

PRÉPAS LICENCE MASTER

ÉCO
SUP

JEAN-PIERRE LECOUTRE

STATISTIQUE ET PROBABILITÉS

7^e édition

- ▶ Cours complet
- ▶ De nombreuses applications
- ▶ 116 exercices progressifs
avec leurs corrigés détaillés

DUNOD

JEAN-PIERRE LECOUTRE

STATISTIQUE ET PROBABILITÉS

7^e édition

DUNOD

Graphisme de couverture : Pierre-André Gualino
Illustration de couverture : © metamorworks / fotolia.fr
Mise en pages : Lumina

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, 2019

11, rue Paul Bert, 92240 Malakoff

www.dunod.com

ISBN 978-2-10-079285-6

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.



Table des matières

Avant-propos	IX
Notations	X
Introduction	1
Chapitre 1 Notion de probabilité	5
1. Modèle probabiliste	5
1.1. Ensemble fondamental	5
1.2. Algèbre et tribu d'événements	7
1.3. Probabilité	9
2. Probabilités conditionnelles	13
3. Théorème de Bayes	15
4. Indépendance en probabilité	17
5. Pour aller plus loin : éléments de combinatoire	19
5.1. Permutations avec répétition	19
5.2. Permutations sans répétition ou arrangements	20
5.3. Permutations avec répétition de n objets, dont k seulement sont distincts	21
5.4. Combinaisons (sans répétition)	21
5.5. Combinaisons avec répétition	23
5.6. Partitions	24
L'essentiel	25
Entraînez-vous	26
Solutions	29
Chapitre 2 Variable aléatoire	35
1. Variable aléatoire réelle discrète	35
1.1. Définition	35
1.2. Loi de probabilité	36
1.3. Fonction de répartition	37
1.4. Moments d'une v.a. discrète	39
2. Variable aléatoire réelle continue	46
2.1. Définition	46
2.2. Loi de probabilité	46
2.3. Propriétés de la fonction de répartition	46
2.4. Loi continue	47
2.5. Loi absolument continue	48

2.6.	Moments d'une v.a. absolument continue	50
2.7.	Changement de variable	53
3.	Compléments	55
3.1.	Application mesurable	55
3.2.	Densité	55
3.3.	Support	55
	L'essentiel	56
	Entraînez-vous	57
	Solutions	60
Chapitre 3	Lois usuelles	67
1.	Lois usuelles discrètes	67
1.1.	Loi de Dirac	67
1.2.	Loi de Bernoulli	68
1.3.	Loi binômiale	69
1.4.	Loi hypergéométrique	71
1.5.	Loi de Poisson	74
1.6.	Loi géométrique ou de Pascal	76
1.7.	Loi binômiale négative	76
2.	Lois usuelles continues	77
2.1.	Loi uniforme	77
2.2.	Loi exponentielle	79
2.3.	Loi normale ou de Laplace-Gauss	81
2.4.	Loi gamma	85
2.5.	Loi du khi-deux	86
2.6.	Loi bêta	87
2.7.	Loi log-normale	88
2.8.	Loi de Pareto	89
3.	Compléments : fonctions génératrices	89
3.1.	Fonction génératrice d'une v.a. discrète positive	89
3.2.	Fonction génératrice d'une loi absolument continue	91
	L'essentiel	93
	Entraînez-vous	94
	Solutions	97
Chapitre 4	Couple et vecteur aléatoires	105
1.	Couple de v.a. discrètes	105
1.1.	Loi d'un couple	105
1.2.	Lois marginales	106
1.3.	Lois conditionnelles	106
1.4.	Moments conditionnels	107

1.5. Moments associés à un couple	108
1.6. Loi d'une somme	109
2. Couple de v.a. continues	112
2.1. Loi du couple	112
2.2. Lois marginales	114
2.3. Lois conditionnelles	115
2.4. Moments associés à un couple	116
2.5. Régression	117
2.6. Loi d'une somme	118
3. Vecteur aléatoire	120
4. Lois usuelles	122
4.1. Loi multinomiale	122
4.2. Loi normale vectorielle	124
5. Compléments	128
5.1. Application mesurable	128
5.2. Changement de variable	129
L'essentiel	130
Entraînez-vous	131
Solutions	134

Chapitre 5 Loi empirique	145
1. Échantillon d'une loi	145
2. Moments empiriques	146
2.1. Moyenne empirique	146
2.2. Variance empirique	147
2.3. Moments empiriques	148
3. Échantillon d'une loi normale	148
3.1. Loi de Student	149
3.2. Loi de Fisher-Snedecor	150
4. Tests d'adéquation	151
4.1. Test du khi-deux	152
4.2. Test de Kolmogorov-Smirnov	154
5. Compléments	156
5.1. Statistique d'ordre	156
5.2. Théorème de Fisher	157
L'essentiel	159
Entraînez-vous	160
Solutions	161

Chapitre 6	Comportement asymptotique	165
1.	Convergence en probabilité	165
1.1.	Inégalité de Markov	166
1.2.	Inégalité de Bienaymé-Tchebychev	166
1.3.	Inégalité de Jensen	167
1.4.	Convergence en probabilité	167
1.5.	Loi des grands nombres	171
2.	Convergence en loi	172
2.1.	Définition	172
2.2.	Lien avec la convergence en probabilité	173
2.3.	Propriété	173
2.4.	Théorème de Slutsky	173
2.5.	Conditions suffisantes de convergence en loi	174
2.6.	Théorème central limite	174
2.7.	Limite d'une suite image	174
2.8.	Convergence des moments empiriques	175
2.9.	Convergence des lois usuelles	176
3.	Compléments	180
3.1.	Convergence presque sûre	180
3.2.	Convergence presque complète	181
	L'essentiel	183
	Entraînez-vous	184
	Solutions	186
Chapitre 7	Estimation	191
1.	Définition d'un estimateur	192
2.	Propriétés d'un estimateur	194
2.1.	Biais d'un estimateur	194
2.2.	Convergence d'un estimateur	195
2.3.	Estimateur optimal	196
3.	Méthodes de construction d'un estimateur	200
3.1.	Méthode du maximum de vraisemblance	200
3.2.	Méthode des moments	202
4.	Estimation par intervalle de confiance	203
4.1.	Exemple introductif	203
4.2.	Principe de construction	204
4.3.	Intervalle pour une proportion	206
4.4.	Intervalles associés aux paramètres de la loi normale	209

5. Compléments	216
5.1. Inégalité de Fréchet-Darmois-Cramer-Rao	216
5.2. Statistique exhaustive	217
5.3. Famille exponentielle	219
5.4. Amélioration d'un estimateur	223
L'essentiel	224
Entraînez-vous	225
Chapitre 8 Tests d'hypothèses	247
1. Concepts principaux en théorie des tests	247
2. Méthode de Bayes	251
3. Méthode de Neyman et Pearson	252
3.1. Principe de la règle de Neyman et Pearson	252
3.2. Hypothèses simples	253
3.3. Hypothèses multiples	255
4. Test d'indépendance du khi-deux	257
5. Compléments	258
L'essentiel	259
Entraînez-vous	260
Solutions	264
Tables statistiques	279
Index	291

Avant-propos

Ce manuel de cours est destiné principalement aux étudiants de la Licence économie et gestion mais peut être utile à toute personne souhaitant connaître et surtout utiliser les principales méthodes de la statistique inférentielle. Il correspond au programme de probabilités et statistique généralement enseigné dans les deux premières années de Licence (L1 et L2). Cette 7^e édition s'est enrichie d'exercices nouveaux. Le niveau mathématique requis est celui de la première année de Licence, avec quelques notions (séries, intégrales multiples...) souvent enseignées seulement en deuxième année.

Si une grande partie de l'ouvrage est consacrée à la théorie des probabilités, l'ordre des termes retenu dans le titre veut signifier qu'il ne s'agit que d'un outil au service de la statistique. Ce n'est qu'un passage obligé pour donner des bases rigoureuses à la méthode statistique. On peut le concevoir comme un ensemble de règles grammaticales, parfois difficiles et fastidieuses à retenir, mais qui permettent de rédiger des textes clairs, rigoureux et sans ambiguïtés, même si l'on n'a pas conscience qu'ils ont été écrits dans le respect de ces règles. La partie statistique correspond aux deux derniers chapitres d'estimation et de tests d'hypothèses.

Les fondements théoriques de la statistique étant parfois délicats, nous avons choisi de présenter sans démonstration les principales propriétés nécessaires à une utilisation judicieuse des méthodes statistiques, en les illustrant systématiquement d'exemples. De même, afin de ne pas alourdir les énoncés de théorèmes, les conditions techniques de leur validité ne sont pas présentées dans leur détail, parfois fastidieux, et qui risque de masquer l'essentiel qui est la propriété énoncée. Notre souci constant a été de faciliter la compréhension, pour pouvoir passer aisément au stade de l'utilisation, sans cependant pour cela sacrifier à la rigueur. La traduction anglaise des termes les plus usuels figure entre parenthèses.

Chaque chapitre se conclut par des exercices corrigés permettant de contrôler l'acquisition des notions essentielles qui y ont été introduites. Faire de nombreux exercices est certainement le meilleur moyen d'arriver à la compréhension de certaines notions quelquefois difficiles. Rappelons cette maxime chinoise : *J'entends et j'oublie. Je vois et je retiens. Je fais et je comprends.* En fin de chapitre se trouvent également quelques compléments ; soit de notions mathématiques utilisées dans celui-ci, la combinatoire par exemple, soit de propriétés comme l'exhaustivité, très importantes et utiles, mais hors du programme d'une Licence d'économie ou de gestion. Avec ces compléments, cet ouvrage peut convenir aussi aux étudiants des écoles de management.

Notations

Ω	Ensemble fondamental
$\mathcal{P}(\Omega)$	Ensemble des parties de Ω
\overline{A}, A^c	Complémentaire de A
\mathcal{A}	Algèbre ou tribu de parties de Ω
$\text{card } A$	Cardinal de A
$\binom{n}{p}$	Coefficient binomial
$[x]$	Partie entière de x
$\ln x$	Logarithme népérien de x
$\mathbf{1}_A$	Indicatrice de A
$\text{Cov}(X, Y)$	Covariance de X et Y
f.r.	Fonction de répartition
v.a.	Variable aléatoire
φ	Densité de la loi $N(0,1)$
Φ	F.r. de la loi $N(0,1)$
\mathbb{C}	Ensemble des nombres complexes
${}^t A$	Matrice transposée de A
I_n	Matrice unité d'ordre n
$X \rightsquigarrow P$	La v.a. X suit la loi de probabilité P
$\mathcal{B}(n, p)$	Loi binomiale de paramètres n et p
$\mathcal{P}(\lambda)$	Loi de Poisson de paramètre λ
$N(m, \sigma)$	Loi normale dans \mathbb{R} , d'espérance m et d'écart type σ
$\mathcal{N}_n(\mu, \Sigma)$	Loi normale dans \mathbb{R}^n , de vecteur espérance μ et de matrice variances-covariances Σ
T_n	Loi de Student à n degrés de liberté
χ_n^2	Loi du khi-deux à n degrés de liberté
$F(n, m)$	Loi de Fisher-Snedecor à n et m degrés de liberté
emv	Estimateur du maximum de vraisemblance

Introduction

La statistique a une origine très ancienne, se réduisant initialement à une collecte d'observations, notamment le dénombrement des hommes (recensement). On mentionne des opérations de recensement il y a plus de 4 000 ans en Chine, en Mésopotamie ou en Égypte et la Bible en cite plusieurs, dans le Livre des Nombres par exemple. Cependant, le terme statistique est apparu assez récemment, vers le milieu du XVII^e siècle ; il vient du latin *statisticus*, relatif à l'état (*status*), et est employé alors dans un sens purement descriptif de recueil ou de collection de faits chiffrés, les **statistiques**. Le mot employé au singulier avec l'article défini, la **statistique**, évoque la méthode utilisée ensuite pour étendre des résultats et dégager des lois (**l'inférence**). Il s'agit donc dans ce sens d'un moyen scientifique d'analyse et de compréhension du phénomène étudié, s'appliquant très largement à l'économie et à toutes les sciences sociales et de la nature.

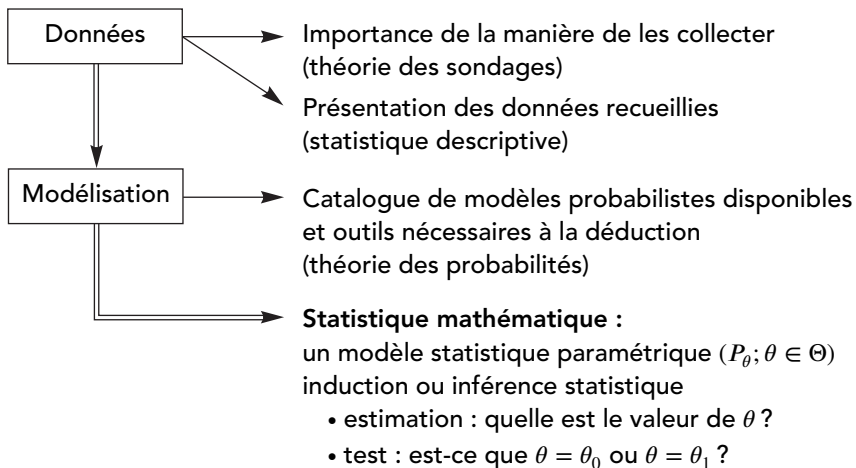
Cette discipline concerne donc tous ceux qui ont à relever, présenter, analyser ou utiliser une information dont la masse peut être volumineuse. On peut la définir comme un ensemble de méthodes dont le but est de traiter des données, les **statistiques**, relatives à un certain domaine d'étude. Elle traite également de la manière de recueillir ces données, auprès de qui et sous quelle forme (théorie des sondages). Son objectif peut se résumer de la façon suivante : dégager, à partir de données observées sur quelques individus d'une population, des résultats valables pour l'ensemble de la population.

Cela consistera par exemple à remplacer des données nombreuses par des indicateurs (résumés) les plus pertinents possibles : résumé clair avec le minimum de perte d'information, permettant de dégager plus facilement un diagnostic. Il s'agit alors de la **statistique descriptive** qui recouvre les moyens de présenter ces données et d'en décrire les principales caractéristiques, en les résumant sous forme de tableaux ou de graphiques. Il s'agira ensuite de les interpréter. La description statistique se propose de mettre en évidence certaines permanences ou **lois statistiques**, qui peuvent éventuellement conduire à des **prévisions** (élément essentiel de l'étude des séries chronologiques). Une règle qui transforme un ensemble de données en une ou plusieurs valeurs numériques se nomme une **statistique**, le terme étant cette fois utilisé avec l'article indéfini.

Le début de la méthodologie statistique peut se situer au XVII^e siècle qui verra également l'éclosion d'un outil fondamental pour une formalisation tout à fait rigoureuse, la **théorie des probabilités**, qui est l'analyse mathématique des phénomènes dans lesquels le hasard intervient. Le **calcul des probabilités** a commencé avec Blaise Pascal, Pierre Fermat, Christian Huygens et Jacques Bernoulli par l'analyse des jeux dits de hasard. Le mot **hasard** est d'ailleurs emprunté à l'arabe *az-zahr* (jeu de dés, *alea* en latin) au XII^e siècle d'où est venue cette expression jeu de hasard au XVI^e siècle. La théorie des probabilités servira ensuite d'outil de base à un ensemble de méthodes ou de règles objectives permettant d'utiliser des données pour fixer la précision avec laquelle on estime

certaines paramètres (théorie de l'estimation) ou on teste certaines hypothèses (théorie des tests) : la **Statistique mathématique** (ou inférentielle). Ceci permet d'obtenir une mesure objective de la distance entre un modèle statistique, traduit par une famille P_θ de lois de probabilité indexée par un paramètre θ parcourant un ensemble donné Θ , et un ensemble de données observées.

Tout ceci peut se synthétiser au moyen du schéma suivant :



Il reste à préciser dans quel cadre cette formalisation à l'aide de modèles aléatoires sera nécessaire. Toute démarche scientifique nécessite la réalisation de certaines expériences que l'on peut regrouper en deux grandes catégories :

- Pour certaines d'entre elles, si elles sont renouvelées dans des conditions totalement identiques, elles produiront le même résultat, qui devient donc prévisible. Il s'agit de **phénomènes déterministes**, où les faits sont régis par des lois universelles physiques (par exemple l'augmentation de la pression d'un gaz provoque une diminution de son volume, ce que traduit la loi de Mariotte : Pression \times Volume = constante ; l'eau portée à 100 degrés Celsius se transforme en vapeur...). Le résultat est entièrement déterminé par les conditions de l'expérience : on peut prévoir le phénomène qui va se produire.
- Par contre, d'autres expériences ont toujours un résultat imprévisible (lancer d'un dé ou d'une pièce de monnaie) : effectuées dans des conditions totalement identiques elles donneront des résultats différents. Le résultat est non prévisible et on dit qu'il est dû **au hasard**, cette expression étant utilisée pour la première fois par Fénelon en 1695, le mot hasard étant compris maintenant au sens absolu et philosophique comme « sans évolution prévisible », à opposer à déterministe. Dans son *Essai philosophique sur les probabilités* (1814), Laplace considère en effet que le déterminisme ne laisse aucune place au hasard : l'état de l'univers à un instant donné détermine son état

à tout autre instant ultérieur. Ainsi, quand on jette une pièce de monnaie en l'air, les lois de la mécanique classique déterminent, en principe, avec certitude si elle retombera sur pile ou face. Le résultat n'est pas dû au hasard, mais à la manière dont elle a été lancée en l'air et à la façon dont elle va retomber sur une certaine surface ; mais la trajectoire décrite par cette pièce avant de retomber sur pile est tellement complexe qu'il n'est pas possible de prévoir son issue. Le phénomène ne relève pas du déterminisme entendu au sens de la possibilité de prédiction, par le calcul ou la loi mathématique.

Dans un mémoire de 1774, Laplace énonce que « le hasard n'a aucune réalité en lui-même : ce n'est qu'un terme propre à désigner notre ignorance... La notion de probabilité tient à cette ignorance ». Retenir un modèle probabiliste est donc simplement un aveu de notre ignorance, de notre incapacité à fournir un modèle physique décrivant une réalité trop complexe. On parle alors d'**épreuve** ou d'**expérience aléatoire** et le résultat obtenu sera un **événement**. Les outils appropriés dans ce cadre sont ceux de la **statistique mathématique**, la base de cette discipline étant la **théorie des probabilités**, que nous devons donc étudier dans les six premiers chapitres de cet ouvrage, comme préalable aux deux chapitres d'estimation et de tests d'hypothèses.

REMARQUE Dans *Science et méthode* publié en 1908, Henri Poincaré exprime que hasard et déterminisme sont rendus compatibles par l'imprédictibilité à long terme. Les relations entre hasard et déterminisme ont été dans les années 1980 l'objet d'une controverse animée entre les mathématiciens René Thom et Ilya Prigogine. L'étude récente des systèmes dynamiques montre que l'on ne peut pas confondre déterminisme et prédictibilité. En effet, une légère perturbation des conditions initiales d'un tel système mathématiquement déterministe peut empêcher de prévoir son évolution future.

Introduction

Au cours de ce chapitre, nous allons donner la définition d'un certain nombre de termes du vocabulaire utilisé dans un contexte non déterministe et indiquer comment construire le modèle adéquat. La notion essentielle introduite étant bien sûr celle de probabilité, avec la notion d'indépendance d'événements qui lui est associée et qui joue un rôle très important en statistique. La représentation formelle du modèle probabiliste sous-jacent est presque toujours absente dans un problème concret de statistique. Cependant, cette formalisation rigoureuse est indispensable pour obtenir les outils théoriques nécessaires à la résolution d'un tel problème statistique.

Objectifs

Montrer que le modèle probabiliste est choisi en fonction du but que l'on poursuit.

Définir de façon théorique (et non intuitive) la notion d'indépendance.

Concepts clés

Probabilité
Probabilité conditionnelle
Indépendance

1 Modèle probabiliste

1.1 Ensemble fondamental

Avant toute formalisation, le résultat d'une expérience aléatoire s'appelle **événement**. La quantification des « chances » qu'un tel événement a de se réaliser correspond à la notion intuitive de **probabilité**. Pour réaliser cette quantification, il est nécessaire de décrire au préalable, très précisément, l'ensemble des résultats possibles, appelés **événements élémentaires**. Cet ensemble expérimental s'appelle **ensemble fondamental** (ou univers) et est noté traditionnellement Ω .

Exemple 1.1

Jet d'un dé à six faces numérotées : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Exemple 1.2

On tire une boule dans une urne contenant une boule noire, deux blanches et cinq rouges et l'ensemble fondamental retenu est $\Omega = \{\text{noire, blanche, rouge}\}$.

Chaque élément $\omega \in \Omega$ représente donc un événement élémentaire, et toute partie $A \subset \Omega$ (ou $A \in \mathcal{P}(\Omega)$) sera un événement. Parfois on dit que Ω est l'ensemble des **éventualités possibles** et les événements élémentaires sont alors les singletons, c'est-à-dire les ensembles réduits à un seul élément $\{\omega\}$, qui sont effectivement en toute rigueur des événements, puisque appartenant à $\mathcal{P}(\Omega)$, ce qui n'est pas le cas du point ω .

Exemple 1.3

À l'expérience du jet de dé on associe $\Omega = \{1, 2, 3, 4, 5, 6\}$; l'événement $A = \{1, 2\}$ traduit, c'est-à-dire représente symboliquement, le résultat « obtenir un résultat inférieur ou égal à 2 ».

L'ensemble Ω dépend évidemment de l'expérience considérée, mais aussi du choix de celui qui construit le modèle, et par là présente donc un certain arbitraire.

Exemple 1.4

Dans l'expérience du jet de dé on peut choisir également $\Omega = \{pair, impair\}$ ou $\Omega = \{\{1, 2, 3, \}, \{4, 5, 6\}\}$.

Exemple 1.5

Si on tire une carte d'un jeu de 32 cartes, on peut retenir comme ensembles fondamentaux $\Omega = \{7, 8, 9, 10, V, D, R, As\}$ ou $\Omega = \{trèfle, carreau, cœur, pique\}$ ou $\Omega = \{rouge, noir\}$.

Cet ensemble Ω peut être fini ou infini, continu ou discret.

Exemple 1.6

On lance une pièce jusqu'à obtenir pile, l'événement retenu étant le nombre de jets effectués :

$$\Omega = \{1, 2, \dots, n, \dots\} = \mathbb{N}^*$$

ensemble infini dénombrable.

Exemple 1.7

On observe la durée de vie d'une lampe :

$$\Omega = [0, +\infty[= \mathbb{R}_+$$

ensemble infini non dénombrable.

L'ensemble retenu est bien sûr une abstraction et peut parfois contenir des événements qui ne se produiront jamais dans la réalité.

Exemple 1.8

On mesure la taille d'un individu choisi au hasard dans une population et on retient $\Omega = \mathbb{R}_+$; cet ensemble contient de très grandes tailles qui n'existent bien sûr pas dans la réalité, mais en raison de la difficulté de fixer une valeur maximale de la taille pour définir l'ensemble fondamental, c'est le choix qui paraît le moins arbitraire.

1.2 Algèbre et tribu d'événements

Un événement étant un élément de $\mathcal{P}(\Omega)$ obéit à la théorie des ensembles. Nous allons indiquer dans le tableau ci-après comment certaines notions ensemblistes s'expriment, ou se traduisent, en termes d'événements.

Ensemble	Événement
On a observé le résultat ω et $\omega \in A$	L'événement A est réalisé.
$A = B$	Les événements A et B sont identiques.
$A \subset B$	L'événement A implique l'événement B .
\emptyset	Événement impossible.
Ω	Événement certain.
$A \cup B$	Un au moins des deux événements est réalisé.
$A \cap B$	Les deux événements A et B sont réalisés.
$A \cap B = \emptyset$	Les événements A et B sont incompatibles.
$\bar{A} = \Omega - A$ ou A^c	L'événement A n'est pas réalisé.

Le couple $(\Omega, \mathcal{P}(\Omega))$ s'appelle un **espace probabilisable**.

Cependant, même si Ω est fini, le cardinal de $\mathcal{P}(\Omega)$ est $2^{card\Omega}$, qui peut être un nombre très grand. Dans ce cas, il peut être souhaitable de ne considérer qu'une famille restreinte \mathcal{A} de parties de Ω , $\mathcal{A} \subset \mathcal{P}(\Omega)$. Pour que le résultat des opérations ensemblistes (union, intersection, complémentaire) soit encore un événement, il est nécessaire que cette famille d'événements retenue soit fermée, ou stable, vis-à-vis de ces opérations, c'est-à-dire qu'il soit bien un élément de la famille (par exemple, si on retient la famille des nombres impairs, elle n'est pas stable pour l'addition puisque le résultat est un nombre pair). De plus, les événements « certain », Ω , et « impossible », \emptyset , doivent également appartenir à cet ensemble. Ainsi, on associera à une épreuve aléatoire un ensemble non vide de parties de Ω , noté \mathcal{A} , qui vérifiera :

- C_1 pour tout $A \in \mathcal{A}$ alors $\bar{A} \in \mathcal{A}$;
- C_2 pour tout $A \in \mathcal{A}$ et tout $B \in \mathcal{A}$ alors $A \cup B \in \mathcal{A}$.

Il y a fermeture pour le complémentaire et l'union. Cet ensemble \mathcal{A} s'appelle une **algèbre** de parties de Ω . Bien entendu, grâce aux lois de Morgan, on a une définition équivalente en remplaçant la condition C_2 par :

- C'_2 pour tout $A \in \mathcal{A}$ et tout $B \in \mathcal{A}$ alors $A \cap B \in \mathcal{A}$.

Exemple 1.9

L'algèbre la plus élémentaire est réduite à $\mathcal{A} = \{\emptyset, \Omega\}$.

Exemple 1.10

À partir d'un événement quelconque A , on peut constituer l'algèbre :

$$\mathcal{A} = \{\emptyset, A, \bar{A}, \Omega\}$$

Exemple 1.11

On peut générer une algèbre à partir d'une partition. À partir de la partition de $\Omega = \{a, b, c, d\}$ en trois ensembles $\{a, b\}, \{c\}, \{d\}$, on construit l'algèbre :

$$\mathcal{A} = \{\emptyset, \{a, b\}, \{c\}, \{d\}, \{a, b, c\}, \{a, b, d\}, \{c, d\}, \Omega\}$$

avec $\text{card } \mathcal{A} = 2^3$.

Exemple 1.12

L'algèbre la plus complète est bien entendu $\mathcal{P}(\Omega)$.

Théorème 1.1

Propriétés d'une algèbre

P_1 La famille étant non vide, on en conclut que :

$$\emptyset \in \mathcal{A}, \Omega \in \mathcal{A}$$

P_2 Si $A_j \in \mathcal{A}$ pour $1 \leq j \leq n$, on démontre par récurrence que :

$$\bigcup_{j=1}^n A_j \in \mathcal{A}$$

P_3 Si $A_j \in \mathcal{A}$ pour $1 \leq j \leq n$, on démontre également par passage au complémentaire que :

$$\bigcap_{j=1}^n A_j \in \mathcal{A}$$

Cependant, certaines expériences peuvent se dérouler indéfiniment (au moins théoriquement), comme par exemple lancer un dé jusqu'à obtenir le chiffre 6. Si A_n représente l'événement « obtenir le chiffre 6 au n -ème lancer », l'événement « obtenir le chiffre 6 »

s'écrira $A = \bigcup_{n=1}^{\infty} A_n$. On a donc besoin de renforcer la propriété P_2 de fermeture pour l'union finie par une condition de fermeture pour l'union dénombrable, soit :

– C_3 Si $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$ alors :

$$\bigcup_{n=0}^{\infty} A_n \in \mathcal{A}$$

condition qui exprime que toute union dénombrable d'événements est encore un événement. L'ensemble \mathcal{A} auquel on impose les conditions C_1 et C_3 s'appelle alors une σ – **algèbre** ou **tribu** d'événements.

REMARQUE On aurait pu remplacer dans la condition C_3 l'union par l'intersection (par passage au complémentaire l'union se transforme en intersection). Bien entendu toute algèbre finie est une σ – **algèbre**.

Pendant, dans les cas simples où Ω est un ensemble fini, ou éventuellement infini dénombrable, on peut retenir comme tribu des événements $\mathcal{P}(\Omega)$ tout entier. Ce n'est que dans les autres cas que l'on est amené à considérer des ensembles \mathcal{A} plus réduits que $\mathcal{P}(\Omega)$ qui est alors trop vaste. Le couple formé de l'ensemble fondamental Ω et de la tribu d'événements associée \mathcal{A} s'appelle un **espace probabilisable**. Cette terminologie signifie que l'on va pouvoir associer une **probabilité**, notée P , à ce modèle (Ω, \mathcal{A}) .

1.3 Probabilité

Une fois défini l'ensemble des événements auxquels on s'intéresse, on va essayer de traduire par un nombre leurs « possibilités » de réalisation. Cela revient à affecter une mesure de « croyance » à chaque événement, c'est-à-dire un degré de certitude que l'on a que l'événement se produise ou non. Afin de correspondre à la notion intuitive, une probabilité sera un nombre associé à un événement, compris entre 0 et 1, pour pouvoir se convertir en pourcentage de « chances » ; l'événement certain Ω se voit attribuer la probabilité 1 et l'événement impossible \emptyset la probabilité 0. Nous verrons dans l'exemple 6.6 que cette définition axiomatique est adaptée à la réalité, puisque la fréquence observée d'un événement a pour limite sa probabilité ainsi définie. D'autre part, si deux événements sont incompatibles, c'est-à-dire ne peuvent pas se réaliser simultanément, la probabilité de réalisation de l'un des deux sera la somme de leurs probabilités respectives (par exemple pour un jet de dé $P(1 \text{ ou } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3$). D'où la définition suivante :

DÉFINITION

On appelle probabilité P sur (Ω, \mathcal{A}) une application $P : \mathcal{A} \rightarrow [0, 1]$ telle que :

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

(ii) pour toute suite A_n d'événements incompatibles, soit $A_n \in \mathcal{A}$ avec $A_m \cap A_n = \emptyset$ pour $m \neq n$:

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n)$$

propriété dite de σ - additivité.

REMARQUE Pour toute suite d'événements quelconques, c'est-à-dire non disjoints, on a l'inégalité de Boole :

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} P(A_n)$$

Une probabilité est donc une application qui à un événement va associer un nombre. Le triplet (Ω, \mathcal{A}, P) s'appelle un **espace probabilisé**. Comme conséquences de la définition on déduit les propriétés suivantes.

Théorème 1.2

P_1 L'événement impossible est de probabilité nulle :

$$P(\emptyset) = 0$$

P_2 La probabilité de l'événement complémentaire d'un événement quelconque A s'obtient par :

$$P(\bar{A}) = 1 - P(A)$$

P_3 Si un événement en implique un autre, sa probabilité est plus petite :

$$A \subset B \Rightarrow P(A) \leq P(B)$$

P_4 La probabilité de l'union de deux événements s'obtient par la formule de Poincaré :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

a) Cas où Ω est fini

Si $\Omega = \{\omega_1, \dots, \omega_n\}$, la donnée de n nombres $p_i, 1 \leq i \leq n$, associés à chacun des événements élémentaires par $p_i = P(\{\omega_i\})$, tels que $0 \leq p_i \leq 1$ et $\sum_{i=1}^n p_i = 1$, suffit à déterminer une probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ par :

$$P(A) = \sum \{p_i / \omega_i \in A\}$$

c'est-à-dire que la probabilité d'un événement quelconque A de $\mathcal{P}(\Omega)$ est définie comme la somme des probabilités de tous les événements élémentaires qui y sont inclus.

b) Cas particulier : équiprobabilité

Il s'agit d'un cas assez fréquent où tous les événements élémentaires ont la même probabilité, ce qui correspond à la loi uniforme discrète définie par :

$$p_i = \frac{1}{n}, \quad 1 \leq i \leq n$$

Cette particularité est souvent sous-entendue ou précisée par l'affirmation que les résultats de l'expérience sont obtenus **au hasard**. On obtient alors :

$$P(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{1}{n} \times \text{card } A = \frac{\text{card } A}{\text{card } \Omega}$$

puisque $n = \text{card } \Omega$. Ce résultat s'énonce souvent sous la forme très dangereuse de la règle énoncée par Laplace au XVIII^e siècle :

$$P(A) = \frac{\text{nbre de cas favorables}}{\text{nbre de cas possibles}}$$

un cas favorable étant un événement élémentaire qui réalise A . On est alors ramené à un simple problème de dénombrement. Mais il faut bien faire attention que cette règle ne s'applique que dans le cas d'**équiprobabilité** des événements élémentaires. Si en jouant au loto il y a deux événements possibles, gagner ou non le gros lot, il n'y a malheureusement pas une chance sur deux pour l'événement favorable qui est de le gagner ! Cette règle ne peut donc en aucun cas servir de définition pour une probabilité.

ATTENTION !

C'est ici qu'il est important de bien préciser l'ensemble fondamental Ω . Si par exemple on lance deux pièces de monnaie identiques, on pourrait être tenté de retenir comme ensemble fondamental $\Omega = \{PP, PF, FF\}$, mais dans ce cas il n'y aurait pas équiprobabilité des événements élémentaires, car PP ou FF ne peut être obtenu que d'une seule façon alors que PF peut être obtenu de deux façons distinctes, le résultat pile pouvant être réalisé sur une pièce ou l'autre. Ces événements ont donc respectivement comme probabilité $1/4$, $1/4$ et $1/2$. Il vaut donc mieux faire comme si les pièces étaient distinctes et retenir $\Omega = \{PP, PF, FP, FF\}$ où les événements élémentaires sont équiprobables.

REMARQUES

1. Dans le cas où Ω est fini, on peut toujours construire un espace probabilisé tel que chaque événement élémentaire, au sens de singleton, ait une probabilité nulle. Par exemple pour $\Omega = \{a, b, c\}$ on choisit $\mathcal{A} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ et la probabilité P définie par $P(\{a\}) = 0$, $P(\{b, c\}) = 1$, ce qui correspond au cas de deux événements indiscernables, avec un seul événement élémentaire.

2. De même, on peut construire un espace probabilisé tel que chaque événement élémentaire, toujours au sens de singleton, ait une probabilité égale à 1. Dans l'ensemble précédent, la probabilité P est alors définie par $P(\{a\}) = 1$ et $P(\{b, c\}) = 0$.

3. Même dans le cas où les événements A et B ne sont pas disjoints, on peut avoir l'égalité $P(A \cup B) = P(A) + P(B)$; il suffit pour cela que $P(A \cap B) = 0$. Par exemple, définissons sur $\Omega = \{a, b, c\}$ la probabilité P par $P(\{a\}) = P(\{b\}) = 1/2$ et $P(\{c\}) = 0$. Avec $A = \{a, c\}$ et $B = \{b, c\}$ on obtient $P(A) = P(B) = 1/2$ et $P(A \cup B) = P(\Omega) = 1 = 1/2 + 1/2 = P(A) + P(B)$ et pourtant $A \cap B = \{c\} \neq \emptyset$. Terminons ce paragraphe par deux exemples, paradoxaux à première vue.

Exemple 1.13

Dans un tournoi exhibition de tennis, Djokovic doit affronter Nadal et Federer au cours de trois sets successifs où ses deux adversaires alterneront. Il remportera ce tournoi s'il gagne deux sets consécutifs. Il considère que la probabilité p de battre Federer est supérieure à celle q de battre Nadal : $p > q$. Quel adversaire va-t-il choisir d'affronter en premier ?

Pour une succession Fed-Nadal-Fed, sa probabilité de gain est :

$$p_1 = pq + (1 - p)qp.$$

Pour une succession Nadal-Fed-Nadal, cette probabilité est :

$$p_2 = qp + (1 - q)pq = p_1 + pq(p - q) > p_1.$$

Il doit donc choisir d'affronter le joueur le plus fort en premier, ce paradoxe s'expliquant par le fait qu'il a plus de chances de le battre une fois sur deux parties que sur une seule.

Exemple 1.14

Déterminons la répartition la plus probable de six atouts entre deux partenaires de bridge. Il y a $\binom{26}{13}$ répartitions possibles des 26 cartes des deux joueurs. Les dénombrements des répartitions de six atouts, et donc des vingt non-atouts, entre les deux joueurs sont indiquées dans le tableau ciaprès :

(1, 5)	(2, 4)	(3, 3)	(0, 6)
$2 \binom{6}{1} \binom{20}{12}$	$2 \binom{6}{2} \binom{20}{11}$	$2 \binom{6}{3} \binom{20}{11}$	$2 \binom{20}{13}$

ce qui correspond aux probabilités :

$$P\{(1, 5)\} = \frac{117}{805}, P\{(2, 4)\} = \frac{390}{805}, P\{(3, 3)\} = \frac{286}{805}, P\{(0, 6)\} = \frac{12}{805}$$

C'est donc la répartition (2, 4) qui est la plus probable.

On demande maintenant de répondre à la même question dans le cas où les deux partenaires ont fourni aux deux premiers tours d'atout. Les seules répartitions initiales possibles sont alors (2, 4) et (3, 3), donnant respectivement les répartitions (0, 2) et (1, 1) après deux tours d'atout. Il reste $\binom{22}{11}$ répartitions équiprobables entre les deux joueurs, les répartitions restantes des deux atouts, et donc des vingt non-atouts, ayant comme probabilités :

$$P\{(0, 2)\} = 2 \frac{\binom{20}{11}}{\binom{22}{11}} = \frac{10}{21}, \quad P\{(1, 1)\} = \frac{\binom{2}{1} \binom{20}{10}}{\binom{22}{11}} = \frac{11}{21}$$

La répartition la plus probable *a posteriori*, c'est-à-dire après les deux tours d'atouts, est donc issue de la moins probable des deux *a priori*.

2 Probabilités conditionnelles

On considère l'espace probabilisé (Ω, \mathcal{A}, P) et un événement particulier B de \mathcal{A} tel que $P(B) > 0$. La connaissance de la réalisation de B modifie la probabilité de réalisation d'un événement élémentaire, puisque l'ensemble des résultats possibles est devenu B et non plus Ω .

Exemple 1.15

À l'issue d'un jet de dé, on sait que le résultat est supérieur à trois et on s'intéresse à l'événement $A = \ll \text{obtenir une face paire} \gg$. Initialement on avait $P(A) = 3/6 = 1/2$; maintenant Ω est devenu $\Omega|B = \{4, 5, 6\}$ et $P(A|B) = 2/3 > 1/2$.

Cette nouvelle probabilité, notée $P(\cdot|B)$, est définie sur la tribu conditionnelle :

$$\mathcal{A}|B = \{A \cap B / A \in \mathcal{A}\}$$

par :

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

Seuls les événements ayant une partie commune avec B peuvent se réaliser et la figure 1.1 visualise cette situation où l'ensemble fondamental est devenu B et donc seule la part de A incluse dans B est prise en compte dans le calcul de la probabilité conditionnelle.

Vérifions que cette application π de \mathcal{A} dans \mathbb{R} , définie par $\pi(A) = P(A \cap B)/P(B)$ est bien une probabilité. On a bien entendu $\pi(A) \geq 0$ et comme $A \cap B \subset B$ on a également $\pi(A) \leq 1$. D'autre part :

$$\pi(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

donc π vérifie la première condition de la définition.

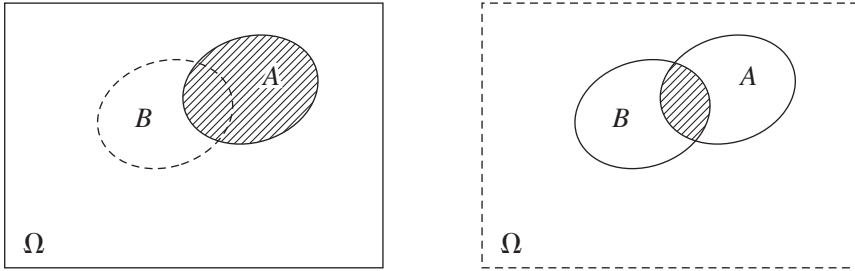


Figure 1.1

Enfin, si $A_n \in \mathcal{A}$ avec $A_m \cap A_n = \emptyset$ pour $m \neq n$, alors :

$$\pi \left(\bigcup_{n=0}^{\infty} A_n \right) = \frac{1}{P(B)} P \left\{ \bigcup_{n=0}^{\infty} (A_n \cap B) \right\} = \sum_{n=0}^{\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=0}^{\infty} \pi(A_n)$$

donc la condition 2 est aussi vérifiée.

Exemple 1.16

On tire sans remise deux cartes successivement d'un jeu de 52 et on cherche la probabilité de tirer un as au deuxième coup sachant que l'on en a obtenu un au premier. Avec des notations évidentes :

$$P(A_2 | A_1) = \frac{P(A_1 A_2)}{P(A_1)} = \frac{(4 \times 3)/(52 \times 51)}{4/52} = \frac{3}{51} = \frac{1}{17}$$

alors que la probabilité non conditionnelle est :

$$P(A_2) = P(A_1 A_2) + P(\bar{A}_1 A_2) = \frac{4 \times 3}{52 \times 51} + \frac{48 \times 4}{52 \times 51} = \frac{4}{52} = P(A_1) = \frac{1}{13}$$

donc valeur plus élevée (avoir tiré un as au premier coup diminue la probabilité d'en tirer un au second).

Exemple 1.17

On lance trois fois une pièce de monnaie et on considère les événements $A =$ « obtenir au moins deux face » et $B =$ « obtenir face au premier coup ». L'ensemble fondamental retenu est $\Omega = \{P, F\}^3$, ensemble des triplets ordonnés, bien que l'ordre des résultats n'intervienne pas, mais pour qu'il y ait équiprobabilité des événements élémentaires, soit $P(\{\omega\}) = 1/8$ puisque $\text{card } \Omega = 2^3 = 8$. Ces événements s'écrivent $A = \{FFP, FPF, PFF, FFF\}$ et $B = \{FFF, FFP, FPF, FPP\}$, avec $\text{card } A = \text{card } B = 4$ donc $P(A) = P(B) = 4/8 = 1/2$. Calculons maintenant la probabilité conditionnelle $P(A|B)$. On a $A \cap B = \{FFP, FPF, FFF\}$ donc $\text{card } A \cap B = 3$ et $P(A \cap B) = 3/8$. Ainsi :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3}{4} > \frac{1}{2}$$

la probabilité conditionnelle a ici augmenté.

La formule de définition de la probabilité conditionnelle peut aussi s'écrire, si $P(A) > 0$:

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

et s'appelle parfois **formule des probabilités composées**; elle se généralise par récurrence :

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ &\quad \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \prod_{k=2}^n P\left(A_k \mid \bigcap_{i=1}^{k-1} A_i\right) \end{aligned}$$

Exemple 1.18

Dans une urne qui contient deux boules rouges et trois noires, quatre personnes tirent successivement une boule sans la remettre; la première qui tire une rouge gagne. Calculons la probabilité de gain de chaque personne A, B, C et D :

$$P(A) = P(R_1) = \frac{2}{5}$$

$$P(B) = P(N_1)P(R_2|N_1) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

$$P(C) = P(N_1)P(N_2|N_1)P(R_3|N_1N_2) = \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3} = \frac{1}{5}$$

$$P(D) = P(N_1)P(N_2|N_1)P(N_3|N_1N_2)P(R_4|N_1N_2N_3) = \frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} \times \frac{2}{2} = \frac{1}{10}$$

3 Théorème de Bayes

Nous allons voir à partir d'un exemple dans quelles conditions on peut être amené à utiliser la formule, ou théorème, de Bayes.

Exemple 1.19

Considérons une expérience aléatoire qui se déroule en deux étapes : on tire au sort entre deux urnes U_1 et U_2 , avec des probabilités respectives de $1/5$ et $4/5$, puis on tire une boule dans l'urne choisie. Leurs compositions respectives sont 6 blanches, 4 noires et 3 blanches, 7 noires. La probabilité *a priori* de U_1 est donc $1/5$. Sa probabilité *a posteriori*, sachant qu'on a obtenu une boule blanche, va être plus élevée car la probabilité de tirer une blanche dans U_1 est plus forte que dans U_2 .

On a :

$$P(U_1|B) = \frac{P(U_1 \cap B)}{P(B)}$$

avec :

$$\begin{aligned}
 P(U_1 \cap B) &= P(U_1)P(B|U_1) = \frac{1}{5} \times \frac{6}{10} = \frac{3}{25} \\
 P(B) &= P(B \cap U_1) + P(B \cap U_2) \\
 &= P(U_1)P(B|U_1) + P(U_2)P(B|U_2) \\
 &= \frac{1}{5} \times \frac{6}{10} + \frac{4}{5} \times \frac{3}{10} = \frac{9}{25}
 \end{aligned}$$

ainsi :

$$P(U_1 | B) = \frac{3/25}{9/25} = \frac{1}{3} > P(U_1) = \frac{1}{5}$$

Cet exemple peut se traiter par application de la formule de Bayes que nous allons établir en considérant un **système complet d'événements**, c'est-à-dire une partition de Ω en événements $\{A_1, \dots, A_n\}$ de probabilités strictement positives, $P(A_i) > 0$ pour $1 \leq i \leq n$, et incompatibles deux à deux, i.e. avec $A_i \cap A_j = \emptyset$ pour $i \neq j$ et $\sum_{i=1}^n P(A_i) = 1$. On suppose que les probabilités des événements inclus dans chacun des A_i sont connues et on va donc décomposer un événement quelconque B sur ce système :

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^n A_i \right) = \bigcup_{i=1}^n (A_i \cap B)$$

On aboutit ainsi à la **formule de la probabilité totale** :

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

Ceci va nous permettre de calculer les probabilités *a posteriori* $P(A_i|B)$, après réalisation d'un événement B , à partir des probabilités *a priori* $P(A_i)$, $1 \leq i \leq n$:

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)}$$

résultat appelé **formule de Bayes** ou parfois **théorème de Bayes**.

Exemple 1.20

On tire au sort entre trois urnes dont les compositions sont indiquées dans le tableau ci-après. Sachant qu'on a obtenu une boule rouge, on se pose la question de savoir quelle est la probabilité qu'elle provienne de l'urne U_2 .

