sum_{i=1}^r n_i = n$?
Indication : on pourra raisonner par récurrence sur r.

▷ *Exercice 253.*

a) Soient $n \in \mathbb{N}^*$, a et b deux réels. Etablir la formule du binôme :

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}.$$

b) Utiliser la formule du binôme pour montrer que si E un ensemble fini, $\mathrm{Card}(\mathcal{P}(E)) = 2^{\mathrm{Card}(E)}$.

c) Montrer que pour tout $p, q \in \mathbb{N}$ et $k \in \{0, \dots, p+q\}$, $C_{p+q}^k = \sum_{i=0}^k C_p^i C_q^{k-i}$.

▷ *Exercice 254.* Soit Ω un ensemble fini de cardinal n. Calculer $\sum_{A \in \mathcal{P}(\Omega)} \mathrm{Card}(A)$.

▷ *Exercice 255.* Soit $n \in \mathbb{N}^*$ et $k \in \{1, \dots, n\}$.

1. Montrer que l'ensemble

$$\mathcal{E}_{n,k} = \{(u_1, \dots, u_k) \in \{1, \dots, n\}^k, \text{ tel que } u_1 < \dots < u_k\}$$

est en bijection avec l'ensemble des parties à k éléments de $\{1, \dots, n\}$. En déduire le cardinal de $\mathcal{E}_{n,k}$.

2. Notons $\mathcal{D}_{n,k}$ l'ensemble défini par

$$\mathcal{D}_{n,k} = \{(u_1, \dots, u_k) \in \{1, \dots, n\}^k \text{ tel que } u_1 \leq \dots \leq u_k\}.$$

On considère l'application f définie sur l'ensemble $\mathcal{D}_{n,k}$ par :

$$f((u_1, \dots, u_k)) = (v_1, \dots, v_k) \text{ avec pour tout } j \in \{1, \dots, k\},\ v_j = u_j + j - 1.$$

Montrer que f définit une bijection de $\mathcal{D}_{n,k}$ sur un ensemble que l'on déterminera. En déduire le cardinal de $\mathcal{D}_{n,k}$.

▷ *Exercice 256.* Soit $n \in \mathbb{N}^*$ et $k \in \{1, \dots, n\}$. L'objectif de cet exercice est de déterminer le cardinal de l'ensemble $\mathcal{X}_{n,k}$ suivant :

$$\mathcal{X}_{n,k} = \{(r_1, \dots, r_k) \in \mathbb{N}^k,\ r_1 + \dots + r_k = n\}.$$

On considère l'application f définie sur $\mathcal{X}_{n,k}$ par

$$f(r_1, \dots, r_k) = (\underbrace{1, \dots, 1}_{r_1\ fois}, 0, \underbrace{1, \dots, 1}_{r_2\ fois}, \dots, 0, \underbrace{1, \dots, 1}_{r_k\ fois}).$$

Montrer que f est une bijection de $\mathcal{X}_{n,k}$ sur un ensemble que l'on déterminera. En déduire le cardinal de $\mathcal{X}_{n,k}$.

▷ *Exercice 257.* Un distributeur de prospectus dispose de p prospectus identiques qu'il distribue au hasard dans b boites aux lettres. Décrire l'ensemble des résultats possibles pour une telle distribution. Quel est le cardinal de cet ensemble ?

▷ *Exercice 258.* Développer $(x_1 + \dots + x_r)^n$. Combien de termes le développement contient-il ?

▷ *Exercice 259.* Soit $(p, n) \in \mathbb{N}^2$ avec $p \leq n$. Montrer que

$$\sum_{i=p}^{n} C_{n+1}^{i+1} = \sum_{i=p}^{n} C_i^p 2^{n-i}$$

en dénombrant de deux façons différentes l'ensemble

$$\Lambda_{p,n} = \{(x_1, \dots, x_{n+1}) \in \{0,1\}^{n+1},\ \sum_{i=1}^{n+1} x_i \geq p+1\}$$

Indications : on pourra introduire, pour $i \in \{p, \dots, n\}$, le sous-ensemble F_i défini par

$$F_i = \{(x_1, \dots, x_{n+1}) \in \Lambda_{p,n},\ x_{i+1} = 1 \text{ et } \sum_{j=1}^{i} x_j = p\}.$$

1.3 Somme d'une famille dénombrable de nombres

1.3.1 Ensembles dénombrable

Un ensemble I est dit *dénombrable* si il existe une bijection de $\mathbb{N}^*$ sur I.
Un ensemble I est dit *au plus dénombrable* si I est un ensemble fini ou dénombrable.

► *Exemple 260.*

- $\mathbb{N}$, $2\mathbb{N}$, $\mathbb{Z}$, $\mathbb{N} \times \mathbb{N}$ sont dénombrables.
- $\mathbb{Q}$ est dénombrable.
- Une réunion au plus dénombrable d'ensembles au plus dénombrables est encore un ensemble au plus dénombrable.

Soit I un ensemble dénombrable et $(x_i)_{i\in I}$ une famille de nombres indexée par I. Pour sommer les valeurs de cette famille, on choisit l'ordre dans lequel on les additionne, c'est-à-dire une bijection $\phi : \mathbb{N}^* \to I$. Mais si on additionne les nombres x_i dans un autre ordre, la somme obtenue est-elle la même ? L'exemple suivant montre que ce n'est pas toujours le cas.

► *Exemple 261.* Soit $(u_n)_{n\in\mathbb{N}^*}$ la famille de réels définie par : $u_n = \frac{(-1)}{n}^{n-1}$ pour tout $n \in \mathbb{N}^*$.

1. La suite des sommes partielles $S_n = \sum\limits_{k=1}^{n} u_k$ converge vers $\ln(2)$.
2. Soit $\phi : \mathbb{N}^* \to \mathbb{N}^*$ l'application définie par $\phi(3k-2) = 2k-1$, $\phi(3k-1) = 4k-2$ et $\phi(3k) = 4k$ pour tout $k \in \mathbb{N}^*$. C'est une bijection et la suite des sommes partielles $\sum\limits_{k=1}^{n} u_{\phi(k)}$ converge vers $\frac{\ln(2)}{2}$

Indications. 1. Utiliser que $\sum\limits_{k=1}^{n} \frac{1}{k} = \ln(n) + \gamma + \epsilon_n$, où γ est un réel appelé la constante d'Euler et où (ϵ_n) est une suite qui tend vers 0, et considérer la suite (S_{2n}).
2. Soit $w_k = u_{\phi(3k-2)} + u_{\phi(3k-1)} + u_{\phi(3k)}$ pour $k \in \mathbb{N}^*$. Utiliser que $w_k = \frac{1}{2}(u_{2k-1} + u_{2k})$ pour tout $k \in \mathbb{N}^*$.