	Rouge	Bleue	Verte
U_1	3	4	1
U_2	1	2	3
U_3	4	3	2

$$P(R) = \sum_{i=1}^3 P(R \cap U_i) = \sum_{i=1}^3 P(U_i)P(R|U_i) = \frac{1}{3} \left(\frac{3}{8} + \frac{1}{6} + \frac{4}{9} \right)$$

La probabilité *a posteriori* de l'urne U_2 étant donc :

$$\begin{aligned} P(U_2 | R) &= \frac{P(U_2)P(R|U_2)}{P(R)} = \frac{1/6}{3/8 + 1/6 + 4/9} \\ &= \frac{12}{71} = \frac{36}{213} < \frac{71}{213} = P(U_2) \end{aligned}$$

4 Indépendance en probabilité

DÉFINITION

Deux événements A et B sont dits indépendants, relativement à la probabilité P , si :

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

La probabilité de réalisation simultanée de deux événements indépendants est égale au produit des probabilités que chacun de ces événements se produise séparément. En conséquence, si $P(B) > 0$:

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

la réalisation d'un événement ne modifie pas la probabilité de réalisation de l'autre.

ATTENTION !

Ne pas confondre indépendance avec incompatibilité, car dans ce dernier cas $A \cap B = \emptyset$ et $P(A \cap B) = 0$.

Exemple 1.21

On jette un dé rouge et un dé vert et on considère les événements $A =$ « le dé vert marque 6 », et $B =$ « le dé rouge marque 5 ». Il nous faut démontrer que ces deux événements sont indépendants (bien entendu ce résultat est évident, il n'y a pas d'influence d'un dé sur l'autre !). L'ensemble fondamental retenu est $\Omega = E \times E$, avec

$E = \{1, 2, 3, 4, 5, 6\}$, sur lequel il y a équiprobabilité : $P(\{\omega\}) = 1/6^2$. Comme $A = \{6\} \times E$, $B = E \times \{5\}$ et $A \cap B = \{(6, 5)\}$ on obtient :

$$P(A) = \frac{\text{card } A}{\text{card } \Omega} = \frac{6}{6^2} = \frac{1}{6}$$

$$P(B) = \frac{\text{card } B}{\text{card } \Omega} = \frac{6}{6^2} = \frac{1}{6}$$

$$P(A \cap B) = \frac{\text{card } (A \cap B)}{\text{card } \Omega} = \frac{1}{6^2} = P(A)P(B)$$

et les événements A et B sont donc bien indépendants.

Indépendance mutuelle

Si l'on considère n événements A_i , avec $n > 2$, il y a lieu de distinguer l'indépendance deux à deux qui impose :

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad 1 \leq i \neq j \leq n$$

de l'indépendance mutuelle, condition plus forte qui s'écrit :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}), \quad 2 \leq k \leq n,$$

pour tous les sous-ensembles $\{i_1, \dots, i_k\}$ de $\{1, 2, \dots, n\}$, ce qui correspond à :

$$\sum_{k=2}^n \binom{n}{k} = 2^n - \binom{n}{0} - \binom{n}{1} = 2^n - 1 - n$$

conditions qui doivent être réalisées.

Exemple 1.22

On lance deux pièces de monnaie distinctes et on s'intéresse aux événements $A = \text{« obtenir pile sur la pièce 1 »}$, $B = \text{« obtenir pile sur la pièce 2 »}$ et $C = \text{« obtenir pile sur une seule pièce »}$. À chaque pièce est attaché le même ensemble fondamental $E = \{P, F\}$ et à l'expérience l'ensemble $\Omega = E \times E$. On peut écrire $A = \{P\} \times E = \{PP, PF\}$ donc $P(A) = 2/4$, $B = E \times \{P\} = \{PP, FP\}$ donc $P(B) = 2/4$ et $C = \{PF, FP\}$ donc $P(C) = 2/4$. On considère maintenant les événements deux à deux : $A \cap B = \{PP\}$ donc $P(A \cap B) = 1/4 = P(A)P(B)$, $A \cap C = \{PF\}$ donc $P(A \cap C) = 1/4 = P(A)P(C)$ et $B \cap C = \{FP\}$ donc $P(B \cap C) = 1/4 = P(B)P(C)$, ces événements sont indépendants deux à deux. Mais $A \cap B \cap C = \emptyset$ donc $P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$ et ces événements ne sont pas indépendants dans leur ensemble.

5 Pour aller plus loin : éléments de combinatoire

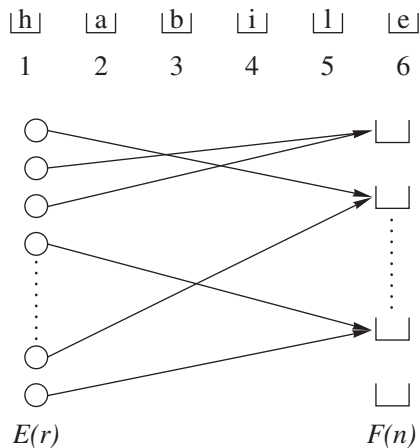
Dans le cas particulier où il y a équiprobabilité sur un ensemble fini d'événements élémentaires, nous avons vu (cf. 1.3) que le calcul d'une probabilité se ramenait à un problème de dénombrement. Ce type de problèmes est parfois très complexe et nécessite de connaître quelques éléments de combinatoire, permettant d'exprimer par une formule le nombre de configurations ayant des propriétés données. Examinons les configurations usuelles.

5.1 Permutations avec répétition

Une **permutation avec répétition** de r objets pris parmi n est une suite **ordonnée** de r éléments choisis parmi n , et pouvant se répéter.

Exemple 1.23

Un mot de six lettres est une permutation avec répétition de six objets choisis parmi un ensemble, l'alphabet, de 26 éléments : coucou, habile, garage...



Une telle permutation peut être représentée par les r objets rangés dans des cases numérotées de 1 à r . Pour chacune de ces r cases, il y a n choix possibles de l'objet à ranger, donc le nombre total de ces permutations est :

$$P_n^r = n^r$$

Exemple 1.24

Le nombre de mots possibles de trois lettres est $26^3 = 17\,576$.

Cela correspond au cardinal de l'ensemble fondamental associé à r tirages avec remise (schéma binomial) dans une urne contenant n objets distincts (ou éventuellement

considérés comme tels de façon justement à obtenir des événements équiprobables) et tenant compte de l'ordre des tirages.

C'est aussi le nombre d'applications quelconques d'un ensemble E à r éléments dans un ensemble F à n éléments, une application quelconque pouvant être définie comme un rangement de r objets dans n boîtes, chaque boîte pouvant contenir zéro, un ou plusieurs objets.

5.2 Permutations sans répétition ou arrangements

Une **permutation sans répétition**, ou **arrangement**, de r objets pris parmi n est une suite **ordonnée** de r éléments choisis parmi n , et qui ne peuvent pas se répéter.

Exemple 1.25

Le quinté est un exemple d'arrangement de cinq chevaux pris parmi tous les partants de la course.

Une telle permutation peut être représentée par les r objets rangés dans des cases numérotées de 1 à r . Pour la première case, il y a n choix possibles, pour la deuxième il n'y en a plus que $n - 1$, et pour la r -ème il n'en reste plus que $n - r + 1$; le nombre d'arrangements est donc :

$$A_n^r = n(n - 1) \dots (n - r + 1) = \frac{n!}{(n - r)!}$$

Cela correspond au cardinal de l'ensemble fondamental associé à r tirages sans remise (schéma hypergéométrique) dans une urne contenant n objets distincts et tenant compte de l'ordre des tirages.

C'est aussi le nombre d'applications injectives d'un ensemble E à r éléments dans un ensemble F à n éléments, une application injective pouvant être définie comme un rangement de r objets dans n boîtes, chaque boîte ne pouvant contenir que zéro ou un objet. Il faut bien sûr que $n = \text{card } F \geq \text{card } E = r$.

Exemple 1.26

Le nombre de tiercés dans l'ordre avec quinze partants est :

$$A_{15}^3 = 15 \times 14 \times 13 = 2730$$

Permutation : il s'agit du cas particulier $n = r$.

Une **permutation** est donc une suite ordonnée de n objets distincts ; le nombre de permutations est :

$$P_n = A_n^n = n!$$

Exemple 1.27

Le classement de cinq candidats à une épreuve forme une permutation, il y en a $5! = 120$.

C'est aussi le nombre de bijections d'un ensemble E à n éléments dans un ensemble F à n éléments.

5.3 Permutations avec répétition de n objets, dont k seulement sont distincts

Il s'agit d'une suite ordonnée de n objets choisis dans k classes distinctes, le nombre d'objets de la classe i étant n_i , $1 \leq i \leq k$, avec bien sûr $n_1 + \dots + n_k = n$. Prenons l'exemple de n boules numérotées, extraites sans remise d'une urne, et de k couleurs distinctes. Il y a $n!$ tirages ordonnés possibles ; mais si on efface les numéros des n_1 boules rouges par exemple, les $n_1!$ permutations de ces boules conduisent à la même permutation, donc le nombre de permutations distinctes devient $n!/n_1!$. Il en est bien sûr de même pour toutes les autres couleurs et par conséquent le nombre de permutations est :

$$\frac{n!}{n_1! \dots n_k!}$$

Exemple 1.28

Cherchons le nombre de mots différents formés avec les lettres du mot barbare. Il y a sept lettres, mais seulement quatre catégories distinctes, soit un nombre de mots distincts égal à :

$$\frac{7!}{2!2!2!1!} = 630$$

C'est aussi le nombre de partitions de n objets en k classes d'effectifs n_i , $1 \leq i \leq k$, fixés. Dans le cas particulier où il n'y a que deux classes, on obtient le coefficient binomial $\binom{n}{n_1}$ qui représente le nombre de sous-ensembles à n_1 éléments que l'on peut extraire d'un ensemble à n éléments.

5.4 Combinaisons (sans répétition)

Une **combinaison** est un sous-ensemble **non ordonné** de r objets choisis dans un ensemble qui en contient n . Ces sous-ensembles sont au nombre de :

$$\binom{n}{r} = \frac{A_n^r}{r!} = \frac{n!}{r!(n-r)!}$$

Exemple 1.29

Le nombre de tiercés dans le désordre avec quinze chevaux au départ est :

$$\binom{15}{3} = 455$$

Le sous-ensemble choisi définit bien sûr le sous-ensemble restant et donc :

$$\binom{n}{r} = \binom{n}{n-r}$$

Exemple 1.30

On joue au poker d'as avec quatre dés identiques et on souhaite calculer la probabilité des différents résultats possibles. Bien que les dés ne soient pas distincts, nous retenons comme ensemble fondamental $\Omega = E^4$, où $E = \{1, 2, 3, 4, 5, 6\}$ est l'ensemble des résultats associés à un dé particulier ; ainsi tous les événements élémentaires sont équiprobables et ont même probabilité $1/6^4$. En effet, un résultat est une permutation de quatre objets choisis parmi six. Un carré est un résultat de la forme $(aaaa)$ avec six choix possibles pour la hauteur :

$$P(\text{carré}) = \frac{6}{6^4} = \frac{1}{6^3}$$

Un brelan est un résultat de la forme $(aaab)$ avec six choix pour la hauteur a , cinq choix pour b et quatre places possibles, soit :

$$P(\text{brelan}) = \frac{6 \times 5 \times 4}{6^4} = \frac{20}{6^3}$$

Une double paire est un résultat de la forme $(aabb)$ avec $\binom{6}{2}$ choix possibles pour les paires et une permutation de quatre objets dont deux seulement sont distincts, soit $\frac{4!}{2!2!}$ et :

$$P(\text{double paire}) = \frac{15}{6^3}$$

Une paire est un résultat de la forme $(aabc)$ avec six choix possibles pour la hauteur, $\binom{5}{2}$ choix possibles pour les hauteurs qui l'accompagnent et un nombre de permutations de quatre objets dont trois distincts $\frac{4!}{2!1!1!}$ soit :

$$P(\text{paire}) = \frac{120}{6^3}$$

Le nombre de résultats quelconques $(abcd)$ est le nombre de permutations sans répétition de quatre objets pris parmi six, d'où :

$$P(\text{quelconque}) = \frac{6 \times 5 \times 4 \times 3}{6^4} = \frac{60}{6^3}$$

On vérifie bien que la somme des probabilités est égale à $\frac{216}{6^3} = 1$.

Tous les sous-ensembles que l'on peut extraire d'un ensemble E à n éléments peuvent contenir $0, 1, \dots, n$ éléments, d'où la relation :

$$\text{card } \mathcal{P}(E) = 2^n = \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n}$$

Enfin, parmi les sous-ensembles à r éléments, il y a ceux qui contiennent un objet particulier, en nombre $\binom{n-1}{r-1}$ et ceux qui ne le contiennent pas, en nombre $\binom{n-1}{r}$ d'où la relation :

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

qui permet de construire le triangle de Pascal.

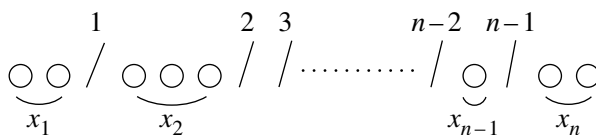
5.5 Combinaisons avec répétition

Une **combinaison avec répétition** est un sous-ensemble **non ordonné** de r objets choisis dans un ensemble qui en contient n et qui peuvent se répéter.

⋮ **Exemple 1.31**

Les r objets tirés avec remise dans une urne qui en contient n distincts forment une combinaison avec répétition puisque seule compte la nature des objets tirés, indépendamment de l'ordre des tirages.

Une telle combinaison sera représentée sous la forme (x_1, \dots, x_n) où $x_i, 1 \leq i \leq n$ représentera le nombre d'objets i appartenant au sous-ensemble, avec bien sûr $x_i \geq 0$ et $x_1 + \dots + x_n = r$. Elle peut être symbolisée par une suite formée d'objets \circ et de séparateurs $/$ avec x_1 objets \circ avant le premier séparateur, x_2 objets entre le premier et le deuxième $/$, ..., x_n objets \circ après le $(n-1)$ -ème $/$. Elle est donc caractérisée par la place des r objets \circ dans la suite des $r+n-1$ symboles.



Par exemple, la combinaison constituée de deux objets de la catégorie 1, d'un objet de la catégorie 3 et de trois objets de la catégorie 4 s'écrit $(2, 0, 1, 3)$ ou $2+0+1+3 = 6$ représentée sous la forme symbolique : $\circ \circ // \circ / \circ \circ \circ$. Réciproquement, la suite de symboles $/ \circ / \circ \circ \circ / \circ \circ$ représente une combinaison d'un objet de la catégorie 2, de trois objets de la catégorie 3 et de deux objets de la catégorie 4. Il y a donc une bijection entre une combinaison et une telle suite, et le nombre de combinaisons avec répétition est donc égal au nombre de rangements des r objets \circ (ou des $n-1$ séparateurs $/$) dans les $r+n-1$ cases, soit :

$$C_n^r = \binom{r+n-1}{r}$$

5.6 Partitions

Le nombre de partitions de n objets en r classes d'effectifs non fixés est appelé **nombre de Stirling de deuxième espèce** et se calcule à partir de la récurrence :

$$S_{n+1}^r = S_n^{r-1} + rS_n^r, \quad 1 < r < n$$

avec bien sûr $S_n^1 = 1$ et $S_n^2 = 2^{n-1} - 1$. Les partitions de $n + 1$ objets en r classes se décomposent en effet en celles où le $(n + 1)$ -ème objet constitue une classe à lui tout seul, il y en a S_n^{r-1} , et celles où on l'intègre à une classe déjà formée, il y en a rS_n^r .

Une surjection d'un ensemble E à n éléments dans un ensemble F à r éléments correspond à un rangement de n objets dans r boîtes dont aucune n'est vide, c'est-à-dire à une partition de ces n objets en r classes, où l'ordre des classes intervient. À une partition donnée correspondent $r!$ surjections distinctes, donc le nombre total de surjections est $r!S_n^r$.

Le nombre de partitions de n objets distincts en k classes dont n_j ont le même effectif j , $1 \leq j \leq k$, (avec bien sûr $\sum_{j=1}^k jn_j = n$) est :

$$\frac{n!}{(1!)^{n_1} (2!)^{n_2} \dots (k!)^{n_k} n_1! \dots n_k!}$$

- 1 Le modèle associé à une expérience aléatoire est arbitraire et doit être choisi le plus simple possible, compte tenu du problème à résoudre. Il est très souvent souhaitable de retenir un ensemble fondamental tel que les événements élémentaires soient équiprobables, même si dans la réalité on ne peut pas les distinguer, car le calcul des probabilités est alors ramené à un problème de dénombrement.
- 2 Une probabilité est une application et non pas un nombre. Il faut toujours vérifier que la somme des probabilités des événements élémentaires est égale à un et, dans le cas d'un nombre fini d'événements, ne pas calculer la probabilité du dernier en retranchant à un la probabilité de tous les autres, car cela exclut cette vérification permettant de déceler éventuellement une erreur dans le calcul de ces probabilités.
- 3 L'indépendance de deux événements est définie relativement à une probabilité. L'indépendance mutuelle d'un nombre quelconque d'événements est une condition plus forte que l'indépendance de ces événements pris seulement deux à deux.
- 4 La notion d'indépendance est à distinguer de celle de non-causalité, la condition $P(A|B \cap C) = P(A|B)$ se traduisant par « C ne cause pas A ».



Entraînez-vous

- 1.1** Une machine fabrique des objets qui sont classés en défectueux, codés 0, et non défectueux, codés 1. On prélève toutes les heures les trois derniers objets produits par cette machine. On demande de préciser l'ensemble fondamental associé à cette expérience et d'écrire les événements suivants : $A = \ll \text{le premier objet est défectueux} \gg$; $B = \ll \text{le dernier objet est non défectueux} \gg$; $C = \ll \text{les premier et dernier objets sont défectueux} \gg$; $D = \ll \text{aucun objet n'est défectueux} \gg$; $E = \ll \text{deux objets sont défectueux} \gg$; $F = \ll \text{au plus un objet est défectueux} \gg$.
- 1.2** Soit $\Omega = \{a, b, c\}$ et considérons les ensembles $\mathcal{A}_1 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ et $\mathcal{A}_2 = \{\emptyset, \{b\}, \{a, c\}, \Omega\}$. Montrer que ce sont des tribus sur Ω . Les ensembles $\mathcal{A}_1 \cap \mathcal{A}_2$ et $\mathcal{A}_1 \cup \mathcal{A}_2$ sont-ils des tribus sur Ω ?
- 1.3** Soit $\Omega = \mathbb{N}$ et \mathcal{A} la famille des parties A de Ω , telles que A est fini ou bien \bar{A} est fini. Montrer que \mathcal{A} est une algèbre mais pas une σ -algèbre. On pourra considérer les ensembles $A_n = \{2n\}, n \in \mathbb{N}$.
- 1.4** On effectue deux tirages successifs dans une urne qui contient une boule blanche et deux boules noires identiques. La première boule tirée n'est pas remise dans l'urne, mais est remplacée par une boule de l'autre couleur (blanche si on a tiré une noire et vice-versa).

1. Construire l'ensemble fondamental Ω associé à cette expérience aléatoire, en tenant compte de l'ordre des tirages.

2. Montrer que l'ensemble des parties de Ω défini par :

$$\mathcal{A} = \{\emptyset, \{(NN)\}, \{(BN), (NB)\}, \Omega\}$$

est une tribu. La notation (BN) représente par exemple l'événement élémentaire « tirer une boule blanche, puis une noire ».

3. Déterminer la probabilité de chacun des événements élémentaires constituant Ω .

- 1.5** On lance un dé quatre fois de suite et le joueur A marque 1 point si le chiffre obtenu est 1 ou 2 ; sinon c'est le joueur B qui marque un point. Le joueur qui gagne la partie est celui qui a le plus de points à l'issue de ces quatre lancers. Calculer la probabilité de gain de chaque joueur, en précisant l'ensemble fondamental Ω retenu pour modéliser ce problème.
- 1.6** Une urne contient une boule blanche et une boule noire. On effectue des tirages avec remise dans cette urne jusqu'à obtenir une boule blanche, ajoutant une boule noire après chaque tirage d'une boule noire. Calculer la probabilité d'effectuer n tirages, $n \in \mathbb{N}^*$.
- 1.7** Un laboratoire a mis au point un alcootest dont les propriétés sont décrites ci-après :
– il se révèle positif pour quelqu'un qui n'est pas en état d'ébriété dans 2 % des cas ;
– il se révèle positif pour quelqu'un qui est en état d'ébriété dans 96 % des cas.
- Dans un département donné, on sait que 3 % des conducteurs sont en état d'ébriété. Si le contrôle se révèle positif, quelle est la probabilité que ce conducteur ne soit pas malgré tout en état d'ébriété ?

- 1.8** Le professeur Tournesol cherche ses gants qu'il pense avoir rangés avec une probabilité p dans sa commode. Si c'est le cas, il les a mis au hasard dans l'un des 4 tiroirs. Sachant qu'il ne les a pas trouvés dans les 3 premiers, calculer la probabilité qu'ils soient dans le quatrième.
- 1.9** On effectue deux tirages successifs avec remise dans une première urne qui contient autant de boules noires que de boules blanches. On place ensuite dans une seconde urne vide deux boules de la même couleur que celles qui ont été tirées.
1. Indiquer les différentes compositions possibles de cette urne et la probabilité associée.
 2. On effectue ensuite des tirages successifs avec remise dans cette seconde urne. On note p_n la probabilité que cette urne contienne deux boules blanches, sachant que l'on n'a obtenu que des boules blanches au cours des n premiers tirages. Calculer p_1, p_2 , puis p_n pour $n \geq 2$ et ensuite la limite de p_n quand n devient infini. Cette limite était-elle prévisible ?
- 1.10** On jette ensemble cinq dés identiques et à chaque coup on enlève les as (chiffre 1) qui sont sortis avant de relancer les autres dés. Quelle est la probabilité d'obtenir un poker d'as, c'est-à-dire cinq as, en trois coups au plus ?
- 1.11** Un document contient quatre erreurs et à chaque relecture la probabilité de détection d'une erreur ayant subsisté est de $1/3$. Quelle est la probabilité p_n qu'il ne subsiste aucune faute après n relectures, $n \in \mathbb{N}^*$?
- 1.12** Un étudiant doit répondre à un questionnaire à choix multiple où cinq réponses sont proposées à une question, une seule étant correcte. Quand l'événement $A = \ll$ l'étudiant a bien travaillé dans l'année \gg est réalisé, la réponse est fournie avec exactitude. Dans le cas contraire, l'étudiant répond au hasard. Si l'événement $B = \ll$ il a fourni la réponse correcte \gg est réalisé, calculer la probabilité $P(A|B)$ en fonction de $p = P(A)$.
- 1.13** Dans la rue principale de Champignac, chacun des deux feux de circulation successifs est vert les deux tiers du temps. Si vous roulez à la vitesse réglementaire de 50 km/h, après avoir franchi le premier feu au vert, vous trouverez le second au vert trois fois sur quatre. Si vous grillez le premier feu au rouge, calculez la probabilité que le second soit aussi au rouge si vous roulez toujours à la vitesse réglementaire de 50 km/h, en précisant l'ensemble fondamental retenu pour ce problème.
- 1.14** Deux joueurs de tennis s'entraînent au service, le joueur A servant une fois sur trois et le joueur B deux fois sur trois. Leurs pourcentages de réussite au service sont respectivement de 90 % et de 60 %.
1. Si vous assistez à un service réussi au cours de cet entraînement, quelle est la probabilité p_B qu'il s'agisse du joueur B ?
 2. Calculer cette probabilité p_B en fonction du pourcentage p de réussite du joueur B . Pour quelle valeur de p_B cette probabilité est-elle égale à $1/2$?
 3. Que vaut p_B dans le cas où les joueurs A et B ont le même pourcentage de réussite au service ? Le résultat était-il prévisible ?
- 1.15** Soit $\Omega = \{a, b, c, d\}$ avec équiprobabilité des événements élémentaires sur $\mathcal{P}(\Omega)$. Les événements $A = \{a, b\}$, $B = \{a, c\}$ et $C = \{a, d\}$ sont-ils indépendants ?

- 1.16** Les événements A et B étant indépendants, montrer que les événements A et \bar{B} ainsi que \bar{A} et \bar{B} sont aussi indépendants. Si l'événement C est indépendant de A et de B , est-il aussi indépendant de $A \cup B$? L'événement \bar{C} est-il indépendant de $A \cap B$?
- 1.17** On effectue trois lancers successifs d'une pièce de monnaie. Les événements $A = \{FFF, PFF, FPP, PPP\}$, $B = \{FFP, FPP, PPF, PPP\}$ et $C = \{FFP, FPF, PFP, PPP\}$ sont-ils indépendants?
- 1.18** Problème de rencontres. Lors d'un bal auquel participent n couples, le choix de sa cavalière pour la première danse se fait au hasard. La probabilité de danser avec sa femme est donc $1/n$, valeur faible si n est grand. Cependant, on demande de montrer que la probabilité qu'il y ait au moins une rencontre entre un danseur et sa femme est supérieure à la probabilité qu'il n'y ait pas.
- 1.19** Avant le tirage au sort des quarts de finale de la Ligue des Champions de football, il reste 4 équipes anglaises. Calculer la probabilité p qu'il y ait au moins un quart de finale qui oppose 2 équipes anglaises.



1.1 On choisit comme ensemble fondamental $\Omega = \{0, 1\}^3$, c'est-à-dire l'ensemble des triplets ordonnés, bien que l'ordre soit sans importance, car cela permet d'obtenir des événements élémentaires équiprobables et facilite donc le calcul ultérieur des probabilités. Les événements indiqués s'écrivent alors : $A = \{0\} \times \{0, 1\}^2$, $B = \{0, 1\}^2 \times \{1\}$, $C = \{0\} \times \{0, 1\} \times \{0\} = \{(000), (010)\}$, $D = \{(111)\}$, $E = \{(001), (010), (100)\}$, $F = \{(111), (110), (101), (011)\}$.

1.2 Si on note $A_1 = \{a\}$ et $A_2 = \{b\}$, les deux ensembles considérés s'écrivent $\mathcal{A}_i = \{\emptyset, A_i, \overline{A_i}, \Omega\}$, avec $i = 1, 2$, et sont donc des tribus sur Ω . L'ensemble $\mathcal{A}_1 \cap \mathcal{A}_2 = \{\emptyset, \Omega\}$ est la tribu grossière et $\mathcal{A}_1 \cup \mathcal{A}_2 = \{\emptyset, \{a\}, \{b\}, \{a, c\}, \{b, c\}, \Omega\}$ n'est pas une tribu car $\{a, c\} \cap \{b, c\} = \{c\} \notin \mathcal{A}_1 \cup \mathcal{A}_2$.

1.3 La famille \mathcal{A} est non vide puisqu'elle contient par exemple les ensembles A_n . Par définition, elle est fermée pour le passage au complémentaire. Si deux éléments A et B de la famille sont finis alors $A \cup B$ l'est aussi ; si l'un des deux n'est pas fini alors $A \cup B$ n'est pas fini mais son complémentaire l'est, comme intersection de deux ensembles dont l'un est fini. Ainsi dans les deux cas $A \cup B$ appartient à \mathcal{A} qui est aussi fermée pour l'union donc est une algèbre.

Mais l'union dénombrable $\bigcup_{n \in \mathbb{N}} A_n = 2\mathbb{N}$ est l'ensemble des entiers pairs qui n'est pas fini. Son complémentaire est l'ensemble des entiers impairs, qui n'est pas fini non plus, et donc cet ensemble n'appartient pas à la famille \mathcal{A} . Il n'y a pas fermeture pour l'union dénombrable donc cette algèbre n'est pas une σ -algèbre.

1.4 1. L'ensemble fondamental est constitué de tous les événements élémentaires possibles, écrits sous la forme de couples dont la première lettre désigne la couleur de la première boule tirée et la seconde la couleur de la boule extraite ensuite. Si la première boule tirée est blanche, l'urne ne contient plus que des boules noires et le seul événement possible est donc représenté par le couple ordonné (BN) . Si la première boule tirée est noire, la composition de l'urne est $\{B, B, N\}$ et il y a alors deux événements possibles, (NB) et (NN) . On a donc :

$$\Omega = \{(BN), (NB), (NN)\}$$

2. L'ensemble \mathcal{A} est un ensemble non vide de parties de Ω qui peut s'écrire $\mathcal{A} = \{\emptyset, A, \overline{A}, \Omega\}$ en ayant posé $A = (NN)$. Il est facile de voir qu'il est fermé pour l'union et le passage au complémentaire, donc est une algèbre. Comme \mathcal{A} comporte un nombre fini d'éléments, cette algèbre est aussi une σ -algèbre, ou tribu.

3. D'après la question 1 :

$$P(BN) = \frac{1}{3} \times 1 \quad P(NB) = \frac{2}{3} \times \frac{2}{3} \quad P(NN) = \frac{2}{3} \times \frac{1}{3}$$

1.5 Si on note A l'événement « le joueur A a marqué un point », l'ensemble fondamental est :

$$\Omega = \{A, \overline{A}\}^4$$

Il n'y a pas équiprobabilité ici car $P(A) = \frac{2}{6}$ et $P(\bar{A}) = \frac{4}{6}$. Pour que le joueur A gagne, il faut qu'il marque 3 ou 4 points et il en est de même pour le joueur B . Les probabilités de gain de chaque joueur sont donc :

$$P(AG) = 4 \left(\frac{1}{3}\right)^3 \frac{2}{3} + \left(\frac{1}{3}\right)^4 = \frac{1}{9}$$

$$P(BG) = 4 \left(\frac{2}{3}\right)^3 \frac{1}{3} + \left(\frac{2}{3}\right)^4 = \frac{16}{27}$$

La somme de ces probabilités n'est pas égale à 1 car les joueurs peuvent marquer 2 points chacun et être ex-aequo :

$$P(EX) = \binom{4}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^2 = \frac{8}{27}$$

- 1.6** Effectuer n tirages correspond à l'événement, noté B_n , « obtenir une boule blanche au n -ème tirage et des boules noires à tous les tirages précédents ». Nous obtenons ainsi :

$$P(B_1) = \frac{1}{2}$$

$$P(B_2) = \frac{1}{2} \times \frac{1}{3}$$

...

$$P(B_n) = \frac{1}{2} \times \frac{2}{3} \times \dots \times \frac{n-1}{n} \times \frac{1}{n+1} = \frac{1}{n(n+1)}$$

- 1.7** Nous noterons A^+ l'événement « l'alcootest est positif » et E l'événement « le conducteur est en état d'ébriété ». Les hypothèses fournies se traduisent en termes de probabilité par : $P(A^+|\bar{E}) = 0,02$, $P(A^+|E) = 0,96$ et $P(E) = 0,03$. La probabilité demandée est :

$$\begin{aligned} P(\bar{E}|A^+) &= \frac{P(\bar{E} \cap A^+)}{P(A^+)} = \frac{P(\bar{E})P(A^+|\bar{E})}{P(\bar{E})P(A^+|\bar{E}) + P(E)P(A^+|E)} \\ &= \frac{0,02 \times 0,97}{0,02 \times 0,97 + 0,96 \times 0,03} = 0,402 \end{aligned}$$

ce qui traduit la non-fiabilité de cet alcootest.

- 1.8** On note C l'événement « les gants sont rangés dans la commode », 4 « ils sont dans le tiroir 4 » et $\bar{3}$ « ils ne sont pas dans les 3 premiers tiroirs ». La probabilité cherchée s'écrit $P(4|\bar{3})$ et se calcule par la formule de Bayes. Comme $P(\bar{3}|C) = 1/4$ et $P(\bar{3}|\bar{C}) = 1$ on obtient :

$$P(4|\bar{3}) = \frac{P(4 \cap \bar{3})}{P(\bar{3})} = \frac{P(4)}{P(\bar{3})} = \frac{p/4}{p/4 + 1 - p} = \frac{p}{4 - 3p}$$

1.9 1. L'ensemble fondamental associé à ces deux tirages indépendants est $\Omega = \{B, N\}^2$. Il y a 3 compositions possibles de cette urne avec $P(BB) = P(NN) = 1/4$ et $P(BN) = 2/4$.

2. On utilise la formule de Bayes pour calculer $p_1 = P(BB|B_1)$, avec $P(B_1|BB) = 1$, $P(B_1|NN) = 0$ et $P(B_1|NB) = 1/2$; ainsi $p_1 = 1/2$. On obtient de la même façon $p_2 = 2/3$ puis pour $n \geq 2$:

$$p_n = \frac{P(BB)}{P(BB) + P(B_1 \dots B_n | BN) / 2} = \frac{1}{1 + 1/2^{n-1}}$$

Quand n devient infini $p_n \rightarrow 1$, résultat prévisible car si on ne tire que des boules blanches, c'est sans doute qu'il n'y a pas de boule d'une autre couleur dans l'urne.

1.10 Soit N_i , $1 \leq i \leq 5$ le nombre de jets nécessaires pour obtenir un as avec le dé i . Le poker d'as est obtenu en moins de trois coups si $\max \{N_i / 1 \leq i \leq 5\} \leq 3$, événement qui est équivalent à :

$$\bigcap_{i=1}^5 \{N_i \leq 3\}$$

Tous ces événements sont indépendants et ont la même probabilité :

$$P(N_i \leq 3) = \sum_{k=1}^3 P(N_i = k) = \sum_{k=1}^3 \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = 1 - \left(\frac{5}{6}\right)^3$$

La probabilité demandée est donc :

$$\left[1 - \left(\frac{5}{6}\right)^3\right]^5 = 0,013$$

1.11 Pour qu'une faute ne soit pas corrigée il faut qu'elle ait subsisté à chaque relecture, ce qui se produit avec une probabilité $(2/3)^n$; pour chaque faute la probabilité d'être corrigée est donc $1 - (2/3)^n$ et la probabilité de correction des quatre erreurs est par conséquent :

$$p_n = \left[1 - \left(\frac{2}{3}\right)^n\right]^4$$

Bien sûr p_n tend vers un quand n devient infini, avec $p_4 = 0,41$; $p_6 = 0,69$; $p_8 = 0,85$ et $p_{10} = 0,93$.

1.12 Il suffit d'appliquer la formule de Bayes :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})} = \frac{p}{p + (1-p)/5} = \frac{5p}{4p+1}$$

On peut noter que $P(A|B) \geq P(A) = p$, avec bien sûr $P(A|B) = 0$ si $p = 0$ et $P(A|B) = 1$ si $p = 1$. Dans tous les autres cas $P(A|B) > P(A)$.

1.13 L'ensemble fondamental est constitué des couples ordonnés dont les éléments symbolisent la couleur des premier et second feux :

$$\Omega = \{V_1V_2, V_1R_2, R_1V_2, R_1R_2\}$$

Nous noterons :

$$p_1 = P(V_1 V_2); p_2 = P(V_1 R_2); p_3 = P(R_1 V_2); p_4 = P(R_1 R_2)$$

Les hypothèses se traduisent par :

$$P(V_1) = p_1 + p_2 = P(V_2) = p_1 + p_3 = 2/3$$

$$P(V_2 | V_1) = 3/4$$

On en déduit :

$$p_1 = P(V_1 V_2) = P(V_1) P(V_2 | V_1) = \frac{1}{2}$$

puis $p_2 = p_3 = 1/6$ et donc $p_4 = 1/6$.

On peut alors calculer :

$$P(R_2 | R_1) = \frac{P(R_1 R_2)}{P(R_1)} = \frac{1}{2}$$

1.14 1. Si on note R l'événement « réussir son service », on obtient :

$$p_B = P(B|R) = \frac{P(B \cap R)}{P(R)} = \frac{2 \times 0,6}{0,9 + 2 \times 0,6} = \frac{4}{7}$$

2. On obtient cette fois :

$$p_B = \frac{2p}{0,9 + 2p}$$

avec $p_B = 1/2$ pour $p = 0,45$.

3. Si les joueurs ont le même taux de réussite :

$$p_B = \frac{2p}{p + 2p} = \frac{2}{3} = P(B)$$

1.15 En raison de l'équiprobabilité :

$$P(A) = P(B) = P(C) = 1/2 \text{ et } P(A \cap B) = P(\{a\}) = 1/4 = P(A)P(B).$$

$$P(A \cap C) = P(\{a\}) = 1/4 = P(A)P(C) \text{ et } P(B \cap C) = P(\{a\}) = 1/4 = P(B)P(C)$$

donc les événements A , B et C sont indépendants deux à deux. Par contre :

$$P(A \cap B \cap C) = P(\{a\}) = \frac{1}{4} \neq P(A)P(B)P(C)$$

et ils ne sont pas mutuellement indépendants.

1.16 On obtient :

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

$$= P(A)[1 - P(B)] = P(A)P(\bar{B})$$

donc les événements A et \bar{B} sont indépendants. De même :

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

$$= P(\bar{B})[1 - P(A)] = P(\bar{A})P(\bar{B})$$

donc \bar{A} et \bar{B} sont aussi indépendants.

On peut écrire :

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

donc pour qu'il y ait indépendance de C avec $A \cup B$ il faudrait que $P(A)P(B)P(C) = P(A \cap B \cap C)$, c'est-à-dire que A , B et C soient mutuellement indépendants, ce qui n'est pas le cas en général. Dans l'exercice précédent on a d'ailleurs $P(A \cup B) = P(\{a, b, c\}) = 3/4$ avec $P\{(A \cup B) \cap C\} = P(\{a\}) = 1/4 \neq P(A \cup B)P(C) = 3/8$ ce qui montre que $A \cup B$ n'est pas indépendant de C . De même, d'après ce qui précède, pour que \overline{C} soit indépendant de $A \cap B$, il faut que C soit aussi indépendant de $A \cap B$, ce qui n'est pas le cas en général comme le montre également l'exercice précédent puisque $P(A \cap B) = 1/4$ et $P(C) = 1/2$ avec $P(A \cap B \cap C) = 1/4 \neq P(A \cap B)P(C)$.

1.17 On obtient $P(A \cap B) = P(\{FPP, PPP\}) = 1/4 = P(A)P(B)$, $P(B \cap C) = P(\{FFP, PPP\}) = 1/4 = P(B)P(C)$ et $P(A \cap C) = P(\{PPP\}) = 1/8 \neq P(A)P(C)$ donc ces événements ne sont pas indépendants deux à deux et *a fortiori* ne sont pas mutuellement indépendants, bien qu'ici $P(A \cap B \cap C) = P(\{PPP\}) = 1/8 = P(A)P(B)P(C)$.

1.18 Notons par A_i l'événement « le danseur i se retrouve avec sa femme ». L'événement « il y a au moins une rencontre » est donc $\cup_{i=1}^n A_i$. Sa probabilité s'obtient à partir de la formule de Poincaré (cf. 1.3) pour n événements :

$$p_n = P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} P(A_{i_1} \cap \dots \cap A_{i_k})$$

Si chaque danseur est représenté par une case numérotée de 1 à n , l'événement $A_{i_1} \cap \dots \cap A_{i_k}$ correspond à une permutation où les femmes de ces danseurs sont placées dans les cases i_1, \dots, i_k . Comme toutes les permutations sont équiprobables :

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}$$

Par ailleurs, le nombre de sous-ensembles $\{i_1, \dots, i_k\}$ extraits de $\{1, \dots, n\}$ est égal à $\binom{n}{k}$, donc :

$$\sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} P(A_{i_1} \cap \dots \cap A_{i_k}) = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}$$

et :

$$p_n = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!}$$

La probabilité qu'il y n'ait aucune rencontre est donc :

$$q_n = 1 - p_n = 1 - \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!} \rightarrow \frac{1}{e} = 0,36788$$

On obtient bien des valeurs plus petites que $1/2$, sauf $q_2 = 0,5$. Les autres valeurs sont :

$$q_3 = \frac{1}{3} = 0,3333; \quad q_4 = \frac{3}{8} = 0,375;$$

$$q_5 = \frac{11}{30} = 0,3667; \quad q_6 = \frac{265}{720} = 0,3681$$

...

1.19 Il y a équiprobabilité de tous les quarts de finale qui sont au nombre de :

$$\frac{8!}{(2!)^4 4!} = 105$$

car il s'agit du nombre de partitions de 8 équipes distinctes en 4 matchs de même effectif 2 (cf. 5.6).

Pour chaque équipe anglaise, le nombre d'associations possibles avec une équipe non anglaise est successivement 4, puis 3, puis 2, puis 1. Il y a donc $4! = 24$ quarts de finale qui ne comportent pas de match entre équipes anglaises. On obtient donc :

$$p = 1 - p_0 = 1 - \frac{24}{105} = \frac{81}{105}$$

On peut calculer la probabilité des autres événements élémentaires. Deux quarts de finale entre équipes anglaises correspondent à 2 partitions de 2 sous-groupes de 4 (équipes anglaises et non anglaises) en matchs de même effectif 2, soit un nombre de possibilités :

$$\frac{4!}{(2!)^2 2!} \times \frac{4!}{(2!)^2 2!} = 9$$

La probabilité de deux quarts de finale entre équipes anglaises est donc :

$$p_2 = \frac{9}{105}$$

Enfin, il y a $\binom{4}{2}$ choix de matchs entre 2 équipes anglaises et $\binom{4}{2}$ choix de matchs entre 2 équipes non anglaises. Il ne reste ensuite que 2 façons de choisir les 2 autres quarts de finale en associant une des 2 équipes anglaises restantes à une des 2 équipes non anglaises restantes. Cela correspond au nombre de quarts de finale :

$$2 \binom{4}{2} \binom{4}{2} = 72$$

La probabilité d'un seul match entre équipes anglaises est donc :

$$p_1 = \frac{72}{105}$$

On retrouve bien $p = p_1 + p_2$.

Introduction

Le résultat ω d'une épreuve aléatoire va être codé numériquement au moyen d'une application $X : \Omega \rightarrow \mathbb{R}$. Pour pouvoir calculer la probabilité de ses différentes valeurs numériques possibles il faut pouvoir revenir à l'ensemble fondamental de départ où est définie une probabilité. On imposera donc une condition à cette application qui sera alors dénommée **variable aléatoire** (v.a., *random variable*). Cette terminologie est source de confusion puisqu'il s'agit bien d'une fonction !

Pour la facilité de l'exposé, il sera utile de distinguer le cas où $X(\Omega)$ est dénombrable, la v.a. X étant alors dite discrète, de celui où $X(\Omega)$ est un ensemble non dénombrable de \mathbb{R} (généralement un intervalle, pouvant être \mathbb{R} tout entier, ou une réunion d'intervalles), la v.a. étant dite continue. La loi de probabilité d'une v.a. sera définie par les probabilités individuelles dans le cas discret et par la densité dans le cas continu. Nous définirons les deux caractéristiques numériques principales d'une distribution, l'espérance, caractéristique de valeur centrale, et la variance, caractéristique de dispersion.

Objectifs

Retenir que dans les cas usuels la loi d'une variable aléatoire est déterminée par les probabilités des points dans le cas discret, par la densité dans le cas continu.

Noter qu'une loi est souvent caractérisée par ses deux premiers moments (l'espérance et la variance).

Concepts clés

- Variable aléatoire
- Fonction de répartition
- Densité de probabilité
- Espérance mathématique
- Variance
- Écart type

1 Variable aléatoire réelle discrète

1.1 Définition

On appelle v.a. discrète définie sur (Ω, \mathcal{A}) une application $X : \Omega \rightarrow \mathbb{R}$ telle que $X(\Omega)$ est dénombrable (en général $X(\Omega)$ est fini ou $X(\Omega) \subset \mathbb{N}$ ou $X(\Omega) \subset \mathbb{Z}$ et dans tous les cas $X(\Omega)$ est en correspondance bijective avec \mathbb{N}) et telle que pour tout x réel :

$$X^{-1}(x) = \{\omega \in \Omega / X(\omega) = x\} \in \mathcal{A}$$

ce qui exprime tout simplement que $X^{-1}(x)$ est un événement.

REMARQUE Si $\mathcal{A} = \mathcal{P}(\Omega)$, toute application X sera une v.a. puisque $X^{-1}(x) \in \mathcal{P}(\Omega)$ pour tout x réel.

1.2 Loi de probabilité

La propriété résultant de la définition d'une v.a. va nous permettre de définir la probabilité de chacune de ses valeurs possibles $x \in X(\Omega)$ par :

$$P_X(X = x) = P\{X^{-1}(x)\} = P\{\omega \in \Omega / X(\omega) = x\}$$

Exemple 2.1

Si on associe la valeur 1 au résultat impair d'un jet de dé : $X^{-1}(1) = \{1, 3, 5\}$, la probabilité de la valeur 1 sera $P(\{1, 3, 5\}) = 1/6 + 1/6 + 1/6 = 1/2$.

Cette nouvelle probabilité, définie cette fois sur $\Omega' = X(\Omega)$ et notée P_X , s'appelle la **probabilité image** de P par X . La relation précédente peut s'écrire schématiquement $P_X(x) = (P \circ X^{-1})(x)$. Nous allons vérifier que P_X est bien une probabilité définie sur Ω' . Tout d'abord, l'ensemble d'arrivée de cette application est celui de P , donc $[0, 1]$. D'autre part :

$$P_X(\Omega') = P\{X^{-1}(\Omega')\} = P(\Omega) = 1;$$

enfin, si $A_n \subset \Omega'$, avec $A_m \cap A_n = \emptyset$ pour $m \neq n$:

$$\begin{aligned} P_X\left(\bigcup_{n=0}^{\infty} A_n\right) &= P\left\{X^{-1}\left(\bigcup_{n=0}^{\infty} A_n\right)\right\} = P\left\{\bigcup_{n=0}^{\infty} X^{-1}(A_n)\right\} = \sum_{n=0}^{\infty} P\{X^{-1}(A_n)\} \\ &= \sum_{n=0}^{\infty} P_X(A_n) \end{aligned}$$

Très souvent dans la suite, par souci de simplification mais par abus de notation, nous noterons P les probabilités images P_X .

L'ensemble $\Omega' = X(\Omega)$ étant dénombrable, il existe une bijection permettant de représenter ses éléments par l'ensemble des $x_i, i \in \mathbb{N}$. La loi de probabilité P_X de X est alors définie par les probabilités individuelles :

$$p_i = P_X(X = x_i) = P\{X^{-1}(x_i)\}, i \in \mathbb{N}$$

On appelle alors **distribution** ou **loi de probabilité** de la v.a. X l'ensemble des couples $(x_i, p_i)_{i \in \mathbb{N}}$. Si X ne prend qu'un petit nombre de valeurs, cette distribution est généralement présentée dans un tableau.

Exemple 2.2

La loi uniforme associée à un lancer de dé à six faces numérotées est présentée dans le tableau ci-après :

x_i	1	2	3	4	5	6	
p_i	1/6	1/6	1/6	1/6	1/6	1/6	1

Cas particuliers

Variable certaine. Il s'agit d'une v.a. qui est constante, i.e. qui prend la même valeur connue a quel que soit le résultat de l'épreuve :

$$P_X(X = a) = 1$$

La masse totale de probabilité est concentrée en a ; on parle de **loi de Dirac** associée à cette variable certaine.

Variable indicatrice. Soit $A \in \mathcal{A}$ un événement quelconque ; on appelle **v.a. indicatrice** de cet événement A , la v.a. définie par :

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in \bar{A} \end{cases}$$

et notée $X = \mathbf{1}_A$. Ainsi :

$$\begin{aligned} P_X(X = 1) &= P\{\omega/\omega \in A\} = P(A) \\ P_X(X = 0) &= P\{\omega/\omega \in \bar{A}\} = P(\bar{A}) = 1 - P(A) \end{aligned}$$

1.3 Fonction de répartition

On appelle **fonction de répartition** de la v.a. X , la fonction F définie pour x réel par :

$$F(x) = P_X\{X < x\} = P\{\omega \in \Omega/X(\omega) < x\}$$

Il faut noter que beaucoup d'ouvrages utilisent la définition anglo-saxonne où $F(x) = P_X\{X \leq x\}$. C'est une fonction en escalier, constante par morceaux, continue à gauche, définie ici par :

$$F(x) = \sum \{p_i/x_i < x\}$$

c'est-à-dire que c'est la somme des poids de tous les points qui sont strictement à gauche de x . Les propriétés de F seront étudiées dans le cas d'une v.a. continue.

Si par exemple X prend les valeurs $x_1 < x_2 < \dots < x_n$, on aura $F(x) = 0$ pour $x \leq x_1$, puis le graphe de F présentera un saut en chaque point x_i , jusqu'à la valeur $F(x) = 1$ pour $x > x_n$.

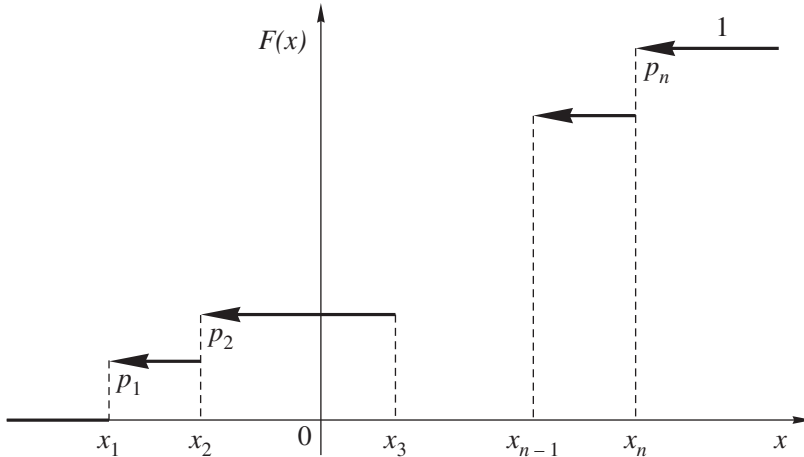


Figure 2.1

On peut déduire de F les probabilités individuelles par : $p_i = F(x_{i+1}) - F(x_i)$ pour $1 \leq i \leq n - 1$ et $p_n = 1 - F(x_n)$.

Exemple 2.3

Variable certaine : $F(x) = 0$ pour $x \leq a$ et $F(x) = 1$ pour $x > a$.

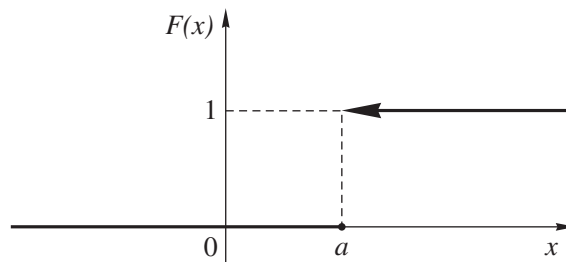


Figure 2.2

Exemple 2.4

Variable indicatrice : $F(x) = 0$ pour $x \leq 0$, saut de hauteur $P_X(X = 0) = 1 - p$ avec $F(x) = 1 - p$ pour $0 < x \leq 1$, puis saut de hauteur $P_X(X = 1) = p$, puis $F(x) = 1$ pour $x > 1$.

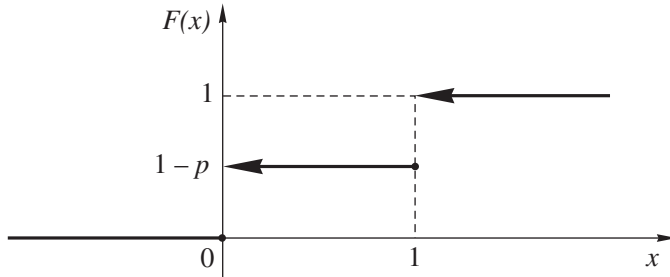


Figure 2.3

Exemple 2.5

Jet de dé : $F(x) = 0$ pour $x \leq 1$, puis sauts de hauteur $P_X(X = i) = 1/6$ aux points $i = 1, \dots, 6$ puis $F(x) = 1$ pour $x > 6$.

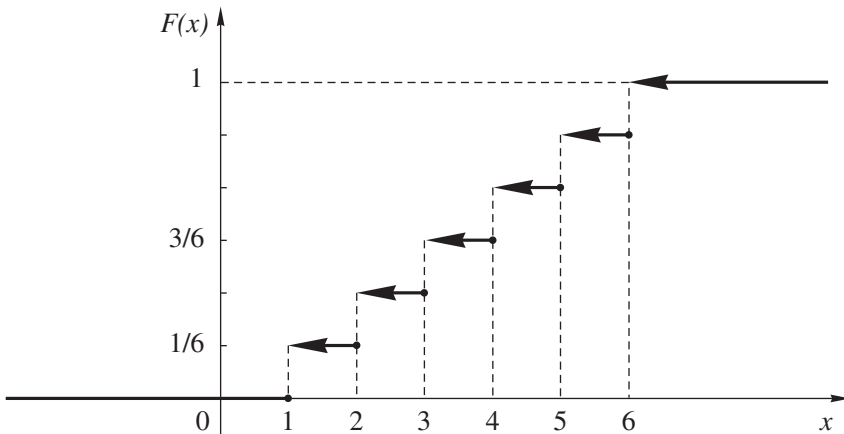


Figure 2.4

1.4 Moments d'une v.a. discrète

a) Espérance mathématique

DÉFINITION

On appelle **espérance mathématique** (*expected value*) de la v.a. X la quantité, si elle existe :

$$E(X) = \sum_{i \in \mathbb{N}} p_i x_i$$

Il s'agit d'une moyenne en probabilité, ou moyenne pondérée, des valeurs x_i que peut prendre la v.a. X , par les probabilités correspondantes p_i . Chaque probabilité p_i peut aussi s'interpréter comme la masse ponctuelle du point ω_i d'abscisse x_i et $E(X)$ est alors

le centre de gravité, ou barycentre, de ces points affectés de masses. C'est une valeur numérique constante qui est la valeur moyenne, en probabilité, de X . Notons que si X prend ses valeurs entre x_1 et x_n , on aura bien sûr $x_1 \leq E(X) \leq x_n$.

Si X ne prend qu'un nombre fini de valeurs, le calcul de $E(X)$ peut se faire à l'aide du tableau de la distribution comme ci-après :

x_i		
p_i	↓	→ 1
$p_i x_i$		→ $E(X)$

Exemple 2.6

Si X est la v.a. qui code 0 le résultat pile et 1 le résultat face d'un lancer de pièce de monnaie :

$$E(X) = 0 \times P_X(X = 0) + 1 \times P_X(X = 1) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$$

la valeur moyenne en probabilité de X est $1/2$, bien que X ne prenne jamais cette valeur.

Exemple 2.7

Pour une v.a. indicatrice :

$$E(X) = 0 \times P_X(X = 0) + 1 \times P_X(X = 1) = P(A) = p$$

Exemple 2.8

Pour le jet de dé :

$$E(X) = \frac{1}{6} \sum_{i=1}^6 i = \frac{7}{2} = 3,5$$

Exemple 2.9

Dans le cas de la loi uniforme sur $X(\Omega) = \{x_1, \dots, x_k\}$, c'est-à-dire avec équiprobabilité de toutes les valeurs $p_i = 1/k$, on obtient :

$$E(X) = \frac{1}{k} \sum_{i=1}^k x_i$$

et dans ce cas $E(X)$ se confond avec la moyenne arithmétique simple \bar{x} des valeurs possibles de X (cependant la notation \bar{x} pour l'espérance est à proscrire en général car ce n'est que dans ce cas particulier que $E(X) = \bar{x}$).

À l'origine des probabilités, la quantité $E(X)$ a été introduite pour traduire la notion de gain moyen, ou espérance de gain, la v.a. X représentant la valeur du gain à un certain jeu.

Exemple 2.10

Deux joueurs A et B jouent avec un dé, le joueur B gagnant les mises si le résultat est supérieur ou égal à trois ; sinon, c'est A qui gagne. La question est de savoir quelles doivent être les mises respectives a et b de ces deux joueurs pour que le jeu soit équitable.

Les gains respectifs de ces deux joueurs sont représentés par les v.a. :

$$X_A = \begin{cases} a + b & 2/6 \\ 0 & 4/6 \end{cases} \quad X_B = \begin{cases} a + b & 4/6 \\ 0 & 2/6 \end{cases}$$

soit les espérances de gain $E(X_A) = (a + b)/3$ et $E(X_B) = 2(a + b)/3$. Le jeu est considéré comme équitable si la mise est égale à l'espérance de gain, ce qui conduit aux conditions $(a + b)/3 = a$ et $2(a + b)/3 = b$ d'où $b = 2a$, résultat intuitivement évident puisque le joueur B a deux fois plus de chances de gagner.

Théorème 2.1

Les propriétés de l'opérateur espérance mathématique sont celles du signe somme.

P_1 Si on ajoute une constante à une v.a., il en est de même pour son espérance :

$$E(X + a) = E(X) + a, \quad a \in \mathbb{R}$$

résultat qui se déduit de :

$$\sum_i p_i (x_i + a) = \sum_i p_i x_i + a$$

P_2 Si on multiplie une v.a. par une constante, il en est de même pour son espérance :

$$E(aX) = aE(X), \quad a \in \mathbb{R}$$

il suffit d'écrire :

$$\sum_i p_i a x_i = a \sum_i p_i x_i$$

P_3 L'espérance d'une somme de deux v.a. est la somme des espérances (bien sûr si elles existent) :

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

La première espérance fait intervenir deux v.a. distinctes et se calcule donc par rapport à la loi du couple (X, Y) qui est définie par les probabilités :

$$p_{ij} = P_{(X,Y)}(X = x_i, Y = y_j), \quad i \in I, j \in J$$

Les deux autres espérances se calculent par rapport aux lois marginales de X et Y qui s'obtiennent par sommations de la loi du couple :

$$p_{i.} = \sum_j p_{ij} = P_X(X = x_i)$$

$$p_{.j} = \sum_i p_{ij} = P_Y(Y = y_j)$$

On calcule alors :

$$E(X + Y) = \sum_{i,j} p_{ij} (x_i + y_j) = \sum_{i,j} p_{ij} x_i + \sum_{i,j} p_{ij} y_j$$

$$= \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \sum_i p_{i.} x_i + \sum_j p_{.j} y_j$$

On peut résumer ces trois propriétés en disant que l'opérateur espérance mathématique est linéaire :

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y), \quad \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}$$

REMARQUES L'espérance d'une constante réelle a est égale à la valeur de cette constante : $E(a) = a$.

Si g est une fonction continue quelconque, alors :

$$E[g(X)] = \sum_{i \in \mathbb{N}} p_i g(x_i)$$

b) Variance

Il s'agit d'un indicateur mesurant la dispersion des valeurs x_i que peut prendre la v.a. X , autour de la moyenne en probabilité $E(X)$ et défini par :

$$V(X) = \sum_{i \in \mathbb{N}} p_i [x_i - E(X)]^2$$

lorsque cette quantité existe.

C'est l'espérance mathématique du carré de la v.a. centrée $X - E(X)$:

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

moment centré d'ordre deux. On note cette quantité $V(X) = \sigma_X^2$, σ_X désignant alors l'**écart type** (standard deviation) de X qui s'exprime dans les mêmes unités de mesure que la variable.

Exemple 2.11

Lancer de pièce :

$$X = \begin{cases} 0 \text{ (pile)} & 1/2 \\ 1 \text{ (face)} & 1/2 \end{cases}$$

On avait calculé $E(X) = 1/2$, d'où :

$$V(X) = \frac{1}{2} \left(0 - \frac{1}{2}\right)^2 + \frac{1}{2} \left(1 - \frac{1}{2}\right)^2 = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

Exemple 2.12

Pour la v.a. indicatrice :

$$\begin{aligned} V(X) &= E(X - p)^2 = p(1 - p)^2 + (1 - p)(0 - p)^2 \\ &= p(1 - p)(1 - p + p) = p(1 - p) \end{aligned}$$

Exemple 2.13

Pour la loi uniforme sur $X(\Omega) = \{x_1, \dots, x_k\}$ nous avons obtenu $E(X) = \bar{x}$, d'où :

$$V(X) = \sum_{i=1}^k p_i (x_i - \bar{x})^2 = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2$$

et la variance se confond ici avec la variance empirique des valeurs possibles de X .

Cet indicateur de dispersion vient compléter l'information sur la distribution, fournie par la valeur moyenne.

Exemple 2.14

Considérons les deux distributions suivantes :

X	2	4	6	Y	-4	3	33
	1/4	1/4	1/2		1/2	1/3	1/6

Elles ont comme valeurs moyennes :

$$E(X) = \frac{1}{2} + 1 + 3 = 4,5 \quad \text{et} \quad E(Y) = -2 + 1 + \frac{33}{6} = 4,5$$

donc même centre de distribution. Par contre :

$$E(X^2) = 1 + 4 + 18 = 23 \quad \text{et} \quad E(Y^2) = 8 + 3 + 121 \times \frac{3}{2} = \frac{385}{2}$$

d'où $V(X) = 11/4$ et $V(Y) = 689/4$, valeur très supérieure qui indique une dispersion de Y autour de sa moyenne beaucoup plus grande que celle de X .

Théorème 2.2

P₁ Par définition :

$$V(X) \geq 0$$

avec :

$$V(X) = 0 \Leftrightarrow p_i[x_i - E(X)] = 0 \forall i \in \mathbb{N} \Leftrightarrow x_i = E(X) \forall i \in \mathbb{N}$$

la variance ne peut être nulle que si $X = E(X)$ avec une probabilité égale à un, i.e. si X est une v.a. certaine : $P_X\{X = E(X)\} = 1$.

P₂ Pour tout réel a :

$$V(X + a) = V(X)$$

c'est-à-dire que le moment centré d'ordre deux est invariant par translation ou changement d'origine.

Ceci provient du fait que $X + a - E(X + a) = X - E(X)$.

P₃ Pour tout réel a :

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

un changement d'échelle modifie la variance ; ceci est dû au fait que $[aX - E(aX)]^2 = a^2[X - E(X)]^2$.

P₄ Pour le calcul de la variance, il est souvent préférable d'utiliser la formule développée :

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

qui se déduit des propriétés de linéarité de l'opérateur espérance :

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

P₅ Si X et Y sont deux v.a. indépendantes, alors :

$$V(X + Y) = V(X) + V(Y)$$

cette propriété se déduisant du fait que dans ce cas $E(XY) = E(X)E(Y)$, que l'on démontre aisément à partir de la condition d'indépendance qui s'écrit :

$$P_{(X,Y)}\{X = x_i, Y = y_j\} = P_X\{X = x_i\}P_Y\{Y = y_j\} \quad \forall i \in I, \forall j \in J$$

ou $p_{ij} = p_i \times p_j$ avec les notations du paragraphe précédent. En effet :

$$E(XY) = \sum_i \sum_j p_{ij}x_iy_j = \sum_i p_i x_i \sum_j p_j y_j = E(X)E(Y)$$

et alors :

$$\begin{aligned} V(X+Y) &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= E(X^2) + E(Y^2) + 2E(XY) - E^2(X) - E^2(Y) - 2E(X)E(Y) \\ &= V(X) + V(Y) + 2[E(XY) - E(X)E(Y)] \end{aligned}$$

On définit la **covariance** par $Cov(X, Y) = E(XY) - E(X)E(Y)$ et on a donc dans le cas général :

$$V(X+Y) = V(X) + V(Y) + 2Cov(X, Y)$$

REMARQUE Considérons la fonction g définie pour t réel par :

$$g(t) = E(X-t)^2 = \sum_i p_i(x_i - t)^2$$

C'est un polynôme du second degré en t , de dérivée :

$$g'(t) = -2 \sum_i p_i(x_i - t) = -2[E(X) - t]$$

qui s'annule en changeant de signe pour $t = \sum_i p_i x_i = E(X)$; comme $g''(t) = 2 > 0$, g est minimum pour $t = E(X)$, ce minimum étant la variance de X . Ainsi, $E(X)$ est solution du problème :

$$\min_{t \in \mathbb{R}} E(X-t)^2$$

On peut également retenir comme caractéristique de valeur centrale la **médiane Md**, notamment dans le cas où l'espérance n'existe pas, nombre qui se définit par :

$$P(X < Md) \leq \frac{1}{2} \leq P(X \leq Md).$$

Elle vérifie aussi :

$$E|X - Md| = \inf_{t \in \mathbb{R}} E|X - t|$$

c) Moments non centrés et centrés

On appelle **moment non centré** d'ordre $r \in \mathbb{N}^*$ la quantité, lorsqu'elle existe :

$$m_r(X) = \sum_{i \in \mathbb{N}} p_i x_i^r = E(X^r).$$

Le **moment centré** d'ordre $r \in \mathbb{N}^*$ est :

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i [x_i - E(X)]^r = E[X - E(X)]^r.$$

Notons que ce moment centré peut s'exprimer, à partir de la formule du binôme, en fonction des moments non centrés m_r, m_{r-1}, \dots, m_1 . Les premiers moments sont :

$$m_1(X) = E(X), \mu_1(X) = 0, \mu_2(X) = m_2(X) - m_1^2(X) = V(X)$$

2 Variable aléatoire réelle continue

2.1 Définition

On appelle v.a. réelle définie sur (Ω, \mathcal{A}) une application $X : \Omega \rightarrow \mathbb{R}$ telle que pour tout intervalle $I \subset \mathbb{R}$ on ait :

$$X^{-1}(I) = \{\omega \in \Omega / X(\omega) \in I\} \in \mathcal{A}$$

Cela exprime que l'image inverse d'un intervalle quelconque est un événement. Il suffit en fait de vérifier que pour tout réel x :

$$X^{-1}(]-\infty, x[) \in \mathcal{A}$$

Exemple 2.15

La durée de vie d'une lampe ou le salaire d'un individu tiré au sort dans une population sont représentés par des v.a. continues.

2.2 Loi de probabilité

Elle est déterminée ici par la fonction de répartition (f.r.) F , définie pour tout x réel par :

$$F(x) = P_X(X < x) = P\{X^{-1}(]-\infty, x[)\} = P\{\omega \in \Omega / X(\omega) < x\}$$

qui représente la probabilité que X soit strictement à gauche de x .

2.3 Propriétés de la fonction de répartition

P₁ Elle est croissante au sens large.

En effet, pour $x < y$ on a $(X < x) \Rightarrow (X < y)$ et par conséquent

$$F(x) = P_X(X < x) \leq P_X(X < y) = F(y).$$

P₂ Elle prend ses valeurs entre 0 et 1 :

$$0 \leq F(x) \leq 1 \text{ avec } \lim_{x \rightarrow -\infty} F(x) = 0 \text{ et } \lim_{x \rightarrow +\infty} F(x) = 1$$

En effet, $F(x)$ est une probabilité et quand $x \rightarrow -\infty$ l'intervalle $]-\infty, x[$ devient \emptyset et $X^{-1}(\emptyset) = \emptyset$; quand $x \rightarrow +\infty$, l'intervalle $]-\infty, x[$ devient \mathbb{R} et $X^{-1}(\mathbb{R}) = \Omega$.

P₃ Elle est continue à gauche :

$$\lim_{h \rightarrow 0^+} F(x - h) = F(x)$$

car l'intervalle $] -\infty, x - h[$ ne contient jamais le point x et devient donc l'intervalle $] -\infty, x[$ quand $h \rightarrow 0^+$.

On peut établir que ces trois propriétés sont caractéristiques d'une fonction de répartition. Si on retient la définition $F(x) = P_X(X \leq x)$, cette fonction F est alors continue à droite.

Par contre, pour tout $h > 0$ l'intervalle $] -\infty, x + h[$ contient toujours le point x et devient l'intervalle $] -\infty, x] =] -\infty, x[\cup \{x\}$ quand $h \rightarrow 0^+$. Par conséquent :

$$\lim_{h \rightarrow 0^+} F(x + h) = F(x) + P_X(X = x)$$

Ainsi, aux points x de discontinuité de F , son graphe présente un saut de hauteur $P_X(X = x)$ (voir figure 2.5) :

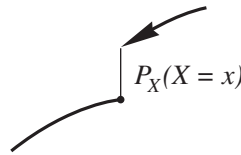


Figure 2.5

P_4 La probabilité de l'intervalle $[a, b[$, pour $a < b$, se calcule par :

$$P_X(a \leq X < b) = F(b) - F(a)$$

On a en effet $] -\infty, b] =] -\infty, a[\cup [a, b[$ donc :

$$P_X(X < b) = P_X(X < a) + P_X(a \leq X < b).$$

2.4 Loi continue

Si la fonction F est continue, i.e. continue à droite, on dit que X est une **variable aléatoire réelle continue**. Dans ce cas, pour tout réel x :

$$P_X(X = x) = 0$$

c'est-à-dire que la probabilité d'un point est toujours nulle, ce qu'on traduit en disant que la loi est **diffuse**. Dans le cas où certains points ont une probabilité non nulle on dit que la loi est **mixte**, comportant une partie continue et une partie discrète correspondant à ces points.

Exemple 2.16

Considérons la f.r. F définie par :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x}{4} & \text{si } 0 < x \leq 1 \\ \frac{x}{2} & \text{si } 1 < x \leq 2 \\ 1 & \text{si } 2 < x \end{cases}$$

Cette fonction est continue pour $x = 0$ et $x = 2$. Par contre, si elle est bien continue à gauche en $x = 1$, comme toute f.r., avec $\lim_{h \rightarrow 0^+} F(1 - h) = F(1) = \frac{1}{4}$, elle n'est pas continue en ce point car :

$$\lim_{h \rightarrow 0^+} F(1 + h) = \lim_{h \rightarrow 0^+} \frac{1}{2}(1 + h) = \frac{1}{2} = F(1) + P_X(X = 1)$$

et par conséquent il s'agit d'une loi mixte, avec une partie continue sur les intervalles $]-\infty, 1[$ et $]1, +\infty[$, et une partie discrète concentrée au point 1 avec $P_X(X = 1) = \frac{1}{4}$.

Exemple 2.17

On peut donner également un exemple plus concret, qui est celui d'une v.a. X représentant la durée de vie d'une ampoule qui suit une loi exponentielle (cf. chap. 3, § 2.2). S'il y a une probabilité $p > 0$ que cette ampoule claque lorsqu'on l'allume, la f.r. de X est alors définie par :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ p + (1 - p)(1 - e^{-x}) & \text{si } x > 0 \end{cases}$$

avec donc une masse non nulle au point $x = 0$:

$$\lim_{h \rightarrow 0^+} F(0 + h) = p = P_X(X = 0)$$

2.5 Loi absolument continue

La valeur moyenne de la probabilité d'un intervalle de longueur $h > 0$ est :

$$\frac{1}{h} P_X(x \leq X < x + h) = \frac{F(x + h) - F(x)}{h}$$

et représente donc une densité moyenne, puisqu'étant le « poids » (probabilité) de l'intervalle, divisé par sa longueur. Si on fait tendre cette longueur vers 0, la limite, si elle existe, représentera la probabilité d'un intervalle de longueur infiniment petite dx . Ce sera le cas si F admet une dérivée f :

$$\lim_{h \rightarrow 0} \frac{1}{h} [F(x + h) - F(x)] = F'(x) = f(x)$$

Dans ce cas, l'équivalent de la probabilité du point x pour une loi discrète peut s'écrire symboliquement $P_X(x \leq X < x + dx) = f(x)dx$, la fonction $f = dF/dx$ étant appelée **densité de probabilité** de la v.a. X dont la loi est alors qualifiée d'**absolument**

continu. Dans ce cas la loi de X est déterminée par sa densité f puisque F est définie comme la primitive de f qui s'annule pour $x = -\infty$:

$$F(x) = \int_{-\infty}^x f(t) dt$$

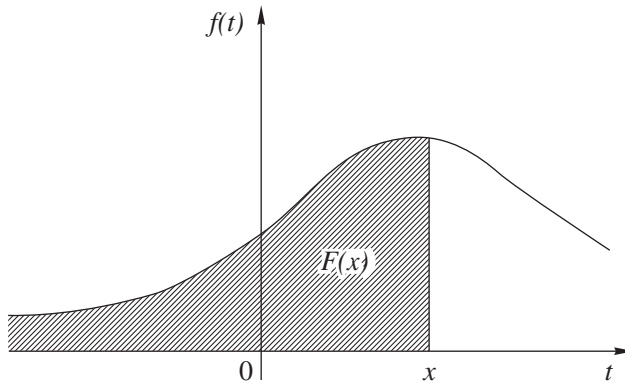


Figure 2.6

NB On peut trouver des exemples, assez complexes, où F est continue et strictement croissante mais n'admet pas de densité, c'est-à-dire de cas où la loi est continue mais pas absolument continue. Dans la pratique, les lois usuelles que nous rencontrerons seront soit discrètes, soit définies par une densité.

Théorème 2.3

P_1 Une densité est positive :

$$f \geq 0 ;$$

propriété immédiate puisque f est la dérivée d'une fonction croissante au sens large.

P_2 Une densité est intégrable sur \mathbb{R} , d'intégrale égale à un :

$$\int_{-\infty}^{+\infty} f(t) dt = 1 ;$$

ceci provient du fait que $F(+\infty) = 1$.

Ces deux propriétés sont caractéristiques d'une densité de probabilité, c'est-à-dire que toute fonction positive et intégrable sur \mathbb{R} définit une loi de probabilité.

P_3 La probabilité d'un intervalle s'obtient en intégrant la densité sur cet intervalle :

$$P_X\{X \in [x_1, x_2]\} = \int_{x_1}^{x_2} f(t) dt.$$

En effet :

$$\begin{aligned}
 P_X(x_1 \leq X \leq x_2) &= F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(t) dt - \int_{-\infty}^{x_1} f(t) dt \\
 &= \int_{-\infty}^{x_2} f(t) dt + \int_{x_1}^{-\infty} f(t) dt
 \end{aligned}$$

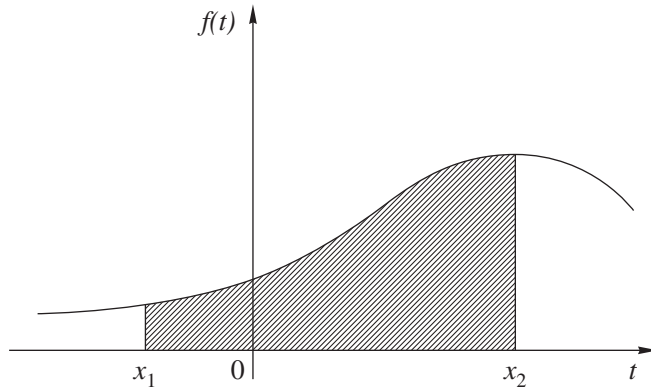


Figure 2.7

2.6 Moments d'une v.a. absolument continue

a) Espérance mathématique

Elle est définie par :

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

lorsque cette intégrale généralisée existe, c'est-à-dire est convergente. Les propriétés de l'espérance sont celles de l'intégrale et sont identiques au cas discret, i.e. il s'agit d'un opérateur linéaire :

Théorème 2.4

P_1 Pour tout réel a :

$$E(X + a) = E(X) + a$$

P_2 Pour tout réel a :

$$E(aX) = aE(X)$$

P_3 Si X et Y sont deux v.a. qui admettent une espérance :

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

b) Variance

Elle est définie par :

$$\begin{aligned} V(X) &= E[X - E(X)]^2 = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx \\ &= E(X^2) - E^2(X) = \sigma^2(X) \end{aligned}$$

lorsque cette intégrale généralisée existe. Ses propriétés sont identiques au cas discret :

Théorème 2.5

P_1 C'est une quantité positive :

$$V(X) \geq 0$$

avec $V(X) = 0 \Leftrightarrow X$ est une v.a. certaine (donc une v.a. discrète !).

P_2 Pour tout réel a :

$$V(X + a) = V(X)$$

P_3 Pour tout réel a :

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

P_4 Si X et Y sont deux v.a. indépendantes admettant une variance :

$$V(X + Y) = V(X) + V(Y)$$

c) Moments non centrés et centrés

Le **moment non centré d'ordre** $r \in \mathbb{N}^*$ de X est la quantité, lorsqu'elle existe :

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

Le **moment centré d'ordre** $r \in \mathbb{N}^*$ de X est la quantité, lorsqu'elle existe :

$$\mu_r(X) = E[X - E(X)]^r = \int_{-\infty}^{+\infty} [x - E(X)]^r f(x) dx$$

Paramètres d'asymétrie. L'asymétrie d'une distribution peut se caractériser par le moment centré d'ordre trois.

La distribution est :

- symétrique si $\mu_3 = 0$;
- dissymétrique étalée vers la droite si $\mu_3 > 0$;
- dissymétrique étalée vers la gauche si $\mu_3 < 0$.

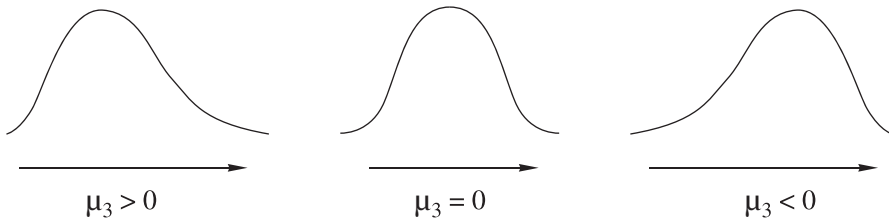


Figure 2.8

Pour obtenir un paramètre indépendant des unités, on considère les coefficients de symétrie (*skewness*) :

- de Pearson :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

- de Fisher :

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

Paramètres d'aplatissement. Ils sont calculés à partir du moment centré d'ordre quatre ; ce sont les coefficients d'aplatissement (*kurtosis*), invariants aussi par changement d'échelle ou d'origine :

- de Pearson :

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

- de Fisher :

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\sigma^4} - 3$$

Le terme de comparaison est ici la loi normale standard pour laquelle $\beta_2 = 3$, avec $\gamma_2 > 0$ pour une distribution plus aplatie que la distribution normale de même moyenne et de même écart type.

2.7 Changement de variable

On cherche à déterminer la loi de probabilité de la v.a. $Y = h(X)$, connaissant la fonction de répartition (f.r.) de la v.a. X . Ceci se fera sans difficulté si h est une application réelle continue et bijective, donc inversible. La f.r. de Y est en effet définie par :

$$G(y) = P_Y(Y < y) = P_X\{h(X) < y\}$$

Nous allons alors distinguer deux cas :

– h est croissante :

$$h(X) < y \Leftrightarrow X < h^{-1}(y)$$

$$\text{et } G(y) = P_X\{X < h^{-1}(y)\} = F[h^{-1}(y)]$$

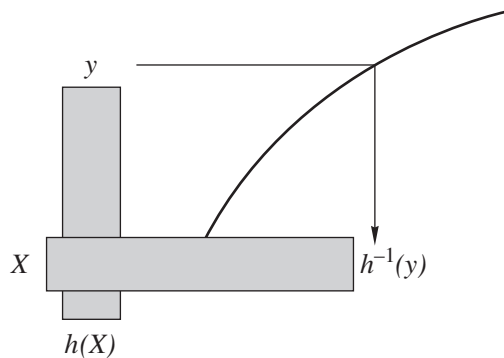


Figure 2.9

– h est décroissante :

$$h(X) < y \Leftrightarrow X > h^{-1}(y)$$

$$\text{et } G(y) = P_X\{X > h^{-1}(y)\} = 1 - F[h^{-1}(y)]$$

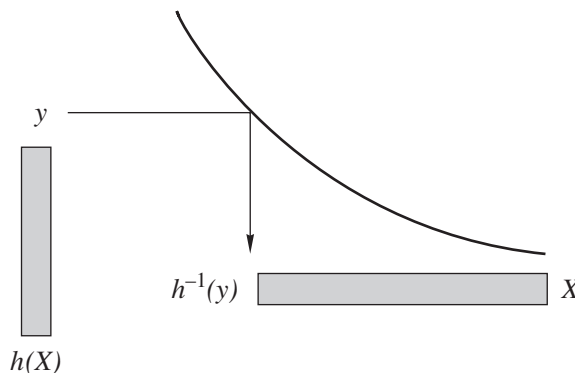


Figure 2.10

Si X admet une densité f et que h est de plus dérivable, on peut déterminer la densité $g = dG/dy$ de Y par dérivation. Dans le premier cas ($h' > 0$) :

$$g(y) = \left[\frac{dF}{du} \right]_{u=h^{-1}(y)} \times \frac{dh^{-1}(y)}{dy} = \frac{f[h^{-1}(y)]}{h'[h^{-1}(y)]}$$

Dans le second cas ($h' < 0$) :

$$g(y) = - \left[\frac{dF}{du} \right]_{u=h^{-1}(y)} \times \frac{dh^{-1}(y)}{dy} = - \frac{f[h^{-1}(y)]}{h'[h^{-1}(y)]}$$

Dans les deux cas la densité de Y peut s'écrire sous la même forme :

$$g(y) = \frac{f[h^{-1}(y)]}{|h'[h^{-1}(y)]|}$$

REMARQUE Même dans le cas où la fonction h n'est pas inversible, on peut parfois déterminer la loi de $Y=h(X)$. Considérons par exemple $h(x)=x^2$; la fonction h n'est pas injective car $h(-x) = h(x)$ pour tout x réel et cependant on peut déterminer la loi de la v.a. positive $Y = X^2$, car pour $y > 0$:

$$X^2 < y \Leftrightarrow -\sqrt{y} < X < \sqrt{y}$$

d'où $G(y) = P_Y(Y < y) = P_X(-\sqrt{y} < X < \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$. Bien entendu pour $y \leq 0$ on a $G(y) = 0$. Si X admet une densité f , alors Y admet comme densité pour $y > 0$:

$$g(y) = \frac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})]$$

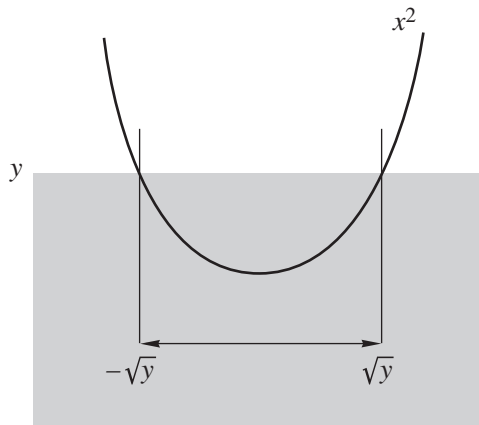


Figure 2.11

3 Compléments

3.1 Application mesurable

La notion de variable aléatoire réelle peut être définie de façon générale, sans distinction entre les cas continu et discret, en introduisant la notion d'**application mesurable**. Pour cela, nous définissons sur l'espace métrique \mathbb{R} sa tribu **borélienne**, notée \mathcal{B} , qui est la tribu engendrée par les intervalles ouverts (ou fermés), c'est-à-dire la plus petite tribu contenant cette famille d'intervalles. Une variable aléatoire réelle X est alors définie comme une application mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B})$, c'est-à-dire telle que :

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

En fait, pour vérifier que X est bien une application mesurable, il suffit de vérifier cette condition pour une famille particulière d'intervalles, comme celle des ouverts de la forme $] -\infty, x[$, c'est-à-dire de vérifier que :

$$\forall x \in \mathbb{R}, X^{-1}(] -\infty, x[) \in \mathcal{A}$$

Notons comme propriété utile que toute application continue de $(\mathbb{R}, \mathcal{B})$ dans $(\mathbb{R}, \mathcal{B})$ est mesurable.

3.2 Densité

Si la loi de probabilité P_X d'une v.a. X admet une densité, celle-ci est la dérivée de la fonction de répartition de cette loi. Cependant, la définition d'une densité ne se fait pas à partir de cette notion analytique de dérivabilité, mais à partir de notions plus complexes que nous allons évoquer. Cela nécessite d'introduire la **mesure de Lebesgue** λ sur $(\mathbb{R}, \mathcal{B})$, qui est la mesure diffuse prolongeant la notion de longueur, la mesure de l'intervalle $]a, b[$ étant définie par $\lambda(]a, b[) = b - a$. Si pour tout borélien $B \in \mathcal{B}$ tel que $\lambda(B) = 0$ on a $P_X(B) = 0$, on dit que la loi de probabilité P_X de X est absolument continue par rapport à la mesure de Lebesgue et alors cette loi admet une densité.

3.3 Support

Le support d'une loi P_X est l'ensemble de tous les points x tels que tout intervalle ouvert contenant x a une probabilité positive :

$$\forall \varepsilon > 0, P_X(]x - \varepsilon, x + \varepsilon[) > 0$$

Exemple 2.18

Soit P_X la loi de probabilité définie sur l'ensemble des rationnels $\mathbb{Q} = \{q_n\}_{n \in \mathbb{N}}$ par $P_X(q_n) = \frac{1}{2^n}$. C'est une loi discrète mais dont le support est pourtant \mathbb{R} car tout intervalle $]x - \varepsilon, x + \varepsilon[$ contient au moins un rationnel, quel que soit $\varepsilon > 0$ et par conséquent $P_X(]x - \varepsilon, x + \varepsilon[) > 0$ pour tout $x \in \mathbb{R}$.

● L'essentiel

- 1 Une variable aléatoire (v.a.) X est une application qui à un événement fait correspondre un nombre.
- 2 La loi de probabilité d'une v.a. peut toujours être définie par sa fonction de répartition (f.r.) F , où $F(x)$ représente la probabilité de toutes les valeurs strictement inférieures au réel x .
- 3 Si l'ensemble des valeurs possibles de X est dénombrable, i.e. s'il existe une bijection permettant de le représenter par l'ensemble des $x_i, i \in \mathbb{N}$, la v.a. est dite discrète et sa loi est définie par l'ensemble des couples $(x_i, p_i)_{i \in \mathbb{N}}$ avec $p_i = P_X(X = x_i)$.
- 4 Si l'ensemble des valeurs possibles de X est un sous-ensemble non dénombrable de \mathbb{R} , la v.a. X est dite continue. Dans le cas où sa f.r. F est dérivable, sa loi est définie par sa densité de probabilité f qui est la fonction dérivée de $F : f = F'$.
- 5 Les deux principales caractéristiques d'une distribution de probabilité sont :

– l'espérance mathématique $E(X)$, caractéristique de valeur centrale :

$$E(X) = \sum_{i \in \mathbb{N}} p_i x_i \text{ ou } E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

– la variance $V(X)$ qui est une caractéristique de dispersion autour du centre :

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

- 6 Le calcul de la variance s'effectue presque toujours à partir de la formule développée : espérance du carré moins carré de l'espérance.

Principales propriétés de ces moments :

$$\forall a \in \mathbb{R}, \forall b \in \mathbb{R}, \quad E(aX + b) = aE(X) + b$$

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

$$\forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}, \quad E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$$

et, si les v.a. X et Y sont **indépendantes**, alors :

$$V(\lambda X + \mu Y) = \lambda^2 V(X) + \mu^2 V(Y)$$

- 7 Pour déterminer la loi de probabilité de la v.a. $Y = h(X)$, même dans le cas où la loi de X est définie par sa densité, il est presque toujours préférable de déterminer au préalable sa fonction de répartition, et éventuellement de dériver ensuite pour obtenir la densité.



- 2.1** Soit l'ensemble fondamental $\Omega = \{a, b, c\}$ et la tribu associée $\mathcal{A} = \{\emptyset, \{a, b\}, \{c\}, \Omega\}$, les événements a et b étant indiscernables. On définit l'application réelle X sur Ω par $X(a) = 1, X(b) = 2$ et $X(c) = 3$.
1. S'il y a équiprobabilité des événements élémentaires, peut-on définir la loi de probabilité de X ?
 2. Si la probabilité P est définie sur (Ω, \mathcal{A}) par $P(\{c\}) = 1$ et si on définit l'application Y sur Ω par $Y(\omega) = 3$ pour tout ω de Ω , calculer $P(X = Y)$. Peut-on en conclure que X et Y ont même loi de probabilité ?
- 2.2** Soit l'ensemble fondamental $\Omega = \{a, b, c, d, e\}$ et la partition $\Pi = \{\{a, b\}, \{c, d\}, \{e\}\}$ sur laquelle on définit la probabilité P par $P(\{a, b\}) = 2/5, P(\{c, d\}) = 1/5$ et $P(\{e\}) = 2/5$. Deux applications f et g sont définies sur Ω par $f(a) = f(b) = g(c) = g(d) = g(e) = -2, f(c) = f(d) = g(a) = g(b) = 2$ et $f(e) = 0$. Peut-on définir la loi de probabilité de l'application $X = f/g$?
- 2.3** Vous participez à un jeu où vous avez la probabilité p de remporter une partie. Si vous gagnez deux parties consécutives le jeu s'arrête et vous emportez un gain de $40 - 4N$ euros, N étant le nombre total de parties jouées. Le nombre maximum de parties jouées est fixé à quatre et vous donnez à votre adversaire dans ce jeu la somme de 25 euros en cas de perte. Ce jeu vous paraît-il équitable ?
- 2.4** Une urne contient cinq boules, deux qui portent le numéro 1, trois qui portent le numéro 2 et quatre qui portent le numéro 3. On effectue deux tirages successifs sans remise dans cette urne. On appelle coïncidence le fait de tirer une boule de numéro i au i -ème tirage, avec $i = 1, 2$. Déterminer la loi de probabilité de la variable aléatoire X qui représente le nombre de coïncidences observées, puis calculer $E(X)$ et $V(X)$.
- 2.5** Une urne contient une boule qui porte le numéro 0, deux qui portent le numéro 1, trois qui portent le numéro 2 et quatre qui portent le numéro 3. On extrait simultanément deux boules dans cette urne. Déterminer la loi de probabilité de la variable aléatoire X qui représente la somme des numéros obtenus puis calculer $E(X)$ et $V(X)$.
- 2.6** On considère n urnes qui contiennent chacune a boules noires et b boules blanches. On tire au hasard une boule de l'urne U_1 que l'on place ensuite dans l'urne U_2 . Puis, on tire au hasard une boule de l'urne U_2 que l'on place ensuite dans l'urne U_3 . Et ainsi de suite, la boule tirée de l'urne U_{n-1} étant placée dans l'urne U_n . On note p_n la probabilité de tirer une boule noire de l'urne U_n . Calculer p_1, p_2 et en déduire par récurrence la valeur de p_n .
- 2.7** La fonction de répartition F d'une v.a. X est définie par :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ 1 - 1/(1 \times 2) & \text{si } 1 < x \leq 2 \\ 1 - 1/(2 \times 3) & \text{si } 2 < x \leq 3 \\ \dots & \dots \\ 1 - 1/n(n+1) & \text{si } n < x \leq n+1 \\ \dots & \dots \end{cases}$$

1. Calculer les probabilités $p_n = P(X = n)$, $n \in \mathbb{N}^*$.
2. Calculer $E(X)$ et $V(X)$.

2.8 Soit X une v.a. de fonction de répartition F définie par :

$$F(x) = \begin{cases} e^x/3 & \text{pour } x \leq 0 \\ 1 & \text{pour } x > 0 \end{cases}$$

La loi de X est-elle continue ?

2.9 Soit X une variable aléatoire de densité nulle en dehors de $[-1, 1]$ et dont l'expression pour $x \in [-1, 1]$ est :

$$f(x) = \frac{3}{4} \sqrt{1 - |x|}$$

Déterminer la fonction de répartition de X .

2.10 Soit X une variable aléatoire de densité $f(x) = x$ pour $x \in [0, 1]$, $f(x) = 2 - x$ pour $x \in [1, 2]$ et nulle en dehors de ces intervalles.

1. Déterminer la fonction de répartition de X et en déduire la médiane de cette loi.
2. Calculer $P\{|X - 1| < x\}$ pour x réel quelconque.

2.11 Soit X une variable aléatoire de densité $f(x) = x + 2$, pour $x \in [-2, -1]$, $f(x) = -x + 2$ pour $x \in [1, 2]$ et nulle en dehors de ces deux intervalles.

1. Déterminer la fonction de répartition de X .
2. Calculer $E(X)$ et $V(X)$.

2.12 Soit X une v.a. de densité :

$$f(x) = \begin{cases} \frac{2}{a} \left(1 - \frac{x}{a}\right) & \text{si } 0 \leq x \leq a \\ 0 & \text{sinon} \end{cases}$$

où a est un réel strictement positif.