On va voir cependant que dans les deux cas suivants :

1. pour tout $i \in I$, $x_i \geq 0$,
2. $\sum_{k=1}^{\infty} x_{\phi(k)}$ est une série absolument convergente,

la somme obtenue en additionnant les nombres x_i dans l'ordre donné par ϕ, ne dépend pas de l'ordre choisi. On pourra alors poser $\sum_{i\in I} x_i = \sum_{k=1}^{\infty} x_{\phi(k)}$.

1.3.2 Familles à termes positifs

Supposons que pour tout $i \in I$, $x_i \geq 0$.
Soit $\phi : \mathbb{N}^* \to I$, une bijection. La suite $\left(\sum_{k=1}^{n} x_{\phi(k)}\right)_n$ est croissante donc elle tend vers un nombre $S \in \mathbb{R}_+$ ou vers $+\infty$. Dans le deuxième cas, on posera $S = +\infty$.

Règles de calculs dans $\mathbb{R}_+ \cup \{+\infty\}$: pour tout $x \in \mathbb{R}_+$, $x + (+\infty) = +\infty$, $(+\infty) + (+\infty) = +\infty$, $x \times (+\infty) = +\infty$ si $x > 0$.

Le théorème suivant montre que la valeur de S ne dépend pas du choix de la bijection ϕ, on notera $S = \sum_{i\in I} x_i$:

Théorème 32 *Soit $(u_i)_{i\in\mathbb{N}^*}$ une suite de réels positifs ou nuls. Il existe un nombre $S \in \mathbb{R}_+ \cup \{+\infty\}$ tel que pour toute bijection $\phi : \mathbb{N}^* \to \mathbb{N}$, $S = \sum_{k=1}^{\infty} u_{\phi(k)}$.*

Preuve : Soit ϕ une bijections de $\mathbb{N}^*$ sur $\mathbb{N}^*$. Posons S la limite des sommes partielles $S_n := \sum_{k=1}^n u_k$ et Σ la limite des sommes partielles $\Sigma_n := \sum_{k=1}^n u_{\phi(k)}$. Pour tout $n \in \mathbb{N}^*$, posons $\rho_n = \max(\phi(k),\ k = 1, \dots, n)$. Alors pour tout $n \in \mathbb{N}$, $\Sigma_n \leq S_{\rho_n}$. Donc, $\Sigma \leq S$. En procédant de la même façon avec la suite $v_k := u_{\phi(k)}$ et l'application ϕ^{-1}, on obtient que pour tout $n \in \mathbb{N}^*$, $S_n \leq \Sigma_{r_n}$ où $r_n = \max(\phi^{-1}(k), k = 1, \dots, n)$. Donc $S = \Sigma$. □

▷ *Exercice 262.* Soit $(x_i)_{i\in I}$ une famille de réels indexée par un ensemble I au plus dénombrable et f une bijection de I sur un ensemble J.
Montrer que $\sum_{i\in I} x_i = \sum_{j\in J} x_{f^{-1}(j)}$.

Pour calculer la somme, on peut regrouper les termes par paquets :

Théorème 33 *Soit $(x_i)_{i\in I}$ une famille de réels positifs ou nuls indexée par un ensemble dénombrable I. Soit $(I_k)_{k\in K}$ une partition de I. Posons $S_k = \sum_{i\in I_k} x_i$ pour tout $k \in K$. On a :*

$$\sum_{i\in I} x_i = \sum_{k\in K} \left(\sum_{i\in I_k} x_i \right).$$

Preuve :

- Démontrons le théorème lorsque $I = \mathbb{N}^*$ et $K = \mathbb{N}^*$. Pour tout $k \in \mathbb{N}^*$, il existe $c_k \in \mathbb{N} \cup \{+\infty\}$ et une bijection $\phi_k : \{1, \dots, c_k\} \to I_k$. Comme $(I_k)_{k\in\mathbb{N}^*}$ forme une partition de $\mathbb{N}^*$, $(\{\phi_k(l)\})_{(k,l)\in\mathbb{N}^*\times\mathbb{N}^*}$ constitue aussi une partition de $\mathbb{N}^*$.
  Pour $K, L \in \mathbb{N}^*$, posons $M_{K,L} = \max(\phi_k(l), k \in \{1, \dots, K\}$ et $l \in \{1, \dots, L\}\}$. On a :
  $$\sum_{k=1}^{K} \sum_{l=1}^{L} x_{\phi_k(l)} \leq \sum_{i=1}^{M_{K,L}} x_i, \text{ pour tout } (K, L) \in \mathbb{N}^* \times \mathbb{N}^*.$$
  Donc en faisant tendre L vers $+\infty$ puis K vers $+\infty$, on obtient $\sum_{k=1}^\infty S_k \leq \sum_{i=1}^\infty x_i$. Pour établir l'inégalité inverse, on considère les entiers
  $$K_N = \max(\{k, \exists n \leq N,\ n \in I_k\}) \text{ pour } N \in \mathbb{N}^*.$$
  Comme pour tout $N \in \mathbb{N}^*$, $\{1, \dots, N\} \subset \cup_{k=1}^{K_N} I_k$, on a $\sum_{i=1}^\infty x_i \leq \sum_{k=1}^{K_N} S_k$. Donc, $\sum_{i=1}^\infty x_i \leq \sum_{k=1}^\infty S_k$.
- Dans le cas général, on se ramène au cas précédent en utilisant les bijections : comme I est dénombrable, il existe une bijection $f : \mathbb{N}^* \to I$. Quitte à rajouter des ensembles I_k vides, on peut supposer K dénombrable : il existe une bijection $g : \mathbb{N}^* \to K$. Posons $u_j = x_{f(j)}$ pour tout $j \in \mathbb{N}^*$. On a $\sum_{i\in I} x_i = \sum_{j=1}^{+\infty} u_j$ et $S_k = \sum_{i\in f^{-1}(I_k)} u_i$ pour tout $k \in K$. On obtient le théorème en appliquant le résultat à la suite de réels positifs $(u_i)_{i\in\mathbb{N}^*}$ et à la partition de $\mathbb{N}^*$ définie par la famille d'ensembles $(f^{-1}(I_{g(l)}))_{l\in\mathbb{N}^*}$.