1. Déterminer la fonction de répartition (f.r.) de X .
2. Calculer $E(X)$ et $V(X)$.
3. Soit X_1, \dots, X_n des v.a. indépendantes, de même loi que X . Déterminer la f.r., puis la densité, de la v.a. $M_n = \max\{X_1, \dots, X_n\}$.

2.13 Soit X une v.a. de densité f continue et strictement positive, de fonction de répartition F . Exprimer à l'aide de f et F la densité et la fonction de répartition de chacune des v.a. suivantes.

1. $Y = aX + b$, $a \in \mathbb{R}^*$, $b \in \mathbb{R}$.
2. $Z = |X|$.
3. $T = \ln |X|$.
4. $U = F(X)$.
5. $V = [X]$, partie entière de X .

2.14 Soit X une variable aléatoire de densité $f(x) = \frac{1}{4} + \frac{3}{4}x^2$ pour $x \in [-1, 1]$ et nulle en dehors de cet intervalle.

1. Déterminer la densité de la v.a. $Y = X^2$.
2. Calculer $E(Y)$.

● Solutions



2.1 1. On a $X^{-1}(1) = \{a\} \in \mathcal{A}$ donc l'application X n'est pas une v.a. et on ne peut donc pas lui associer une loi de probabilité.

2. On a :

$$P(X = Y) = P(\{\omega \in \Omega / X(\omega) = Y(\omega)\}) = P(\{c\}) = 1$$

donc ces deux applications sont égales avec une probabilité égale à 1 (événement certain). L'application Y est constante, avec $Y^{-1}(3) = \Omega$, donc c 'est une variable aléatoire certaine : $P(Y = 3) = 1$. Cependant on ne peut pas en conclure que la loi de X est la même, puisque X n'est pas une variable aléatoire et que cette notion n'a donc pas de sens.

2.2 L'algèbre engendrée par la partition est :

$$\mathcal{A} = \{\emptyset, \{a, b\}, \{c, d\}, \{e\}, \{a, b, e\}, \{c, d, e\}, \{a, b, c, d\}, \Omega\}$$

L'application X ne peut prendre que les valeurs -1 , 0 et 1 avec :

$$\begin{aligned} X^{-1}(-1) &= \{\omega \in \Omega / f(\omega) = -g(\omega)\} = \{a, b\} \cup (\{c, d\} \cap \{c, d, e\}) \\ &= \{a, b, c, d\} \in \mathcal{A} \end{aligned}$$

$$X^{-1}(0) = \{\omega \in \Omega / f(\omega) = 0\} = \{e\} \in \mathcal{A}$$

$$\begin{aligned} X^{-1}(1) &= \{\omega \in \Omega / f(\omega) = g(\omega)\} = (\{a, b\} \cap \{c, d, e\}) \cup (\{c, d\} \cap \{a, b\}) \\ &= \emptyset \in \mathcal{A} \end{aligned}$$

Donc X est bien une variable aléatoire dont la loi de probabilité est définie par :

$$P_X(X = 0) = P(\{e\}) = \frac{2}{5} \text{ et } P_X(X = -1) = P(\{a, b, c, d\}) = \frac{3}{5}$$

2.3 Si le nombre de parties jouées est deux ou trois, c'est que vous avez gagné les deux dernières parties et donc $P(N = 2) = p^2$ et $P(N = 3) = qp^2$ où $q = 1 - p$. Vous avez gagné, événement noté E^+ , à l'issue de quatre parties avec une probabilité qui est :

$$P(\{N = 4\} \cap E^+) = pqp^2 + q^2p^2 = qp^2$$

La probabilité de perdre à l'issue des quatre parties jouées est :

$$P^- = 1 - P(N = 2) - P(N = 3) - P(\{N = 4\} \cap E^+)$$

L'espérance de gain à ce jeu est alors :

$$\begin{aligned} E(G) &= \sum_{k=2}^4 (40 - 4k) P(\{N = k\} \cap E^+) - 25P^- \\ &= 32p^2 + 28qp^2 + 24qp^2 - 25(1 - p^2 - 2qp^2) \\ &= 57p^2 + 102qp^2 - 25 = -102p^3 + 159p^2 - 25 \end{aligned}$$

Le jeu est considéré comme équitable si cette espérance est nulle puisqu'il n'y a pas de mise initiale. L'espérance est bien sûr une fonction croissante de p , de valeur -25 pour $p = 0$ et 32 pour $p = 1$. Pour $p = 1/2$ on obtient $E(G) = 2$, donc le jeu sera considéré comme équitable pour une valeur de p un peu inférieure à $1/2$.

- 2.4** Le nombre possible de coïncidences est 0, 1 ou 2. Aucune coïncidence correspond au tirage d'une boule 2, puis d'une boule 1 :

$$P(X = 0) = P(21) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

Une coïncidence unique peut se produire au premier ou au second tirage :

$$P(X = 1) = P(11) + P(22) = \frac{2}{5} \times \frac{1}{4} + \frac{3}{5} \times \frac{2}{4} = \frac{4}{10}$$

Pour deux coïncidences :

$$P(X = 2) = P(12) = \frac{2}{5} \times \frac{3}{4} = \frac{3}{10}$$

On calcule ensuite aisément :

$$E(X) = \frac{4}{10} + \frac{6}{4} = 1 \quad E(X^2) = \frac{4}{10} + \frac{12}{10} = \frac{16}{10} \quad V(X) = \frac{6}{10} = \frac{3}{5}$$

- 2.5** On note (i, j) l'événement « avoir tiré les boules i et j » ; pour $i = j$ il y a une seule façon d'obtenir ce couple et deux pour $i \neq j$. On obtient ainsi :

$$P(X = 0) = P(0, 0) = 0,$$

$$P(X = 1) = P(0, 1) = 2 \times \frac{1}{10} \times \frac{2}{9} = \frac{2}{45},$$

$$P(X = 2) = P(0, 2) + P(1, 1) = \frac{4}{45},$$

$$P(X = 3) = P(0, 3) + P(2, 1) = \frac{10}{45},$$

$$P(X = 4) = P(2, 2) + P(3, 1) = \frac{11}{45},$$

$$P(X = 5) = P(3, 2) = \frac{12}{45}, \quad P(X = 6) = \frac{6}{45}$$

On en déduit :

$$E(X) = 4 \quad E(X^2) = \frac{160}{9} \quad V(X) = \frac{16}{9}$$

- 2.6** On obtient :

$$p_1 = \frac{a}{a+b}$$

$$\begin{aligned} p_2 &= P(N_2 N_1) + P(N_2 B_1) = P(N_1)P(N_2 | N_1) + P(B_1)P(N_2 | B_1) \\ &= p_1 \frac{a+1}{a+b+1} + (1-p_1) \frac{a}{a+b+1} = \frac{a}{a+b} \end{aligned}$$

De la même manière :

$$p_n = p_{n-1} \frac{a+1}{a+b+1} + (1-p_{n-1}) \frac{a}{a+b+1}$$

Par récurrence, si on suppose que $p_{n-1} = p_1$ on en déduit alors $p_n = p_1$.

2.7 1. On obtient :

$$p_1 = P(X = 1) = P(X < 2) = F(2) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$p_2 = P(X = 2) = P(X < 3) - P(X < 2) = F(3) - F(2) = \frac{1}{3}$$

et plus généralement pour $n \geq 2$:

$$\begin{aligned} p_n &= P(X < n+1) - P(X < n) = \left[1 - \frac{1}{n(n+1)}\right] - \left[1 - \frac{1}{n(n-1)}\right] \\ &= \frac{2}{n(n^2-1)} \end{aligned}$$

2. On obtient comme espérance :

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} np_n = \frac{1}{2} + \sum_{n=2}^{\infty} \frac{2}{(n-1)(n+1)} = \frac{1}{2} + \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n+1}\right) \\ &= \frac{1}{2} + \left(1 - \frac{1}{3}\right) + \left(\frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{5}\right) + \dots \\ &\quad + \left(\frac{1}{n-3} - \frac{1}{n-1}\right) + \left(\frac{1}{n-2} - \frac{1}{n}\right) + \left(\frac{1}{n-1} - \frac{1}{n+1}\right) \\ &\quad + \left(\frac{1}{n} - \frac{1}{n+2}\right) + \dots \\ &= \frac{1}{2} + 1 + \frac{1}{2} = 2 \end{aligned}$$

Par définition :

$$E(X^2) = \sum_{n=1}^{\infty} n^2 p_n = \frac{1}{2} + \sum_{n=2}^{\infty} \frac{2n}{(n-1)(n+1)}$$

Or le terme général de cette série est équivalent à $2/n$, terme de la série harmonique divergente, donc cette variable aléatoire n'admet pas de moment d'ordre deux.

2.8 Par définition $F(0) = 1/3$ et, comme toute fonction de répartition, on vérifie que F est continue à gauche en ce point :

$$F(0-h) = \frac{1}{3}e^{-h} \rightarrow \frac{1}{3} = F(0) \text{ quand } h \rightarrow 0^+$$

Par contre $F(0+0) = 1$, donc F est discontinue à droite. La loi est mixte, avec un seul point de probabilité non nulle :

$$P(X = 0) = F(0+0) - F(0) = 2/3$$

2.9 On a $F(x) = 0$ pour $x \leq -1$ et $F(x) = 1$ pour $x \geq 1$. Pour $-1 \leq x \leq 0$:

$$F(x) = \frac{3}{4} \int_{-1}^x (1+t)^{1/2} dt = \frac{1}{2} (1+x)^{3/2}$$

Pour $0 \leq x \leq 1$:

$$F(x) = F(0) + \frac{3}{4} \int_{-1}^x (1-t)^{1/2} dt = 1 - \frac{1}{2} (1-x)^{3/2}$$

2.10 1. On a $F(x) = 0$ pour $x \leq 0$ et $F(x) = 1$ pour $x \geq 2$. Pour $0 \leq x \leq 1$:

$$F(x) = \int_0^x t dt = \frac{x^2}{2}$$

Pour $1 \leq x \leq 2$:

$$F(x) = F(1) + \int_1^x (2-t) dt = -\frac{x^2}{2} + 2x - 1$$

La médiane est le centre de symétrie de la distribution soit $Med = 1$.

2. On a $h(x) = P(|X-1| < x) = 0$ pour $x \leq 0$ et pour $x > 0$:

$$h(x) = P(1-x < X < 1+x) = F(1+x) - F(1-x)$$

Ainsi $h(x) = 1$ si $x \geq 1$ et pour $0 \leq x \leq 1$:

$$h(x) = 2x - x^2$$

2.11 On a $F(x) = 0$ pour $x \leq -2$ et $F(x) = 1$ pour $x \geq 2$. Pour $-2 \leq x \leq -1$:

$$F(x) = \int_{-2}^x (t+2) dt = \frac{x^2}{2} + 2x + 2$$

Pour $-1 \leq x \leq 1$:

$$F(x) = F(-1) = 1/2$$

Pour $1 \leq x \leq 2$:

$$F(x) = F(1) + \int_1^x (-t+2) dt = -\frac{x^2}{2} + 2x - 1$$

2.12 1. La f.r. F de X est nulle pour $x \leq 0$, a pour valeur 1 pour $x \geq a$ et a pour expression, quand $0 \leq x \leq a$:

$$F(x) = \frac{2}{a} \int_0^x \left(1 - \frac{t}{a}\right) dt = \frac{2}{a} \left[t - \frac{t^2}{2a} \right]_0^x = \frac{2x}{a} \left(1 - \frac{x}{2a}\right)$$

2. On obtient :

$$E(X) = \frac{2}{a} \int_0^a \left(x - \frac{x^2}{a}\right) dx = \frac{2}{a} \left(\frac{a^2}{2} - \frac{a^3}{3a}\right) = \frac{a}{3}$$

On calcule de même :

$$E(X^2) = \frac{2}{a} \int_0^a \left(x^2 - \frac{x^3}{a}\right) dx = \frac{2}{a} \left(\frac{a^3}{3} - \frac{a^4}{4a}\right) = \frac{a^2}{6}$$

On en déduit la variance :

$$V(X) = E(X^2) - E^2(X) = \frac{a^2}{18}$$

3. On calcule la f.r. de M_n :

$$G(x) = P(M_n < x) = P\left\{\bigcap_{i=1}^n (X_i < x)\right\} = \prod_{i=1}^n P(X_i < x) = F^n(x)$$

On obtient la densité par dérivation :

$$g(x) = G'(x) = nF^{n-1}(x) f(x)$$

Donc pour $0 \leq x \leq a$:

$$g(x) = n \left(\frac{x}{a}\right)^{n-1} \left(1 - \frac{x}{a}\right)^{n-1} \left(1 - \frac{x}{a}\right)$$

2.13 La méthode la plus simple consiste à déterminer la fonction de répartition (f.r.) de la nouvelle variable, puis à la dériver pour obtenir la densité.

1. Par définition :

$$G(y) = P(Y < y) = P(aX + b < y) = P(aX < y - b)$$

Pour pouvoir exprimer cette probabilité à l'aide de la f.r. de X , il faut distinguer deux cas suivant le signe de a .

Si $a > 0$:

$$G(y) = P\left(X < \frac{y-b}{a}\right) = F\left(\frac{y-b}{a}\right)$$

$$g(y) = G'(y) = \frac{1}{a} f\left(\frac{y-b}{a}\right)$$

Pour $a < 0$:

$$G(y) = P\left(X > \frac{y-b}{a}\right) = 1 - F\left(\frac{y-b}{a}\right)$$

$$g(y) = G'(y) = -\frac{1}{a} f\left(\frac{y-b}{a}\right)$$

Ainsi, pour $a \in \mathbb{R}^*$, la densité de Y peut s'écrire :

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$$

2. La f.r. de Z est définie par $H(z) = P(Z < z) = P(|X| < z)$, donc $H(z) = 0$ pour $z \leq 0$. Pour $z > 0$:

$$H(z) = P(-z < X < z) = F(z) - F(-z)$$

$$h(z) = H'(z) = f(z) + f(-z)$$

Remarquons que le changement de variable s'est effectué à l'aide de la fonction $x \mapsto |x|$ qui n'est pas injective et cependant nous avons pu déterminer la loi de $|X|$. La densité de Z présente une discontinuité pour $z = 0$ puisque $h(z) \rightarrow 2f(0) > 0$ quand $z \rightarrow 0^+$, alors que bien sûr H est continue, avec $H(z) \rightarrow F(0^+) - F(0^-) = 0 = H(0)$.

3. On obtient :

$$K(t) = P(T < t) = P(\ln |X| < t) = P(|X| < e^t) = H(e^t) = F(e^t) - F(-e^t)$$

$$k(t) = K'(t) = e^t [f(e^t) + f(-e^t)]$$

4. La fonction F étant dérivable est bien sûr continue; comme de plus elle est strictement monotone, elle admet une fonction réciproque F^{-1} définie sur $[0,1]$ et aussi strictement monotone, donc pour $0 \leq u \leq 1$:

$$P(U < u) = P\{F(X) < u\} = P\{X < F^{-1}(u)\} = F[F^{-1}(u)] = u$$

La variable $U = F(X)$ suit donc une loi uniforme sur $[0,1]$.

5. La variable V est bien sûr à valeurs entières, avec pour $v \in \mathbb{N}$:

$$P(V = v) = P([X] = v) = P(v \leq X < v + 1) = F(v + 1) - F(v).$$

Pour v entier négatif, on a $P([X] = v) = P(v < X \leq v + 1)$, ce qui ne change rien pour la valeur de cette probabilité puisque la loi de X est continue. Ainsi, pour tout $v \in \mathbb{Z}$, $P(V = v) = F(v + 1) - F(v)$.

2.14 1. La fonction de répartition G de Y est nulle pour $y \leq 0$ et pour $y > 0$:

$$G(y) = P(-\sqrt{y} < X < \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$$

La densité s'obtient par dérivation, soit pour $y > 0$:

$$g(y) = \frac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})] = \frac{1 + 3y}{4\sqrt{y}}$$

2. On obtient :

$$E(Y) = \frac{1}{4} \int_0^1 (\sqrt{y} + 3y\sqrt{y}) dy = \frac{7}{15}$$

Introduction

Si on améliore la compréhension et l'analyse d'un phénomène complexe par l'introduction d'un modèle qui la simplifie, celui-ci ne doit cependant pas être trop éloigné de la réalité. Nous allons présenter ici les principaux modèles qui peuvent être retenus pour une modélisation aléatoire. Ce « catalogue » des lois usuelles distingue encore entre lois discrètes et lois admettant une densité de probabilité.

Objectifs

Présenter les principales lois de probabilité pouvant être retenues dans la modélisation statistique.

Noter que pour certaines lois la convolution donne le même type de loi

Concepts clés

- Schéma binomial ou hypergéométrique
- Loi de Poisson
- Loi uniforme
- Loi exponentielle
- Loi normale
- Loi du khi-deux

1 Lois usuelles discrètes

1.1 Loi de Dirac

Soit $a \in \mathbb{R}$ un point fixé. On appelle **loi de Dirac**, notée δ_a , la loi de la v.a. certaine X qui est constante, prenant la même valeur a quel que soit le résultat de l'épreuve :

$$X(\omega) = a, \quad \forall \omega \in \Omega$$

Ainsi :

$$X(\Omega) = \{a\}, \quad P_X(X = a) = P\{\omega \in \Omega / X(\omega) = a\} = P(\Omega) = 1$$

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ 1 & \text{si } x > a \end{cases}$$

Le graphe de F présente un saut de valeur 1 au point de discontinuité a , qu'on appelle échelon de Heaviside.

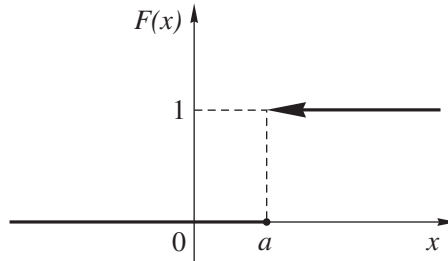


Figure 3.1

Bien entendu, on obtient comme moments :

$$E(X) = a \quad \text{et} \quad V(X) = 0.$$

Rappelons que c'est la seule v.a. dont la variance est nulle.

1.2 Loi de Bernoulli

Soit $A \in \mathcal{A}$ un événement quelconque ; on appelle v.a. indicatrice de l'événement A , la v.a. définie par $X = \mathbf{1}_A$, c'est-à-dire :

$$X(\omega) = \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in \bar{A} \end{cases}$$

Ainsi $X(\Omega) = \{0, 1\}$ avec :

$$P_X(X = 0) = P\{\omega \in \Omega / X(\omega) = 0\} = P(\bar{A}) = 1 - P(A)$$

$$P_X(X = 1) = P\{\omega \in \Omega / X(\omega) = 1\} = P(A)$$

On dit que X suit une loi de Bernoulli de paramètre $p = P(A)$, ce qu'on écrit symboliquement $X \rightsquigarrow \mathcal{B}(1, p)$ ou sous la forme suivante :

$$X = \begin{cases} 1 & p \\ 0 & q = 1 - p \end{cases}$$

La fonction de répartition est définie par :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ q & \text{si } 0 < x \leq 1 \\ 1 & \text{si } 1 < x \end{cases}$$

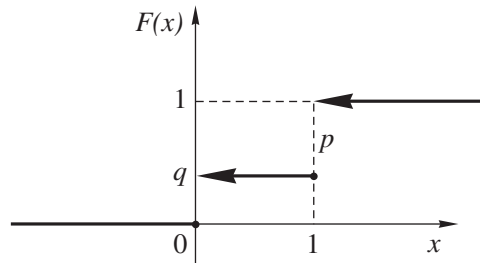


Figure 3.2

Exemple 3.1

Dans une population de n individus, on associe à chacun d'eux une v.a. de Bernoulli, indicatrice de possession d'un certain caractère A :

$$X_i = \begin{cases} 1 & \text{si } i \text{ possède le caractère } A \\ 0 & \text{si } i \text{ ne possède pas le caractère } A \end{cases}$$

Le paramètre $p = P(A)$ représente la proportion d'individus de la population qui possèdent ce caractère A .

Les moments de cette loi sont :

$$E(X) = 1 \times P(A) + 0 \times P(\bar{A}) = P(A) = p$$

$$E(X^2) = 1^2 \times P(A) + 0^2 \times P(\bar{A}) = P(A)$$

$$V(X) = E(X^2) - E^2(X) = P(A) - P^2(A) = P(A)P(\bar{A}) = pq$$

$$\mu_3(X) = p(1-p)^3 + (1-p)(0-p)^3 = p(1-p)(1-2p)$$

1.3 Loi binômiale

Si on effectue n épreuves successives indépendantes où on note à chaque fois la réalisation ou non d'un certain événement A , on obtient une suite de la forme $AA\bar{A}\bar{A}\bar{A}\bar{A}\dots\bar{A}\bar{A}\bar{A}$. À cet événement élémentaire ω on associe le nombre $X(\omega)$ de réalisations de A . On définit ainsi une v.a. X qui suit une loi binômiale de paramètres n et $p = P(A)$, caractérisée par $X(\Omega) = \{0, 1, \dots, n\}$ et pour $k \in X(\Omega)$:

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

car $\binom{n}{k}$ est le nombre d'échantillons de taille n comportant exactement k événements A , de probabilité p^k , indépendamment de l'ordre, et donc $n-k$ événements \bar{A} , de probabilité $(1-p)^{n-k}$. On écrit $X \rightsquigarrow B(n, p)$. Pour calculer facilement les moments de cette loi, nous allons associer à chaque épreuve i , $1 \leq i \leq n$, une v.a. de Bernoulli :

$$X_i = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } \bar{A} \text{ est réalisé} \end{cases}$$

On peut alors écrire :

$$X = X_1 + X_2 + \dots + X_n$$

d'où on déduit aisément :

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = np$$

et :

$$V(X) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = npq$$

car les v.a. X_i sont indépendantes.

On vérifie bien que c'est une loi de probabilité :

$$\sum_{k=0}^n P_X(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

Le calcul direct des moments de X peut s'effectuer à partir de la définition générale, mais de façon beaucoup plus laborieuse :

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

Pour obtenir $E(X^2)$ par un procédé de calcul identique, on passe par l'intermédiaire du moment factoriel $E[X(X-1)] = E(X^2) - E(X)$:

$$\begin{aligned} E[X(X-1)] &= \sum_{k=0}^n k(k-1) \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} \\ &= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j q^{n-2-j} = n(n-1)p^2(p+q)^{n-2} \\ &= n(n-1)p^2 \end{aligned}$$

on en déduit alors $E(X^2) = n(n-1)p^2 + np$, puis :

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

Exemple 3.2

Le nombre de résultats pile apparus au cours de n jets d'une pièce de monnaie suit une loi $\mathcal{B}(n, 1/2)$:

$$P_X(X = K) = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \frac{\binom{n}{k}}{2^n}, 0 \leq k \leq n$$

avec $E(X) = n/2$ et $V(X) = n/4$.

Exemple 3.3

Le nombre N de boules rouges apparues au cours de n tirages avec remise dans une urne contenant deux rouges, trois vertes et une noire suit une loi binômiale $\mathcal{B}(n, 1/3)$:

$$P_N(N = K) = \binom{n}{k} \left(\frac{2}{6}\right)^k \left(\frac{1}{6}\right)^{n-k} = \binom{n}{k} \frac{2^{n-k}}{3^n}, 0 \leq k \leq n$$

avec $E(N) = n/3$ et $V(N) = 2n/9$.

REMARQUES Si $X_1 \rightsquigarrow \mathcal{B}(n_1, p)$ et $X_2 \rightsquigarrow \mathcal{B}(n_2, p)$, les v.a X_1 et X_2 étant indépendantes, alors $X_1 + X_2 \rightsquigarrow \mathcal{B}(n_1 + n_2, p)$. Ceci résulte de la définition d'une loi binômiale puisqu'on totalise ici le résultat de $n_1 + n_2$ épreuves indépendantes.

Les tables de la loi binômiale ne sont fournies que pour $p < 1/2$. Pour une valeur $p > 1/2$, on utilise le résultat suivant :

$$\begin{aligned} P_X(X = k) &= P_X(n - X = n - k) = \binom{n}{k} p^k (1 - p)^{n-k} \\ &= \binom{n}{n - k} (1 - p)^{n-k} p^k \end{aligned}$$

qui exprime que $n - X \rightsquigarrow \mathcal{B}(n, 1 - p)$, loi qui est tabulée puisque $1 - p < 1/2$.

1.4 Loi hypergéométrique

On effectue n tirages **sans remise** dans une urne contenant N objets dont N_A objets A . On note $X(\omega)$ le nombre d'objets A tirés à l'issue de l'événement élémentaire ω . Les tirages successifs sont ici **dépendants** puisque la composition de l'urne est différente après chaque tirage, dépendant des tirages précédents. Dans le schéma binomial du paragraphe précédent on peut aussi considérer que l'on effectue n tirages **avec remise** dans une urne dont la composition est telle que $N_A/N = p$. Les épreuves successives sont alors **indépendantes**.

Dans le schéma hypergéométrique ici, ces n tirages sans remise sont équivalents à un seul tirage de n objets et il y a donc équiprobabilité de chacun des $\binom{N}{n}$ échantillons possibles. Pour calculer la probabilité d'obtenir k objets A il faut donc dénombrer tous

les échantillons qui contiennent exactement k des N_A objets A , il y en a $\binom{N_A}{k}$ chacun d'eux contenant simultanément $n - k$ objets \bar{A} , il y en a $\binom{N - N_A}{n - k}$.

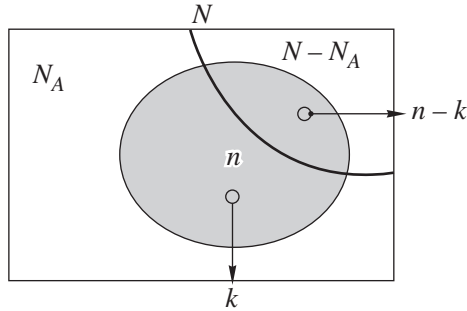


Figure 3.3

Ainsi, pour tout entier k tel que $0 \leq k \leq n$:

$$P_X(X = k) = \frac{\binom{N_A}{k} \binom{N - N_A}{n - k}}{\binom{N}{n}}$$

il faut bien entendu aussi que $k \leq N_A$ (nombre total d'objets A) et $n - k \leq N - N_A$ (nombre d'objets \bar{A}) d'où les conditions :

$$\max\{0, n - (N - N_A)\} \leq k \leq \min\{n, N_A\}.$$

Pour vérifier qu'il s'agit bien d'une loi de probabilité, on utilise le résultat suivant de combinatoire :

$$\sum_{k=0}^m \binom{r}{k} \binom{s}{m - k} = \binom{r + s}{m}$$

obtenu en effectuant le produit :

$$(1 + x)^r (1 + x)^s = \sum_{k=0}^r \binom{r}{k} x^k \sum_{j=0}^s \binom{s}{j} x^j = \sum_{k=0}^r \sum_{j=0}^s \binom{r}{k} \binom{s}{j} x^{j+k}$$

et en identifiant avec les termes de :

$$(1 + x)^{r+s} = \sum_{m=0}^{r+s} \binom{r+s}{m} x^m$$

On en déduit que $\sum_{k=0}^n \binom{N_A}{k} \binom{N - N_A}{n - k} = \binom{N}{n}$, ce qui permet de conclure que $\sum_{k=0}^n P_X(X = k) = 1$. La loi hypergéométrique dépend des trois paramètres N , n et N_A et on écrit symboliquement $X \rightsquigarrow \mathcal{H}(N, n, N_A)$.

Pour calculer plus facilement les moments de cette loi, nous allons supposer que chacun des objets A est numéroté et nous allons leur associer une v.a. indicatrice de tirage :

$$X_i = \begin{cases} 1 & \text{si l'objet } A_i \text{ est tiré} \\ 0 & \text{sinon} \end{cases}, \quad 1 \leq i \leq N_A$$

Ces variables permettent d'écrire :

$$X = \sum_{i=1}^{N_A} X_i$$

mais notons qu'ici les v.a. X_i ne sont pas indépendantes. Elles suivent la même loi de Bernoulli dont le paramètre est la probabilité de tirage d'un objet A particulier. Chaque échantillon a la même probabilité $1/\binom{N}{n}$; les échantillons qui contiennent l'objet A_i sont au nombre de $\binom{N-1}{n-1}$, donc :

$$P(X_i = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

On en déduit facilement que :

$$E(X) = \sum_{i=1}^{N_A} E(X_i) = N_A \frac{n}{N} = np$$

ayant posé $p = P(A) = N_A/N$.

Pour calculer $V(X)$, il est nécessaire de déterminer la loi des couples (X_i, X_j) puisque ces variables ne sont pas indépendantes. Le nombre d'échantillons qui contiennent simultanément les objets A_i et A_j est $\binom{N-2}{n-2}$, donc :

$$P(X_i = 1, X_j = 1) = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

Ainsi :

$$E(X_i X_j) = P(X_i = 1, X_j = 1) = \frac{n(n-1)}{N(N-1)}$$

et :

$$\begin{aligned} \text{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) = \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} \\ &= -\frac{n(N-n)}{N^2(N-1)} < 0 \end{aligned}$$

Par conséquent :

$$\begin{aligned} V(X) &= V\left(\sum_{i=1}^{N_A} X_i\right) = \sum_{i=1}^{N_A} V(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &= N_A \frac{n}{N} \frac{N-n}{N} - N_A(N_A-1) \frac{n(N-n)}{N^2(N-1)} \\ &= n \frac{N-n}{N-1} \frac{N_A}{N} \left(1 - \frac{N_A}{N}\right) \end{aligned}$$

Soit en posant $q = 1 - p = 1 - N_A/N$:

$$V(X) = npq \frac{N-n}{N-1}$$

Si la taille N de la population est grande vis-à-vis de la taille n de l'échantillon, on a l'approximation :

$$\frac{N-n}{N-1} = \frac{1-n/N}{1-1/N} \simeq 1$$

et :

$$V(X) \simeq npq$$

qui est l'expression de la variance de la loi binômiale $B(n, p)$, c'est-à-dire du cas de tirages indépendants (cas limite d'une population de taille infinie). Pour n/N petit et N grand on peut utiliser la loi binômiale (plus simple) comme approximation de la loi hypergéométrique (cf. chap. 6, 2.9 §2).

1.5 Loi de Poisson

Une v.a. X suit une loi de Poisson de paramètre $\lambda > 0$ si c'est une variable à valeurs entières, $X(\Omega) = \mathbb{N}$, donc avec une infinité de valeurs possibles, de probabilité :

$$P_X(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}$$

loi qui ne dépend que d'un seul paramètre réel positif, avec l'écriture symbolique $X \rightsquigarrow \mathcal{P}(\lambda)$.

Le développement en série entière de l'exponentielle :

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

permet de vérifier que :

$$\sum_{k=0}^{\infty} P_X(X = k) = 1$$

On peut déterminer quel est l'entier le plus probable en formant le rapport :

$$\frac{P_X(X = k)}{P_X(X = k - 1)} = \frac{\lambda^k}{k!} \times \frac{(k - 1)!}{\lambda^{k-1}} = \frac{\lambda}{k}, \quad k \geq 1$$

ce qui montre que pour tous les entiers k inférieurs à λ on a $P_X(X = k) > P_X(X = k - 1)$, donc $P_X(X = k)$ est croissant, puis décroissant pour les entiers $k > \lambda$, le maximum étant atteint pour l'entier $k = [\lambda]$. Pour $[\lambda] = 0$, les valeurs de $P_X(X = k)$ sont décroissantes à partir de la valeur maximale qui est $P_X(X = 0)$. Dans le cas particulier où λ est entier, il y a deux valeurs de probabilité maximale qui est $P_X(X = \lambda) = P_X(X = \lambda - 1)$.

Le calcul de l'espérance mathématique se déduit du développement en série entière de l'exponentielle :

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

Pour calculer la variance nous n'allons pas calculer $E(X^2)$ mais le moment factoriel $E[X(X - 1)]$ qui s'obtient plus facilement, selon la méthode précédente :

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

On en déduit :

$$V(X) = E(X^2) - E^2(X) = E[X(X - 1)] + E(X) - E^2(X) = \lambda$$

REMARQUES Si deux variables suivent des lois de Poisson et sont indépendantes, $X \rightsquigarrow \mathcal{P}(\lambda)$ et $Y \rightsquigarrow \mathcal{P}(\mu)$, alors leur somme suit aussi une loi de Poisson :

$$X + Y \rightsquigarrow \mathcal{P}(\lambda + \mu).$$

Le moment factoriel d'ordre trois s'obtient aisément et permet d'en déduire le moment centré d'ordre trois :

$$E[X(X - 1)(X - 2)] = \lambda^3 = E(X^3) - 3E(X^2) + 2E(X)$$

d'où on déduit :

$$\mu_3 = E[X - E(X)]^3 = E(X^3) - 3E(X^2)E(X) + 2E^3(X) = \lambda > 0$$

donc loi dissymétrique à droite.

1.6 Loi géométrique ou de Pascal

On effectue des épreuves successives indépendantes jusqu'à la réalisation d'un événement particulier A et on note X le nombre (aléatoire) d'épreuves effectuées. On définit ainsi une v.a. à valeurs entières de loi géométrique, ou de Pascal. À chaque épreuve est associé l'ensemble fondamental $\Omega = \left\{ A, \overline{A} \right\}$ et l'événement $\{X = k\}$ pour $k \in \mathbb{N}^*$ est représenté par une suite de $k - 1$ événements \overline{A} , terminée par l'événement A :

$$\underbrace{\overline{A} \overline{A} \dots \overline{A} A}_{k-1}$$

Si on pose $p = P(A)$, la probabilité de cet événement est :

$$P_X(X = k) = (1 - p)^{k-1} p$$

En utilisant la série entière $\sum_{k=0}^{\infty} x^k = 1/(1 - x)$ pour $|x| < 1$, puis en dérivant, on en déduit $\sum_{k=1}^{\infty} kx^{k-1} = 1/(1 - x)^2$, ce qui permet de vérifier que $\sum_{k=0}^{\infty} P_X(X = k) = 1$. Ceci permet également d'obtenir l'espérance :

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

où on a posé $q = 1 - p$. Le calcul de la variance se fait à partir du moment factoriel :

$$\begin{aligned} E[X(X - 1)] &= \sum_{k=2}^{\infty} k(k - 1) p q^{k-1} = p q \sum_{k=2}^{\infty} k(k - 1) q^{k-2} \\ &= \frac{2 p q}{(1 - q)^3} = \frac{2 q}{p^2} \end{aligned}$$

d'où on déduit :

$$V(X) = E[X(X - 1)] + E(X) - E^2(X) = \frac{q}{p^2}$$

1.7 Loi binômiale négative

On effectue cette fois des épreuves successives indépendantes jusqu'à ce que n événements A soient réalisés et on note Y le nombre (aléatoire) d'épreuves effectuées. L'événement $\{Y = y\}$, pour tout entier $y \geq n$, est représenté par une suite de la forme :

$$\underbrace{\overline{A} \overline{A} A \dots \overline{A} \overline{A} \dots \overline{A} A}_{y-1}$$

qui comporte $n - 1$ réalisations de l'événement A au cours des $y - 1$ premières épreuves et qui se conclut par un événement A . On en déduit la probabilité individuelle :

$$P_Y(Y = y) = \binom{y - 1}{n - 1} p^n (1 - p)^{y-n}, \quad y \geq n$$

Pour obtenir sans calculs les moments de Y , nous allons décomposer la suite des épreuves en n séquences se terminant toutes par un événement A , associant à chacune de ces séquences une v.a. de Pascal $X_i, 1 \leq i \leq n$, qui représente le nombre d'épreuves nécessaires pour que le i -ème événement A soit réalisé, en comptant à partir de la réalisation du précédent A :

$$\underbrace{\bar{A}\bar{A}\dots\bar{A}A}_{X_1} \underbrace{\bar{A}\bar{A}\dots\bar{A}A}_{X_2} \dots \underbrace{\bar{A}\bar{A}\dots\bar{A}A}_{X_n}$$

Ceci permet de définir la loi de Y , dite loi binômiale négative, comme somme de lois de Pascal indépendantes et de même paramètre p :

$$Y = X_1 + \dots + X_n$$

On en déduit alors facilement :

$$E(Y) = nE(X_1) = \frac{n}{p} \quad \text{et} \quad V(Y) = nV(X_1) = \frac{nq}{p^2}$$

Pour la loi binômiale, le nombre d'épreuves est fixé et on observe le nombre aléatoire d'événements réalisés. Pour la loi binômiale négative, au contraire, le nombre de réalisations d'événements est fixé et c'est le nombre d'épreuves nécessaires pour les obtenir qui devient aléatoire.

2 Lois usuelles continues

2.1 Loi uniforme

Une v.a. X suit une loi uniforme continue si sa densité est constante sur un intervalle fini $[a, b]$, étant donc de la forme :

$$f(x) = \begin{cases} k & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

On écrit $X \rightsquigarrow \mathcal{U}([a, b])$. Compte tenu du graphe de la densité, on appelle aussi cette loi distribution rectangulaire. La constante k doit être telle que $\int_{-\infty}^{+\infty} f(x)dx = 1$ soit $\int_a^b kdx = k(b-a) = 1$, ce qui impose $k = 1/(b-a)$. La densité a pour valeur sur l'intervalle $[a, b]$ l'inverse de sa longueur :

$$f(x) = \frac{1}{b-a}, \forall x \in [a, b]$$

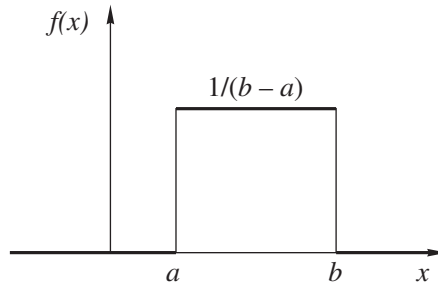


Figure 3.4

Déterminons la fonction de répartition de X :

–si $x < a$:

$$F(x) = \int_{-\infty}^x 0 dt = 0$$

–si $a \leq x < b$:

$$F(x) = \int_{-\infty}^a 0 dt + \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}$$

–si $b \leq x$:

$$F(x) = \int_{-\infty}^a 0 dt + \int_a^b \frac{1}{b-a} dt + \int_b^x 0 dt = \frac{b-a}{b-a} = 1$$

On obtient donc :

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x < b \\ 1 & \text{si } b \leq x \end{cases}$$

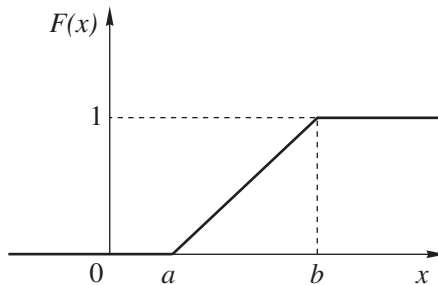


Figure 3.5

Le fractile d'ordre $p \in]0, 1[$, qui est défini par $F(x_p) = p$, a pour valeur ici $x_p = a + (b-a)p$.

La densité est discontinue en a et b mais la loi est absolument continue et la fonction de répartition est bien sûr continue en ces points.

Dans le cas particulier où $a = 0$ et $b = 1$, on a $X \rightsquigarrow \mathcal{U}([0, 1])$ avec :

$$f(x) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{sinon} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{cases}$$

La probabilité d'un intervalle $[x_1, x_2]$ inclus dans $[a, b]$ est proportionnelle à sa longueur :

$$P_X(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx = \frac{1}{b-a} \int_{x_1}^{x_2} dx = \frac{x_2 - x_1}{b-a}$$

Calculons l'espérance :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{b+a}{2}$$

c'est-à-dire que le centre de la distribution est le milieu de l'intervalle, résultat prévisible puisque la loi est uniforme sur cet intervalle. Dans le cas particulier où $a = 0$ et $b = 1$ on obtient $E(X) = 1/2$.

Calculons maintenant la variance :

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b$$

$$= \frac{1}{3}(b^2 + ab + a^2)$$

$$V(X) = E(X^2) - E^2(X) = \frac{(b-a)^2}{12}$$

Donc $V(X) = 1/12$ dans le cas de la loi uniforme sur $[0, 1]$.

2.2 Loi exponentielle

La loi exponentielle de paramètre $\theta > 0$ est celle d'une variable positive de densité :

$$f(x) = \begin{cases} \theta e^{-\theta x} & \text{si } 0 \leq x \\ 0 & \text{si } x < 0 \end{cases}$$

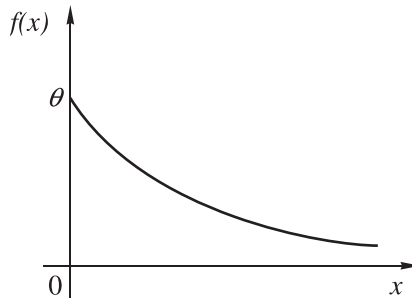


Figure 3.6

La variable associée X est souvent utilisée pour représenter une durée de vie (durée de vie d'un matériel donné, durée de chômage, durée d'hospitalisation ...). On écrit $X \rightsquigarrow \mathcal{E}(\theta)$. Sa fonction de répartition est bien sûr nulle pour $x \leq 0$, et pour $x > 0$ on obtient :

$$F(x) = \theta \int_0^x e^{-\theta t} dt = [-e^{-\theta t}]_0^x = 1 - e^{-\theta x}$$

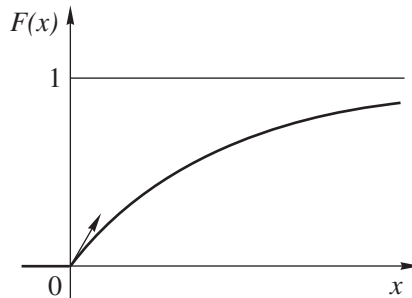


Figure 3.7

On calcule l'espérance en intégrant par parties :

$$E(X) = \theta \int_0^{+\infty} x e^{-\theta x} dx = [-x e^{-\theta x}]_0^{+\infty} + \int_0^{+\infty} e^{-\theta x} dx = \frac{1}{\theta}$$

On calcule de même, en intégrant par parties, le moment d'ordre deux :

$$E(X^2) = \theta \int_0^{+\infty} x^2 e^{-\theta x} dx = [-x^2 e^{-\theta x}]_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\theta x} dx = \frac{2}{\theta} E(X) = \frac{2}{\theta^2}$$

d'où on déduit :

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

2.3 Loi normale ou de Laplace-Gauss

C'est la loi d'une variable aléatoire X à valeurs dans \mathbb{R} , de densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-m)^2}{2\sigma^2}$$

qui est définie par deux paramètres m et $\sigma > 0$ dont nous verrons l'interprétation un peu plus loin. On note $X \rightsquigarrow N(m, \sigma)$. Bien entendu, s'agissant d'une densité de probabilité, on en déduit la valeur de l'intégrale suivante :

$$\int_{-\infty}^{+\infty} e^{-(x-m)^2/2\sigma^2} dx = \sigma\sqrt{2\pi}$$

Les remarques suivantes vont nous permettre de préciser le graphe de f .

REMARQUES

1. On peut constater que $f(2m-x) = f(x)$, ce qui indique que le graphe de f est symétrique par rapport à la droite verticale $x = m$.
2. L'expression $(x-m)^2$ est minimum pour $x = m$, ce qui va correspondre à un maximum pour f de valeur :

$$f(m) = \frac{1}{\sigma\sqrt{2\pi}}$$

3. Pour calculer facilement la dérivée, considérons :

$$\ln f(x) = -\ln\sigma\sqrt{2\pi} - \frac{1}{2\sigma^2}(x-m)^2$$

D'où en dérivant :

$$\frac{f'(x)}{f(x)} = -\frac{1}{\sigma^2}(x-m) \quad \text{et} \quad \sigma^2 f'(x) = (m-x)f(x)$$

Et en dérivant à nouveau :

$$\sigma^2 f''(x) = -f(x) + (m-x)f'(x)$$

d'où on déduit :

$$\sigma^4 f''(x) = (m-x)^2 f(x) - \sigma^2 f(x) = (m-x-\sigma)(m-x+\sigma)f(x)$$

donc f'' s'annule en changeant de signe pour $x = m-\sigma$ et $x = m+\sigma$, ce qui correspond à deux points d'inflexion pour le graphe de f .

4. Enfin, quand x devient infini, alors $f(x) \rightarrow 0$ donc l'axe des abscisses est asymptote au graphe.

Toutes ces remarques permettent de tracer le graphe en cloche de la densité f .

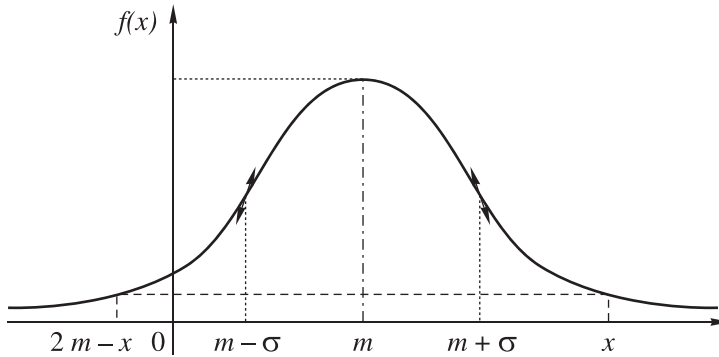


Figure 3.8

L'intégrale $\int_{-\infty}^{+\infty} x f(x) dx$ est convergente en raison de la présence de l'exponentielle, donc $E(X)$ existe et sa valeur ne peut être que m en raison de la symétrie de la densité par rapport à cette valeur. Vérifions-le en écrivant :

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} (x - m + m) f(x) dx \\ &= m \int_{-\infty}^{+\infty} f(x) dx + \int_{-\infty}^{+\infty} (x - m) f(x) dx \\ &= m + \int_{-\infty}^{+\infty} (x - m) f(x) dx \end{aligned}$$

l'intégrande étant une fonction impaire de $x - m = u$, l'intégrale est nulle puisqu'elle est convergente et qu'on intègre une fonction impaire sur un intervalle centré à l'origine. On retrouve bien l'interprétation du premier paramètre comme moyenne de la loi :

$$E(X) = m.$$

Pour calculer le moment d'ordre deux, on fait le changement de variable $u = (x - m)/\sigma$:

$$\begin{aligned} E(X^2) &= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma^2 u^2 + 2m\sigma u + m^2) e^{-u^2/2} du \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{-u^2/2} du + m^2 \end{aligned}$$

que l'on intègre par parties :

$$E(X^2) = \left[-\frac{\sigma^2 u e^{-u^2/2}}{\sqrt{2\pi}} \right]_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2} du + m^2 = \sigma^2 + m^2$$

d'où :

$$V(X) = \sigma^2$$

le second paramètre étant donc l'écart type σ . On obtient aussi $E(X - m)^3 = 0$, comme d'ailleurs tous les autres moments centrés impairs qui sont nuls, et $E(X - m)^4 = 3\sigma^4$.

a) Loi normale centrée réduite (loi normale standard)

En faisant le changement de variable $U = (X - m)/\sigma$, c'est-à-dire en centrant et en réduisant, on obtient une v.a. de loi standard, de moyenne nulle $E(U) = 0$ et de variance unité $V(U) = E(U^2) = E(X - m)^2/\sigma^2 = V(X)/\sigma^2 = 1$, donc de densité :

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

On peut donc en déduire la valeur de l'intégrale $\int_{-\infty}^{+\infty} e^{-u^2/2} du = \sqrt{2\pi}$.