□

▷ *Exercice 263.* Soit $(x_{i,j})_{(i,j)\in\mathbb{N}^2}$ une famille de réels positifs. Montrer que

$$\sum_{(i,j)\in\mathbb{N}^2} x_{i,j} = \sum_{i=0}^{\infty} \left(\sum_{j=0}^{\infty} x_{i,j} \right) = \sum_{j=0}^{\infty} \left(\sum_{i=0}^{\infty} x_{i,j} \right) = \sum_{k=0}^{\infty} \left(\sum_{i+j=k}^{\infty} x_{i,j} \right).$$

1.3.3 Familles sommables

Soit $(x_i)_{i\in I}$ une famille de nombres réels indexée par un ensemble dénombrable I.
La famille $(x_i)_{i\in I}$ est dite sommable si $\sum_{i\in I}|x_i| < +\infty$.
Le théorème suivant montre que pour une famille sommable $(x_i)_{i\in I}$, l'ordre dans lequel on additionne les nombres x_i de la famille n'a pas d'importance, on note alors la somme obtenue $\sum_{i\in I} x_i$:

Théorème 34 *Soit $(x_i)_{i\in I}$ une famille sommable. Il existe un réel S tel que pour toute bijection $\phi : \mathbb{N}^* \to I$, $S = \sum_{k=1}^{\infty} u_{\phi(k)}$.*

Preuve : La preuve consiste à appliquer les résultats pour les sommes à termes positifs aux familles $(x_i^+)_{i\in I}$ et $(x_i^-)_{i\in I}$. Rappelons que pour $x \in \mathbb{R}$, $x^+ = \max(x, 0)$ (partie positive de x) et $x^- = \max(-x, 0)$ (partie négative de x). On a les égalités suivantes : $x = x^+ - x^-$ et $|x| = x^+ + x^-$.
Comme la famille $(x_i)_{i\in I}$ est sommable, les familles $(x_i^+)_{i\in I}$ et $(x_i^-)_{i\in I}$ ont des sommes finies que l'on notera respectivement S^+ et S^-. Pour toute bijection ϕ de $\mathbb{N}^*$ dans I, $S^+ = \sum_{i=1}^{+\infty} x_{\phi(i)}^+$ et $S^- = \sum_{i=1}^{+\infty} x_{\phi(i)}^-$. Donc, la série $\sum_{i=1}^{+\infty} x_{\phi(i)}$ converge et a pour somme $S = S^+ - S^-$. □

Pour une famille sommable, on peut aussi additionner les termes par paquet :

Théorème 35 *Soit $(x_i)_{i\in I}$ une famille de réels indexés par un ensemble dénombrable I. Soit $(I_k)_{k\in K}$ une partition de I.*

1. *La famille $(x_i)_{i\in I}$ est sommable si et seulement si les deux conditions suivantes sont réalisées :*
   - *pour tout $k \in K$, la famille $(x_i)_{i\in I_k}$ est sommable.*
   - *la famille $(\sum_{i\in I_k}|x_i|)_{k\in K}$ est sommable.*
2. *Si $(x_i)_{i\in I}$ est sommable, alors $\sum_{i\in I} x_i = \sum_{k\in K}\left(\sum_{i\in I_k} x_i\right)$.*

Preuve :

1. D'après le théorème 32, on a $\sum_{i\in I}|x_i| = \sum_{k\in K}\left(\sum_{i\in I_k}|x_i|\right)$. Cela donne la condition nécessaire et suffisante pour que $(x_i)_{i\in I}$ soit sommable.

2. Toujours d'après le théorème 32,

$$\sum_{i\in I} x_i^+ = \sum_{k\in K}\left(\sum_{i\in I_k} x_i^+\right) \text{ et } \sum_{i\in I} x_i^- = \sum_{k\in K}\left(\sum_{i\in I_k} x_i^-\right).$$

   Si $(x_i)_{i\in I}$ est sommable, ces sommes sont finies, on peut donc soustraire les deux égalités, ce qui donne :

$$\sum_{i\in I} x_i = \sum_{k\in K}\left(\sum_{i\in I_k} x_i\right).$$

□

En appliquant le théorème aux familles indexées par un ensemble produit $I \times J$, on obtient un théorème de type "Fubini" pour les familles sommables.

Corollaire 36 *Soit $(x_{i,j})_{(i,j)\in I\times J}$ une famille dénombrable de réels.*
Si $\sum_{i\in I}\left(\sum_{j\in J}|x_{i,j}|\right) < \infty$, alors la famille $(x_{i,j})$ est sommable et

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

▷ *Exercice 264.* Etablir le corollaire 36.

▷ *Exercice 265.* Soit u et v deux nombres réels tels que $-1 < u < 1$ et $-1 < v < 1$. Montrer que la famille $(u^k v^l)_{(k,l)\in\mathbb{N}^2}$ est sommable et calculer sa somme.

▷ *Exercice 266.* Déterminer l'ensemble des réels x pour lesquels la famille $(\frac{(p+q)!}{p!q!}x^{p+q})_{(p,q)\in\mathbb{N}^2}$ est sommable. Pour ces valeurs, calculer la somme.

▷ *Exercice 267.* Montrer que la famille $\left(\left(\frac{-1}{p}\right)^q\right)_{p\geq 2;\ q\geq 2}$ est sommable, mais que $\left(\left(\frac{-1}{p}\right)^q\right)_{p\geq 1;\ q\geq 2}$ n'est pas sommable.

▷ *Exercice 268.* Considérons la famille $\left(x_{p,q} = \frac{(-p)^q}{q!}e^{-p}u_p\right)_{p\in\mathbb{N};\ q\in\mathbb{N}}$, où $(u_p)_{p\in\mathbb{N}}$ est une suite de nombres réels.

1. Montrer que la famille $(x_{p,q})_{(p,q)\in\mathbb{N}^2}$ est sommable si et seulement si la famille $(u_p)_{p\in\mathbb{N}}$ est sommable.
2. Dans le cas où $u_p = x^p$, avec $x \in]-1, 1[$, calculer $\sum_{(p,q)\in\mathbb{N}^2} x_{p,q}$.

▷ *Exercice 269.* Montrer que pour tout $x \in \mathbb{R}$, la famille $(\frac{x^n}{n!})_n$ est sommable. Soit $f(x)$ sa somme. Etablir que pour tout $x, y \in \mathbb{R}$, $f(x+y) = f(x)f(y)$.

▷ *Exercice 270.* Soit $(a_k)_{k\in\mathbb{N}}$ une famille de réels telle que la famille $(ka_k)_{k\in\mathbb{N}}$ soit sommable. Montrer que

$$\sum_{k=0}^{+\infty} ka_k = \sum_{i=0}^{+\infty}\sum_{j=i+1}^{+\infty} a_j.$$

Appendix B

Tables de valeurs numériques

Loi binomiale

Le tableau donne la fonction de répartition de la loi $\mathcal{B}(n, 1/2)$ en des valeurs entières k, et ceci pour différentes valeurs de n :

$k \setminus n$	5	6	7	8	9	10	11	12	15	20	30
0	0.0312	0.0156	0.0078	0.0039	0.0020	0.0010	0.0005	0.0002	0.0000	0.0000	0.0000
1	0.1875	0.1094	0.0625	0.0352	0.0195	0.0107	0.0059	0.0032	0.0005	0.0000	0.0000
2	0.5000	0.3438	0.2266	0.1445	0.0898	0.0547	0.0327	0.0193	0.0037	0.0002	0.0000
3	0.8125	0.6563	0.5000	0.3633	0.2539	0.1719	0.1133	0.0730	0.0176	0.0013	0.0000
4	0.9687	0.8906	0.7734	0.6367	0.5000	0.3770	0.2744	0.1938	0.0592	0.0059	0.0000
5	1.0000	0.9844	0.9375	0.8555	0.7461	0.6230	0.5000	0.3872	0.1509	0.0207	0.0002
6	1.0000	1.0000	0.9922	0.9648	0.9102	0.8281	0.7256	0.6128	0.3036	0.0577	0.0007
7	1.0000	1.0000	1.0000	0.9961	0.9805	0.9453	0.8867	0.8062	0.5000	0.1316	0.0026
8	1.0000	1.0000	1.0000	1.0000	0.9980	0.9893	0.9673	0.9270	0.6964	0.2517	0.0081
9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9990	0.9941	0.9807	0.8491	0.4119	0.0214
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9968	0.9408	0.5881	0.0494
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9824	0.7483	0.1002
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9963	0.8684	0.1808
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9423	0.2923
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9793	0.4278
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9941	0.5722
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.7077
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.8192
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8998
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9506
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9786
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9919
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9974
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998

Les tableaux suivants donnent la fonction de répartition de la loi $\mathcal{B}(n, p)$ en des valeurs entières k et ceci pour différentes valeurs de p et de n :

$n = 10$											
$k \setminus p$	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0	0.9511	0.9044	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025
1	0.9989	0.9957	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233
2	1.0000	0.9999	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996
3	1.0000	1.0000	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660
4	1.0000	1.0000	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044
5	1.0000	1.0000	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384
6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980
7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997

$n = 12$											
$k \setminus p$	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0	0.9416	0.8864	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008
1	0.9984	0.9938	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083
2	1.0000	0.9998	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421
3	1.0000	1.0000	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345
4	1.0000	1.0000	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044
5	1.0000	1.0000	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269
6	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393
7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$n = 15$											
$k \setminus p$	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0	0.9276	0.8601	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001
1	0.9975	0.9904	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017
2	0.9999	0.9996	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107
3	1.0000	1.0000	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424
4	1.0000	1.0000	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204
5	1.0000	1.0000	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608
6	1.0000	1.0000	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522
7	1.0000	1.0000	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535
8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182
9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

$n = 20$											
$k \setminus p$	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0	0.9046	0.8179	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000
1	0.9955	0.9831	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001
2	0.9999	0.9990	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009
3	1.0000	1.0000	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049
4	1.0000	1.0000	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189
5	1.0000	1.0000	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553
6	1.0000	1.0000	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299
7	1.0000	1.0000	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520
8	1.0000	1.0000	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143
9	1.0000	1.0000	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914
10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997

						$n = 30$						
$k \setminus p$	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.5
0	0.8604	0.7397	0.2146	0.0424	0.0076	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9901	0.9639	0.5535	0.1837	0.0480	0.0105	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000
2	0.9995	0.9967	0.8122	0.4114	0.1514	0.0442	0.0106	0.0021	0.0003	0.0000	0.0000	0.0000
3	1.0000	0.9998	0.9392	0.6474	0.3217	0.1227	0.0374	0.0093	0.0019	0.0003	0.0000	0.0000
4	1.0000	1.0000	0.9844	0.8245	0.5245	0.2552	0.0979	0.0302	0.0075	0.0015	0.0002	0.0000
5	1.0000	1.0000	0.9967	0.9268	0.7106	0.4275	0.2026	0.0766	0.0233	0.0057	0.0011	0.0002
6	1.0000	1.0000	0.9994	0.9742	0.8474	0.6070	0.3481	0.1595	0.0586	0.0172	0.0040	0.0007
7	1.0000	1.0000	0.9999	0.9922	0.9302	0.7608	0.5143	0.2814	0.1238	0.0435	0.0121	0.0026
8	1.0000	1.0000	1.0000	0.9980	0.9722	0.8713	0.6736	0.4315	0.2247	0.0940	0.0312	0.0081
9	1.0000	1.0000	1.0000	0.9995	0.9903	0.9389	0.8034	0.5888	0.3575	0.1763	0.0694	0.0214
10	1.0000	1.0000	1.0000	0.9999	0.9971	0.9744	0.8943	0.7304	0.5078	0.2915	0.1350	0.0494
11	1.0000	1.0000	1.0000	1.0000	0.9992	0.9905	0.9493	0.8407	0.6548	0.4311	0.2327	0.1002
12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9969	0.9784	0.9155	0.7802	0.5785	0.3592	0.1808
13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9918	0.9599	0.8737	0.7145	0.5025	0.2923
14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9973	0.9831	0.9348	0.8246	0.6448	0.4278
15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9936	0.9699	0.9029	0.7691	0.5722
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9979	0.9876	0.9519	0.8644	0.7077
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9955	0.9788	0.9286	0.8192
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9917	0.9666	0.8998
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9971	0.9862	0.9506
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9950	0.9786
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9919
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9974
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993