On obtient les dérivées à partir du logarithme, $\ln\varphi(u) = -\ln\sqrt{2\pi} - u^2/2$:

$$\varphi'(u) = -u\varphi(u) \quad \text{et} \quad \varphi''(u) = -\varphi(u) - u\varphi'(u) = (u^2 - 1)\varphi(u)$$

La valeur maximale de φ est $\varphi(0) = 1/\sqrt{2\pi} = 0,3989$ et les valeurs décroissent rapidement, avec par exemple $\varphi(2) = 0,0540$ et $\varphi(4) = 0,0001$.

La fonction de répartition est définie par :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$$

et n'est pas exprimable au moyen d'une fonction usuelle. Les valeurs de Φ sont fournies dans les tables statistiques (table 1) pour $x \geq 0$. Pour $x < 0$, on utilise le fait que φ est une fonction paire, $\varphi(u) = \varphi(-u)$, c'est-à-dire que la loi est symétrique par rapport au centre de distribution 0, soit : $P(U < -x) = P(U > x)$, ce qui se traduit pour la f.r. par $\Phi(-x) = 1 - \Phi(x)$.

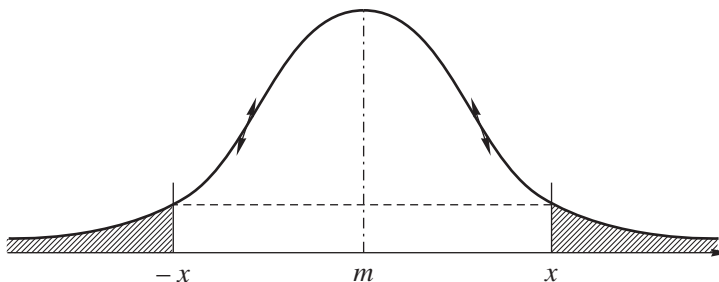


Figure 3.9

De cette symétrie découle également la probabilité d'un intervalle centré à l'origine :

$$P(|U| < a) = P(-a < U < a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1, a > 0$$

Ceci permet par exemple de déterminer l'intervalle interquartile, c'est-à-dire l'intervalle centré à l'origine et qui contient 50 % de la masse totale. On doit trouver la valeur de a telle que $P(|U| < a) = 1/2$, soit $2\Phi(a) - 1 = 1/2$ ou $\Phi(a) = 0,75$. Il s'agit donc du fractile d'ordre 0,75 de la loi normale standard, dont la valeur est lue dans la table 2 des fractiles : $a = Q_3 = 0,6745 \simeq 2/3$. Rappelons que le fractile d'ordre $p \in]0, 1[$ est la valeur u_p telle que $\Phi(u_p) = p$, soit $u_p = \Phi^{-1}(p)$. De même, on peut calculer la probabilité $P(|U| < a)$ des intervalles centrés à l'origine, pour les valeurs entières de a . Ainsi : $2\Phi(1) - 1 = 0,68268$, $2\Phi(2) - 1 = 0,95450$, $2\Phi(3) - 1 = 0,9973$; il n'y a pratiquement plus de masse de probabilité au-delà de la valeur 4 puisque : $P(|U| > 4) \simeq 6 \times 10^{-5}$. La figure 3.10 ci-après synthétise ces résultats pour une loi quelconque, à partir de l'équivalence :

$$-a < \frac{x - m}{\sigma} < a \Leftrightarrow m - \sigma a < x < m + \sigma a$$

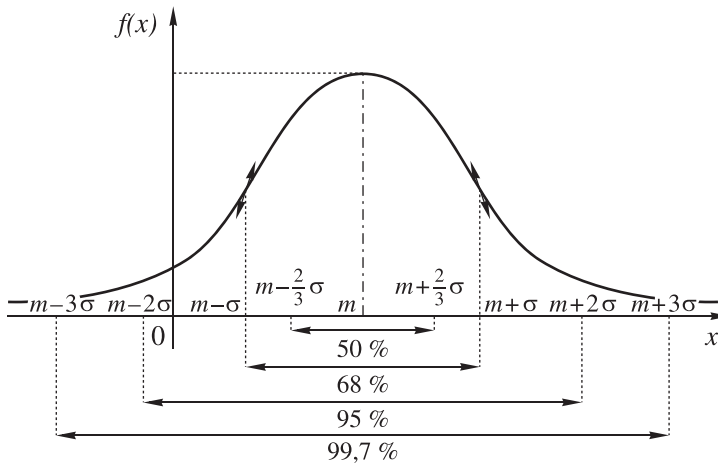


Figure 3.10

Pour déterminer les valeurs de la f.r. d'une loi normale quelconque, on se ramène à la loi standard qui est tabulée, à partir de :

$$\begin{aligned} F(x) &= P(X < x) = P\left(\frac{X - m}{\sigma} < \frac{x - m}{\sigma}\right) = P\left(U < \frac{x - m}{\sigma}\right) \\ &= \Phi\left(\frac{x - m}{\sigma}\right) \end{aligned}$$

b) Convolution de lois normales

La convolution (somme) de deux lois normales indépendantes est encore une loi normale : si $X \rightsquigarrow \mathcal{N}(m_1, \sigma_1)$ et $Y \rightsquigarrow \mathcal{N}(m_2, \sigma_2)$ sont des v.a. indépendantes, alors

$$X + Y \rightsquigarrow \mathcal{N}\left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right).$$

2.4 Loi gamma

Une v.a. X suit une loi gamma de paramètres $p > 0$ et $\theta > 0$ si c'est une v.a. positive dont la densité est de la forme :

$$f(x) = \frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1}, x \geq 0$$

la fonction gamma étant définie pour tout $p > 0$ par :

$$\Gamma(p) = \int_0^{+\infty} e^{-x} x^{p-1} dx$$

On écrit $X \rightsquigarrow \gamma(p, \theta)$. Parmi les nombreuses propriétés de la fonction Γ , on montre en intégrant par parties que pour tout $p > 1$:

$$\Gamma(p) = (p-1)\Gamma(p-1)$$

Donc, pour p entier strictement positif on en déduit que :

$$\Gamma(p) = (p-1)!$$

Le calcul des moments s'effectue aisément par le changement de variable $y = \theta x$

$$E(X) = \frac{\theta^p}{\Gamma(p)} \int_0^{+\infty} e^{-\theta x} x^p dx = \frac{1}{\Gamma(p)} \int_0^{+\infty} e^{-y} y^p \frac{dy}{\theta} = \frac{1}{\theta} \frac{\Gamma(p+1)}{\Gamma(p)} = \frac{p}{\theta}$$

$$E(X^2) = \frac{\theta^p}{\Gamma(p)} \int_0^{+\infty} e^{-\theta x} x^{p+1} dx = \frac{1}{\Gamma(p)} \int_0^{+\infty} e^{-y} y^{p+1} \frac{dy}{\theta^2} = \frac{1}{\theta^2} \frac{\Gamma(p+2)}{\Gamma(p)} = \frac{p(p+1)}{\theta^2}$$

$$V(X) = \frac{p}{\theta^2}.$$

Étudions la loi de la v.a. $Y = \theta X$:

$$G(y) = P(Y < y) = P(\theta X < y) = P\left(X < \frac{y}{\theta}\right) = F\left(\frac{y}{\theta}\right)$$

où F est la f.r. de X ; la densité de Y est obtenue par dérivation :

$$g(y) = \frac{1}{\theta} f\left(\frac{y}{\theta}\right) = \frac{\theta^{p-1}}{\Gamma(p)} e^{-y} \left(\frac{y}{\theta}\right)^{p-1} = \frac{1}{\Gamma(p)} e^{-y} y^{p-1}, \quad y > 0$$

et on reconnaît la densité d'une loi $\gamma(p, 1)$ que nous noterons simplement $\gamma(p)$.

Les moments s'obtiennent à partir de ceux de X en faisant $\theta = 1$:

$$E(Y) = V(Y) = p$$

On a donc les équivalences suivantes entre les lois gammas à 1 et 2 paramètres :

$$X \rightsquigarrow \gamma(p, \theta) \Leftrightarrow Y = \theta X \rightsquigarrow \gamma(p) \Leftrightarrow X = \frac{Y}{\theta} \rightsquigarrow \gamma(p, \theta)$$

Convolution de lois gammas

La convolution (somme) de deux lois gammas indépendantes est encore une loi gamma : si $X \rightsquigarrow \gamma(p, \theta)$ et $Y \rightsquigarrow \gamma(q, \theta)$ sont des v.a. indépendantes, alors la v.a. somme $X + Y \rightsquigarrow \gamma(p + q, \theta)$.

La famille des lois gammas contient comme cas particuliers deux lois usuelles, l'une déjà étudiée qui est la loi exponentielle de paramètre $\theta > 0$, obtenue pour $p = 1$ et qui est donc la loi $\gamma(1, \theta)$ de densité $\theta e^{-\theta x}$ pour $x > 0$. Cette remarque peut être utile en statistique dans le cas de v.a. X_1, \dots, X_n indépendantes et de même loi exponentielle de paramètre θ , avec :

$$S_n = \sum_{i=1}^n X_i \rightsquigarrow \gamma(n, \theta) \quad \text{et} \quad \theta S_n \rightsquigarrow \gamma(n)$$

La seconde est la loi du khi-deux, très utilisée en statistique et que nous allons étudier maintenant.

2.5 Loi du khi-deux

La loi du khi-deux à n degrés de liberté, notée χ_n^2 , est la loi $\gamma(n/2, 1/2)$ où n est un entier positif, donc de densité pour $x > 0$:

$$f(x) = \frac{1}{2^{n/2} \Gamma(n/2)} e^{-x/2} x^{n/2-1}$$

Ses moments se déduisent de ceux de la loi gamma :

$$E(\chi_n^2) = \frac{n/2}{1/2} = n \quad \text{et} \quad V(\chi_n^2) = \frac{n/2}{1/4} = 2n$$

REMARQUES En appliquant les équivalences ci-dessus entre lois gammas pour $\theta = \frac{1}{2}$, on peut passer de la loi gamma non tabulée à la loi du khi-deux qui est tabulée :

$$Y = \frac{1}{2} X \rightsquigarrow \gamma(p) \Leftrightarrow X = 2Y \rightsquigarrow \gamma\left(p, \frac{1}{2}\right) \equiv \chi_{2p}^2$$

Pour des v.a. exponentielles comme ci-dessus on en déduit $2\theta S_n \rightsquigarrow \chi_{2n}^2$.

Il existe également un lien avec la loi normale qui explique son importance en statistique. Si $X \rightsquigarrow N(m, \sigma)$ alors :

$$\left(\frac{X - m}{\sigma}\right)^2 \rightsquigarrow \chi_1^2$$

En effet, la v.a. $U = (X - m)/\sigma$ suit une loi $N(0, 1)$ et si on pose $Y = U^2$, sa f.r. est définie pour $y > 0$ par :

$$G(y) = P(Y < y) = P(-\sqrt{y} < U < \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

et sa densité est donc :

$$g(y) = \frac{1}{2\sqrt{2}} [\varphi(\sqrt{y}) + \varphi(-\sqrt{y})] = \frac{\varphi(\sqrt{y})}{\sqrt{y}} = \frac{e^{-y/2}}{\sqrt{2\pi y}}$$

c'est-à-dire, puisque $\Gamma(1/2) = \sqrt{\pi}$, la densité d'une loi $\gamma(1/2, 1/2)$ ou loi χ_1^2 .

Ceci permet notamment de retrouver l'expression du moment centré d'ordre quatre de la loi normale, puisque de $V[(X - m)/\sigma]^2 = 2$ on en déduit par la formule développée de la variance $E(X - m)^4 - E^2(X - m)^2 = 2\sigma^4$ puis $E(X - m)^4 = 3\sigma^4$.

Convolution de lois du khi-deux

La convolution (somme) de deux lois du khi-deux indépendantes est encore une loi du khi-deux : si $X \rightsquigarrow \chi_n^2$ et $Y \rightsquigarrow \chi_m^2$ sont des v.a. indépendantes alors la v.a. $X + Y \rightsquigarrow \chi_{n+m}^2$. Il s'agit simplement de la propriété relative aux lois gammas.

Des deux propriétés précédentes nous allons déduire une autre propriété qui indique comment s'introduit cette loi dans les échantillons gaussiens : si X_1, \dots, X_n sont des v.a. indépendantes et de même loi $N(0, 1)$, alors $X_1^2 + \dots + X_n^2$ suit une loi du khi-deux à n degrés de liberté. Cette propriété pourrait d'ailleurs servir de définition de cette loi, sans référence à la famille des lois gammas. Le nombre de degrés de liberté correspond au nombre de variables indépendantes qui sont intervenues dans sa construction ; si ces variables étaient liées par k relations, le nombre de degrés de liberté serait alors $n - k$.

2.6 Loi bêta

Il existe deux familles de lois bêtas qui se déduisent de la famille des lois gammas.

a) Loi bêta de seconde espèce

Si X et Y sont deux v.a. indépendantes de lois respectives $\gamma(p)$ et $\gamma(q)$, alors la v.a. $Z = X/Y$ suit une loi bêta de seconde espèce de paramètres $p > 0$ et $q > 0$, notée $\beta_{II}(p, q)$, et de densité pour $z > 0$:

$$f(z) = \frac{1}{B(p, q)} \frac{z^{p-1}}{(1+z)^{p+q}}$$

où :

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Des propriétés de la fonction Γ on déduit aisément les moments de Z :

$$E(Z) = \frac{1}{B(p, q)} \int_0^{+\infty} \frac{z^p}{(1+z)^{p+q}} dz = \frac{B(p+1, q-1)}{B(p, q)} = \frac{p}{q-1}, \quad q > 1$$

$$E(Z^2) = \frac{1}{B(p, q)} \int_0^{+\infty} \frac{z^{p+1}}{(1+z)^{p+q}} dz = \frac{B(p+2, q-2)}{B(p, q)} = \frac{p(p+1)}{(q-1)(q-2)}, \quad q > 2$$

$$V(Z) = \frac{p(p+q-1)}{(q-1)^2(q-2)}, \quad q > 2$$

b) Loi bêta de première espèce

La loi bêta de première espèce est également définie par un rapport de lois gammas ou à partir de la loi précédente ; c'est celle de la v.a. à valeurs dans $[0, 1]$:

$$T = \frac{X}{X+Y} = \frac{Z}{1+Z}$$

sa densité pour $0 \leq t \leq 1$ est :

$$f(t) = \frac{1}{B(p, q)} t^{p-1} (1-t)^{q-1}$$

on écrit $T \rightsquigarrow \beta_I(p, q)$. Notons que :

$$\int_0^1 t^{p-1} (1-t)^{q-1} dt = B(p, q)$$

On obtient comme moments :

$$E(T) = \frac{1}{B(p, q)} \int_0^1 t^p (1-t)^{q-1} dt = \frac{B(p+1, q)}{B(p, q)} = \frac{p}{p+q}$$

$$E(T^2) = \frac{1}{B(p, q)} \int_0^1 t^{p+1} (1-t)^{q-1} dt = \frac{B(p+2, q)}{B(p, q)} = \frac{p(p+1)}{(p+q)(p+q+1)}$$

$$V(T) = \frac{pq}{(p+q)^2(p+q+1)}$$

2.7 Loi log-normale

La v.a. positive X suit une loi log-normale de paramètres m et $\sigma > 0$ si la v.a. $\ln X$ suit une loi $N(m, \sigma)$; sa f.r. vérifie donc pour $x > 0$:

$$F(x) = P(X < x) = P(\ln X < \ln x) = P\left(\frac{\ln X - m}{\sigma} < \frac{\ln x - m}{\sigma}\right)$$

$$= \Phi\left(\frac{\ln x - m}{\sigma}\right)$$

sa densité est donc pour $x > 0$:

$$f(x) = \frac{1}{\sigma x} \varphi\left(\frac{\ln x - m}{\sigma}\right) = \frac{1}{\sigma x \sqrt{2\pi}} \exp -\frac{1}{2\sigma^2} (\ln x - m)^2$$

2.8 Loi de Pareto

C'est une loi qui est utilisée notamment dans la modélisation de la distribution des revenus d'une population ou en théorie des assurances. Sa densité est définie pour $x \geq x_0 > 0$, x_0 pouvant s'interpréter comme le revenu minimum, en fonction d'un paramètre $\alpha > 0$:

$$f(x) = \frac{\alpha}{x_0} \left(\frac{x_0}{x}\right)^{\alpha+1}$$

3 Compléments : fonctions génératrices

La notion de fonction génératrice peut être utile parfois pour calculer plus facilement les moments de certaines lois de probabilité.

3.1 Fonction génératrice d'une v.a. discrète positive

Si X est une v.a. discrète positive dont la loi de probabilité est définie par l'ensemble des couples (x_k, p_k) , où k parcourt un ensemble d'indices $K \subset \mathbb{N}$, on peut définir sur $[0, 1]$ sa fonction génératrice par :

$$G_X(u) = E(u^X) = \sum_{k \in K} p_k u^{x_k}$$

Pour $0 \leq u \leq 1$, la série qui définit G_X est uniformément convergente puisque $|p_k u^{x_k}| \leq p_k$ et que $\sum_{k \in K} p_k = 1$. Par ailleurs, $G_X(0) = 0$ et $G_X(1) = 1$. Un cas particulier intéressant est celui d'une variable à valeurs entières, soit $X(\Omega) = K \subset \mathbb{N}$, où G_X est définie pour tout $u \in \mathbb{C}$ tel que $|u| \leq 1$ par :

$$G_X(u) = E(u^X) = \sum_{k \in K} p_k u^k$$

Exemple 3.4

Loi de Dirac (masse en a) : $G_X(u) = u^a$.

Exemple 3.5

Loi de Bernoulli : $G_X(u) = pu + q$.

Exemple 3.6

Loi binômiale : $G_X(u) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} u^k = (pu + q)^n$.

Exemple 3.7

Loi de Poisson : $G_X(u) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} u^k = e^{\lambda(u-1)}$.

Exemple 3.8

Loi géométrique : $G_X(u) = \sum_{k=1}^{\infty} pq^{k-1} u^k = pu \sum_{k=1}^{\infty} (qu)^{k-1} = \frac{pu}{1-qu}$.

Il y a une correspondance biunivoque entre loi de probabilité et fonction génératrice, due à l'unicité du développement en série entière de Taylor :

$$G_X(u) = \sum_{k \in \mathbb{N}} p_k u^k = \sum_{k \in \mathbb{N}} \frac{G_X^{(k)}(0)}{k!} u^k$$

ce qui permet d'obtenir par exemple la loi de probabilité à partir de la fonction génératrice, par les relations :

$$G_X(0) = p_0 \text{ et } G_X^{(k)}(0) = k! p_k \text{ pour } k \in \mathbb{N}^*$$

Si X_1, \dots, X_n sont des v.a. mutuellement indépendantes, alors :

$$\begin{aligned} G_{X_1+\dots+X_n}(u) &= E(u^{X_1+\dots+X_n}) = E\left(\prod_{i=1}^n u^{X_i}\right) \\ &= \prod_{i=1}^n E(u^{X_i}) = \prod_{i=1}^n G_{X_i}(u) \end{aligned}$$

Exemple 3.9

La loi binômiale négative de paramètres n et p peut s'écrire comme somme de n lois géométriques indépendantes, donc, d'après l'exemple 3.8 :

$$G_X(u) = \left(\frac{pu}{1-qu}\right)^n$$

La fonction génératrice est surtout utile pour déterminer les moments factoriels :

$$\mu_{[k]} = E[X(X-1)\dots(X-k+1)], \quad k \in \mathbb{N}^*$$

puisqu'on obtient $G_X^{(k)}(u) = \sum_{j=k}^{\infty} j(j-1)\dots(j-k+1) p_j u^{j-k}$ et par conséquent $G_X^{(k)}(1) = \mu_{[k]}$.

Exemple 3.10

Pour la loi de Poisson, $G_X(u) = e^{-\lambda} e^{\lambda u}$ donc $G_X^{(k)}(u) = \lambda^k e^{-\lambda} e^{\lambda u}$ et $\mu_{[k]} = \lambda^k$.

La fonction génératrice des moments factoriels est donc définie par :

$$G_X(1+u) = \sum_{k=0}^{\infty} \mu_{[k]} \frac{u^k}{k!}$$

La **fonction génératrice des moments** pour une v.a. discrète X est définie par :

$$H_X(u) = E(e^{uX}) = G_X(e^u) = \sum_{k \in K} p_k e^{u x_k}$$

En développant en série entière e^{uX} , on obtient en effet :

$$H_X(u) = E(e^{uX}) = \sum_{k=0}^{\infty} \frac{u^k}{k!} E(X^k) = \sum_{k=0}^{\infty} m_k \frac{u^k}{k!}$$

d'où on déduit $m_k = E(X^k) = H_X^{(k)}(0)$ pour $k \in \mathbb{N}^*$, avec $m_0 = 1$.

Exemple 3.11

Pour la loi de Bernoulli, $H_X(u) = E(e^{uX}) = G_X(e^u) = pe^u + 1 - p$, donc $H_X^{(k)}(u) = pe^u$ et $m_k = p$ pour tout $k \in \mathbb{N}^*$.

3.2 Fonction génératrice d'une loi absolument continue

La **fonction génératrice des moments**, pour une v.a. X de densité f , est définie par :

$$H_X(u) = E(e^{uX}) = \int_{-\infty}^{+\infty} e^{ux} f(x) dx$$

lorsque cette intégrale généralisée existe.

Exemple 3.12

Si $X \rightsquigarrow \mathcal{E}(\theta)$:

$$H_X(u) = \theta \int_0^{+\infty} e^{(u-\theta)x} dx = \frac{\theta}{\theta - u}, \quad \text{pour } u < \theta.$$

Pour $|u| < \theta$ on obtient donc :

$$H_X(u) = \frac{1}{1 - u/\theta} = \sum_{k=0}^{\infty} \frac{u^k}{\theta^k} = \sum_{k=0}^{\infty} \frac{k!}{\theta^k} \times \frac{u^k}{k!}$$

donc $m_k = \frac{k!}{\theta^k}$ pour tout $k \in \mathbb{N}$.

Si X_1, \dots, X_n sont des v.a. mutuellement indépendantes, alors :

$$\begin{aligned} H_{X_1 + \dots + X_n}(u) &= E(e^{u(X_1 + \dots + X_n)}) = E\left(\prod_{i=1}^n e^{uX_i}\right) \\ &= \prod_{i=1}^n E(e^{uX_i}) = \prod_{i=1}^n H_{X_i}(u) \end{aligned}$$

Exemple 3.13

Dans l'exemple 3.12, on a calculé la fonction génératrice des moments de la loi exponentielle de paramètre θ , qui est aussi la loi $\gamma(1, \theta)$. En additionnant p variables indépendantes de ce type, on obtient la loi $\gamma(p, \theta)$ dont la fonction génératrice est donc :

$$H_X(u) = \left(\frac{1}{1 - u/\theta} \right)^p$$

La fonction génératrice des moments centrés est définie par :

$$M_X(u) = e^{-um_1} H_X(u) = E[e^{u(X-m_1)}]$$

que l'on détermine à partir du développement en série entière :

$$e^{u(X-m_1)} = \sum_{k=0}^{\infty} \frac{u^k}{k!} (X - m_1)^k$$

pour obtenir :

$$M_X(u) = \sum_{k=0}^{\infty} \mu_k \frac{u^k}{k!}$$

qui permet de déduire les moments centrés à partir de cette fonction par :

$$\mu_k = E(X - m_1)^k = M_X^{(k)}(0)$$

Exemple 3.14

Pour la loi exponentielle de paramètre θ on a $m_1 = 1/\theta$ et donc :

$$\begin{aligned} M_X(u) &= e^{-u/\theta} \times \frac{1}{1 - u/\theta} = 1 + \sum_{k=2}^{\infty} \frac{(-u/\theta)^k}{k!} \times \sum_{j=0}^{\infty} \left(\frac{u}{\theta}\right)^j \\ &= 1 + \sum_{k=2}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^k}{k!} \times \left(\frac{u}{\theta}\right)^{j+k} \\ &= 1 + \sum_{n=2}^{\infty} \frac{n!}{\theta^n} \sum_{k=2}^n \frac{(-1)^k}{k!} \times \frac{u^n}{n!} \\ &= 1 + \frac{u^2}{2\theta^2} + \frac{u^3}{3\theta^3} + \sum_{n=4}^{\infty} \left(\sum_{k=2}^n \frac{(-1)^k}{k!} \right) \frac{u^n}{\theta^n} \end{aligned}$$

On obtient ainsi :

$$\mu_2 = \frac{1}{\theta^2}, \mu_3 = \frac{2}{\theta^3} \text{ et } \mu_n = \frac{n!}{\theta^n} \sum_{k=2}^n \frac{(-1)^k}{k!}, \quad n \geq 2$$

Le choix de la loi à retenir pour modéliser un phénomène réel peut paraître délicat compte tenu du grand nombre de lois usuelles. Cependant, on peut savoir sans ambiguïté s'il s'agit d'un phénomène discret ou continu. Dans le cas discret, il faut bien noter les conditions précises de l'expérience conduisant à telle ou telle loi. Dans le cas continu, on peut aussi savoir sans ambiguïté s'il s'agit de valeurs finies ou non, qui peuvent être éventuellement négatives ou non, d'un phénomène qui est de nature symétrique ou non. Enfin, l'expérience permet de savoir que certains types de lois sont particulièrement adaptés à certains phénomènes (exemple loi exponentielle pour les modèles de durée de vie ou loi de Pareto pour les modèles de distribution de revenus dans une population).



Entraînez-vous

- 3.1**
1. Vous effectuez un voyage en avion à bord d'un biréacteur qui peut poursuivre son vol avec un seul réacteur qui fonctionne. Les réacteurs fonctionnent de façon indépendante et ont chacun une probabilité p de tomber en panne au cours du vol. Calculer en fonction de p la probabilité π_B que votre vol ait pu se poursuivre jusqu'à sa destination.
 2. Dans le cas d'un quadriréacteur, qui peut poursuivre son vol avec au moins deux réacteurs qui fonctionnent, calculer en fonction de p la probabilité π_Q que votre vol ait pu se poursuivre jusqu'à sa destination.
 3. Pour quelles valeurs de p le biréacteur est-il plus sûr que le quadriréacteur ? Calculer π_B et π_Q pour $p = \frac{1}{2}$.
- 3.2** Au casino, un joueur décide de miser sur un même numéro (ou série de numéros), jusqu'à ce qu'il gagne. Sa mise initiale est $a > 0$, le numéro qu'il joue a la probabilité p de sortir à chaque partie et il rapporte k fois la mise, $k \in \mathbb{N}^*$. Calculer l'espérance mathématique du gain G de ce joueur qui double sa mise à chaque partie.
- 3.3** Une urne contient une boule blanche et une boule noire.
1. On effectue des tirages avec remise jusqu'à obtention d'une boule blanche. Déterminer la loi de probabilité du nombre N de tirages, puis calculer $E(N)$ et $V(N)$.
 2. Mêmes questions si on remet une boule noire en plus après chaque tirage d'une boule noire. Calculer alors $P(N > n)$, $n \in \mathbb{N}^*$.
- 3.4** Vous avez besoin d'une personne pour vous aider à déménager. Quand vous téléphonez à un ami, il y a une chance sur quatre qu'il accepte. Soit X la variable aléatoire qui représente le nombre d'amis que vous devrez contacter pour obtenir cette aide. Déterminer la loi de probabilité de X puis calculer $P(X \leq 3)$ et $E(X)$.
- 3.5** Lors d'un examen oral, on vous demande de tirer les trois sujets que vous aurez à traiter dans une urne qui en contient dix. Parmi ces dix sujets, il y en a 3 que vous ne connaissez pas. Soit X la variable aléatoire qui représente le nombre de sujets qui vous seront inconnus à l'issue de ce tirage. Calculer les probabilités des différentes valeurs possibles de X et en déduire $E(X)$.
- 3.6** Pour être sélectionné aux Jeux olympiques, un athlète doit réussir deux fois à dépasser les minima fixés par sa fédération. Il a une chance sur trois de réussir à chaque épreuve à laquelle il participe. On note X la variable aléatoire qui représente le nombre d'épreuves auxquelles il devra participer pour être sélectionné.
1. Déterminer la loi de probabilité de X .
 2. Si cet athlète ne peut participer qu'à quatre épreuves maximum, quelle est la probabilité qu'il soit sélectionné ?
- 3.7** Soit X une v.a. de loi binômiale de paramètres $n = 20$ et $p = 0,1$.
1. Calculer les probabilités suivantes : $P(X = 5)$, $P(X \leq 2)$, $P(X < 4)$, $P(X = 1, 5)$, $P(3 \leq X \leq 4)$ et $P(2 < X \leq 8)$.

2. Déterminer les valeurs de x telles que $P(X \geq x) \leq 0,75$.
3. Calculer $P(X = 16)$ dans le cas où $p = 0,9$.
- 3.8** Si X est une v.a. de loi de Poisson de paramètre $\lambda = 5$, calculer les probabilités $P(X = 6)$, $P(X < 4)$, $P(X \geq 5)$ et $P(\pi/2 < X < 2\pi)$ puis déterminer les valeurs de x telles que $P(X < x) \geq 0,95$.
- 3.9** Un commentateur sportif affirmait que le gain du match de double en coupe Davis (événement noté D), était généralement synonyme de victoire. Le pays gagnant est celui qui remporte le plus de matchs, la rencontre comportant 4 matchs en simple et un match en double. On fait l'hypothèse que pour ces 5 matchs chaque pays a la même probabilité de l'emporter. Déterminer la loi de probabilité de la v.a. X qui représente le nombre de matchs gagnés par une équipe. En déduire la probabilité que le pays gagnant ait effectivement remporté le match de double. Calculer alors la probabilité qu'un pays ait remporté le match de double, sachant qu'il a gagné. Que penser de l'affirmation de ce commentateur ?
- 3.10** Si U est une v.a. de loi normale standard, calculer $P(U < -2)$, $P(-1 < U < 0,5)$ et $P(4U \geq -3)$ puis déterminer u_0 et v_0 tels que $P(|U| < u_0) = 0,82$ et $P(U < -v_0) = 0,61$.
- 3.11** Soit X une v.a. de loi normale telle que $P(X < 3) = 0,1587$ et $P(X > 12) = 0,0228$. Calculer $P(1 < X < 10)$.
- 3.12** Si X est une v.a. de loi normale telle que $P(X < 2) = 0,0668$ et $P(X \geq 12) = 0,1587$ calculer la valeur de a telle que $P\{[X - E(X)]^2 < a\} = 0,95$.
- 3.13** Une v.a. X suit une loi uniforme dans l'intervalle $[0,1]$.
Exprimer la probabilité que X appartienne à l'intervalle $[x_1, x_2]$ en fonction des réels x_1 et x_2 tels que $x_1 < x_2$.
- 3.14** Soit X une variable aléatoire dont la densité a pour expression, pour $x > 0$:

$$f(x) = \frac{1}{x\sqrt{2\pi\theta}} \exp -\frac{(\ln x)^2}{2\theta} \text{ avec } \theta > 0$$

Déterminer la loi de probabilité de la variable aléatoire $Y = \ln X$.

- 3.15** Déterminer la f.r., puis calculer l'espérance et la variance, d'une v.a. qui suit une **loi de Laplace** de densité $f(x) = \frac{1}{2}e^{-|x|}$.
- 3.16** Si T est une v.a. positive représentant une durée de vie, on dit qu'elle vérifie la propriété de **non-vieillessement** si pour tout $t > 0$ et $h > 0$:

$$P(T > t + h | T > t) = P(T > h)$$

Montrer que la loi de Pascal et la loi exponentielle de paramètre $\theta > 0$ vérifient cette propriété, c'est-à-dire sont des lois **sans mémoire**.

- 3.17** Si F et f sont respectivement la f.r. et la densité d'une v.a. positive, on désigne par **taux de panne** la fonction h définie pour $x > 0$ par :

$$h(x) = \frac{f(x)}{1 - F(x)}$$

Déterminer cette fonction h pour la loi exponentielle de paramètre $\theta > 0$ et pour la **loi de Weibull** de densité $f(x) = \alpha\theta x^{\alpha-1}\exp(-\theta x^\alpha)$ pour $x > 0$, où α et θ sont deux paramètres positifs.

- 3.18** Calculer l'espérance et la variance d'une loi log-normale de paramètres m et $\sigma > 0$.
- 3.19** Soit X_1, \dots, X_n des v.a. indépendantes de densité f et de f.r. F . Déterminer les f.r. puis les densités des v.a. $m_n = \min \{X_i/1 \leq i \leq n\}$ et $M_n = \max \{X_i/1 \leq i \leq n\}$. Appliquer ce résultat au cas particulier de la loi uniforme sur $[0,1]$, puis calculer dans ce cas $E(m_n)$ et $E(M_n)$.
- 3.20** Déterminer les moments non centrés d'ordre $k, k \in \mathbb{N}^*$, de la loi de Pareto de paramètres $\alpha > 0$ et $x_0 > 0$.



- 3.1** 1. On note P_i l'événement « le réacteur i est tombé en panne au cours du vol », avec $i = 1, 2$. On obtient :

$$\Pi_B = P(\overline{P_1 \cap P_2}) = 1 - P(P_1 \cap P_2) = 1 - p^2$$

2. Le nombre de réacteurs qui tombent en panne au cours du vol est une variable aléatoire X qui suit une loi binômiale de paramètres 4 et p . Ainsi :

$$\begin{aligned} \pi_Q = P(X \leq 2) &= 1 - P(X = 3) - P(X = 4) = 1 - 4p^3(1-p) - p^4 \\ &= 1 - 4p^3 + 3p^4 \end{aligned}$$

3. On obtient :

$$\pi_Q - \pi_B = p^2(1-p)(1-3p)$$

donc le biréacteur est plus sûr que le quadriréacteur pour $p > 1/3$.

On vérifie que $\pi_B\left(\frac{1}{2}\right) = \frac{12}{16} > \pi_Q\left(\frac{1}{2}\right) = \frac{11}{16}$.

- 3.2** Si N est le nombre de parties jouées, l'événement $\{N = n\}$, $n \in \mathbb{N}^*$, signifie que le joueur a perdu les $n - 1$ premières parties et que son numéro est sorti à la dernière ; cette v.a. suit donc une loi de Pascal de paramètre p :

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

À l'issue de ces n parties, le joueur reçoit k fois sa mise, soit $k2^{n-1}a$, après avoir misé au cours de ces parties : $a + 2a + 2^2a + \dots + 2^{n-1}a = a(1 + 2 + 2^2 + \dots + 2^{n-1}) = a(2^n - 1)$. Son gain est alors : $g_n = k2^{n-1}a - (2^n - 1)a = a + (k - 2)2^{n-1}a$.

L'espérance de gain est donc :

$$\begin{aligned} E(G) &= \sum_{n=1}^{\infty} g_n P(N = n) = a + (k-2)ap \sum_{n=1}^{\infty} 2^{n-1}(1-p)^{n-1} \\ &= a + (k-2)ap \sum_{n=1}^{\infty} (2q)^{n-1} \end{aligned}$$

où on a posé $q = 1 - p$. Si $q \geq 1/2$, la série est divergente et cette espérance est infinie. Si $q < 1/2$, c'est-à-dire $p > 1/2$, on obtient :

$$E(G) = a + (k-2)a \frac{p}{1-2q} = \frac{kp-1}{2p-1}a$$

Pour $k = 2$ par exemple, $E(G) = a$, c'est-à-dire que l'espérance de gain du joueur est égale à sa mise initiale.

- 3.3** 1. Les tirages sont effectués jusqu'à ce que l'on obtienne une boule blanche, donc la variable N suit une loi de Pascal de paramètre $p = 1/2$ puisque c'est la probabilité de tirer une boule blanche :

$$P(N = n) = \frac{1}{2^n}$$

D'après les résultats du cours : $E(N) = V(N) = 2$.

2. Si on note respectivement B_i et N_i les événements tirer une boule blanche et tirer une boule noire au i -ème tirage, $i \in \mathbb{N}^*$, la loi de la variable entière N est définie par :

$$P(N = 1) = P(B_1) = \frac{1}{2}$$

$$P(N = 2) = P(N_1 B_2) = \frac{1}{2} \times \frac{1}{3}$$

...

$$P(N = n) = P(N_1 \dots N_{n-1} B_n) = \frac{1}{2} \times \frac{2}{3} \times \dots \times \frac{n-1}{n} \times \frac{1}{n+1} = \frac{1}{n(n+1)}$$

Ainsi :

$$E(N) = \frac{1}{2} + \sum_{n=2}^{\infty} n \frac{1}{n(n+1)} = \sum_{n=2}^{\infty} \frac{1}{n}$$

série harmonique divergente, donc l'espérance est infinie. *A fortiori* la variance n'existe pas non plus.

On obtient :

$$\begin{aligned} P(N > n) &= 1 - P(N \leq n) = 1 - \sum_{k=1}^n P(N = k) \\ &= 1 - \frac{1}{2} - \sum_{k=2}^n \left(\frac{1}{k} - \frac{1}{k+1} \right) = \frac{1}{n+1} \end{aligned}$$

3.4 La v. a. X suit une loi géométrique (de Pascal) ; pour tout entier $k \geq 1$:

$$P(X = k) = \frac{3^{k-1}}{4^k}$$

On obtient ensuite :

$$P(X \leq 3) = \sum_{k=1}^3 P(X = k) = \frac{37}{64}$$

Le paramètre de cette loi est $1/4$ donc $E(X) = 4$.

3.5 La v. a. X suit une loi hypergéométrique ; pour tout entier $0 \leq k \leq 3$:

$$P(X = k) = \frac{\binom{4}{k} \binom{6}{3-k}}{\binom{10}{3}}$$

On obtient ensuite :

$$P(X = 0) = \frac{1}{6} \quad P(X = 1) = \frac{1}{2} \quad P(X = 2) = \frac{3}{10} \quad P(X = 3) = \frac{1}{30}$$

$$E(X) = \frac{1}{2} + \frac{3}{5} + \frac{1}{10} = 1,2$$

En utilisant la formule du cours on retrouve $E(X) = 3 \frac{4}{10} = 1,2$.

3.6 1. Il s'agit de la loi binômiale négative ; pour tout entier $k \geq 2$:

$$P(X = k) = (k - 1) \frac{2^{k-2}}{3^k}$$

2. La probabilité d'être sélectionné est

$$P(X \leq 4) = \sum_{k=2}^4 P(X = k) = \frac{11}{27}$$

3.7 1. Par lecture de la table 3 on obtient : $P(X = 5) = 0,0319$, $P(X \leq 2) = 0,6769$, $P(X < 4) = 0,8670$, $P(X = 1,5) = 0$, $P(3 \leq X \leq 4) = 0,2799$ et $P(2 < X \leq 8) = 0,9999 - 0,6769 = 0,3230$.

2. La condition $P(X < x) \geq 0,25$ équivalente à $P(X \leq x - 1) \geq 0,25$ conduit à $x - 1 \geq 1$ ou $x \geq 2$.

3. Si $X \rightsquigarrow B(n, p)$ on a $P_p(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ et $P_p(X = n - x) = \binom{n}{n-x} p^{n-x} (1 - p)^x = \binom{n}{x} p^x (1 - p)^{n-x}$ où on a poussé $q = 1 - p$. Par conséquent $P_p(X = x) = P_{1-p}(X = n - x)$, soit en appliquant ce résultat : $P_{0,9}(X = 16) = P_{0,1}(X = 4) = 0,0898$.

3.8 Par lecture de la table 4 on obtient : $P(X = 6) = 0,1462$, $P(X < 4) = 0,2650$, $P(X \geq 5) = 0,5595$ et $P(\pi/2 < X < 2\pi) = P(X \leq 6) - P(X \leq 1) = 0,7218$. On lit $P(X < 10) = 0,9682$ donc $x \geq 10$.

3.9 La v.a. X suit une loi binômiale de paramètres 5 et 1/2. Un pays est gagnant si $X \geq 3$; la probabilité demandée est donc :

$$P\{D \cap (X \geq 3)\} = \sum_{k=3}^5 P\{D \cap (X = k)\} = \sum_{k=3}^5 P(X = k) P\{D|X = k\}$$

Si un pays a gagné k matchs, tous les choix parmi les cinq rencontres sont équiprobables, donc :

$$P\{D|X = k\} = \frac{\binom{4}{k-1}}{\binom{5}{k}}$$

Ainsi :

$$P\{D \cap (X \geq 3)\} = \sum_{k=3}^5 \frac{\binom{5}{k}}{2^5} \frac{\binom{4}{k-1}}{\binom{5}{k}} = \frac{11}{2^5}$$

On en déduit :

$$P\{D|G\} = \frac{P\{D \cap (X \geq 3)\}}{P(X \geq 3)} = \frac{11/2^5}{1/2} = \frac{11}{16}$$

L'affirmation de ce commentateur paraît fondée puisque cette probabilité est supérieure à 1/2. Cependant, s'il affirmait que le gain du premier match en coupe Davis est généralement synonyme de victoire on obtiendrait la même probabilité. C'est le fait d'avoir emporté un match (quel qu'il soit) qui renforce évidemment la probabilité de gagner !

- 3.10** Par lecture des tables 1 et 2 on obtient $P(U < -2) = 1 - \Phi(2) = 0,0228$, $P(-1 < U < 0,5) = \Phi(0,5) - [1 - \Phi(1)] = 0,6915 - 0,1587 = 0,5328$, $P(4U \geq -3) = 1 - \Phi(-0,75) = \Phi(0,75) = 0,7734$;
 $P(|U| < u_0) = \Phi(u_0) - \Phi(-u_0) = 2\Phi(u_0) - 1 = 0,82$ d'où $\Phi(u_0) = 0,91$ et $u_0 = 1,3408$; $P(U < -v_0) = \Phi(-v_0) = 0,61$ et $v_0 = -0,2793$.

- 3.11** Nous allons d'abord déterminer les valeurs de $m = E(X)$ et $\sigma = \sqrt{V(X)}$. La première valeur étant inférieure à 0,5 on considère son complément à 1, soit ici $1 - 0,1587 = 0,8413$, valeur dans la table 1 de $\Phi(1)$. De même $1 - 0,0228 = 0,9772 = \Phi(2)$. En centrant et réduisant on a donc :

$$P\left(\frac{X - m}{\sigma} < \frac{3 - m}{\sigma}\right) = 1 - \Phi(1) = \Phi(-1)$$

$$P\left(\frac{X - m}{\sigma} < \frac{12 - m}{\sigma}\right) = \Phi(2)$$

soit :

$$\frac{3 - m}{\sigma} = -1 \text{ et } \frac{12 - m}{\sigma} = 2$$

ce qui conduit à $m = 6$ et $\sigma = 3$, puis :

$$P(1 < X < 10) = P\left(-\frac{5}{3} < \frac{X - 6}{3} < \frac{4}{3}\right)$$

$$= \Phi(1,33) - [1 - \Phi(1,67)] = 0,8607.$$

- 3.12** Dans la table 1 on constate que $1 - 0,0668 = 0,9332 = \Phi(1,5)$ donc $P(X < 2) = \Phi(-1,5)$; de même $1 - 0,1587 = 0,8413 = \Phi(1)$. Donc en centrant sur $E(X) = m$ et réduisant par $\sqrt{V(X)} = \sigma$:

$$P\left(\frac{X - m}{\sigma} < \frac{2 - m}{\sigma}\right) = \Phi(-1,5)$$

$$P\left(\frac{X - m}{\sigma} < \frac{12 - m}{\sigma}\right) = \Phi(1)$$

soit $\frac{2 - m}{\sigma} = -1,5$ et $\frac{12 - m}{\sigma} = 1$ d'où $m = 8$ et $\sigma = 4$. On sait que la v.a.

$\left(\frac{X - m}{\sigma}\right)^2$ suit une loi χ_1^2 donc a/σ^2 est le fractile d'ordre 0,95 de cette loi χ_1^2 , lu dans la table 5, soit $a/\sigma^2 = 3,841$ et $a = 61,44$.

- 3.13** Si f est la densité de cette loi uniforme, cette probabilité se calcule par :

$$p = P\{X \in [x_1, x_2]\} = \int_{x_1}^{x_2} f(t) dt$$

La densité a pour valeur 1 dans l'intervalle $[0,1]$ et 0 en dehors. La valeur de p est donc égale à la longueur de la partie commune des intervalles $[0,1]$ et $[x_1, x_2]$. Elle est indiquée dans le tableau suivant, en fonction de la position de x_1 et x_2 :

$x_2 < 0$	$x_1 < 0 < x_2 < 1$	$x_1 < 0 < 1 < x_2$	$0 < x_1 < x_2 < 1$	$0 < x_1 < 1 < x_2$	$1 < x_1$
0	x_2	1	$x_2 - x_1$	$1 - x_1$	0

3.14 La v.a. Y a pour fonction de répartition :

$$G(y) = P(\ln X < y) = P(X < e^y) = F(e^y)$$

où F est la f.r. de X . La densité obtenue par dérivation est :

$$g(y) = e^y f(e^y) = \frac{1}{\sqrt{2\pi\theta}} \exp -\frac{y^2}{2\theta}$$

qui est la densité de la loi normale centrée de variance θ .

3.15 La f.r. de la loi de Laplace est définie pour $x \leq 0$ par :

$$F(x) = \frac{1}{2} \int_{-\infty}^x e^t dt = \frac{1}{2} e^x$$

et pour $x > 0$ par :

$$F(x) = F(0) + \frac{1}{2} \int_0^x e^{-t} dt = \frac{1}{2} + \frac{1}{2} [-e^{-t}]_0^x = 1 - \frac{1}{2} e^{-x}$$

L'espérance est définie par l'intégrale généralisée :

$$E(X) = \frac{1}{2} \int_{-\infty}^{+\infty} x e^{-|x|} dx$$

qui est convergente d'après le théorème des croissances comparées ; sa valeur est donc nulle puisque la fonction à intégrer sur \mathbb{R} est impaire : la loi est symétrique et $E(X) = 0$.

La variance se calcule en intégrant deux fois par parties :

$$\begin{aligned} V(X) = E(X^2) &= \frac{1}{2} \int_{-\infty}^{+\infty} x^2 e^{-|x|} dx = \int_0^{+\infty} x^2 e^{-x} dx \\ &= [-x^2 e^{-x}]_0^{+\infty} + 2 \int_0^{+\infty} x e^{-x} dx = 2 [-x e^{-x}]_0^{+\infty} + 2 \int_0^{+\infty} e^{-x} dx = 2 \end{aligned}$$

3.16 Pour la loi de Pascal on a pour $t \in \mathbb{N}$:

$$P(T > t) = \sum_{k=t+1}^{\infty} P(T = k) = p \sum_{k=t+1}^{\infty} q^{k-1} = pq^t \sum_{k=0}^{\infty} q^k = \frac{pq^t}{1-q} = q^t$$

Par conséquent, pour $h \in \mathbb{N}$:

$$\begin{aligned} P(T > t+h | T > t) &= \frac{P\{(T > t+h) \cap (T > t)\}}{P(T > t)} = \frac{P(T > t+h)}{P(T > t)} \\ &= q^h = P(T > h) \end{aligned}$$

ce qui établit que c'est une loi sans mémoire.

Pour la loi exponentielle de paramètre θ , on a pour $t > 0$:

$$P(T > t) = \int_t^{+\infty} \theta e^{-\theta x} dx = [-e^{-\theta x}]_t^{+\infty} = e^{-\theta t}$$

d'où pour $h > 0$:

$$P(T > t + h | T > t) = \frac{P(T > t + h)}{P(T > t)} = \frac{e^{-\theta(t+h)}}{e^{-\theta t}} = e^{-\theta h} = P(T > h)$$

ce qui prouve la propriété de non-vieillessement.

3.17 La f.r. de la loi exponentielle vérifie $1 - F(x) = e^{-\theta x}$ pour $x > 0$, donc :

$$h(x) = \frac{\theta e^{-\theta x}}{e^{-\theta x}} = \theta$$

la loi exponentielle est à taux de panne constant, en plus d'être sans mémoire comme nous l'avons vu dans l'exercice précédent.

Pour la loi de Weibull, si $x > 0$:

$$F(x) = \int_0^x \alpha \theta t^{\alpha-1} e^{-\theta t^\alpha} dt = [-\exp(-\theta t^\alpha)]_0^x = 1 - \exp(-\theta x^\alpha)$$

et :

$$h(x) = \alpha \theta x^{\alpha-1}$$

on retrouve bien le résultat obtenu pour la loi exponentielle en faisant $\alpha = 1$.

3.18 On calcule l'espérance de X de loi log-normale de paramètres m et σ en faisant le changement de variable $y = \ln x$:

$$\begin{aligned} E(X) &= \frac{1}{\sigma \sqrt{2\pi}} \int_0^{+\infty} \exp -\frac{1}{2\sigma^2} (\ln x - m)^2 dx \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^y \exp -\frac{1}{2\sigma^2} (y - m)^2 dy \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp -\frac{1}{2\sigma^2} [y^2 - 2(m + \sigma^2)y + m^2] dy \\ &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ \frac{1}{2\sigma^2} [(m + \sigma^2)^2 - m^2] \right\} \int_{-\infty}^{+\infty} \exp -\frac{1}{2\sigma^2} [y - (m + \sigma^2)]^2 dy \\ &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ \frac{1}{2\sigma^2} (2m\sigma^2 + \sigma^4) \right\} \sigma \sqrt{2\pi} = \exp(m + \sigma^2/2) \end{aligned}$$

On obtient de la même façon :

$$E(X^2) = \exp(2m + 2\sigma^2) \text{ et } V(X) = (e^{\sigma^2} - 1) \exp(2m + \sigma^2)$$

3.19 Pour déterminer la f.r. de m_n , on considère l'événement :

$$\{m_n \geq y\} = \bigcap_{i=1}^n \{X_i \geq y\}$$

De même, pour que M_n soit inférieur à z il faut et il suffit que tous les X_i soient inférieurs à z , ce qui s'écrit :

$$\{M_n < z\} = \bigcap_{i=1}^n \{X_i < z\}$$

On en déduit, du fait de l'indépendance des X_i et de l'identité de leur loi :

$$G(y) = P(m_n < y) = 1 - P\left\{\bigcap_{i=1}^n \{X_i \geq y\}\right\} = 1 - [1 - F(y)]^n$$

$$H(z) = P(M_n < z) = \prod_{i=1}^n P(X_i < z) = F^n(z)$$

Les densités correspondantes s'obtiennent par dérivation :

$$g(y) = nf(y)[1 - F(y)]^{n-1}$$

$$h(z) = nf(z)F^{n-1}(z)$$

Dans le cas particulier de la loi uniforme sur $[0,1]$, on obtient :

$$G(y) = 1 - (1 - y)^n \quad \text{et} \quad g(y) = n(1 - y)^{n-1}, \quad 0 \leq y \leq 1$$

$$H(z) = z^n \quad \text{et} \quad h(z) = nz^{n-1}, \quad 0 \leq z \leq 1$$

On calcule alors l'espérance :

$$E(m_n) = n \int_0^1 y(1 - y)^{n-1} dy = n \int_0^1 (1 - u)u^{n-1} du$$

$$= n \int_0^1 u^{n-1} du - n \int_0^1 u^n du = 1 - \frac{n}{n+1} = \frac{1}{n+1}$$

obtenue en faisant le changement de variable $u = 1 - y$. De même :

$$E(M_n) = n \int_0^1 z^n dz = \frac{n}{n+1}$$

Remarquons ici que :

$$G(y) = P(m_n < y) = 1 - H(1 - y) = P(M_n > 1 - y) = P(1 - M_n < y)$$

ce qui montre que les v.a. m_n et $1 - M_n$ ont la même loi de probabilité et donc que $E(m_n) = 1 - E(M_n)$.

3.20 Le moment non centré d'ordre k de la loi de Pareto de paramètres α et x_0 est :

$$E(X^k) = \alpha x_0^\alpha \int_{x_0}^{+\infty} x^{k-\alpha-1} dx = \alpha x_0^\alpha \left[\frac{x^{k-\alpha}}{k-\alpha} \right]_{x_0}^{+\infty} = \frac{\alpha}{\alpha - k} x_0^k$$

à condition que $k < \alpha$. Seuls existent les moments d'ordre inférieur à α . Par exemple, pour $\alpha = 2$, la loi admet une espérance mais pas de variance. Pour $\alpha > 2$, on obtient :

$$V(X) = \frac{\alpha}{(\alpha - 2)(\alpha - 1)^2} x_0^2$$

Couple et vecteur aléatoires

Introduction

Comme nous avons associé un nombre à une expérience aléatoire, dans certains cas nous pouvons être amenés à en associer plusieurs. Par exemple, le jet de deux dés distincts ne peut pas être codé avec une seule valeur numérique. De même, à un individu d'une population donnée, on peut associer son revenu et sa consommation. On est alors amené à associer à de telles épreuves aléatoires deux, voire plusieurs valeurs numériques, au moyen donc de plusieurs applications qui seront des v.a. pouvant être regroupées dans un vecteur, ce qui conduit à la généralisation en multidimensionnel de la notion de variable aléatoire réelle : un **vecteur aléatoire**. Pour la commodité de l'exposé nous commencerons par étudier le cas bidimensionnel, celui d'un couple aléatoire, en distinguant toujours les cas discret et continu. Nous verrons ensuite comment se généralisent les moments associés à un vecteur aléatoire et présenterons deux lois particulières, la loi multinomiale et la loi normale vectorielle.

Objectifs

Généraliser les notions de variable aléatoire, d'espérance et de variance au cas multidimensionnel.

Définir les lois conditionnelles, la fonction de régression et la convolution de deux lois.

Concepts clés

Loi marginale et loi conditionnelle
 Régression
 Covariance
 Indépendance
 Matrice de variances/covariances

1 Couple de v.a. discrètes

1.1 Loi d'un couple

Un couple de v.a. discrètes est constitué de deux v.a. discrètes X et Y dont l'ensemble des valeurs possibles peut s'écrire respectivement sous la forme $\{x_i\}_{i \in I}$ et $\{y_j\}_{j \in J}$, où I et J sont des ensembles d'indices inclus dans \mathbb{N} , pouvant d'ailleurs être \mathbb{N} tout entier. On convient de ne faire figurer que des valeurs de probabilité strictement positive. Comme dans le cas unidimensionnel, la loi d'un couple discret est définie par l'ensemble des

valeurs possibles, soit ici $\{(x_i, y_j) ; (i, j) \in I \times J\}$, et par les probabilités associées :

$$p_{ij} = P_{(X,Y)} (X = x_i, Y = y_j)$$

1.2 Lois marginales

À la loi d'un couple sont associées deux lois marginales qui sont les lois de chacun des éléments du couple pris séparément, définies par l'ensemble des valeurs possibles et les probabilités associées obtenues par sommation, soit :

$$P_X (X = x_i) = \sum_{j \in J} P_{(X,Y)} (X = x_i, Y = y_j) = \sum_{j \in J} p_{ij} = p_{i.}$$

$$P_Y (Y = y_j) = \sum_{i \in I} P_{(X,Y)} (X = x_i, Y = y_j) = \sum_{i \in I} p_{ij} = p_{.j}$$

Si la loi du couple est présentée dans un tableau, ces lois sont obtenues dans les marges, par sommation de ligne ou de colonne.

Y \ X	x_i	
y_j	$\cdots \cdots \cdots p_{ij} \rightarrow$	$p_{.j}$
	\downarrow	\downarrow
	$p_{i.} \rightarrow$	1

1.3 Lois conditionnelles

On peut également associer deux lois conditionnelles à la loi d'un couple, c'est-à-dire la loi d'une variable, l'autre ayant une valeur fixée (loi dans une ligne ou dans une colonne donnée). Par exemple, pour $Y = y_j$ fixé, la loi conditionnelle de X est définie par l'ensemble des valeurs possibles et les probabilités associées :

$$P (X = x_i | Y = y_j) = \frac{P (X = x_i, Y = y_j)}{P (Y = y_j)} = \frac{p_{ij}}{p_{.j}} = p_i^j$$

on vérifie que c'est bien une loi de probabilité sur $\Omega_X = \{x_i ; i \in I\}$:

$$\sum_{i \in I} p_i^j = \frac{1}{p_{.j}} \sum_{i \in I} p_{ij} = 1$$

Exemple 4.1

La loi d'un couple (X, Y) est donnée par le tableau suivant :

Y \ X	-2	0	2	
-1	0,1	0,2	0,1	0,4
2	0,2	0,2	0,2	0,6
	0,3	0,4	0,3	1

La loi conditionnelle de X pour $Y = -1$ figure dans le tableau ci-après :

$X Y = -1$	-2	0	2	
	$\frac{0.1}{0.4}$	$\frac{0.2}{0.4}$	$\frac{0.1}{0.4}$	1

Rappelons que les deux v. a. X et Y sont indépendantes si pour tout $i \in I$ et tout $j \in J$:

$$P(X = x_i, Y = y_j) = P(X = x_i) P(Y = y_j)$$

Dans ce cas, bien entendu, les lois conditionnelles sont confondues avec les lois marginales ; par exemple :

$$P(X = x_i | Y = y_j) = p_i^j = \frac{p_i \cdot p_j}{p_j} = p_i.$$

C'est l'un des seuls cas où la donnée des lois marginales permet de reconstituer la loi du couple.

1.4 Moments conditionnels

Aux lois conditionnelles sont associés des moments conditionnels, comme par exemple l'espérance conditionnelle de Y pour $X = x_i$ fixé, qui est l'espérance de la loi définie par les couples $\left\{ (y_j, p_j^i) ; j \in J \right\}$, soit :

$$E(Y|X = x_i) = \sum_{j \in J} y_j P(Y = y_j | X = x_i) = \sum_{j \in J} y_j p_j^i$$

Le graphe de cette espérance conditionnelle en fonction de x_i s'appelle courbe de régression (non linéaire) de Y en X .

Exemple 4.2

Dans l'exemple 4.1 précédent, la loi conditionnelle de Y pour $X = 2$ est donnée par le tableau suivant :

$Y X = 2$	-1	2	
	$\frac{0.1}{0.3}$	$\frac{0.2}{0.3}$	1

On peut calculer, à partir de ce tableau, l'espérance conditionnelle :

$$E(Y|X = 2) = (-1) \times \frac{1}{3} + 2 \times \frac{2}{3} = 1$$

Notons que $E(Y|X)$ est une fonction de X , donc une variable aléatoire discrète dont la loi de probabilité est définie par l'ensemble des valeurs possibles, soit ici $\{E(Y|X = x_i); i \in I\}$, et les probabilités associées $p_i = P(X = x_i)$. On peut donc calculer la valeur moyenne de cette v.a., soit :

$$\begin{aligned} E[E(Y|X)] &= \sum_{i \in I} p_i \cdot E(Y|X = x_i) \\ &= \sum_{i \in I} p_i \cdot \sum_{j \in J} y_j P(Y = y_j | X = x_i) \\ &= \sum_{i \in I} p_i \cdot \sum_{j \in J} y_j p_{ij} = \sum_{i \in I} \sum_{j \in J} y_j p_{ij} \frac{p_{ij}}{p_i} = \sum_{j \in J} y_j \sum_{i \in I} p_{ij} \\ &= \sum_{j \in J} p_{.j} y_j \\ &= E(Y) \end{aligned}$$

On peut également calculer la variance conditionnelle :

$$\begin{aligned} V(Y|X = x_i) &= E\left\{[Y - E(Y|X = x_i)]^2 | X = x_i\right\} \\ &= E(Y^2 | X = x_i) - E^2(Y|X = x_i) \\ &= \sum_{j \in J} p_{ij}^i [y_j - E(Y|X = x_i)]^2 \end{aligned}$$

On peut établir que la variance totale de Y peut se décomposer à l'aide de ces deux premiers moments conditionnels :

$$V(Y) = E[V(Y|X)] + V[E(Y|X)]$$

1.5 Moments associés à un couple

Si $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une application continue, elle définit une variable aléatoire réelle dont on peut calculer les moments, comme par exemple l'espérance :

$$E[h(X, Y)] = \sum_{i \in I} \sum_{j \in J} p_{ij} h(x_i, y_j)$$

Dans le cas particulier où $h(X, Y) = [X - E(X)][Y - E(Y)]$ on définit ainsi la **covariance** de X et Y :

$$Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

Nous avons vu dans le chap. 2, 1.4, 2 que si les v.a. X et Y sont indépendantes, alors $E(XY) = E(X)E(Y)$ et par conséquent $Cov(X, Y) = 0$. Cependant, il faut faire attention au fait que la réciproque est en général fautive, c'est-à-dire que si la covariance de deux v.a. est nulle, elles ne sont pas forcément indépendantes.

Exemple 4.3

Considérons le couple (X, Y) dont la loi est définie par le tableau ci-après :

$Y \backslash X$	-1	0	1
-1	1/8	1/8	1/8
0	1/16	1/8	1/16
1	1/8	1/8	1/8

Les lois de X et Y sont symétriques par rapport à 0, donc $E(X) = E(Y) = 0$ et :

$$\text{Cov}(X, Y) = E(XY) = 1 \times \frac{1}{8} + (-1) \times \frac{1}{8} + (-1) \times \frac{1}{8} + 1 \times \frac{1}{8} = 0$$

et cependant ces deux variables ne sont pas indépendantes puisque par exemple :

$$P(X = -1, Y = -1) = 1/8 \neq P(X = -1)P(Y = -1) = 5/16 \times 3/8.$$

On appelle *coefficient de corrélation linéaire* de deux v.a. X et Y le nombre réel :

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

c'est un nombre tel que $-1 \leq \rho \leq 1$, avec :

$$|\rho| = 1 \Leftrightarrow \exists a \in \mathbb{R}^*, \exists b \in \mathbb{R} : Y = aX + b$$

c'est-à-dire qu'un coefficient de corrélation de valeur absolue égale à un est l'indication d'une liaison affine entre les deux variables.

1.6 Loi d'une somme

Si X et Y sont deux v.a. discrètes de lois respectives $\{(x_i, p_i); i \in I\}$ et $\{(y_j, q_j); j \in J\}$, la v.a. $Z = X + Y$ est aussi une v.a. discrète dont la loi de probabilité est définie par l'ensemble des valeurs possibles, soit ici $\{(x_i + y_j); i \in I, j \in J\}$, et les probabilités associées :

$$P(Z = z_k) = \sum \{P(X = x_i, Y = y_j) / x_i + y_j = z_k\}$$

Dans le cas général cette loi ne peut donc être définie que par la donnée de la loi du couple (X, Y) .

Cas particulier X et Y sont indépendantes.

On parle alors de convolution des lois de X et Y , qui est définie par :

$$\begin{aligned} P(Z = z_k) &= \sum_{i \in I} P(X = x_i) P(Y = z_k - x_i) \\ &= \sum_{j \in J} P(Y = y_j) P(X = z_k - y_j) \end{aligned}$$

En général, beaucoup des termes des sommations précédentes sont nuls car il faut tenir compte des restrictions $z_k - x_i \in \Omega_Y = \{y_j; j \in J\}$ et $z_k - y_j \in \Omega_X = \{x_i; i \in I\}$.

Exemple 4.4

Convolution de deux lois binômiales.

Soit X et Y deux v.a. indépendantes, de lois respectives $B(n_1, p)$ et $B(n_2, p)$. La v.a. $S = X + Y$ est à valeurs dans $\Omega_S = \{0, 1, \dots, n_1 + n_2\}$ avec pour $s \in \Omega_S$:

$$\{S = s\} = \bigcup_{x=0}^s \{(X = x) \cap (Y = s - x)\}$$

Ainsi :

$$\begin{aligned} P(S = s) &= \sum_{x=0}^s P(X = x) P(Y = s - x) \\ &= \sum_{x=0}^s \binom{n_1}{x} p^x q^{n_1-x} \binom{n_2}{s-x} p^{s-x} q^{n_2-s+x} \\ &= p^s q^{n_1+n_2-s} \sum_{x=0}^s \binom{n_1}{x} \binom{n_2}{s-x} = \binom{n_1 + n_2}{s} p^s q^{n_1+n_2-s} \end{aligned}$$

ce qui indique que $S = X + Y$ suit une loi binômiale $B(n_1 + n_2, p)$. Le résultat :

$$\sum_{x=0}^s \binom{n_1}{x} \binom{n_2}{s-x} = \binom{n_1 + n_2}{s}$$

exprime que le nombre de sous-ensembles de s éléments extraits d'un ensemble comportant n_1 éléments d'une sorte et n_2 éléments d'une autre est obtenu à partir des sous-ensembles associés de x éléments de la première sorte et de $s - x$ éléments de l'autre, avec x variant de 0 à s (voir figure 4.1).

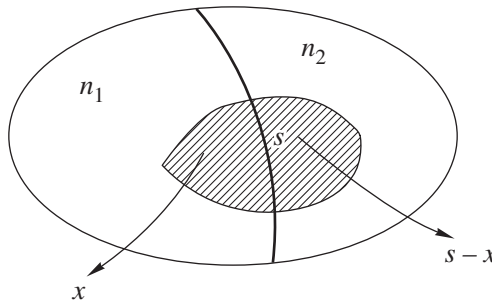


Figure 4.1

Nous allons établir maintenant que la loi d'une variable conditionnellement à une valeur fixée de la somme est une loi hypergéométrique. La loi de X , conditionnellement à $S = s$ par exemple, est définie pour $x \in \{0, 1, \dots, n_1\}$ par :

$$\begin{aligned} P(X = x | S = s) &= \frac{P(X = x) P(S = s | X = x)}{P(S = s)} \\ &= \frac{P(X = x) P(Y = s - x)}{P(S = s)} \\ &= \frac{\binom{n_1}{x} p^x q^{n_1-x} \binom{n_2}{s-x} p^{s-x} q^{n_2-s+x}}{\binom{n_1+n_2}{s} p^s q^{n_1+n_2-s}} \\ &= \frac{\binom{n_1}{x} \binom{n_2}{s-x}}{\binom{n_1+n_2}{s}} \end{aligned}$$

Exemple 4.5

Convolution de lois de Poisson

Si X et Y sont deux v.a. indépendantes de lois de Poisson respectives $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$, alors $Z = X + Y$ est une v.a. à valeurs dans \mathbb{N} , avec pour tout entier k :

$$\begin{aligned} P(Z = k) &= \sum_{x+y=k} P(X = x) P(Y = y) = \sum_{x \in \mathbb{N}} P(X = x) P(Y = k - x) \\ &= \sum_{x=0}^k P(X = x) P(Y = k - x) = \sum_{x=0}^k e^{-\lambda} \frac{\lambda^x}{x!} e^{-\mu} \frac{\mu^{k-x}}{(k-x)!} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{x=0}^k \frac{k!}{x!(k-x)!} \lambda^x \mu^{k-x} = \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k \end{aligned}$$

On retrouve bien comme résultat :

$$X + Y \rightsquigarrow \mathcal{P}(\lambda + \mu).$$

2 Couple de v.a. continues

2.1 Loi du couple

Si X et Y sont deux v.a. réelles continues, la loi de probabilité du couple (X, Y) est déterminée par sa fonction de répartition F , définie sur \mathbb{R}^2 par :

$$F(x, y) = P(X < x, Y < y)$$

La valeur $F(x, y)$ représente la probabilité de la zone hachurée indiquée dans la figure 4.2.

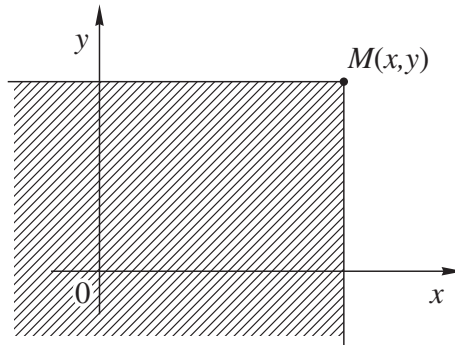


Figure 4.2

L'application F est croissante au sens large par rapport à chacune des deux variables et telle que $0 \leq F(x, y) \leq 1$, avec pour valeurs limites $F(-\infty, y) = F(x, -\infty) = 0$ pour tout x réel et pour tout y réel, et $F(+\infty, +\infty) = 1$.

Si F est deux fois dérivable par rapport aux deux variables, alors la loi de (X, Y) est dite absolument continue, de densité f définie par :

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

La fonction de répartition se calcule alors par intégration :

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, du \, dv$$

Exemple 4.6

La loi de (X, Y) est définie par la densité :

$$f(x, y) = \begin{cases} e^{-y} & \text{si } 0 \leq x \leq y \\ 0 & \text{sinon} \end{cases}$$

Les valeurs de la f.r. F vont être déterminées selon la zone d'appartenance du point (x_0, y_0) , en intégrant la densité f sur l'intersection du domaine où elle est non nulle avec la partie du plan à gauche et au-dessous du point (voir figure 4.3). On obtient de façon évidente :

$$F(x_0, y_0) = 0 \quad \text{si } x_0 \leq 0 \quad \text{ou} \quad y_0 \leq 0$$

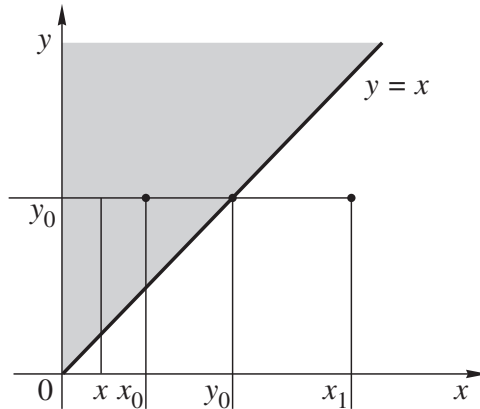


Figure 4.3

Si maintenant le point (x_0, y_0) est dans la zone où la densité f est strictement positive, soit $0 \leq x_0 \leq y_0$:

$$\begin{aligned} F(x_0, y_0) &= \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x, y) dx dy = \int_{x=0}^{x=x_0} \left(\int_{y=x}^{y=y_0} e^{-y} dy \right) dx \\ &= \int_0^{x_0} [-e^{-y}]_x^{y_0} dx = \int_0^{x_0} (e^{-x} - e^{-y_0}) dx = [-e^{-x}]_0^{x_0} - x_0 e^{-y_0} \\ &= 1 - e^{-x_0} - x_0 e^{-y_0} \end{aligned}$$

Enfin, si $0 \leq y_0 \leq x_1$, on peut voir sur la figure 4.3 que :

$$F(x_1, y_0) = F(y_0, y_0) = 1 - e^{-y_0} - y_0 e^{-y_0}$$

En résumé :

$$F(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \quad \text{ou} \quad y \leq 0 \\ 1 - e^{-x} - x e^{-y} & \text{si } 0 \leq x \leq y \\ 1 - e^{-y} - y e^{-y} & \text{si } 0 \leq y \leq x \end{cases}$$

Comme dans cet exemple, la densité se présente souvent sous la forme suivante :

$$f(x, y) = \begin{cases} \dots & \text{si } (x, y) \in D \\ 0 & \text{si } (x, y) \notin D \end{cases}$$

c'est-à-dire que la densité est non nulle seulement à l'intérieur d'un certain domaine D . L'expression de la f.r. du couple (X, Y) se détermine alors en déplaçant le point $M(x, y)$ du bas à gauche vers le haut à droite, conformément à la figure 4.4 ci-après.

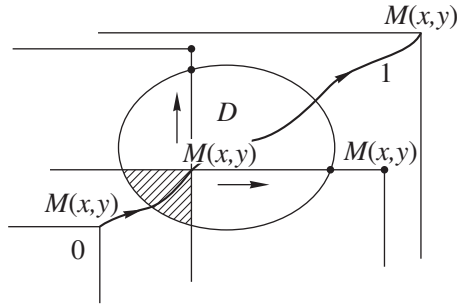


Figure 4.4

Lorsque l'on quitte le domaine sur l'horizontale de M (même ordonnée) ou sur la verticale de M (même abscisse), la valeur de F en ces points extérieurs à D est la même qu'au point d'intersection de l'horizontale ou de la verticale avec la frontière de D .

2.2 Lois marginales

Les fonctions de répartition marginales de X et Y sont définies à partir de la f.r. du couple par :

$$F_X(x) = P(X < x) = F(x, +\infty)$$

$$F_Y(y) = P(Y < y) = F(+\infty, y)$$

c'est-à-dire en faisant tendre y , respectivement x , vers plus l'infini. Dans le cas d'une loi absolument continue, les densités marginales sont alors obtenues par dérivation de ces f.r. marginales.

Cependant, si la loi du couple est définie par sa densité, les densités marginales sont obtenues plutôt par intégration :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Exemple 4.7

Si nous reprenons la loi de l'exemple 4.6 précédent, en faisant tendre y vers plus l'infini dans l'expression de F , on obtient :

$$F_X(x) = F(x, +\infty) = 1 - e^{-x} \text{ pour } x > 0$$

et par dérivation : $f_X(x) = e^{-x}$ pour $x > 0$, densité de la loi exponentielle ou γ (1). Cette densité peut également être obtenue par intégration de la densité du couple, pour $x > 0$:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_x^{+\infty} e^{-y} dy = [-e^{-y}]_x^{+\infty} = e^{-x}$$

La f.r. de Y s'obtient à partir de F en faisant tendre x vers plus l'infini :

$$F_Y(y) = F(+\infty, y) = 1 - e^{-y} - ye^{-y} \text{ pour } y > 0$$

d'où par dérivation, pour $y > 0$: $f_Y(y) = ye^{-y}$ qui est la densité de la loi γ (2).

Ou par intégration de f , pour $y > 0$:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \int_0^y e^{-y} dx = ye^{-y}$$

2.3 Lois conditionnelles

Si l'une des deux variables X ou Y a une valeur fixée, on peut définir la loi conditionnelle de l'autre variable. Pour des lois absolument continues, les lois conditionnelles sont définies par les densités conditionnelles :

$$f_X(x|Y=y) = \frac{f(x, y)}{f_Y(y)}$$

$$f_Y(y|X=x) = \frac{f(x, y)}{f_X(x)}$$

à condition bien sûr que $f_Y(y) > 0$ et $f_X(x) > 0$.

Exemple 4.8

Si nous reprenons la loi précédente, nous obtenons pour $0 \leq x \leq y$:

$$f_X(x|Y=y) = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y}$$

donc la loi conditionnelle $\mathcal{L}(X|Y=y)$ est la loi uniforme sur $[0, y]$, pour $y > 0$ fixé.

De même, pour $y \geq x \geq 0$:

$$f_Y(y|X=x) = \frac{e^{-y}}{e^{-x}} = e^{x-y}$$

qui est la loi exponentielle sur $[x, +\infty[$, pour $x > 0$ fixé.

L'indépendance des v. a. X et Y se définit alors par :

$$f(x, y) = f_X(x) f_Y(y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}$$

On a bien entendu dans ce cas :

$$f_X(x|Y = y) = f_X(x) \quad \text{et} \quad f_Y(y|X = x) = f_Y(y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}$$

2.4 Moments associés à un couple

Si $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une application continue, l'espérance de $h(X, Y)$ se calcule pour une loi de densité f par l'intégrale :

$$E[h(X, Y)] = \int \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy$$

Dans le cas particulier où $h(X, Y) = [X - E(X)][Y - E(Y)]$, ceci définit la **covariance** :

$$Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

Dans le cas particulier où les v.a. X et Y sont indépendantes :

$$\begin{aligned} E(XY) &= \int \int_{\mathbb{R}^2} xyf(x, y) dx dy \\ &= \int_{\mathbb{R}} xf_X(x) dx \int_{\mathbb{R}} yf_Y(y) dy = E(X)E(Y) \end{aligned}$$

et par conséquent :

$$Cov(X, Y) = 0$$

Il faut faire attention à la réciproque, généralement fautive, c'est-à-dire que si deux v.a. ont une covariance nulle elles ne sont pas forcément indépendantes, sauf dans le cas particulier où (X, Y) est un couple normal (cf. Propriété 4.2).

Exemple 4.9

Soit X une v.a. de loi $N(0, 1)$ et définissons $Y = X^2$. On a :

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X^3) = 0$$

car tous les moments d'ordre impair sont nuls ; donc les v.a. X et X^2 ont une covariance nulle alors qu'elles sont dépendantes (mais non linéairement) puisque la seconde est fonction de la première.

On peut définir également le coefficient de corrélation linéaire par :

$$\rho = corr(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

ayant posé $V(X) = \sigma^2(X)$ et $V(Y) = \sigma^2(Y)$.

Nous allons établir que ce coefficient est compris entre -1 et $+1$, en raison de l'inégalité de Schwarz :

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

que l'on obtient en considérant le polynôme en λ , toujours positif :

$$E(X - \lambda Y)^2 = \lambda^2 E(Y^2) - 2\lambda E(XY) + E(X^2) \geq 0$$

ce qui implique $E^2(XY) - E(X^2) E(Y^2) \leq 0$, soit en appliquant cette inégalité aux variables centrées :

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

et par conséquent $|\rho| \leq 1$. Le cas $|\rho| = 1$ correspond à l'existence d'une relation affine entre les deux variables :

$$|\rho| = 1 \Leftrightarrow \exists a \in \mathbb{R}^*, \exists b \in \mathbb{R} : Y = aX + b$$

Cette relation affine, quand elle existe, s'écrit d'ailleurs précisément :

$$Y = E(Y) + \frac{Cov(X, Y)}{V(X)} [X - E(X)]$$

2.5 Régression

Les densités conditionnelles permettent de calculer les moments conditionnels, comme par exemple les espérances ; on peut définir notamment la fonction :

$$x \mapsto E(Y|X = x)$$

qui s'appelle **fonction de régression** (non linéaire), son graphe étant la courbe de **régression** de Y en X .

Pour une loi absolument continue, on obtient :

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y f_Y(y|X = x) dy = \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)} dy$$

Exemple 4.10

En reprenant toujours la même loi des exemples précédents, on obtient ici :

$$\begin{aligned} E(Y|X = x) &= \int_x^{+\infty} y e^{x-y} dy = e^x \int_x^{+\infty} y d(-e^{-y}) \\ &= x + e^x \int_x^{+\infty} e^{-y} dy = x + 1 \end{aligned}$$

Notons que $E(Y|X)$ est une v.a., comme fonction de la v.a. X , et que ses réalisations sont les valeurs $E(Y|X = x)$ pour tout événement élémentaire ω tel que $X(\omega) = x$. On peut donc calculer l'espérance mathématique de cette v.a. :

$$\begin{aligned}
 E[E(Y|X)] &= \int_{-\infty}^{+\infty} E(Y|X=x) f_X(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x,y) dx dy \\
 &= \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f(x,y) dx \right) dy = \int_{-\infty}^{+\infty} y f_Y(y) dy = E(Y)
 \end{aligned}$$

Ainsi, en prenant l'espérance de $E(Y|X)$ par rapport à la loi de X on retrouve $E(Y)$.

Exemple 4.11

Dans l'exemple précédent on a obtenu la v.a. $E(Y|X) = X + 1$ et $E[E(Y|X)] = E(X) + 1 = 2$ qui est bien l'espérance de Y qui suit une loi $\gamma(2)$.

On peut également calculer la variance conditionnelle :

$$\begin{aligned}
 V(Y|X=x) &= E\{[Y - E(Y|X=x)]^2 | X=x\} \\
 &= E(Y^2 | X=x) - E^2(Y|X=x)
 \end{aligned}$$

De même $V(Y|X)$ est une v.a., étant fonction de X , dont on peut calculer l'espérance. On démontre que la variance de Y se décompose à partir des deux premiers moments conditionnels de la façon suivante :

$$V(Y) = E[V(Y|X)] + V[E(Y|X)]$$

2.6 Loi d'une somme

La loi de la v.a. $Z = X + Y$ se détermine par sa f.r G , définie par :

$$G(z) = P(Z < z) = P(X + Y < z)$$

qui ne peut se calculer que si l'on connaît la loi du couple (X, Y) .

Dans le cas particulier où ce couple admet une densité f , on obtient :

$$G(z) = \int \int_D f(x,y) dx dy$$

où le domaine d'intégration est $D = \{(x,y) / x + y < z\}$. On effectue le changement de variable suivant :

$$\begin{cases} x = x \\ s = x + y \end{cases} \quad \begin{cases} x = x \\ y = s - x \end{cases}$$

Le jacobien de cette transformation est :

$$\frac{D(x,y)}{D(x,s)} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

d'où l'intégrale :

$$G(z) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x,y) dy = \int_{-\infty}^{+\infty} dx \int_{-\infty}^z f(x,s-x) ds$$

qui peut s'écrire sous la forme :

$$G(z) = \int_{-\infty}^z g(s) ds$$

ce qui permet de mettre en évidence la densité de Z , soit :

$$g(s) = \int_{-\infty}^{+\infty} f(x, s-x) dx$$

Dans le cas particulier où X et Y sont indépendantes, la loi de $Z = X + Y$ se détermine à partir des densités marginales f_X et f_Y :

$$G(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^z f_X(x) f_Y(s-x) ds dx = \int_{-\infty}^z g(s) ds$$

où la densité g de Z est définie par :

$$g(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy$$

et g s'appelle alors le **produit de convolution** de f_X et f_Y .

Exemple 4.12

Soit X et Y deux v.a. indépendantes et de même loi de Laplace, ou double exponentielle, de densité $\frac{1}{2}e^{-|x|}$. On a vu que la densité du couple $(X, Z = X + Y)$ était $f_X(x)f_Y(z-x)$ et la densité marginale de Z s'obtient donc par intégration :

$$\begin{aligned} g(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} \frac{1}{2}e^{-|x|} \frac{1}{2}e^{-|z-x|} dx \\ &= \frac{1}{4} \int_{-\infty}^z e^{-z+x-|x|} dx + \frac{1}{4} \int_z^{+\infty} e^{z-x-|x|} dx \end{aligned}$$

Pour pouvoir retirer la valeur absolue, il faut connaître le signe sur l'intervalle d'intégration et pour cela distinguer deux cas :

– si $z < 0$:

$$\begin{aligned} 4g(z) &= \int_{-\infty}^z e^{2x-z} dx + \int_z^0 e^z dx + \int_0^{+\infty} e^{z-2x} dx \\ &= e^{-z} \left[\frac{1}{2} e^{2x} \right]_{-\infty}^z - z e^z + e^z \left[-\frac{1}{2} e^{-2x} \right]_0^{+\infty} \\ &= \frac{1}{2} e^z - z e^z + \frac{1}{2} e^z = (1-z) e^z \end{aligned}$$

—si $z > 0$:

$$\begin{aligned}
 4g(z) &= \int_{-\infty}^0 e^{2x-z} dx + \int_0^z e^{-z} dx + \int_z^{+\infty} e^{z-2x} dx \\
 &= e^{-z} \left[\frac{1}{2} e^{2x} \right]_{-\infty}^0 + ze^{-z} + e^z \left[-\frac{1}{2} e^{-2x} \right]_z^{+\infty} \\
 &= \frac{1}{2} e^{-z} + ze^{-z} + \frac{1}{2} e^{-z} = (1+z) e^{-z}
 \end{aligned}$$

La densité de $Z = X + Y$ pouvant s'écrire dans tous les cas :

$$g(z) = \frac{1}{4} (1 + |z|) e^{-|z|}$$

3 Vecteur aléatoire

Nous allons étendre maintenant les notions précédentes au cas d'un vecteur aléatoire ayant n composantes, avec $n \geq 2$. Un vecteur aléatoire X de \mathbb{R}^n est une application de Ω dans \mathbb{R}^n qui s'écrit sous la forme d'un vecteur colonne :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

dont toutes les composantes $X_i, 1 \leq i \leq n$, sont des v.a., c'est-à-dire des applications de Ω dans \mathbb{R} . On définit alors l'espérance de ce vecteur aléatoire comme le vecteur dont les n composantes sont les espérances des v.a. réelles composantes de ce vecteur et on conserve la même notation qu'en unidimensionnel :

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

Dans le cas d'un couple aléatoire, nous avons introduit la covariance comme moment centré d'ordre deux ; s'agissant ici d'un vecteur à n composantes aléatoires, il faut considérer toutes les associations deux à deux de ses composantes. La généralisation de la variance définie en unidimensionnel, indicateur numérique de dispersion autour du centre, sera ici une matrice, notée encore $V(X)$, et qui contiendra toutes les variances des composantes de X , ainsi que les covariances de tous les couples associant deux composantes. Cette matrice est aussi un indicateur de dispersion autour du centre $E(X)$ qui renseigne sur la forme de l'ellipsoïde d'inertie ; c'est une matrice carrée symétrique d'ordre n , appelée **matrice de variances-covariances**, et définie par :

$$V(X) = E\{[X - E(X)]^t [X - E(X)]\}$$

c'est-à-dire comme l'espérance d'une matrice M dont les éléments sont des v.a. M_{ij} et qui est définie comme la matrice d'éléments $E(M_{ij})$. L'élément ligne i , colonne j , de cette matrice, $1 \leq i, j \leq n$, est d'ailleurs :

$$v_{ij} = E\{[X_i - E(X_i)][X_j - E(X_j)]\} = Cov(X_i, X_j)$$

et par conséquent on trouve les variances sur la diagonale principale de cette matrice :

$$v_{ii} = E\{[X_i - E(X_i)]^2\} = V(X_i)$$

Comme propriétés de ces deux moments généralisés, nous allons examiner les effets d'un changement d'échelles, traduit par une application linéaire de \mathbb{R}^n dans \mathbb{R}^m de représentation matricielle A de format (m, n) , puis d'un changement d'origine, traduit par une translation de vecteur b de \mathbb{R}^m . Le vecteur X devient après cette transformation affine le vecteur $Y = AX + b$. L'opérateur espérance mathématique est toujours linéaire en multidimensionnel et le vecteur espérance va donc subir la même transformation :

$$E(Y) = E(AX + b) = AE(X) + b$$

alors que la matrice de variances-covariances n'est pas modifiée par le changement d'origine, étant un moment centré :

$$V(Y) = V(AX + b) = V(AX) = AV(X)'A$$

En effet :

$$Y - E(Y) = AX + b - [AE(X) + b] = A[X - E(X)]$$

et par conséquent :

$$[Y - E(Y)]'[Y - E(Y)] = A[X - E(X)]'[X - E(X)]'A$$

On peut opérer une transformation particulière permettant d'obtenir le vecteur Y centré et réduit. La matrice $\Sigma = V(X)$ étant définie-positive admet une inverse Σ^{-1} définie-positive et il existe une matrice symétrique S telle que $S^2 = \Sigma^{-1}$, qui sera notée $S = \Sigma^{-1/2}$. On effectue alors la transformation :

$$Y = \Sigma^{-1/2}[X - E(X)]$$

où le vecteur Y est centré puisque $E(Y) = 0$ et réduit, au sens où ses composantes sont non corrélées et de variance un :

$$V(Y) = \Sigma^{-1/2}\Sigma\Sigma^{-1/2} = S(\Sigma S) = SS^{-1} = I_n$$

car $\Sigma\Sigma^{-1} = \Sigma S^2 = I_n$ d'où $\Sigma S = S^{-1}$.

4 Lois usuelles

4.1 Loi multinomiale

Il s'agit en quelque sorte d'une loi binômiale multidimensionnelle où on ne s'intéresse pas seulement à la réalisation d'un événement particulier A , mais à celle de k événements distincts A_1, \dots, A_k de probabilités respectives p_1, \dots, p_k au cours de n épreuves successives indépendantes. On a bien entendu $p_j > 0$ pour $1 \leq j \leq k$ et $\sum_{j=1}^k p_j = 1$. On associe alors à cette expérience aléatoire le vecteur N à valeurs dans $\{0, 1, \dots, n\}^k$ dont les composantes N_j , $1 \leq j \leq k$, représentent le nombre d'événements A_j réalisés. Le vecteur aléatoire N suit une **loi multinomiale** de paramètres n, p_1, \dots, p_k , notée $\mathcal{M}(n; p_1, \dots, p_k)$ et définie par :

$$P(N_1 = n_1, \dots, N_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

où $(n_1, \dots, n_k) \in \{0, 1, \dots, n\}^k$. Il est important de noter que les v.a. composantes de N ne sont pas indépendantes puisque :

$$\sum_{j=1}^k N_j = \sum_{j=1}^k n_j = n$$

Le coefficient qui figure devant $p_1^{n_1} \dots p_k^{n_k}$, qui représente la probabilité d'un événement élémentaire réalisant l'événement $(N_1 = n_1, \dots, N_k = n_k)$, est le nombre de partitions de n objets en k classes d'effectifs donnés n_i , avec $n_1 + \dots + n_k = n$, ou le nombre de suites ordonnées de n objets comportant n_i objets identiques, $1 \leq i \leq k$, choisis dans k classes distinctes (cf. chap. 1 Compléments § 5).

Si on pose ${}^t p = (p_1, \dots, p_k)$, on peut écrire $N \rightsquigarrow \mathcal{M}(n; p)$ et ce vecteur aléatoire admet comme espérance :

$$E(N) = \begin{pmatrix} E(N_1) \\ \vdots \\ E(N_k) \end{pmatrix} = n \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} = np$$

En effet, toutes les lois marginales vont être des lois binômiales puisque les v.a. N_j représentent le nombre de réalisations de l'événement A_j au cours de n épreuves indépendantes : $N_j \rightsquigarrow \mathcal{B}(n, p_j)$ avec $p_j = P(A_j)$, donc $E(N_j) = np_j$ et $V(N_j) = np_j(1 - p_j)$.

Pour calculer les covariances :

$$\text{Cov}(N_i, N_j) = E(N_i N_j) - E(N_i) E(N_j)$$

qui ne sont pas nulles puisque les composantes de N sont liées par une relation, nous avons besoin de la loi du couple (N_i, N_j) qui est une loi $\mathcal{M}(n; p_i, p_j, 1 - p_i - p_j)$ ou loi trinômiale associée aux trois événements A_i, A_j et $\overline{A_i} \cap \overline{A_j}$:

$$P(N_i = n_i, N_j = n_j) = \frac{n!}{n_i!n_j!(n - n_i - n_j)!} p_i^{n_i} p_j^{n_j} (1 - p_i - p_j)^{n - n_i - n_j}$$

avec n_i et n_j entiers positifs tels que $0 \leq n_i + n_j \leq n$.