	$n = 50$											
$k \setminus p$	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.5
0	0.7783	0.6050	0.0769	0.0052	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9739	0.9106	0.2794	0.0338	0.0029	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9979	0.9862	0.5405	0.1117	0.0142	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.9999	0.9984	0.7604	0.2503	0.0460	0.0057	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
4	1.0000	0.9999	0.8964	0.4312	0.1121	0.0185	0.0021	0.0002	0.0000	0.0000	0.0000	0.0000
5	1.0000	1.0000	0.9622	0.6161	0.2194	0.0480	0.0070	0.0007	0.0001	0.0000	0.0000	0.0000
6	1.0000	1.0000	0.9882	0.7702	0.3613	0.1034	0.0194	0.0025	0.0002	0.0000	0.0000	0.0000
7	1.0000	1.0000	0.9968	0.8779	0.5188	0.1904	0.0453	0.0073	0.0008	0.0001	0.0000	0.0000
8	1.0000	1.0000	0.9992	0.9421	0.6681	0.3073	0.0916	0.0183	0.0025	0.0002	0.0000	0.0000
9	1.0000	1.0000	0.9998	0.9755	0.7911	0.4437	0.1637	0.0402	0.0067	0.0008	0.0001	0.0000
10	1.0000	1.0000	1.0000	0.9906	0.8801	0.5836	0.2622	0.0789	0.0160	0.0022	0.0002	0.0000
11	1.0000	1.0000	1.0000	0.9968	0.9372	0.7107	0.3816	0.1390	0.0342	0.0057	0.0006	0.0000
12	1.0000	1.0000	1.0000	0.9990	0.9699	0.8139	0.5110	0.2229	0.0661	0.0133	0.0018	0.0002
13	1.0000	1.0000	1.0000	0.9997	0.9868	0.8894	0.6370	0.3279	0.1163	0.0280	0.0045	0.0005
14	1.0000	1.0000	1.0000	0.9999	0.9947	0.9393	0.7481	0.4468	0.1878	0.0540	0.0104	0.0013
15	1.0000	1.0000	1.0000	1.0000	0.9981	0.9692	0.8369	0.5692	0.2801	0.0955	0.0220	0.0033
16	1.0000	1.0000	1.0000	1.0000	0.9993	0.9856	0.9017	0.6839	0.3889	0.1561	0.0427	0.0077
17	1.0000	1.0000	1.0000	1.0000	0.9998	0.9937	0.9449	0.7822	0.5060	0.2369	0.0765	0.0164
18	1.0000	1.0000	1.0000	1.0000	0.9999	0.9975	0.9713	0.8594	0.6216	0.3356	0.1273	0.0325
19	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9861	0.9152	0.7264	0.4465	0.1974	0.0595
20	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9937	0.9522	0.8139	0.5610	0.2862	0.1013
21	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9974	0.9749	0.8813	0.6701	0.3900	0.1611
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9990	0.9877	0.9290	0.7660	0.5019	0.2399
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9944	0.9604	0.8438	0.6134	0.3359
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9976	0.9793	0.9022	0.7160	0.4439
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9900	0.9427	0.8034	0.5561
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9955	0.9686	0.8721	0.6641
27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9981	0.9840	0.9220	0.7601
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9924	0.9556	0.8389
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9966	0.9765	0.8987
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9986	0.9884	0.9405
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9947	0.9675
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9978	0.9836
33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9923
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9967
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9987
36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998

Loi de Poisson

Les tableaux donnent quelques valeurs de la fonction de répartition F_λ, d'une loi de Poisson de paramètre λ, pour différentes valeurs de λ : $F_\lambda(k) = \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^i}{i!}$ pour tout $k \in \mathbb{N}$.

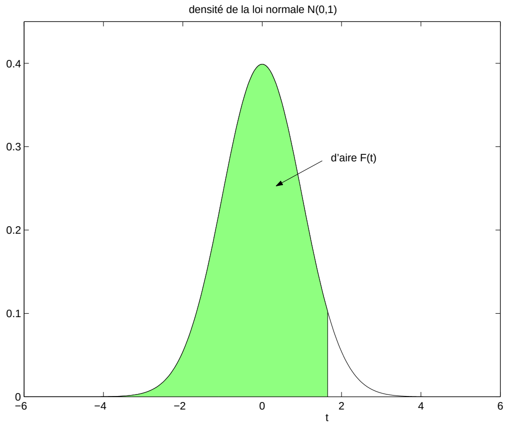
$k \setminus \lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

$k \setminus \lambda$	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5
0	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067	0.0041	0.0025	0.0015
1	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404	0.0266	0.0174	0.0113
2	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247	0.0884	0.0620	0.0430
3	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650	0.2017	0.1512	0.1118
4	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405	0.3575	0.2851	0.2237
5	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160	0.5289	0.4457	0.3690
6	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622	0.6860	0.6063	0.5265
7	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666	0.8095	0.7440	0.6728
8	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319	0.8944	0.8472	0.7916
9	1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682	0.9462	0.9161	0.8774
10	1.0000	0.9999	0.9997	0.9990	0.9972	0.9933	0.9863	0.9747	0.9574	0.9332
11	1.0000	1.0000	0.9999	0.9997	0.9991	0.9976	0.9945	0.9890	0.9799	0.9661
12	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980	0.9955	0.9912	0.9840
13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9983	0.9964	0.9929
14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9970
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9988
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

$k \setminus \lambda$	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5
0	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
1	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005	0.0003	0.0002	0.0001
2	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0018	0.0012	0.0008
3	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0071	0.0049	0.0034
4	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293	0.0211	0.0151	0.0107
5	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0504	0.0375	0.0277
6	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301	0.1016	0.0786	0.0603
7	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202	0.1785	0.1432	0.1137
8	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328	0.2794	0.2320	0.1906
9	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579	0.3971	0.3405	0.2888
10	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830	0.5207	0.4599	0.4017
11	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968	0.6387	0.5793	0.5198
12	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916	0.7420	0.6887	0.6329
13	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645	0.8253	0.7813	0.7330
14	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165	0.8879	0.8540	0.8153
15	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513	0.9317	0.9074	0.8783
16	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730	0.9604	0.9441	0.9236
17	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911	0.9857	0.9781	0.9678	0.9542
18	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957	0.9928	0.9885	0.9823	0.9738
19	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980	0.9965	0.9942	0.9907	0.9857
20	1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9972	0.9953	0.9925
21	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9987	0.9977	0.9962
22	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9994	0.9990	0.9982
23	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9995	0.9992
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

Loi Normale $\mathcal{N}(0,1)$

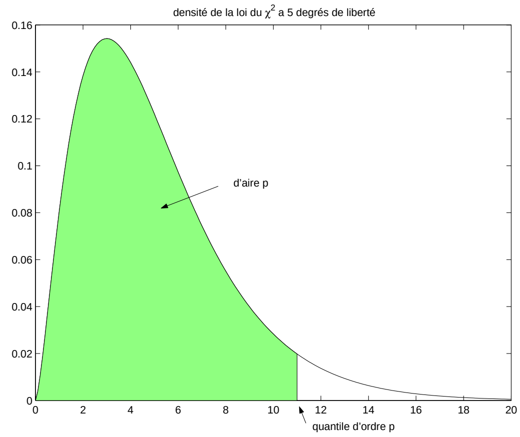
Le tableau donne les valeurs de la fonction de répartition F de la loi normale $\mathcal{N}(0,1)$ en différents points t.
Par exemple, à l'intersection de la ligne 17 et de la colonne 6, on lit $F(1.65) = 0.9505$.



t	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Loi du χ^2

Le tableau donne différents quantiles de la loi du χ^2 à n degrés de liberté : si F_n est la fonction de répartition de la loi du $\chi^2(n)$ et $p \in]0,1[$, alors le quantile d'ordre p de la loi du χ^2 à n degrés de liberté est le réel t tel que $F_n(t) = p$.