Nous pouvons écrire :

$$E(N_i N_j) = E[N_i E(N_j | N_i)]$$

en effet :

$$E(N_i N_j) = \sum_{n_i} \sum_{n_j} n_i n_j P(N_i = n_i, N_j = n_j) = \sum_{n_i} n_i E_{n_i} P(N_i = n_i)$$

où on a posé :

$$\begin{aligned} E_{n_i} &= \sum_{n_j} n_j \frac{P(N_i = n_i, N_j = n_j)}{P(N_i = n_i)} \\ &= \sum_{n_j} n_j P(N_j = n_j | N_i = n_i) = E(N_j | N_i = n_i) \end{aligned}$$

et $n_i E_{n_i}$ représente bien une réalisation de la v.a. $N_i E(N_j | N_i)$. Pour calculer cette espérance conditionnelle, nous allons déterminer la loi conditionnelle de $N_j | N_i = n_i$ qui est la loi binômiale de paramètres $n - n_i$ (on retire les n_i réalisations des A_i) et de probabilité individuelle de réalisation :

$$P(A_j | \bar{A}_i) = \frac{P(A_j \cap \bar{A}_i)}{P(\bar{A}_i)} = \frac{P(A_j)}{P(\bar{A}_i)} = \frac{p_j}{1 - p_i}$$

L'espérance de cette loi binômiale est donc :

$$E(N_j | N_i) = (n - N_i) \frac{p_j}{1 - p_i}$$

et ainsi :

$$E(N_i N_j) = \frac{p_j}{1 - p_i} E[N_i (n - N_i)] = \frac{p_j}{1 - p_i} [n^2 p_i - E(N_i^2)]$$

Par ailleurs :

$$E(N_i^2) = V(N_i) + E^2(N_i) = n p_i (1 - p_i) + n^2 p_i^2$$

d'où :

$$E(N_i N_j) = n p_i p_j \left(\frac{n}{1 - p_i} - 1 - \frac{n p_i}{1 - p_i} \right) = n(n - 1) p_i p_j$$

et enfin :

$$Cov(N_i, N_j) = n p_i p_j (n - 1 - n) = -n p_i p_j$$

La matrice de variances-covariances de N s'écrit donc :

$$V(N) = n \left[p_i (\delta_j^i - p_j) \right]_{1 \leq i, j \leq k}$$

où δ_j^i est le symbole de Kronecker qui vaut 1 quand $i = j$ et 0 sinon.

4.2 Loi normale vectorielle

Nous allons définir la loi normale multidimensionnelle à partir de la loi normale qui a été définie dans \mathbb{R} au chapitre 3.

DÉFINITION

Un vecteur aléatoire X à valeurs dans \mathbb{R}^n suit une **loi normale** si toute combinaison linéaire de ses composantes (qui sont des v.a. réelles) suit une loi normale dans \mathbb{R} . Si a est un vecteur de \mathbb{R}^n qui définit une telle combinaison linéaire, ceci s'écrit :

$$X \rightsquigarrow \mathcal{N}_n \Leftrightarrow \forall a \in \mathbb{R}^n, \quad {}^t a X = \sum_{i=1}^n a_i X_i \rightsquigarrow \mathcal{N}_1$$

REMARQUE Pour ne pas introduire de restrictions sur les coefficients de la combinaison linéaire, on adoptera comme convention que la loi de Dirac δ_m (cf. chap. 3, § 1.1) se confond avec la loi normale $N(m, 0)$ d'écart type nul.

Il découle immédiatement de la définition que toutes les composantes d'un vecteur normal sont des v.a. réelles normales, puisqu'il suffit de choisir toutes les coordonnées de a nulles, sauf $a_i = 1, 1 \leq i \leq n$ et ainsi ${}^t a X = X_i \rightsquigarrow \mathcal{N}_1$. Cependant, il faut faire attention que la réciproque de cette propriété est fautive en général, c'est-à-dire que si toutes les v.a. X_i suivent des lois normales, il n'est pas certain que X soit un vecteur normal. C'est pour cela que l'on ne peut pas définir un vecteur normal comme ayant des composantes normales, sauf dans un cas particulier que nous verrons ultérieurement.

Comme dans le cas unidimensionnel, la loi normale en multidimensionnel est entièrement déterminée par deux paramètres qui sont ici le vecteur espérance $\mu = E(X)$ et la matrice variances-covariances $\Sigma = V(X)$, la densité au point $x = (x_1, \dots, x_n)$ ayant pour expression :

$$f(x) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \Sigma}} \exp -\frac{1}{2} {}^t (x - \mu) \Sigma^{-1} (x - \mu)$$

il s'agit de la généralisation de l'expression de la densité de la loi $\mathcal{N}_1(\mu, \sigma^2)$ que l'on peut écrire sous une forme équivalente, dans le cas où $x \in \mathbb{R}$:

$$f(x) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2}} \exp -\frac{1}{2} {}^t (x - \mu) (\sigma^2)^{-1} (x - \mu)$$

Le terme de l'exponentielle fait intervenir la métrique définie par la matrice Σ^{-1} : $\|x - \mu\|_{\Sigma^{-1}}^2$. On écrit symboliquement : $X \rightsquigarrow \mathcal{N}_n(\mu, \Sigma)$. Remarquons que cette notation n'est pas homogène avec celle retenue en unidimensionnel, puisqu'en faisant $n = 1$ le second paramètre est ici la variance alors que figurait dans le cas réel l'écart type comme second paramètre. On peut remédier à cet inconvénient en distinguant dans le cas réel les notations $\mathcal{N}_1(\mu, \sigma^2)$ et $N(\mu, \sigma)$.

Exemple 4.13

Étudions le cas particulier de la loi normale d'un couple (X, Y) dont la densité s'écrit pour $|\rho| < 1$:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp -\frac{1}{2(1-\rho^2)} \times \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \frac{x-\mu_X}{\sigma_X} \frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]$$

Par identification avec l'expression générale de la densité de la loi normale, nous allons voir comment interpréter les cinq paramètres qui apparaissent ici. Tout d'abord on voit que $\det \Sigma = \sigma_X^2 \sigma_Y^2 (1 - \rho^2)$. Par ailleurs, on remarque aussi que le crochet de l'exponentielle est une forme quadratique de $x - \mu_X$ et $y - \mu_Y$ qui peut s'écrire :

$$\frac{1}{\sigma_X^2 \sigma_Y^2} [\sigma_Y^2 (x - \mu_X)^2 - 2\rho \sigma_X \sigma_Y (x - \mu_X)(y - \mu_Y) + \sigma_X^2 (y - \mu_Y)^2]$$

Le terme de l'exponentielle peut donc s'écrire :

$$-\frac{1}{2 \det \Sigma} \begin{pmatrix} x - \mu_X & y - \mu_Y \end{pmatrix} \begin{pmatrix} \sigma_Y^2 & -\rho \sigma_X \sigma_Y \\ -\rho \sigma_X \sigma_Y & \sigma_X^2 \end{pmatrix} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}$$

On voit bien alors sous cette forme que (X, Y) est un couple de v.a. normales, d'espérances respectives $\mu_X = E(X)$ et $\mu_Y = E(Y)$ et de matrices variances-covariances Σ vérifiant :

$$\Sigma^{-1} = \frac{1}{\det \Sigma} \begin{pmatrix} \sigma_Y^2 & -\rho \sigma_X \sigma_Y \\ -\rho \sigma_X \sigma_Y & \sigma_X^2 \end{pmatrix}$$

d'où :

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix}$$

ce qui permet d'en déduire que $\sigma_X^2 = V(X)$, $\sigma_Y^2 = V(Y)$ et $Cov(X, Y) = \rho \sigma_X \sigma_Y$ donc $\rho = Corr(X, Y)$.

Nous allons déterminer maintenant la régression de Y sur X :

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y f_Y(y|X = x) dy$$

avec ici :

$$f_Y(y|X = x) = \frac{f(x, y)}{f_X(x)} = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_Y}} \exp \left\{ -\frac{1}{2(1-\rho^2)} g(x, y) \right\}$$

où :

$$\begin{aligned}
 g(x, y) &= \left(\frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \frac{x - \mu_X}{\sigma_X} \times \frac{y - \mu_Y}{\sigma_Y} + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 \\
 &\quad - (1 - \rho^2) \left(\frac{x - \mu_X}{\sigma_X} \right)^2 \\
 &= \left(\frac{y - \mu_Y}{\sigma_Y} - \rho \frac{x - \mu_X}{\sigma_X} \right)^2
 \end{aligned}$$

Pour calculer l'intégrale, on fait donc le changement de variable :

$$\frac{y - \mu_Y}{\sigma_Y} - \rho \frac{x - \mu_X}{\sigma_X} = u\sqrt{1 - \rho^2} \quad \text{et} \quad dy = \sigma_Y \sqrt{1 - \rho^2} du.$$

On obtient alors :

$$\begin{aligned}
 E(Y|X = x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left[u\sigma_Y \sqrt{1 - \rho^2} + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) + \mu_Y \right] e^{-u^2/2} du \\
 &= \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) + \mu_Y
 \end{aligned}$$

la fonction de régression d'une variable d'un couple normal sur l'autre est donc une fonction affine, c'est-à-dire que la courbe de régression est ici une droite. Dans le cas particulier où les variables sont centrées et réduites on obtient la régression linéaire $E(Y|X) = \rho X$.

a) Cas particulier

Après la définition de la loi normale vectorielle, nous avons noté que n v.a. réelles normales ne constituaient pas nécessairement les composantes d'un vecteur normal. Cependant, si X_1, \dots, X_n sont des v.a. normales **indépendantes**, de lois respectives $N(\mu_i, \sigma_i)$, $1 \leq i \leq n$, alors elles constituent les composantes d'un vecteur X normal, toute combinaison linéaire de v.a. normales indépendantes suivant une loi normale. Ses paramètres sont :

$$E(X) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{et} \quad V(X) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}$$

avec pour densité au point $x = (x_1, \dots, x_n)$:

$$\begin{aligned}
 f(x) &= \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \exp -\frac{1}{2\sigma_i^2} (x_i - \mu_i)^2 \\
 &= \frac{1}{(\sqrt{2\pi})^n \prod_{i=1}^n \sigma_i} \exp -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2
 \end{aligned}$$

Dans le cas particulier de variables centrées et réduites, c'est-à-dire pour $\mu_i = 0$ et $\sigma_i = 1$, $1 \leq i \leq n$, on obtient la loi normale standard vectorielle, $X \rightsquigarrow \mathcal{N}_n(0, I_n)$, de densité :

$$f(x) = \frac{1}{(\sqrt{2\pi})^n} \exp -\frac{1}{2} \sum_{i=1}^n x_i^2 = \frac{1}{(2\pi)^{n/2}} \exp -\frac{1}{2} \|x\|^2$$

la norme étant ici la norme euclidienne classique.

Théorème 4.1

Si X_i et X_j sont les composantes d'un vecteur normal, elles sont indépendantes si et seulement si leur covariance est nulle : $Cov(X_i, X_j) = 0$.

Nous avons ici un cas particulier où la covariance nulle implique l'indépendance. Il faut cependant bien faire attention à l'énoncé de cette propriété, car si X_i et X_j sont deux variables normales dont la covariance est nulle, cela n'implique pas forcément qu'elles soient indépendantes. Pour s'en convaincre examinons l'exemple suivant.

Exemple 4.14

Soit X de loi $N(0, 1)$ et ε de loi définie par $P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$ deux v.a. indépendantes à partir desquelles on définit la v.a. $Y = \varepsilon X$. La loi de probabilité de cette nouvelle variable est définie par :

$$\begin{aligned} P(Y < y) &= P(\varepsilon X < y) \\ &= P\{(\varepsilon X < y) \cap (\varepsilon = 1)\} + P\{(\varepsilon X < y) \cap (\varepsilon = -1)\} \\ &= \frac{1}{2}P(X < y) + \frac{1}{2}P(-X < y) \\ &= \frac{1}{2}[\Phi(y) + 1 - \Phi(-y)] = \Phi(y) \end{aligned}$$

où Φ est la f.r. de la loi $N(0, 1)$ qui est donc la loi suivie par Y . Par ailleurs :

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(\varepsilon X^2) = E(\varepsilon)E(X^2) = 0$$

Si on en concluait que X et Y étaient deux variables indépendantes, la v.a. $X + Y$ suivrait alors une loi $N(0, \sqrt{2})$. Or :

$$P(X + Y = 0) = P\{(\varepsilon + 1)X = 0\} = P(\varepsilon = -1) = \frac{1}{2}$$

ce qui est incompatible avec le fait que la loi de $X + Y$ soit continue. On a donc établi que X et Y sont deux variables normales de covariance nulle et cependant dépendantes.

b) Transformation affine d'un vecteur normal

Soit X un vecteur de loi $\mathcal{N}_n(\mu, \Sigma)$ et de composantes $X_i, 1 \leq i \leq n$. Nous allons effectuer une transformation affine à partir d'une matrice A de format (m, n) , traduction matricielle d'une application linéaire de \mathbb{R}^n dans \mathbb{R}^m , et d'un vecteur de translation b de \mathbb{R}^m : $Y = AX + b$. Nous allons établir tout d'abord que le vecteur Y est aussi normal. En effet, si a est un vecteur quelconque de \mathbb{R}^m , la combinaison linéaire :

$${}^t aY = ({}^t aA)X + {}^t ab = \sum_{i=1}^n \alpha_i X_i + \beta \rightsquigarrow \mathcal{N}_1$$

ayant posé ${}^t \alpha = {}^t aA$, vecteur de composantes $\alpha_i, 1 \leq i \leq n$, et $\beta = {}^t ab \in \mathbb{R}$. En effet, par définition de la loi normale vectorielle, la combinaison linéaire ${}^t \alpha X$ est une v.a. réelle normale et donc aussi ${}^t \alpha X + \beta$. Ainsi, toute combinaison linéaire des composantes de Y suit une loi normale, ce qui signifie par définition que Y est un vecteur normal de \mathbb{R}^m . D'après les propriétés des opérateurs espérance et variance vues au §3, ses moments sont $E(Y) = AE(X) + b = A\mu + b = \mu'$ et $V(Y) = A\Sigma^t A = \Sigma'$, et ainsi $Y \rightsquigarrow \mathcal{N}_m(\mu', \Sigma')$.

Dans le cas particulier d'une transformation linéaire, c'est-à-dire avec $b = 0$, et pour un vecteur X de loi normale standard, $X \rightsquigarrow \mathcal{N}_n(0, I_n)$, alors :

$$Y = AX \rightsquigarrow \mathcal{N}_m(0, A^t A).$$

c) Loi du khi-deux

Soit un vecteur X de loi normale standard $\mathcal{N}_n(0, I_n)$ et A une matrice symétrique d'ordre n . On établit le résultat suivant :

$${}^t XAX \rightsquigarrow \chi_p^2 \iff A \text{ est une matrice idempotente de rang } p.$$

Rappelons que A est idempotente si $A^2 = A$ et que dans ce cas $\text{rang } A = \text{trace } A$.

Si A et B sont deux matrices symétriques et idempotentes, on a également le résultat suivant :

$${}^t XAX \text{ et } {}^t XBX \text{ sont des v.a. indépendantes} \iff AB = 0.$$

5 Compléments

5.1 Application mesurable

Les notions vues dans le chapitre 2 se généralisent en multidimensionnel, où \mathbb{R}^n est muni de sa tribu borélienne \mathcal{B} , engendrée par les ouverts ou les pavés de la forme $\prod_{i=1}^n]-\infty, x_i[$. Un vecteur aléatoire X est alors défini comme une application mesu-

table de (Ω, \mathcal{A}) dans $(\mathbb{R}^n, \mathcal{B})$. La loi de X admettra une densité si elle est absolument continue par rapport à la mesure de Lebesgue λ de \mathbb{R}^n , définie à partir de :

$$\lambda \left(\prod_{i=1}^n]a_i, b_i] \right) = \prod_{i=1}^n (b_i - a_i).$$

5.2 Changement de variable

Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une application bijective de $X(\Omega)$, ouvert de \mathbb{R}^n , dans un ouvert de \mathbb{R}^n , admettant des dérivées partielles continues ainsi que son inverse g^{-1} . Si X est un vecteur aléatoire de \mathbb{R}^n de densité f , alors $Y = g(X)$ est un vecteur aléatoire de \mathbb{R}^n de densité h définie par :

$$h(y) = f[g^{-1}(y)] \left| \frac{D(x)}{D(y)} \right|$$

où $\frac{D(x)}{D(y)}$ représente le jacobien de la transformation inverse, c'est-à-dire le déterminant des dérivées partielles des anciennes variables exprimées en fonction des nouvelles sous la forme $x_i = g_i^{-1}(y)$, $1 \leq i \leq n$, donc ayant comme élément ligne i , colonne j , $1 \leq j \leq n$:

$$\frac{\partial g_i^{-1}(y)}{\partial y_j}$$

Si on retient comme application la convolution, pour obtenir la loi de la somme de n v.a. indépendantes X_1, \dots, X_n on fait le changement $X_1 = X_1, \dots, X_{n-1} = X_{n-1}$, $Y = X_1 + \dots + X_n$. Les anciennes coordonnées s'écrivent sous la forme $x_1 = x_1, \dots, x_{n-1} = x_{n-1}, x_n = y - x_1 - \dots - x_{n-1}$ d'où un jacobien égal à un et une densité de Y qui se détermine par :

$$h(y) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_{n-1}, y - x_1 - \dots - x_{n-1}) dx_1 \dots dx_{n-1}$$

Dans le cas d'un couple aléatoire (X, Y) transformé par l'application g en couple (U, V) on détermine les anciennes coordonnées en fonction des nouvelles sous la forme $x = x(u, v)$ et $y = y(u, v)$ et l'élément différentiel $f(x, y) dx dy$ est transformé en $f[g^{-1}(u, v)] \left| \frac{D(x, y)}{D(u, v)} \right| du dv$ où :

$$\frac{D(x, y)}{D(u, v)} = \begin{vmatrix} \frac{\partial x(u, v)}{\partial u} & \frac{\partial x(u, v)}{\partial v} \\ \frac{\partial y(u, v)}{\partial u} & \frac{\partial y(u, v)}{\partial v} \end{vmatrix}$$

● L'essentiel

- 1 Deux v.a. sont indépendantes si la loi du couple s'obtient comme produit des lois marginales : produit des probabilités des points dans le cas discret et produit des densités dans le cas continu.
- 2 La covariance est une mesure de la dépendance linéaire entre deux v.a. Deux v.a. indépendantes ont une covariance nulle. La réciproque est fautive en général. Cependant, si deux v.a. sont des composantes d'un vecteur normal et que leur covariance est nulle, alors elles sont indépendantes.
- 3 La fonction de régression d'une variable Y sur une variable X associée à tout réel x l'espérance (moyenne) de Y lorsque la valeur de X est fixée, égale à x .
- 4 Un vecteur aléatoire admet comme composantes des variables aléatoires réelles. On lui associe comme moments son espérance, caractéristique de valeur centrale, vecteur dont les composantes sont les espérances des v.a. composantes, et sa matrice de variances-covariances, caractéristique de dispersion dans l'espace, qui rassemble les variances des composantes sur la diagonale principale et les covariances des composantes associées deux à deux pour les autres éléments.
- 5 Par une transformation affine (changement d'échelles et d'origine), les moments deviennent :

$$E(AX + b) = AE(X) + b$$

$$V(AX + b) = AV(X)^t A$$

pour toute matrice A de format (m, n) et tout vecteur b de \mathbb{R}^m .



- 4.1** Soit (X, Y) un couple de v.a. discrètes dont la loi de probabilité est donnée par le tableau ci-après :

$X \backslash Y$	1	2	3	4
1	0,08	0,04	0,16	0,12
2	0,04	0,02	0,08	0,06
3	0,08	0,04	0,16	0,12

1. Déterminer les lois marginales de X et Y et préciser si ces v.a. sont indépendantes.
 2. Calculer $Cov(X, Y)$.
 3. Déterminer la loi du couple $(\min\{X, Y\}, \max\{X, Y\})$.
- 4.2** On lance successivement trois pièces de monnaie et les v.a. X et Y désignent respectivement le nombre de faces apparues sur les deux premières pièces et le nombre de piles sur les deux dernières.
1. Déterminer la loi du couple (X, Y) , puis les lois marginales de X et Y et indiquer si ces v.a. sont indépendantes.
 2. Calculer $Cov(X, Y)$.
- 4.3** Soit (X, Y) un couple de v.a. discrètes dont la loi de probabilité est donnée par le tableau ci-après :

$Y \backslash X$	0	1	2
1	1/12	0	1/12
2	2/12	1/12	1/12
3	3/12	2/12	1/12

1. Déterminer les lois marginales de X et Y et préciser si ces v.a. sont indépendantes.
 2. Calculer $Cov(X, Y)$.
 3. En déduire la loi de probabilité de la variable aléatoire $E(Y|X)$. Calculer $E[E(Y|X)]$; le résultat était-il prévisible ?
- 4.4** Soit (X, Y) un couple de v.a. discrètes dont la loi de probabilité est donnée par le tableau ci-après :

$X \backslash Y$	1	2	3	4
1	0	0	0	0,3
2	0,2	0	0	0
3	0	0	0,1	0
4	0,3	0,1	0	0

Déterminer les lois conditionnelles de X sachant que $Y = 1$ puis de Y sachant que $X \in \{3, 4\}$.

4.5 Soit X et Y deux v. a. de loi normale centrée et réduite et de coefficient de corrélation linéaire $\rho = \sin \theta$. On montre que $P(X > 0, Y > 0) = \frac{1}{4} + \frac{\theta}{2\pi} = p$.

1. Déterminer la loi de probabilité des v. a. $-X$ et $-Y$ et en déduire la valeur de $P(X < 0, Y < 0)$.

2. Soit X_1 et X_2 deux v. a. qui ont la même loi que X et Y_1, Y_2 deux v. a. qui ont la même loi que Y , les couples (X_1, Y_1) et (X_2, Y_2) étant indépendants. Déterminer les lois de probabilité des v. a. $U = X_1 - X_2$ et $V = Y_1 - Y_2$. Calculer le coefficient de corrélation linéaire de U et V et en déduire $P(U > 0, V > 0)$.

3. On définit les v. a. $W = 1$ si $U > 0$, $W = -1$ si $U < 0$, $Z = 1$ si $V > 0$ et $Z = -1$ si $V < 0$. Calculer $E(WZ)$.

4.6 Soit X et Y deux v.a. indépendantes de lois de Poisson de paramètres respectifs λ et μ . Déterminer la loi conditionnelle de X lorsque la somme $S = X + Y$ a une valeur fixée $S = s$. En déduire l'expression de la fonction de régression de X sur S puis la valeur de $E[E(X|S)]$.

4.7 Un couple (X, Y) de variables aléatoires réelles admet pour densité $f(x, y) = \exp(y-x)$ pour $0 \leq y \leq 1$ et $y \leq x$, avec $f(x, y) = 0$ sinon.

Déterminer les densités marginales de X et Y . Les variables aléatoires X et Y sont-elles indépendantes ?

4.8 Paul et Virginie se fixent un rendez-vous entre 1 h et midi, sans autre précision. Comme ils sont impatients tous les deux, celui qui est arrivé part au bout d'un quart d'heure si l'autre n'arrive pas pendant ce laps de temps. Calculer la probabilité p qu'ils se rencontrent effectivement à ce rendez-vous. On formalisera le problème en considérant que l'arrivée de chacun se produit au hasard pendant l'heure fixée du rendez-vous.

4.9 Soit (X, Y) un couple de v.a. dont la loi est déterminée par la densité :

$$f(x, y) = \begin{cases} xy/2 & \text{si } 0 \leq x \leq 2 \text{ et } 0 \leq y \leq x \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer la fonction de répartition F de ce couple.

2. Déterminer les lois marginales de X et Y . Ces variables sont-elles indépendantes ?

3. Déterminer la loi conditionnelle de Y sachant que $X = x$.

4.10 Soit (X, Y) un couple de v.a. de densité :

$$f(x, y) = \begin{cases} \frac{k}{\sqrt{xy}} & \text{si } 0 < x \leq y < 1 \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer la valeur de la constante k puis la fonction de répartition F de ce couple.

2. Déterminer les lois marginales de X et Y . Ces variables sont-elles indépendantes ?

3. Déterminer les lois conditionnelles de $X|Y = y$ et de $Y|X = x$. En déduire l'expression de la fonction de régression $x \mapsto E(Y|X = x)$ puis calculer $E[E(Y|X)]$.

4.11 Soit X et Y deux v.a. indépendantes et de même loi uniforme sur $] -1, 1[$. Déterminer la loi de probabilité de la v.a. $Z = Y - X$.

4.12 Soit (X, Y) un couple de v.a. de densité :

$$f(x, y) = \frac{1}{2\pi\sqrt{a}} \exp -\frac{1}{2a} [(1+a)x^2 + 2(1+2a)xy + (1+4a)y^2]$$

avec $a > 0$.

1. Déterminer les lois marginales de X et Y . Ces variables sont-elles indépendantes ?

2. Déterminer la loi de la v.a. $Z = \frac{1}{a}(X + Y)^2$.

4.13 Si X est un vecteur de loi $\mathcal{N}_n(\mu, \Sigma)$, montrer que la v.a. réelle ${}^t(X - \mu)\Sigma^{-1}(X - \mu)$ suit une loi du khi-deux à n degrés de liberté.

4.14 Soit X_1, X_2 et X_3 trois v.a. indépendantes de même loi $N(0, 1)$. Déterminer la loi de probabilité des v.a. suivantes :

$$U = \frac{1}{\sqrt{2}}(X_2 - X_1), \quad V = \frac{(X_1 + X_2)^2}{(X_1 - X_2)^2}, \quad W = \frac{X_1 + X_2}{\sqrt{(X_1 - X_2)^2}}$$

$$Y = \frac{1}{6}(5X_1^2 + 2X_2^2 + 5X_3^2 + 4X_1X_2 - 2X_1X_3 + 4X_2X_3)$$

$$Z = X_1^2 + X_2^2 + X_3^2 - Y$$

et montrer que Y et Z sont indépendantes.

● Solutions



- 4.1** 1. Les lois marginales s'obtiennent par sommation de lignes et de colonnes et figurent dans le tableau ci-après :

X \ Y	1	2	3	4	
1	0,08	0,04	0,16	0,12	0,4
2	0,04	0,02	0,08	0,06	0,2
3	0,08	0,04	0,16	0,12	0,4
	0,2	0,1	0,4	0,3	1

On constate que toutes les probabilités des couples (x, y) s'obtiennent en faisant le produit des probabilités marginales, donc X et Y sont indépendantes.

2. On déduit de la question précédente que $Cov(X, Y) = 0$.

3. La loi du couple $(\min\{X, Y\}, \max\{X, Y\})$ est donnée dans le tableau ci-après :

$\min\{X, Y\} \backslash \max\{X, Y\}$	1	2	3	4
1	0,08	0,08	0,24	0,12
2	0	0,02	0,12	0,06
3	0	0	0,16	0,12

Donnons un exemple de la façon dont ce tableau a été constitué :

$$\begin{aligned}
 P(\min\{X, Y\} = 2, \max\{X, Y\} = 3) &= P(X = 2, Y = 3) + P(X = 3, Y = 2) \\
 &= 0,08 + 0,04 = 0,12
 \end{aligned}$$

- 4.2** 1. La loi du couple et les lois marginales sont données dans le tableau suivant :

Y \ X	0	1	2	
0	0	1/8	1/8	1/4
1	1/8	2/8	1/8	1/2
2	1/8	1/8	0	1/4
	1/4	1/2	1/4	1

On voit dans le tableau que par exemple $P(X = 0, Y = 0) = 0 \neq P(X = 0)P(Y = 0) = \frac{1}{16}$ donc les variables X et Y ne sont pas indépendantes.

2. On obtient aisément $E(X) = E(Y) = 1$ et $E(XY) = \frac{3}{4}$ d'où $Cov(X, Y) = -\frac{1}{4}$.

- 4.3** 1. Les lois marginales s'obtiennent par addition des lignes et des colonnes et figurent dans le tableau ci-après :

Y \ X	0	1	2	
1	1/12	0	1/12	1/6
2	2/12	1/12	1/12	1/3
3	3/12	2/12	1/12	1/2
	1/2	1/4	1/4	1

On voit par exemple que $P(X = 1, Y = 1) = 0 \neq P(X = 1)P(Y = 1) = 1/24$ donc les v.a. X et Y ne sont pas indépendantes.

2. On calcule à partir du tableau $E(X) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}$, $E(Y) = \frac{1}{6} + \frac{2}{3} + \frac{3}{2} = \frac{7}{3}$ puis $E(XY) = \frac{1}{12}(2 + 6 + 2 + 4 + 6) = \frac{5}{3}$. On en déduit $Cov(X, Y) = \frac{5}{3} - \frac{7}{4} = -\frac{1}{12}$ ce qui confirme que les variables ne peuvent pas être indépendantes.

3. Il faut au préalable déterminer les lois conditionnelles de Y qui sont données dans le tableau ci-après :

Y	1	2	3
$X = 0$	1/6	1/3	1/2
$X = 1$	0	1/3	2/3
$X = 2$	1/3	1/3	1/3

On calcule alors à partir du tableau précédent :

$$E(Y | X = 0) = \frac{7}{3} \quad E(Y | X = 1) = \frac{8}{3} \quad E(Y | X = 2) = 2$$

On en déduit la loi :

$E(Y X)$	2	7/3	8/3
	1/4	1/2	1/4

puis la valeur :

$$E[E(Y | X)] = \frac{2}{4} + \frac{7}{6} + \frac{2}{3} = \frac{7}{3}$$

On retrouve bien la valeur de $E(Y)$.

4.4 Comme $P(Y = 1) = 0,5$ la loi de $X|Y = 1$ s'obtient en multipliant la première colonne par deux et est donnée par le tableau ci-après :

$X Y=1$	1	2	3	4
	0	0,4	0	0,6

On a $P(X \in \{3, 4\}) = 0,5$; on additionne les deux dernières lignes et ensuite on multiplie aussi par deux pour obtenir la loi de $Y|X \in \{3, 4\}$, donnée dans le tableau ci-dessous :

$Y X \in \{3, 4\}$	1	2	3	4
	0,6	0,2	0,2	0

4.5 1. On a :

$$P(-X < x) = P(X > -x) = 1 - \Phi(-x) = \Phi(x) = P(X < x)$$

donc $-X$ et $-Y$ suivent aussi des lois normales centrées et réduites. On en déduit :

$$P(X < 0, Y < 0) = P(-X > 0, -Y > 0) = p$$

2. Toute combinaison linéaire de v. a. normales indépendantes suit une loi normale, donc U et V suivent des lois normales centrées et de variance 2. On calcule :

$$\text{Cov}(U, V) = E(UV) = E(X_1Y_1) - E(X_1Y_2) - E(X_2Y_1) + E(X_2Y_2) = 2\rho$$

On en déduit $\text{corr}(U, V) = \rho$ puis :

$$P(U > 0, V > 0) = P\left(\frac{U}{\sqrt{2}} > 0, \frac{V}{\sqrt{2}} > 0\right) = p$$

3. On calcule :

$$E(WZ) = P(WZ = 1) - P(WZ = -1) = 2P(WZ = 1) - 1.$$

Ensuite :

$$\begin{aligned} P(WZ = 1) &= P(W = Z = 1) + P(W = Z = -1) \\ &= P(U > 0, V > 0) + P(U < 0, V < 0) = 2p \end{aligned}$$

Ainsi :

$$E(WZ) = 4p - 1 = \frac{2\theta}{\pi}$$

4.6 Nous savons que S suit une loi de Poisson de paramètre $\lambda + \mu$ et par conséquent, pour $0 \leq x \leq s$:

$$\begin{aligned} P(X = x | S = s) &= \frac{P(X = x, S = s)}{P(S = s)} = \frac{P(X = x)P(S = s | X = x)}{P(S = s)} \\ &= \frac{P(X = x)P(Y = s - x)}{P(S = s)} = \frac{s!}{x!(s-x)!} \times \frac{\lambda^x \mu^{s-x}}{(\lambda + \mu)^s} \\ &= \binom{s}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{s-x} \end{aligned}$$

ce qui montre que la loi de $X|S = s$ est une loi binômiale de paramètres s et $\frac{\lambda}{\lambda + \mu}$, donc d'espérance :

$$E(X | S = s) = s \frac{\lambda}{\lambda + \mu}$$

Ainsi, $E(X|S) = S \frac{\lambda}{\lambda + \mu}$ et $E[E(X | S)] = E(S) \frac{\lambda}{\lambda + \mu} = \lambda = E(X)$.

4.7 Les densités marginales s'obtiennent par intégration de la densité du couple :

$$f_X(x) = \int_0^x e^{y-x} dy = 1 - e^{-x} \quad \text{si } 0 \leq x \leq 1$$

$$f_X(x) = \int_0^1 e^{y-x} dy = (e - 1)e^{-x} \quad \text{si } 1 \leq x$$

avec $f_X(x) = 0$ si $x \leq 0$;

et :

$$f_Y(y) = \int_y^{+\infty} e^{y-x} dx = 1 \quad \text{si } 0 \leq y \leq 1$$

la densité étant nulle en dehors de cet intervalle.

Les variables aléatoires X et Y sont dépendantes car $f(x, y) \neq f_X(x) f_Y(y)$

- 4.8** On note X et Y les v.a. qui représentent les dates d'arrivée de Paul et Virginie à leur rendez-vous, en prenant 11h comme origine et l'heure comme unité de temps. Leur arrivée au hasard se traduit par l'hypothèse que ces deux v.a. suivent une loi uniforme sur $[0,1]$ et sont indépendantes. Leur rencontre correspond alors à l'événement $|X - Y| < \frac{1}{4}$ et la probabilité p s'obtient par l'intégrale suivante :

$$p = P \left\{ |X - Y| < \frac{1}{4} \right\} = \int \int_D dx dy$$

où le domaine d'intégration est $D = \left\{ (x, y) / 0 \leq x \leq 1, 0 \leq y \leq 1, |x - y| < \frac{1}{4} \right\}$.

Compte tenu de la forme de ce domaine et de sa symétrie par rapport à la première bissectrice (cf. figure 4.5), il est préférable de calculer la probabilité du complémentaire :

$$1 - p = 2 \int_0^{3/4} dx \int_{x+1/4}^1 dy = 2 \int_0^{3/4} \left(\frac{3}{4} - x \right) dx = \frac{9}{16}$$

Ainsi $p = \frac{7}{16}$.

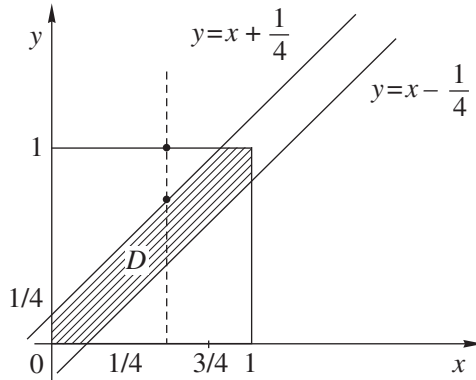


Figure 4.5

- 4.9** 1. Si $x \leq 0$ ou $y \leq 0$ on a bien sûr $F(x, y) = 0$ puisque la densité est nulle dans toute cette zone du plan. Pour un point $M(x, y)$ situé dans le triangle où $f > 0$, c'est-à-dire pour $0 < y \leq x \leq 2$ on obtient (voir figure 4.6) :

$$\begin{aligned} F(x, y) &= \frac{1}{2} \int_0^y v dv \int_v^x u du = \frac{1}{4} \int_0^y v(x^2 - v^2) dv \\ &= \frac{1}{4} \left(\frac{x^2 y^2}{2} - \frac{y^4}{4} \right) = \frac{y^2}{16} (2x^2 - y^2) \end{aligned}$$

Nous allons maintenant sortir du triangle, sur l'horizontale à droite de $M(x, y)$ pour nous situer en un point $M_1(x, y)$ de coordonnées telles que $0 < y < 2 \leq x$. La valeur

de F en ce point est la même qu'au point d'intersection de l'horizontale avec la frontière du triangle puisqu'au-delà $f = 0$, soit :

$$F(x, y) = F(2, y) = \frac{y^2}{16}(8 - y^2)$$

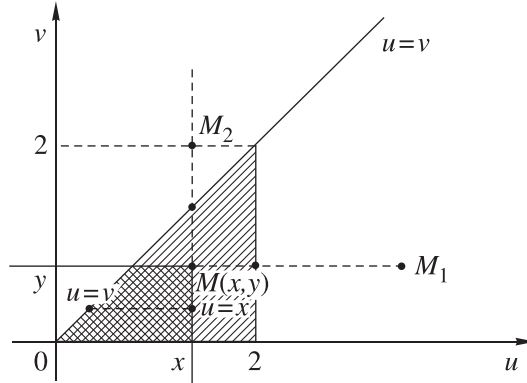


Figure 4.6

Si on sort du triangle sur la verticale de M pour atteindre un point $M_2(x, y)$ de coordonnées telles que $0 < x \leq y$ et $x \leq 2$, on obtient de même :

$$F(x, y) = F(x, x) = \frac{x^4}{16}$$

Enfin, si $x \geq 2$ et $y \geq 2$, tout le triangle est recouvert et $F(x, y) = 1$. En résumé :

$$F(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ou } y \leq 0 \\ \frac{1}{16}y^2(2x^2 - y^2) & \text{si } 0 < y \leq x < 2 \\ \frac{1}{16}x^4 & \text{si } 0 \leq x \leq 2 \text{ et } x \leq y \\ \frac{1}{16}y^2(8 - y^2) & \text{si } 0 \leq y \leq 2 \leq x \\ 1 & \text{si } x \geq 2 \text{ et } y \geq 2 \end{cases}$$

2. Les densités marginales s'obtiennent par intégration de la densité du couple ; pour $0 \leq x \leq 2$:

$$f_X(x) = \frac{x}{2} \int_0^x y dy = \frac{x^3}{4}$$

et pour $0 \leq y \leq 2$:

$$f_Y(y) = \frac{y}{2} \int_y^2 x dx = \frac{y}{4}(4 - y^2)$$

Le produit $f_X(x)f_Y(y)$ est différent de $f(x, y)$ donc les variables X et Y ne sont pas indépendantes.

3. La densité de Y sachant que $X = x$ a pour expression quand $0 \leq y \leq x$:

$$f(y|X = x) = \frac{xy}{2} \times \frac{4}{x^3} = \frac{2y}{x^2}$$

4.10 1. L'intégrale de f sur \mathbb{R}^2 doit être égale à 1 ; elle se réduit à l'intégrale sur le domaine D où elle est non nulle :

$$1 = k \int \int_D \frac{dxdy}{\sqrt{xy}} = k \int_0^1 \frac{dx}{\sqrt{x}} \int_x^1 \frac{dy}{\sqrt{y}} = 2k \int_0^1 \left(\frac{1}{\sqrt{x}} - 1 \right) dx = 2k$$

donc $k = \frac{1}{2}$. Remarquons que si l'expression de f est symétrique en x et y , il n'en est pas de même du domaine D et il aurait été plus rapide ici d'intégrer d'abord par rapport à x .

Si on note Δ le domaine défini par $x < x_0$ et $y < y_0$, par définition $F(x_0, y_0) = \int \int_{\Delta} f(x, y)dxdy$. Bien entendu, si $x_0 \leq 0$ ou $y_0 \leq 0$, Δ n'a aucun point commun avec D où $f > 0$, donc $F(x_0, y_0) = 0$. Pour un point $M(x_0, y_0)$ situé dans D maintenant (voir figure 4.7) c'est-à-dire tel que $0 < x_0 \leq y_0 < 1$:

$$\begin{aligned} F(x_0, y_0) &= \int \int_{D \cap \Delta} \frac{dxdy}{2\sqrt{xy}} = \frac{1}{2} \int_0^{x_0} \frac{dx}{\sqrt{x}} \int_x^{y_0} \frac{dy}{\sqrt{y}} = \frac{1}{2} \int_0^{x_0} \frac{dx}{\sqrt{x}} (2\sqrt{y_0} - 2\sqrt{x}) \\ &= \int_0^{x_0} \left(\frac{\sqrt{y_0}}{\sqrt{x}} - 1 \right) dx = \sqrt{x_0} (2\sqrt{y_0} - \sqrt{x_0}) \end{aligned}$$

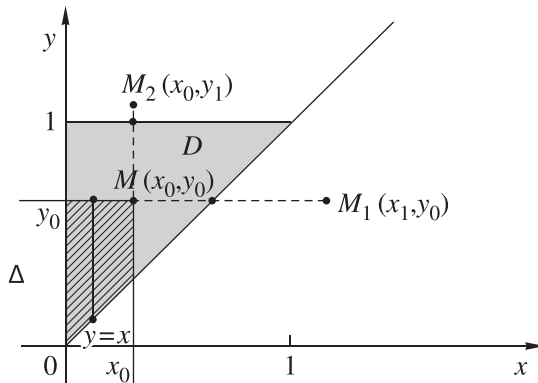


Figure 4.7

Notons qu'ici il fallait d'abord intégrer par rapport à y , car sinon il aurait fallu séparer $D \cap \Delta$ en deux domaines d'intégration. Pour les autres zones du plan, si nous quittons D sur une horizontale, à droite du point précédent, la valeur de F au point $M_1(x_1, y_0)$

est la même qu'au point d'intersection avec la frontière de D , puisqu'au-delà $f = 0$, soit pour un point de coordonnées telles que $x_1 > y_0 > 0$ et $y_0 < 1$:

$$F(x_1, y_0) = F(y_0, y_0) = y_0$$

Si maintenant nous quittons le domaine D sur une verticale au-dessus du point $M(x_0, y_0)$, nous atteignons un point M_2 de coordonnées x_0 et y_1 , avec $0 < x_0 < 1 \leq y_1$ et en ce point :

$$F(x_0, y_1) = F(x_0, 1) = \sqrt{x_0}(2 - \sqrt{x_0})$$

Enfin, si $x_0 > 1$ et $y_0 > 1$, alors $D \cap \Delta = D$ et $F(x_0, y_0) = 1$. En résumé :

$$F(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ou } y \leq 0 \\ 2\sqrt{xy} - x & \text{si } 0 < x \leq y < 1 \\ 2\sqrt{x} - x & \text{si } 0 < x < 1 \leq y \\ y & \text{si } 0 < y < x \text{ et } y \leq 1 \\ 1 & \text{si } x \geq 1 \text{ et } y \geq 1 \end{cases}$$

2. On peut en déduire les f.r. marginales :

$$F_X(x) = F(x + \infty) = \begin{cases} 0 & \text{si } x \leq 0 \\ 2\sqrt{x} - x & \text{si } 0 < x \leq 1 \\ 1 & \text{si } 1 \leq x \end{cases}$$

$$F_Y(y) = F(+\infty, y) = \begin{cases} 0 & \text{si } y \leq 0 \\ y & \text{si } 0 \leq y \leq 1 \\ 1 & \text{si } 1 \leq y \end{cases}$$

On en déduit les densités par dérivation ; pour $0 < x \leq 1$:

$$f_X(x) = \frac{1}{\sqrt{x}} - 1,$$

résultat que l'on peut retrouver par intégration de f :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_x^1 \frac{dy}{2\sqrt{xy}} = \frac{1}{\sqrt{x}} [\sqrt{y}]_x^1 = \frac{1}{\sqrt{x}} - 1$$

La loi de Y est la loi uniforme sur $[0,1]$, avec $f_Y(y) = 1$ si $0 \leq y \leq 1$.

Comme $f_X(x)f_Y(y) \neq f(x, y)$ on en conclut que X et Y ne sont pas indépendantes.

3. Les lois conditionnelles peuvent se définir par leurs densités. Pour $0 < y_0 < 1$, quand $0 < x \leq y_0$:

$$f_X(x|Y = y_0) = \frac{f(x, y_0)}{f_Y(y_0)} = \frac{1}{2\sqrt{xy_0}}$$

Pour $0 < x_0 < 1$ et quand $x_0 < y \leq 1$:

$$f_Y(y|X = x_0) = \frac{f(x_0, y)}{f_X(x_0)} = \frac{1}{2(1 - \sqrt{x_0})\sqrt{y}}$$

On peut alors déterminer la régression :

$$E(Y | X = x_0) = \frac{1}{2(1 - \sqrt{x_0})} \int_{x_0}^1 \frac{y}{\sqrt{y}} dy = \frac{1 - x_0 \sqrt{x_0}}{3(1 - \sqrt{x_0})} = \frac{1}{3} (1 + \sqrt{x_0} + x_0)$$

On calcule alors :

$$E[E(Y | X)] = \int_{-\infty}^{+\infty} E(Y | X = x) f_X(x) dx = \frac{1}{3} \int_0^1 \left(\frac{1}{\sqrt{x}} - x \right) dx = \frac{1}{2}$$

qui est bien la valeur de $E(Y)$.

- 4.11** Les v.a. X et Y ont la même densité $f(t) = \frac{1}{2} 1_{]-1,1[}(t)$ et comme elles sont indépendantes, la densité du couple (X, Y) est $f(x)f(y)$. Pour obtenir la densité du couple (X, Z) on effectue le changement de variable suivant :

$$\begin{cases} x = x \\ z = y - x \end{cases}$$

Les anciennes variables s'expriment alors par :

$$\begin{cases} x = x \\ y = z + x \end{cases}$$

ce qui permet de déterminer le jacobien de la transformation :

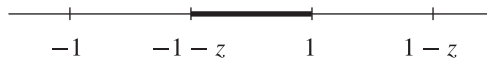
$$\frac{D(x, y)}{D(x, z)} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

Le couple (X, Z) a donc pour densité $f(x)f(x+z)$ et la densité marginale de Z s'obtient par intégration :

$$g(z) = \int_{-\infty}^{+\infty} f(x)f(x+z) dx$$

L'intégrande sera non nulle si on a simultanément $-1 < x < 1$ et $-1 - z < x < 1 - z$, soit $\max\{-1, -1 - z\} < x < \min\{1, 1 - z\}$. Nous allons distinguer deux cas :

-si $z < 0$:



la densité est non nulle pour $-1 - z < x < 1$, condition qui ne peut être réalisée que si $-1 - z < 1$, soit $z > -2$; dans ce cas :

$$g(z) = \int_{-1-z}^1 \frac{1}{4} dx = \frac{1}{4}(2 + z)$$

-si $z > 0$:



la densité est non nulle pour $-1 < x < 1-z$, condition qui ne peut être réalisée que si $-1 < 1-z$, soit $z < 2$; dans ce cas :

$$g(z) = \int_{-1}^{1-z} \frac{1}{4} dx = \frac{1}{4}(2-z)$$

On obtient ainsi :

$$g(z) = \begin{cases} \frac{1}{4}(2-|z|) & \text{si } -2 < z < 2 \\ 0 & \text{sinon} \end{cases}$$

4.12 1. L'argument de l'exponentielle dans la densité peut aussi s'écrire :

$$-\frac{(1+a)(1+4a)}{2a} \left[\frac{x^2}{1+4a} + 2 \frac{1+2a}{(1+4a)(1+a)} xy + \frac{y^2}{1+a} \right]$$

L'exemple 4.13 montre que le couple (X, Y) suit une loi normale, donc les lois marginales de X et Y sont des lois normales centrées de variances respectives $\sigma_X^2 = 1+4a$ et $\sigma_Y^2 = 1+a$. Enfin, par identification on voit que :

$$-\frac{\rho}{\sigma_X \sigma_Y} = \frac{1+2a}{(1+4a)(1+a)}$$

donc $\rho = -\frac{1+2a}{\sigma_X \sigma_Y}$ et par conséquent $Cov(X, Y) = -1-2a$. Les variables X et Y sont donc dépendantes.