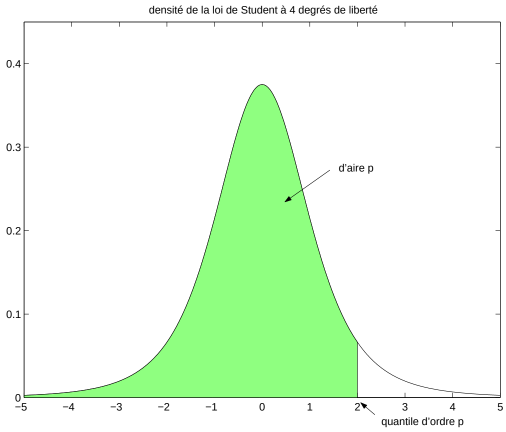


$n \setminus p$	0.001	0.005	0.010	0.020	0.025	0.050	0.1	0.15	0.2
1	0.0000	0.0000	0.0002	0.0006	0.0010	0.0039	0.0158	0.0358	0.0642
2	0.0020	0.0100	0.0201	0.0404	0.0506	0.1026	0.2107	0.3250	0.4463
3	0.0243	0.0717	0.1148	0.1848	0.2158	0.3518	0.5844	0.7978	1.0052
4	0.0908	0.2070	0.2971	0.4294	0.4844	0.7107	1.0636	1.3665	1.6488
5	0.2102	0.4117	0.5543	0.7519	0.8312	1.1455	1.6103	1.9938	2.3425
6	0.3811	0.6757	0.8721	1.1344	1.2373	1.6354	2.2041	2.6613	3.0701
7	0.5985	0.9893	1.2390	1.5643	1.6899	2.1673	2.8331	3.3583	3.8223
8	0.8571	1.3444	1.6465	2.0325	2.1797	2.7326	3.4895	4.0782	4.5936
9	1.1519	1.7349	2.0879	2.5324	2.7004	3.3251	4.1682	4.8165	5.3801
10	1.4787	2.1559	2.5582	3.0591	3.2470	3.9403	4.8652	5.5701	6.1791
11	1.8339	2.6032	3.0535	3.6087	3.8157	4.5748	5.5778	6.3364	6.9887
12	2.2142	3.0738	3.5706	4.1783	4.4038	5.2260	6.3038	7.1138	7.8073
13	2.6172	3.5650	4.1069	4.7654	5.0088	5.8919	7.0415	7.9008	8.6339
14	3.0407	4.0747	4.6604	5.3682	5.6287	6.5706	7.7895	8.6963	9.4673
15	3.4827	4.6009	5.2293	5.9849	6.2621	7.2609	8.5468	9.4993	10.3070
16	3.9416	5.1422	5.8122	6.6142	6.9077	7.9616	9.3122	10.3090	11.1521
17	4.4161	5.6972	6.4078	7.2550	7.5642	8.6718	10.0852	11.1249	12.0023
18	4.9048	6.2648	7.0149	7.9062	8.2307	9.3905	10.8649	11.9463	12.8570
19	5.4068	6.8440	7.6327	8.5670	8.9065	10.1170	11.6509	12.7727	13.7158
20	5.9210	7.4338	8.2604	9.2367	9.5908	10.8508	12.4426	13.6039	14.5784
21	6.4467	8.0337	8.8972	9.9146	10.2829	11.5913	13.2396	14.4393	15.4446
22	6.9830	8.6427	9.5425	10.6000	10.9823	12.3380	14.0415	15.2788	16.3140
23	7.5292	9.2604	10.1957	11.2926	11.6886	13.0905	14.8480	16.1219	17.1865
24	8.0849	9.8862	10.8564	11.9918	12.4012	13.8484	15.6587	16.9686	18.0618
25	8.6493	10.5197	11.5240	12.6973	13.1197	14.6114	16.4734	17.8184	18.9398
26	9.2221	11.1602	12.1981	13.4086	13.8439	15.3792	17.2919	18.6714	19.8202
27	9.8028	11.8076	12.8785	14.1254	14.5734	16.1514	18.1139	19.5272	20.7030
28	10.3909	12.4613	13.5647	14.8475	15.3079	16.9279	18.9392	20.3857	21.5880
29	10.9861	13.1211	14.2565	15.5745	16.0471	17.7084	19.7677	21.2468	22.4751
30	11.5880	13.7867	14.9535	16.3062	16.7908	18.4927	20.5992	22.1103	23.3641
31	12.1963	14.4578	15.6555	17.0423	17.5387	19.2806	21.4336	22.9762	24.2551
32	12.8107	15.1340	16.3622	17.7827	18.2908	20.0719	22.2706	23.8442	25.1478
33	13.4309	15.8153	17.0735	18.5271	19.0467	20.8665	23.1102	24.7143	26.0422
34	14.0567	16.5013	17.7891	19.2754	19.8063	21.6643	23.9523	25.5864	26.9383
35	14.6878	17.1918	18.5089	20.0274	20.5694	22.4650	24.7967	26.4604	27.8359
36	15.3241	17.8867	19.2327	20.7829	21.3359	23.2686	25.6433	27.3362	28.7350

$n \setminus p$	0.800	0.850	0.900	0.950	0.975	0.980	0.990	0.995	0.999
1	1.6424	2.0723	2.7055	3.8415	5.0239	5.4119	6.6349	7.8794	10.8276
2	3.2189	3.7942	4.6052	5.9915	7.3778	7.8240	9.2103	10.5966	13.8155
3	4.6416	5.3170	6.2514	7.8147	9.3484	9.8374	11.3449	12.8382	16.2662
4	5.9886	6.7449	7.7794	9.4877	11.1433	11.6678	13.2767	14.8603	18.4668
5	7.2893	8.1152	9.2364	11.0705	12.8325	13.3882	15.0863	16.7496	20.5150
6	8.5581	9.4461	10.6446	12.5916	14.4494	15.0332	16.8119	18.5476	22.4577
7	9.8032	10.7479	12.0170	14.0671	16.0128	16.6224	18.4753	20.2777	24.3219
8	11.0301	12.0271	13.3616	15.5073	17.5345	18.1682	20.0902	21.9550	26.1245
9	12.2421	13.2880	14.6837	16.9190	19.0228	19.6790	21.6660	23.5894	27.8772
10	13.4420	14.5339	15.9872	18.3070	20.4832	21.1608	23.2093	25.1882	29.5883
11	14.6314	15.7671	17.2750	19.6751	21.9200	22.6179	24.7250	26.7568	31.2641
12	15.8120	16.9893	18.5493	21.0261	23.3367	24.0540	26.2170	28.2995	32.9095
13	16.9848	18.2020	19.8119	22.3620	24.7356	25.4715	27.6882	29.8195	34.5282
14	18.1508	19.4062	21.0641	23.6848	26.1189	26.8728	29.1412	31.3193	36.1233
15	19.3107	20.6030	22.3071	24.9958	27.4884	28.2595	30.5779	32.8013	37.6973
16	20.4651	21.7931	23.5418	26.2962	28.8454	29.6332	31.9999	34.2672	39.2524
17	21.6146	22.9770	24.7690	27.5871	30.1910	30.9950	33.4087	35.7185	40.7902
18	22.7595	24.1555	25.9894	28.8693	31.5264	32.3462	34.8053	37.1565	42.3124
19	23.9004	25.3289	27.2036	30.1435	32.8523	33.6874	36.1909	38.5823	43.8202
20	25.0375	26.4976	28.4120	31.4104	34.1696	35.0196	37.5662	39.9968	45.3147
21	26.1711	27.6620	29.6151	32.6706	35.4789	36.3434	38.9322	41.4011	46.7970
22	27.3015	28.8225	30.8133	33.9244	36.7807	37.6595	40.2894	42.7957	48.2679
23	28.4288	29.9792	32.0069	35.1725	38.0756	38.9683	41.6384	44.1813	49.7282
24	29.5533	31.1325	33.1962	36.4150	39.3641	40.2704	42.9798	45.5585	51.1786
25	30.6752	32.2825	34.3816	37.6525	40.6465	41.5661	44.3141	46.9279	52.6197
26	31.7946	33.4295	35.5632	38.8851	41.9232	42.8558	45.6417	48.2899	54.0520
27	32.9117	34.5736	36.7412	40.1133	43.1945	44.1400	46.9629	49.6449	55.4760
28	34.0266	35.7150	37.9159	41.3371	44.4608	45.4188	48.2782	50.9934	56.8923
29	35.1394	36.8538	39.0875	42.5570	45.7223	46.6927	49.5879	52.3356	58.3012
30	36.2502	37.9903	40.2560	43.7730	46.9792	47.9618	50.8922	53.6720	59.7031
31	37.3591	39.1244	41.4217	44.9853	48.2319	49.2264	52.1914	55.0027	61.0983
32	38.4663	40.2563	42.5847	46.1943	49.4804	50.4867	53.4858	56.3281	62.4872
33	39.5718	41.3861	43.7452	47.3999	50.7251	51.7429	54.7755	57.6484	63.8701
34	40.6756	42.5140	44.9032	48.6024	51.9660	52.9952	56.0609	58.9639	65.2472
35	41.7780	43.6399	46.0588	49.8018	53.2033	54.2438	57.3421	60.2748	66.6188
36	42.8788	44.7641	47.2122	50.9985	54.4373	55.4889	58.6192	61.5812	67.9852

Loi de Student

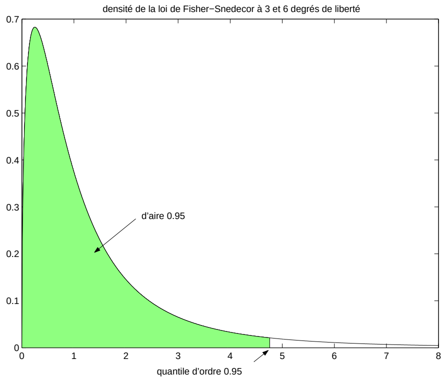
Le tableau donne différents quantiles de la loi de Student à n degrés de liberté $\mathcal{T}(n)$: si F_n est la fonction de répartition de la loi de Student $\mathcal{T}(n)$ et $p \in]0,1[$, alors le quantile d'ordre p de la loi $\mathcal{T}(n)$ est le réel t tel que $F_n(t) = p$.



$n \setminus p$	0.550	0.650	0.750	0.850	0.900	0.950	0.975	0.990	0.995	0.999	0.9995
1	0.158	0.510	1.000	1.963	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.142	0.445	0.816	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.137	0.424	0.765	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.134	0.414	0.741	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.132	0.408	0.727	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.131	0.404	0.718	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.130	0.402	0.711	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.130	0.399	0.706	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.129	0.398	0.703	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.129	0.397	0.700	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.129	0.396	0.697	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.128	0.395	0.695	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.128	0.394	0.694	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.128	0.393	0.692	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.128	0.393	0.691	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.128	0.392	0.690	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.128	0.392	0.689	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.127	0.392	0.688	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.127	0.391	0.688	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.127	0.391	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.127	0.391	0.686	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.127	0.390	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.127	0.390	0.685	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.127	0.390	0.685	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.127	0.390	0.684	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.127	0.390	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.127	0.389	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.127	0.389	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.127	0.389	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.127	0.389	0.683	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.126	0.388	0.681	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.126	0.388	0.679	1.047	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	0.126	0.387	0.679	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
70	0.126	0.387	0.678	1.044	1.294	1.667	1.994	2.381	2.648	3.211	3.435
80	0.126	0.387	0.678	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
90	0.126	0.387	0.677	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.126	0.386	0.677	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390

Loi de Fisher-Snedecor

Le tableau suivant donne les quantiles d'ordre $q = 0,95$ et $q = 0,975$ de la loi de Fisher-Snedecor à n_1 et n_2 degrés de liberté $\mathcal{F}(n_1, n_2)$.



n_2	$q \setminus n_1$	1	2	3	4	5	6	7	8	9	10
2	0.950	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
	0.975	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398
3	0.950	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
	0.975	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419
4	0.950	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
	0.975	12.218	10.649	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844
5	0.950	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
	0.975	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619
6	0.950	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
	0.975	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461
7	0.950	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
	0.975	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761
8	0.950	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
	0.975	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295
9	0.950	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
	0.975	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964
10	0.950	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
	0.975	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717
11	0.950	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
	0.975	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526
12	0.950	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
	0.975	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374
13	0.950	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
	0.975	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250
14	0.950	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
	0.975	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147
15	0.950	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
	0.975	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060
16	0.950	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
	0.975	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986
17	0.950	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
	0.975	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922
18	0.950	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
	0.975	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866
19	0.950	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
	0.975	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817
20	0.950	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
	0.975	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774
30	0.950	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
	0.975	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511
40	0.950	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
	0.975	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388
50	0.950	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
	0.975	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317
60	0.950	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
	0.975	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270
70	0.950	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969
	0.975	5.247	3.890	3.309	2.975	2.754	2.595	2.474	2.379	2.302	2.237

n_2	$q \setminus n_1$	12	14	16	18	20	30	40	50	60	70
2	0.950	19.413	19.424	19.433	19.440	19.446	19.462	19.471	19.476	19.479	19.481
	0.975	39.415	39.427	39.435	39.442	39.448	39.465	39.473	39.478	39.481	39.484
3	0.950	8.745	8.715	8.692	8.675	8.660	8.617	8.594	8.581	8.572	8.566
	0.975	14.337	14.277	14.232	14.196	14.167	14.081	14.037	14.010	13.992	13.979
4	0.950	5.912	5.873	5.844	5.821	5.803	5.746	5.717	5.699	5.688	5.679
	0.975	8.751	8.684	8.633	8.592	8.560	8.461	8.411	8.381	8.360	8.346
5	0.950	4.678	4.636	4.604	4.579	4.558	4.496	4.464	4.444	4.431	4.422
	0.975	6.525	6.456	6.403	6.362	6.329	6.227	6.175	6.144	6.123	6.107
6	0.950	4.000	3.956	3.922	3.896	3.874	3.808	3.774	3.754	3.740	3.730
	0.975	5.366	5.297	5.244	5.202	5.168	5.065	5.012	4.980	4.959	4.943
7	0.950	3.575	3.529	3.494	3.467	3.445	3.376	3.340	3.319	3.304	3.294
	0.975	4.666	4.596	4.543	4.501	4.467	4.362	4.309	4.276	4.254	4.239
8	0.950	3.284	3.237	3.202	3.173	3.150	3.079	3.043	3.020	3.005	2.994
	0.975	4.200	4.130	4.076	4.034	3.999	3.894	3.840	3.807	3.784	3.768
9	0.950	3.073	3.025	2.989	2.960	2.936	2.864	2.826	2.803	2.787	2.776
	0.975	3.868	3.798	3.744	3.701	3.667	3.560	3.505	3.472	3.449	3.433
10	0.950	2.913	2.865	2.828	2.798	2.774	2.700	2.661	2.637	2.621	2.610
	0.975	3.621	3.550	3.496	3.453	3.419	3.311	3.255	3.221	3.198	3.182
11	0.950	2.788	2.739	2.701	2.671	2.646	2.570	2.531	2.507	2.490	2.478
	0.975	3.430	3.359	3.304	3.261	3.226	3.118	3.061	3.027	3.004	2.987
12	0.950	2.687	2.637	2.599	2.568	2.544	2.466	2.426	2.401	2.384	2.372
	0.975	3.277	3.206	3.152	3.108	3.073	2.963	2.906	2.871	2.848	2.831
13	0.950	2.604	2.554	2.515	2.484	2.459	2.380	2.339	2.314	2.297	2.284
	0.975	3.153	3.082	3.027	2.983	2.948	2.837	2.780	2.744	2.720	2.703
14	0.950	2.534	2.484	2.445	2.413	2.388	2.308	2.266	2.241	2.223	2.210
	0.975	3.050	2.979	2.923	2.879	2.844	2.732	2.674	2.638	2.614	2.597
15	0.950	2.475	2.424	2.385	2.353	2.328	2.247	2.204	2.178	2.160	2.147
	0.975	2.963	2.891	2.836	2.792	2.756	2.644	2.585	2.549	2.524	2.506
16	0.950	2.425	2.373	2.333	2.302	2.276	2.194	2.151	2.124	2.106	2.093
	0.975	2.889	2.817	2.761	2.717	2.681	2.568	2.509	2.472	2.447	2.429
17	0.950	2.381	2.329	2.289	2.257	2.230	2.148	2.104	2.077	2.058	2.045
	0.975	2.825	2.753	2.697	2.652	2.616	2.502	2.442	2.405	2.380	2.362
18	0.950	2.342	2.290	2.250	2.217	2.191	2.107	2.063	2.035	2.017	2.003
	0.975	2.769	2.696	2.640	2.596	2.559	2.445	2.384	2.347	2.321	2.303
19	0.950	2.308	2.256	2.215	2.182	2.155	2.071	2.026	1.999	1.980	1.966
	0.975	2.720	2.647	2.591	2.546	2.509	2.394	2.333	2.295	2.270	2.251
20	0.950	2.278	2.225	2.184	2.151	2.124	2.039	1.994	1.966	1.946	1.932
	0.975	2.676	2.603	2.547	2.501	2.464	2.349	2.287	2.249	2.223	2.205
30	0.950	2.092	2.037	1.995	1.960	1.932	1.841	1.792	1.761	1.740	1.724
	0.975	2.412	2.338	2.280	2.233	2.195	2.074	2.009	1.968	1.940	1.920
40	0.950	2.003	1.948	1.904	1.868	1.839	1.744	1.693	1.660	1.637	1.621
	0.975	2.288	2.213	2.154	2.107	2.068	1.943	1.875	1.832	1.803	1.781
50	0.950	1.952	1.895	1.850	1.814	1.784	1.687	1.634	1.599	1.576	1.558
	0.975	2.216	2.140	2.081	2.033	1.993	1.866	1.796	1.752	1.721	1.698
60	0.950	1.917	1.860	1.815	1.778	1.748	1.649	1.594	1.559	1.534	1.516
	0.975	2.169	2.093	2.033	1.985	1.944	1.815	1.744	1.699	1.667	1.643
70	0.950	1.893	1.836	1.790	1.753	1.722	1.622	1.566	1.530	1.505	1.486
	0.975	2.136	2.059	1.999	1.950	1.910	1.779	1.707	1.660	1.628	1.604

Quantiles d'ordre $q = 0.95$ et $q = 0.975$ pour la loi de Fisher-Snedecor $\mathcal{F}(n_1, 1)$:

$q \setminus n_1$	1	2	3	4	5	6	7	8	9	10
0.950	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
0.975	647.789	799.500	864.163	899.583	921.848	937.111	948.217	956.656	963.285	968.627
$q \setminus n_1$	12	14	16	18	20	30	40	50	60	70
0.950	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
0.975	647.789	799.500	864.163	899.583	921.848	937.111	948.217	956.656	963.285	968.627