2. La définition d'un vecteur normal implique que la combinaison linéaire $X+Y$ suive une loi normale, avec :

$$E(X+Y) = 0 \text{ et } V(X+Y) = V(X) + 2Cov(X, Y) + V(Y) = a$$

On en déduit que Z suit une loi χ_1^2 étant le carré d'une v.a. de loi normale centrée et réduite.

4.13 La matrice Σ est symétrique réelle, régulière et définie positive, donc il existe une matrice orthogonale S (i.e. telle que $S^{-1} = {}^t S$) telle que $S^{-1} \Sigma S = D$, où D est une matrice diagonale d'éléments $\lambda_i, 1 \leq i \leq n$. Ainsi :

$$Z = {}^t(X - \mu) \Sigma^{-1} (X - \mu) = {}^t(X - \mu) S D^{-1} S^{-1} (X - \mu)$$

Le vecteur $Y = S^{-1}(X - \mu) = {}^t S(X - \mu)$ suit une loi normale centrée, de matrice variances-covariances $V(Y) = S^{-1} \Sigma S = D$ qui est diagonale, donc ses composantes $Y_i, 1 \leq i \leq n$, sont indépendantes, de loi $N(0, \sqrt{\lambda_i})$. On obtient alors :

$$Z = {}^t Y D^{-1} Y = \sum_{i=1}^n \frac{Y_i^2}{\lambda_i}$$

qui est la somme des carrés de n v.a. normales indépendantes centrées et réduites, donc suit une loi χ_n^2 .

On aurait pu également introduire la matrice $A = D^{-1/2} S^{-1}$, où la matrice $D^{-1/2}$ est la matrice diagonale d'éléments $1/\sqrt{\lambda_i}$, le vecteur $U = D^{-1/2} Y = A(X - \mu)$ suivant

une loi normale standard, avec $Z = {}^t U U = \sum_{i=1}^n U_i^2 \rightsquigarrow \chi_n^2$. Par ailleurs, si on pose $B = A^{-1}$:

$$B^t B = S D^{1/2} D^{1/2} S^{-1} = S D S^{-1} = \Sigma$$

donc la matrice Σ peut s'écrire sous la forme $B^t B$ et le vecteur X sous la forme $X = B U + \mu$, avec $U \rightsquigarrow \mathcal{N}_n(0, I_n)$.

4.14 La variable U est une combinaison linéaire de deux v.a. normales indépendantes et centrées, donc suit aussi une loi normale centrée, de variance $V(U) = \frac{1}{2} [V(X_1) + V(X_2)] = 1$.

De même, la v.a. $U' = \frac{1}{\sqrt{2}}(X_1 + X_2)$ suit une loi normale centrée réduite. D'autre part :

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} X_1 + X_2 \\ X_2 - X_1 \end{pmatrix} = \begin{pmatrix} U' \\ U \end{pmatrix}$$

suit une loi normale dans \mathbb{R}^2 , comme image linéaire du vecteur normal de composantes X_1 et X_2 . Les composantes de ce vecteur normal sont telles que :

$$Cov(U, U') = E(U U') = \frac{1}{2} [E(X_2^2) - E(X_1^2)] = 0$$

donc elles sont indépendantes. Ainsi, $V = U'^2/U^2$ est le rapport de deux lois χ_1^2 indépendantes, donc suit une loi de Fisher-Snedecor $F(1, 1)$ (cf. chap. 5, 3.1). Le rapport $W = U'/\sqrt{U^2}$ suit une loi de Student (cf. chap. 5, 3.1) à un degré de liberté, c'est-à-dire la loi de Cauchy.

Si X est le vecteur normal de composantes X_1, X_2 et X_3 , on peut écrire $Y = {}^t X A X$ en ayant posé :

$$A = \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$$

La matrice A étant symétrique et idempotente ($A^2 = A$), avec $\text{rang } A = \text{trace } A = 2$, on en conclut que Y suit une loi χ_2^2 .

D'autre part, $Z = {}^t X X - Y = {}^t X (I - A) X \rightsquigarrow \chi_1^2$ car $I - A$ est une matrice symétrique idempotente, de rang égal à sa trace, c'est-à-dire un. De plus : $A(I - A) = A - A^2 = A - A = 0$, donc $Y = {}^t X A X$ et $Z = {}^t X (I - A) X$ sont indépendantes.

Introduction

Si le hasard conduit à un résultat non prévisible, l'observation de plusieurs résultats d'une même expérience aléatoire permettra cependant de choisir judicieusement le modèle aléatoire à retenir. En jetant un dé plusieurs fois consécutivement et en notant le résultat à l'issue de chaque jet, nous obtenons une suite de nombres entiers compris entre un et six, obtenus dans les mêmes conditions et de façon indépendante, que nous pourrions appeler **échantillon** de la loi associée à un jet de dé, qui est on le sait la loi uniforme sur l'ensemble $\{1, 2, 3, 4, 5, 6\}$. Si on calcule la fréquence (pourcentage) observée de chacun de ces chiffres, pour un nombre suffisamment élevé de lancers, on obtiendra une distribution empirique (c'est-à-dire obtenue par l'observation) de valeurs proches les unes des autres et proches de la valeur 16,7 %. Ceci nous oriente donc vers la loi théorique qui est la loi uniforme attribuant la même probabilité $1/6$ aux chiffres 1, 2, 3, 4, 5 et 6. C'est donc à partir d'observations qui vont constituer ce qu'on appelle un échantillon qu'on pourra déterminer la distribution empirique nous permettant de retenir la distribution théorique qui lui ressemble le plus, dans la liste des lois usuelles vues au chapitre 3. Les tests d'adéquation nous donnent des critères précis permettant de retenir une loi théorique pour le modèle, qui paraisse raisonnable compte tenu des observations recueillies.

Objectifs

Introduire la notion d'échantillon.

Définir les lois de Student et de Fisher-Snedecor associées à un échantillon gaussien.

Présenter deux tests d'adéquation à une loi donnée.

Concepts clés

Échantillon

Loi empirique

Moments empiriques

Test du khi-deux

Test de Kolmogorov-Smirnov

1 Échantillon d'une loi

On appelle échantillon de **taille** n d'une loi de probabilité P , une suite (X_1, \dots, X_n) de variables aléatoires indépendantes et de même loi de probabilité P . On dit que P est la loi **parente** de l'échantillon. Pour tout ω de Ω , la réalisation correspondante de l'échantillon se note (x_1, \dots, x_n) , ayant posé $x_i = X_i(\omega)$, $1 \leq i \leq n$. Cette réalisation particulière étant parfois aussi appelée échantillon, on peut alors préciser dans le premier

cas en disant que (X_1, \dots, X_n) est un **échantillon aléatoire**. Si X est une v.a. de loi P , on dit également qu'il s'agit d'un échantillon de la variable aléatoire X .

À partir d'un échantillon (X_1, \dots, X_n) de X , nous définirons la **loi de probabilité empirique** P_n , qui sera la distribution uniforme sur l'ensemble fini des valeurs (X_1, \dots, X_n) , c'est-à-dire qui attribue la même masse de probabilité $1/n$ à chacun des points X_i , $1 \leq i \leq n$, et qui peut se noter sous la forme :

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

où δ_a est la masse de Dirac au point a . Ainsi, pour tout intervalle I de \mathbb{R} , sa probabilité est égale au pourcentage de points de l'échantillon qui appartiennent à cet intervalle :

$$P_n(I) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_I(X_i)$$

Cette loi de probabilité admet une fonction de répartition, notée F_n , appelée **fonction de répartition empirique** et définie pour tout réel x par :

$$F_n(x) = \int_{-\infty}^x dP_n = P_n(]-\infty, x]) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i)$$

qui représente le pourcentage de points observés qui sont situés avant x . Les v.a. X_1, \dots, X_n étant indépendantes, la v.a. $nF_n(x) = \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i)$ suit une loi binômiale de paramètres n et $F(x)$ puisque c'est la somme de n variables indépendantes de Bernoulli, de paramètre $P(\mathbf{1}_{]-\infty, x]}(X_i) = 1) = P(X_i \in]-\infty, x]) = F(x)$. Ainsi :

$$E[F_n(x)] = F(x) \text{ et } V[F_n(x)] = \frac{1}{n} F(x)[1 - F(x)] \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

2 Moments empiriques

La loi de probabilité empirique P_n admet des moments de tous ordres que l'on nomme **moments empiriques** ou moments de l'échantillon et qui seront tous des variables aléatoires.

2.1 Moyenne empirique

La moyenne de l'échantillon aléatoire est la moyenne de la loi empirique, c'est-à-dire l'espérance d'une loi uniforme discrète qui affecte le même poids $1/n$ à chacune des valeurs X_i et notée :

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

Cette v.a. admet comme espérance :

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X) = m$$

et comme variance :

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$$

Ainsi, \bar{X}_n a la même moyenne que X , mais avec une variance qui est divisée par la taille de l'échantillon. Les autres moments centrés de \bar{X}_n ont pour expressions :

$$E(\bar{X}_n - m)^3 = \frac{\mu_3}{n^2} \quad \text{et} \quad E(\bar{X}_n - m)^4 = \frac{\mu_4}{n^3} + 3 \frac{n-1}{n^3} \sigma^4$$

2.2 Variance empirique

La variance de la loi empirique est celle d'une loi de moyenne \bar{X}_n et qui attribue le même poids $1/n$ à chacune des valeurs X_i , soit :

$$S_n'^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Pour calculer les moments de cette variable aléatoire, nous allons l'exprimer à l'aide des v.a. centrées en écrivant $X_i - \bar{X}_n = (X_i - m) - (\bar{X}_n - m)$:

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - m)^2 + n(\bar{X}_n - m)^2 - 2(\bar{X}_n - m) \sum_{i=1}^n (X_i - m)$$

d'où on déduit :

$$S_n'^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2 - (\bar{X}_n - m)^2$$

Ceci permet d'obtenir aisément l'espérance de ce moment empirique :

$$E(S_n'^2) = E(X - m)^2 - E(\bar{X}_n - m)^2 = V(X) - V(\bar{X}_n) = \frac{n-1}{n} \sigma^2$$

La valeur moyenne de la variance empirique n'est pas exactement égale à la variance théorique, c'est pourquoi on introduit la variance empirique modifiée (ou corrigée) où on divise non pas par le nombre de termes de la somme, mais par le nombre de termes indépendants, puisqu'ils sont liés par la relation $\sum_{i=1}^n (X_i - \bar{X}_n) = 0$:

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

on a alors $E(S_n^2) = \sigma^2$. Le calcul de la variance est plus délicat et on obtient :

$$V(S_n^2) = \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4$$

où $\mu_4 = E(X - m)^4$. Pour la variance empirique on obtient :

$$V(S_n'^2) = \frac{(n-1)^2}{n^3} (\mu_4 - \sigma^4) + 2\frac{n-1}{n^3}\sigma^4$$

On peut également préciser le lien existant entre moyenne et variance empiriques par le calcul de la covariance :

$$Cov(\bar{X}_n, S_n'^2) = \frac{n-1}{n^2}\mu_3 \quad \text{et} \quad Cov(\bar{X}_n, S_n^2) = \frac{\mu_3}{n}$$

Ainsi, pour toute loi symétrique, on a : $Cov(\bar{X}_n, S_n^2) = 0$.

2.3 Moments empiriques

Le moment non centré d'ordre $k \in \mathbb{N}^*$ de cette loi empirique est :

$$m_{kn} = \frac{1}{n} \sum_{i=1}^n X_i^k$$

avec $E(m_{kn}) = E(X^k) = m_k$ et :

$$V(m_{kn}) = \frac{1}{n} V(X^k) = \frac{1}{n} [E(X^{2k}) - E^2(X^k)] = \frac{1}{n} (m_{2k} - m_k^2)$$

Le moment centré d'ordre $k \in \mathbb{N}^*$ de cette loi empirique est :

$$\mu_{kn} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k$$

3 Échantillon d'une loi normale

Nous allons étudier le cas particulier où la v.a. X suit une loi normale $N(m, \sigma)$.

La moyenne empirique étant une combinaison linéaire de v.a. normales indépendantes suit aussi une loi normale :

$$\bar{X}_n \rightsquigarrow N\left(m, \frac{\sigma}{\sqrt{n}}\right)$$

La variance empirique est construite à partir des variables :

$$X_i - \bar{X}_n = -\frac{1}{n} \sum_{j \neq i} X_j + \left(1 - \frac{1}{n}\right) X_i$$

qui sont des combinaisons linéaires de variables normales indépendantes donc sont aussi des variables normales, d'espérance $E(X_i - \bar{X}_n) = E(X_i) - E(\bar{X}_n) = 0$ et de variance $V(X_i - \bar{X}_n) = E(X_i - \bar{X}_n)^2 = E(S_n'^2)$ car toutes ces variables ont la même loi. La variable $nS_n'^2$ est la somme des carrés de n variables normales centrées qui sont liées par la relation $\sum_{i=1}^n (X_i - \bar{X}_n) = 0$ et on peut démontrer que :

$$n \frac{S_n'^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$$

On retrouve bien le résultat général $E(S_n'^2) = \frac{n-1}{n}\sigma^2$ et on obtient ici, d'après les moments de la loi du khi-deux, $V(S_n'^2) = 2\frac{n-1}{n^2}\sigma^4$. Pour la variance empirique modifiée : $E(S_n^2) = \sigma^2$ et $V(S_n^2) = \frac{2\sigma^4}{n-1}$. La loi normale étant symétrique, moyenne et variance empiriques sont bien entendu non corrélées, mais de plus elles sont indépendantes et c'est une propriété caractéristique de la loi normale. Ce résultat s'énonce de la façon suivante :

Théorème 5.1

Théorème de Fisher

Les v.a. X_1, \dots, X_n forment un échantillon d'une loi normale si et seulement si les v.a. \bar{X}_n et $S_n'^2$ sont indépendantes.

3.1 Loi de Student

Nous avons vu que :

$$\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \rightsquigarrow N(0, 1)$$

Dans le cas où σ est un paramètre inconnu, on peut le remplacer par l'écart type empirique modifié, ce qui amène à considérer la variable :

$$\sqrt{n} \frac{\bar{X}_n - m}{S_n} = \frac{(\bar{X}_n - m) / (\sigma / \sqrt{n})}{\sqrt{S_n^2 / \sigma^2}}$$

Le numérateur suit une loi normale centrée réduite et le dénominateur est la racine carrée de la variable :

$$\frac{S_n^2}{\sigma^2} = \frac{(n-1)S_n'^2 / \sigma^2}{n-1}$$

qui est donc une variable de loi χ_{n-1}^2 , divisée par son nombre de degrés de liberté. D'après le théorème de Fisher, le numérateur et le dénominateur sont des v.a. indépendantes et leur rapport définit une nouvelle loi de probabilité, usuelle en statistique, appelée **loi de Student** à $n - 1$ degrés de liberté. Au-delà de ce cas particulier lié à un échantillon gaussien, on peut définir cette loi à partir d'une v.a. U de loi $N(0, 1)$ et d'une autre v.a. indépendante Y de loi χ_n^2 . Le rapport $U / \sqrt{Y/n}$ suit une loi de Student à n degrés de liberté, notée T_n . Comme le numérateur U suit une loi symétrique par rapport à 0, il en est de même de T_n , avec $E(T_n) = 0$ pour $n > 1$. On obtient aussi $V(T_n) = \frac{n}{n-2}$ pour $n > 2$.

Le cas particulier $n = 1$ correspond au rapport de deux variables normales indépendantes qui suit une **loi de Cauchy**, qui possède la particularité assez rare de n'admettre aucun moment. En effet, sa densité est $\frac{1}{\pi(1+x^2)}$ et le calcul de l'espérance conduit à l'intégrale généralisée :

$$E(X) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x dx}{1+x^2} = \frac{1}{2\pi} [\ln(1+x^2)]_{-\infty}^{+\infty}$$

intégrale divergente car cette quantité n'est pas définie, donc l'espérance d'une loi de Cauchy n'existe pas.

3.2 Loi de Fisher-Snedecor

En présence de deux échantillons (X_1, \dots, X_n) et (Y_1, \dots, Y_m) auxquels sont associées les variances empiriques $S_n'^2$ et $S_m'^2$, on peut se poser la question de savoir s'ils proviennent de deux lois normales ayant la même variance, et pour cela former le rapport $S_n'^2/S_m'^2$. Si effectivement ces deux lois ont la même variance, ce rapport de deux lois du khi-deux indépendantes, réduites (divisées) par leur nombre de degrés de liberté (car s'écrivant $[U/(n-1)]/[V/(m-1)]$ avec $U = (n-1)S_n'^2/\sigma^2$ et $V = (m-1)S_m'^2/\sigma^2$, définit une nouvelle loi usuelle en statistique, appelée **loi de Fisher-Snedecor**. Plus généralement, si U et V sont deux v.a. indépendantes de lois respectives χ_n^2 et χ_m^2 , alors le rapport $(U/n)/(V/m)$ suit une loi de Fisher-Snedecor à n et m degrés de liberté, notée $F(n, m)$. On obtient comme moments :

$$E[F(n, m)] = \frac{m}{m-2}, m > 2 \text{ et } V[F(n, m)] = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}, m > 4$$

On peut remarquer que : $\frac{n}{m}F(n, m) \rightsquigarrow \beta_{II}\left(\frac{n}{2}, \frac{m}{2}\right)$ et $T_n^2 \rightsquigarrow F(1, n)$.

Les fractiles de cette loi sont tabulés (table 7) pour certains couples (n, m) ; si le couple (n, m) cherché ne figure pas, mais si le couple (m, n) est dans la table, on utilise le fait que $(V/m)/(U/n) \rightsquigarrow F(m, n)$.

Exemple 5.1

Pour α donné, cherchons le fractile $f_\alpha(n, m)$ de $F(n, m)$ défini par :

$$\begin{aligned}\alpha &= P \left\{ F(n, m) < f_\alpha(n, m) \right\} = P \left\{ \frac{U/n}{V/m} < f_\alpha(n, m) \right\} \\ &= P \left\{ \frac{V/m}{U/n} > \frac{1}{f_\alpha(n, m)} \right\} = P \left\{ F(m, n) > \frac{1}{f_\alpha(n, m)} \right\} \\ &= 1 - P \left\{ F(m, n) < \frac{1}{f_\alpha(n, m)} \right\}\end{aligned}$$

on voit ainsi que $1 - \alpha = P \{ F(m, n) < 1/f_\alpha(n, m) \}$ et par conséquent $1/f_\alpha(n, m)$ est le fractile $f_{1-\alpha}(m, n)$, d'ordre $1 - \alpha$, de la loi $F(m, n)$. Appliquons ceci à la recherche du fractile d'ordre 0,025 de la loi $F(15, 10)$, défini par :

$$0,025 = P \left\{ F(15, 10) < f_{0,025}(15, 10) \right\} = P \left\{ F(10, 15) > \frac{1}{f_{0,025}(15, 10)} \right\}$$

d'où on déduit $0,975 = P \left\{ F(10, 15) < 1/f_{0,025}(15, 10) \right\}$; on lit dans la table 7 la valeur du fractile $f_{0,975}(10, 15) = 3,06$ d'où la valeur cherchée :

$$f_{0,025}(15, 10) = \frac{1}{3,06} = 0,327.$$

4 Tests d'adéquation

L'examen de la loi de probabilité empirique associée à un échantillon dont la loi parente est inconnue permet de choisir parmi les lois usuelles celle qui lui « ressemble » le plus. Si notre choix s'oriente vers une certaine loi P de fonction de répartition (f.r.) F , on pourra retenir l'hypothèse que l'échantillon provient de cette loi si la distance entre la f.r. théorique F et la f.r. empirique F_n est faible. Ayant fait le choix d'une certaine distance d entre fonctions de répartition, on se fixera une règle de décision qui s'énonce ainsi : « Si l'événement $d(F_n, F) < C$ est réalisé, alors je retiens l'hypothèse qu'il s'agit d'un échantillon de la loi de f.r. F ». On peut cependant se tromper en rejetant cette hypothèse alors que F est bien la f.r. des variables de l'échantillon ; cette erreur se produit avec une probabilité qui est $\alpha = P \{ d(F_n, F) > C \}$. Si on veut que ce risque d'erreur soit faible, on fixera une valeur α faible à cette probabilité (par exemple 5 % ou 1 %) et cette valeur permettra alors de préciser la valeur de la constante C qui apparaît dans la règle de décision, si on connaît la loi de probabilité de la v.a. $d(F_n, F)$. Nous aurons ainsi réalisé un test d'adéquation, ou d'ajustement, entre une loi théorique donnée et une loi empirique associée à un échantillon d'observations. La fixation du risque α déterminera alors la valeur du seuil d'acceptation, ou seuil critique C .

Nous allons présenter maintenant deux tests, associés à deux distances entre f.r., permettant de déterminer la loi approchée de la v.a. $d(F_n, F)$ pour toute f.r. F , le premier étant plutôt destiné aux lois discrètes et le second réservé aux lois continues.

4.1 Test du khi-deux

Ce test est à retenir si les données sont discrètes, avec des valeurs possibles notées x_i , de probabilité p_i pour $1 \leq i \leq k$, ou si les données individuelles ne sont pas fournies, mais ont été réparties en classes (a_i, a_{i+1}) dont les fréquences théoriques sont calculées à partir de la loi théorique postulée :

$$p_i = P \{ X \in (a_i, a_{i+1}) \} = F(a_{i+1}) - F(a_i)$$

Si N_i est le nombre (aléatoire) d'observations x_i , ou appartenant à la classe (a_i, a_{i+1}) , nous allons le comparer à l'effectif théorique qui est np_i . La distance euclidienne classique entre F_n , représentée par les k effectifs observés N_i , et la f.r. F , représentée par les k effectifs théoriques np_i , serait $\sum_{i=1}^k (N_i - np_i)^2$.

Cependant, comme cette distance ne permet pas de déterminer la loi asymptotique de cette v.a., on préfère retenir une autre distance. Cette dernière sera déterminée à partir de la remarque que les v.a. N_i suivent des lois binômiales de paramètres n et p_i et que les variables centrées $(N_i - np_i) / \sqrt{np_i}$ convergent vers la loi $N(0, \sqrt{1 - p_i})$. On retient donc la distance :

$$d(F_n, F) = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$$

et cette somme de carrés de v.a. centrées qui sont asymptotiquement normales et liées par la relation $\sum_{i=1}^k (N_i - np_i) = 0$ converge vers une loi χ_{k-1}^2 . La valeur de C est alors déterminée approximativement, en utilisant cette loi asymptotique, comme le fractile d'ordre $1 - \alpha$ de la loi du khi-deux à $k - 1$ degrés de liberté. Cette approximation est justifiée si n est assez grand et p_i pas trop petit, avec comme règle empirique $np_i \geq 5$. Si ce n'est pas le cas à cause d'une valeur de p_i trop petite on doit regrouper des classes (ou des valeurs) contiguës. Pour le calcul de la distance, il est préférable d'utiliser la formule développée :

$$d(F_n, F) = \sum_{i=1}^k \frac{N_i^2}{np_i} - n$$

Exemple 5.2

Ayant demandé à dix personnes de fournir chacune dix chiffres choisis au hasard, on souhaite savoir si effectivement les cent chiffres obtenus forment bien une distribution au hasard. L'hypothèse à tester ici est donc celle d'une loi uniforme discrète sur l'ensemble des dix chiffres $\{0, 1, \dots, 9\}$, donc avec des probabilités égales pour chacune de ces valeurs, $p_i = 1/10$ pour $i = 0, 1, \dots, 9$. Le tableau ci-après contient la distribution empirique (observée) sur la deuxième ligne et la distribution théorique sur la troisième.

x_i	0	1	2	3	4	5	6	7	8	9
N_i	10	8	9	14	8	9	11	9	12	10
np_i	10	10	10	10	10	10	10	10	10	10

On obtient comme distance :

$$d(F_n, F) = \frac{1}{10} (0 + 4 + 1 + 16 + 4 + 1 + 4 + 1 + 4 + 0) = 3,2$$

le fractile d'ordre 0,95 de la loi du khi-deux à neuf degrés de liberté est $C = 16,9$ donc la valeur de la distance est très inférieure et on peut accepter l'hypothèse d'une répartition uniforme.

Cependant, dans beaucoup de cas, si le type de loi est précisé, la loi dépend de paramètres dont la valeur n'est pas spécifiée et qu'il va falloir estimer pour pouvoir calculer les fréquences théoriques p_i . Si on doit estimer r paramètres, cela diminue d'autant le nombre de degrés de liberté qui devient alors $k - 1 - r$.

Exemple 5.3

On a observé pendant deux heures le nombre de voitures arrivées par minute à un poste de péage. Si X est la v.a. représentant le nombre de voitures arrivant dans une minute à ce poste de péage, on fait l'hypothèse qu'elle suit une loi de Poisson. Le tableau ci-après contient les valeurs observées x_i de cette variable et le nombre d'observations correspondantes N_i . Pour calculer les probabilités théoriques $p_i = P(X = x_i)$ il faut spécifier complètement la loi, c'est-à-dire indiquer le paramètre de cette loi de Poisson. Le calcul de la moyenne empirique donne $\bar{x} = 3,7$ et la variance empirique a pour valeur 4,41. On retient alors la valeur entière 4 comme paramètre de cette loi de Poisson. Les valeurs de np_i sont alors obtenues par lecture de la table 4 et arrondies à l'entier le plus proche, en vérifiant bien que le total est égal à $n = 120$; par exemple $nP(X = 3) = 120 \times 0,1954 = 23,448$ est arrondi à 23.

x_i	0	1	2	3	4	5	6	7	8	9	10	11
N_i	4	9	24	25	22	18	6	5	3	2	1	1
np_i	2	9	18	23	23	19	13	7	3	2	1	0

Les valeurs des effectifs théoriques inférieures à 5 nécessitent de regrouper les deux premières et les quatre dernières valeurs, ramenant à 8 le nombre de valeurs retenues, soit $8 - 1 - 1 = 6$ degrés de liberté puisqu'un paramètre a été estimé. Le fractile d'ordre 0,95 de la loi du khi-deux à 6 degrés de liberté est $C = 12,6$. On obtient $d(F_n, F) = 7,14$ ce qui conduit donc à accepter l'hypothèse que X suit une loi de Poisson de paramètre 4.

4.2 Test de Kolmogorov-Smirnov

Dans le cas d'une variable continue pour laquelle on dispose des données individuelles, il est préférable d'utiliser toute l'information disponible et de ne pas regrouper les observations en classes. On retient alors la distance de Kolmogorov, ou distance de la convergence uniforme, définie par :

$$K_n = d(F_n, F) = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

Là encore, on retiendra l'hypothèse que la loi parente admet F comme f.r. si cette distance est faible, c'est-à-dire plus précisément si l'événement $d(F_n, F) < C$ est réalisé. La valeur de C sera déterminée par la fixation du risque d'erreur $\alpha = P\{d(F_n, F) > C\}$ et en utilisant la loi limite de la v.a. $\sqrt{n}K_n$ qui admet pour f.r. la fonction K définie pour $x > 0$ par :

$$K(x) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 x^2} = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 x^2}$$

Les valeurs de K sont tabulées, permettant de déterminer les fractiles de la loi. Les valeurs de C sont données en fonction de α dans la table suivante :

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
5	0,509	0,563	0,669
10	0,369	0,409	0,486
15	0,304	0,338	0,404
20	0,265	0,294	0,352
25	0,238	0,264	0,317
30	0,218	0,242	0,290
40	0,189	0,210	0,252
$n > 40$	$1,22/\sqrt{n}$	$1,36/\sqrt{n}$	$1,63/\sqrt{n}$

Pour le calcul pratique de cette distance, on utilise la définition de F_n faisant intervenir l'échantillon ordonné $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. L'expression de $F_n(x) = P_n\{]-\infty, x[\}$ s'écrit alors :

$$F_n(x) = \begin{cases} 0 & \text{si } x \leq X_{(1)} \\ \frac{i-1}{n} & \text{si } X_{(i-1)} < x \leq X_{(i)} \\ 1 & \text{si } x > X_{(n)} \end{cases}$$

On calcule au préalable les statistiques :

$$d^+(F_n, F) = \sup_{x \in \mathbb{R}} [F_n(x) - F(x)] = \max_{1 \leq i \leq n} \left\{ \frac{1}{n} - F(X_{(i)}) \right\}$$

$$d^-(F, F_n) = \sup_{x \in \mathbb{R}} [F(x) - F_n(x)] = \max_{1 \leq i \leq n} \left\{ F(X_{(i)}) - \frac{i-1}{n} \right\}$$

car F_n est constante sur chacun des intervalles délimités par les points de l'échantillon ordonné :

$$\begin{aligned} \sup_{x \in]X_{(i)}, X_{(i+1)}]} [F_n(x) - F(x)] &= \frac{i}{n} - \inf_{X_{(i)} < x \leq X_{(i+1)}} F(x) = \frac{i}{n} - F(X_{(i)} + 0) \\ &= \frac{i}{n} - F(X_{(i)}) \end{aligned}$$

On calcule ensuite :

$$d(F_n, F) = \max \{d^+(F_n, F), d^+(F, F_n)\}$$

Exemple 5.4

On se pose la question de savoir si les données suivantes peuvent provenir d'une loi normale centrée et réduite :

6,42; -5,23; -1,25; 0,12; -0,01; -1,02; 18,54; 0,06; -7,64; 2,85; -1,84; 0,74; -0,65; 0,24

Disposant des données individuelles, nous allons utiliser le test de Kolmogorov-Smirnov et pour cela ordonner les observations par valeurs croissantes $x_{(i)}$ puis calculer les valeurs $\Phi(x_{(i)})$ à l'aide de la f.r. Φ de la loi $N(0, 1)$. Les valeurs permettant le calcul de la distance entre loi empirique et loi théorique figurent dans le tableau suivant :

$x_{(i)}$	$\Phi(x_{(i)})$	$\frac{i}{n} - \Phi(x_{(i)})$	$\Phi(x_{(i)}) - \frac{i-1}{n}$
-7,64	0	0,0667	0
-5,23	0	0,1333	-0,0667
-1,84	0,0329	0,1671	-0,1004
-1,25	0,1056	0,1611	-0,0944
-1,02	0,1539	0,1794	-0,1128
-0,65	0,2578	0,1422	-0,0755
-0,01	0,4960	-0,0293	0,0960
0,06	0,5239	0,009	0,0572
0,12	0,5478	0,0616	0,0145
0,24	0,5948	0,0719	-0,005
0,74	0,7704	-0,0371	0,1037
2,18	0,9854	-0,1854	0,2521
2,85	0,9978	-0,1311	0,1978
6,42	1	-0,0667	0,1333
18,54	1	0	0,0667

On lit dans les deux dernières colonnes du tableau les valeurs maximales $d^+(F_n, F) = 0,1794$ et $d^+(F, F_n) = 0,2521$ d'où on déduit $d(F_n, F) = 0,2521$. Pour un risque $\alpha = 0,10$ on lit dans la table 5.1, pour $n = 15$, le seuil critique $C = 0,304$ donc on accepte l'hypothèse d'une loi parente $N(0, 1)$, bien que la valeur de la distance soit proche du seuil de rejet de cette hypothèse.

5 Compléments

5.1 Statistique d'ordre

La statistique d'ordre associée à un échantillon (X_1, \dots, X_n) d'une loi de f.r. F et de densité f , les observations classées ordre croissant : $X_{(1)} \leq \dots \leq X_{(n)}$. Elle a pour densité :

$$n! \mathbf{1}_{C_n}(x_1, \dots, x_n) \prod_{i=1}^n f(x_i)$$

où $C_n = \{(x_1, \dots, x_n) / x_1 < \dots < x_n\}$.

Les valeurs extrêmes sont $X_{(1)}$, de f.r. définie par $G_1(x) = 1 - [1 - F(x)]^n$, de densité $g_1(x) = nf(x)[1 - F(x)]^{n-1}$, et $X_{(n)}$, de f.r. définie par $G_n(x) = F^n(x)$, de densité $g_n(x) = nf(x)F^{n-1}(x)$.

La loi de l'**étendue** (range) $W_n = X_{(n)} - X_{(1)}$ se déduit de la loi du couple $(X_{(1)}, X_{(n)})$, de f.r. définie pour $x < y$ par :

$$G_{1n}(x, y) = F^n(x) - [F(y) - F(x)]^n$$

au moyen du changement de variable $x = x$ et $w = y - x$. On obtient comme f.r. pour $w > 0$:

$$H(w) = n \int_{\mathbb{R}} [F(x+w) - F(x)]^{n-1} f(x) dx$$

et comme densité :

$$h(w) = n(n-1) \int_{\mathbb{R}} [F(x+w) - F(x)]^{n-2} f(x+w) f(x) dx$$

Si le support de la loi est \mathbb{R} , l'expression des f.r. des valeurs extrêmes montre que :

$$X_{(1)} \xrightarrow{p.co.} -\infty \quad \text{et} \quad X_{(n)} \xrightarrow{p.co.} +\infty$$

Si le support est l'intervalle (a, b) on obtient de même :

$$X_{(1)} \xrightarrow{p.co.} a \quad \text{et} \quad X_{(n)} \xrightarrow{p.co.} b$$

Si la loi admet une espérance, on établit également que :

$$\frac{X_{(1)}}{n} \xrightarrow[p]{} 0 \quad \text{et} \quad \frac{X_{(n)}}{n} \xrightarrow[p]{} 0$$

(voir le chapitre 6 pour la définition de ces convergences).

Pour obtenir la loi de $X_{(k)}$, $1 < k < n$, il suffit de remarquer que le nombre $N_n(x)$ de variables de l'échantillon qui sont inférieures au réel x vérifie :

$$N_n(x) \geq k \Leftrightarrow X_{(k)} < x$$

Comme $N_n(x) \rightsquigarrow \mathcal{B}(n, F(x))$ on en déduit la f.r. G_k de $X_{(k)}$ définie par :

$$G_k(x) = \sum_{j=k}^n \binom{n}{j} F^j(x) [1 - F(x)]^{n-j}$$

et sa densité peut être obtenue de façon « intuitive » comme probabilité qu'un élément de l'échantillon soit compris entre x et $x + dx$, soit $f(x)$, que $k - 1$ soient situés avant x , soit $F^{k-1}(x)$, que $n - k$ soient situés après x , soit $[1 - F(x)]^{n-k}$, avec un nombre de permutations $n!$ des variables de l'échantillon qui doit être divisé par le nombre $(k - 1)!$ de variables avant et $(n - k)!$ de variables après, soit :

$$g_k(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x) [1 - F(x)]^{n-k} f(x)$$

On peut également exprimer la f.r. de $X_{(k)}$ au moyen de l'intégrale bêta incomplète :

$$G_k(x) = \frac{1}{B(k, n - k + 1)} \int_0^{F(x)} t^{k-1} (1 - t)^{n-k} dt$$

et retrouver ainsi par dérivation l'expression de la densité.

En utilisant le résultat de l'exercice 2.13 qui a montré que $F(X) \rightsquigarrow U([0, 1])$ et en faisant le changement de variable $y = F(x)$, on établit alors que :

$$F(X_{(k)}) \rightsquigarrow \beta_I(k, n - k + 1).$$

5.2 Théorème de Fisher

À partir d'un échantillon de n v.a. de la loi $N(0, 1)$, on construit le vecteur X de coordonnées (X_1, \dots, X_n) dans la base canonique (e_1, \dots, e_n) de \mathbb{R}^n . Sa projection sur le vecteur bissecteur $u = (1, \dots, 1)$ est le vecteur $\bar{X} = (\bar{X}_n, \dots, \bar{X}_n)$. Si on effectue un changement de base orthonormée au moyen de la matrice orthogonale P , en choisissant dans la nouvelle base (f_1, \dots, f_n) le vecteur unitaire $f_n = \frac{u}{\sqrt{n}}$, le vecteur des nouvelles coordonnées de X est $Y = {}^t P X$, avec $Y_n = \sqrt{n} \bar{X}_n$.

On obtient ainsi :

$$\begin{aligned} X - \bar{X} &= X - Y_n f_n = \sum_{j=1}^{n-1} Y_j f_j = \sum_{i=1}^n X_i e_i - \sqrt{n} \bar{X}_n \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \\ &= \sum_{i=1}^n (X_i - \bar{X}_n) e_i \end{aligned}$$

D'autre part :

$$V(Y) = V({}^t P X) = {}^t P V(X) P = {}^t P P = I_n$$

donc Y_n est indépendant de Y_1, \dots, Y_{n-1} et par conséquent de $\sum_{j=1}^{n-1} Y_j^2 = \|X - \bar{X}\|^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$, ce qui établit l'indépendance de \bar{X}_n avec $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ et le fait que cette dernière variable suit la loi χ_{n-1}^2 , comme somme des carrés des $n - 1$ variables Y_1, \dots, Y_{n-1} qui sont indépendantes et de même loi $N(0, 1)$.

- ❶ Un échantillon d'une loi P , ou d'une v.a. X , est un ensemble de v.a. indépendantes et de même loi P , ou de même loi que la v.a. X .
- ❷ La loi uniforme qui attribue la même probabilité $\frac{1}{n}$ à chacune des n observations d'un échantillon est appelée loi empirique et ses moments, moments empiriques. On étudie plus particulièrement la moyenne et la variance empiriques. Les moments empiriques sont des fonctions de l'échantillon, donc à ce titre sont aussi des v.a. qui admettent des moments théoriques.
- ❸ À partir de la fonction de répartition empirique on peut effectuer un test permettant de savoir si on peut retenir pour le modèle théorique une certaine loi de probabilité donnée. Dans le cas de données discrètes ou regroupées en classes, on utilise comme test d'ajustement le test du khi-deux. Pour des données individuelles issues d'une loi continue, il est préférable d'utiliser le test d'adéquation de Kolmogorov-Smirnov.



Entraînez-vous

- 5.1** Déterminer les fractiles d'ordres 0,90 et 0,05 de la loi de Student T_{25} .
- 5.2** Déterminer le fractile d'ordre 0,05 de la loi de Fisher-Snedecor $F(30, 20)$.
- 5.3** Le tableau ci-après contient les fréquences absolues N_i d'apparition des entiers x_i de 0 à 9 dans les 10 000 premiers chiffres de la partie décimale du nombre π . On demande de tester l'hypothèse d'une répartition uniforme de ces entiers.

x_i	0	1	2	3	4	5	6	7	8	9
N_i	968	1025	1021	974	1014	1045	1021	970	948	1 014

- 5.4** Pour vérifier si un ensemble de deux dés n'est pas truqué, on les jette 108 fois. La somme des chiffres obtenus a pris les valeurs 6, 9 et 11 respectivement 12, 15 et 8 fois. Peut-on accepter l'hypothèse de l'équilibre de ces dés, en prenant comme risque de première espèce $\alpha = 0,05$?
- 5.5** Le tableau ci-après fournit la distribution du nombre X d'accidents graves par semaine à un carrefour dangereux, relevés pendant une certaine période. On demande de tester l'hypothèse que la loi parente de X est une loi de Poisson de paramètre $\lambda = 2$.

x_i	0	1	2	3	4	5
N_i	5	10	7	4	3	1

- 5.6** Un organisme de défense des consommateurs a prélevé au hasard 200 boîtes de conserve de haricots verts pour tester la validité de l'étiquetage indiquant un poids net égoutté de 560 g. La distribution observée du poids égoutté X en grammes figure dans le tableau suivant :

X	N_i
[530,540[14
[540,545[15
[545,550[29
[550,555[40
[555,560[37
[560,565[27
[565,570[20
[570,580[18

On demande de tester l'hypothèse que X suit une loi $N(555, 10)$.

- 5.7** Peut-on accepter l'hypothèse que les données suivantes proviennent d'une loi $N(1,5; 1)$?
2,51 ; 1,45 ; 0,71 ; 0,25 ; 2,35 ; 0,00 ; 2,06 ; 1,95 ; 0,41 ; 2,75 ; 0,78 ; 1,01 ; 1,75 ; 2,32 ; 1,36.



- 5.1** La table 6 fournit la valeur 1,316 dans la colonne 0,90 et, du fait de la symétrie de cette loi, le fractile d'ordre 0,05 est obtenu dans la colonne 0,95 en changeant le signe, soit $-1,708$.
- 5.2** La table 7 ne comporte pas le couple de degrés de liberté (30, 20) mais comporte le couple (20, 30); nous allons voir comment nous y ramener. Par définition, la loi $F(n, m)$ est la loi du rapport $U = \frac{X/n}{Y/m}$ où X et Y suivent des lois du khi-deux à respectivement n et m degrés de liberté. Par conséquent $\frac{1}{U} = \frac{Y/m}{X/n}$ suit une loi de Fisher-Snedecor $F(m, n)$. La valeur u cherchée est définie par : $0,05 = P(U < u) = P(1/U > 1/u)$, avec U de loi $F(30, 20)$, ou $P(1/U < 1/u) = 0,95$ donc $1/u$ est le fractile d'ordre 0,95 de la loi $F(20, 30)$, valeur lue dans la table 7, soit $1/u = 1,93$ et $u = 0,518$.
- 5.3** Si la répartition des entiers est uniforme, chaque entier x_i a la même probabilité $p_i = 1/10$ et tous les effectifs théoriques ont la même valeur $np_i = 1000$. On calcule la distance du khi-deux $d(F_n, F) = 9,37$ que l'on compare au seuil critique $C = 16,92$ associé au risque $\alpha = 0,05$. Même pour un risque plus élevé $\alpha = 0,10$ le fractile d'ordre 0,90 de la loi χ_9^2 est $C = 14,68$ et on accepte encore l'hypothèse d'une répartition uniforme de ces entiers.
- 5.4** Tester l'équilibre de cet ensemble de deux dés revient à tester l'adéquation de la loi empirique avec la loi théorique en cas d'équilibre. Comme seulement trois valeurs de la somme S sont données, on va regrouper toutes les autres valeurs possibles. Les 4 valeurs théoriques retenues sont donc :

$$p_1 = P(S = 6) = \frac{5}{36}, p_2 = P(S = 9) = \frac{4}{36}, p_3 = P(S = 11) = \frac{2}{36},$$

$$p_4 = 1 - p_1 - p_2 - p_3 = \frac{25}{36}$$

On peut donc constituer le tableau de comparaison des distributions empirique et théorique :

x_i	6	9	11	autre
n_i	12	15	8	73
np_i	15	12	6	75

La distance entre ces deux distributions a pour valeur :

$$D_n = \frac{(12 - 15)^2}{15} + \frac{(15 - 12)^2}{12} + \frac{(8 - 6)^2}{6} + \frac{(73 - 75)^2}{75} = 2,07$$

Le fractile d'ordre 0,95 de la loi du khi-deux à $4 - 1 = 3$ degrés de liberté a pour valeur 7,8 donc on accepte l'hypothèse de l'équilibre de ces dés, en prenant comme risque de première espèce $\alpha = 0,05$.

- 5.5** On calcule les effectifs théoriques np_i arrondis à partir des probabilités de la loi de Poisson $\mathcal{P}(2)$. Pour obtenir des effectifs suffisants on doit regrouper les deux dernières classes ; on obtient le tableau suivant :

x_i	0	1	2	3	4 ou plus
N_i	5	10	7	4	4
np_i	4	8	8	5,5	4,5

On obtient $d(F_n, F) = 1,49$ valeur qui est inférieure au fractile d'ordre 0,95 de la loi χ_4^2 qui vaut $C = 9,49$ donc on accepte l'hypothèse que ces observations forment un échantillon d'une loi de Poisson $\mathcal{P}(2)$.

5.6 Les données étant regroupées en classes, on effectue le test du khi-deux en calculant les effectifs théoriques np_i où $n = 200$ et $p_i = P(X \in [a_i, a_{i+1}[)$ se calcule à partir de la f.r. Φ de la loi $N(0, 1)$ en centrant et réduisant les observations. Par exemple :

$$P(560 \leq X < 565) = P\left(\frac{5}{10} \leq \frac{X - 555}{10} < \frac{10}{10}\right) = \Phi(1) - \Phi(0,5) = 0,1498$$

et $200 \times 0,1498 = 29,96$ est arrondi à 30. Les classes extrêmes $[530,540[$ et $[570,580[$ sont assimilées respectivement aux classe $]-\infty, 540[$ et $[570, +\infty[$.

Le calcul de la distance du khi-deux se fait à partir du tableau suivant :

X	N_i	np_i
$[530,540[$	14	13
$[540,545[$	15	18,5
$[545,550[$	29	30
$[550,555[$	40	38,5
$[555,560[$	37	38,5
$[560,565[$	27	30
$[565,570[$	20	18,5
$[570,580[$	18	13

On obtient $d(F_n, F) = 3,23$ que l'on compare au fractile d'ordre 0,95 de la loi χ_7^2 qui vaut $C = 14,07$. On accepte donc l'hypothèse de données provenant d'une loi $N(555, 10)$, donc d'une loi dont la moyenne est inférieure de 5 g à la valeur indiquée sur la boîte.

5.7 Disposant des données individuelles, nous allons utiliser le test de Kolmogorov-Smirnov et pour cela ordonner les observations par valeurs croissantes $x_{(i)}$ puis calculer les valeurs centrées (et réduites) $u_{(i)} = x_{(i)} - 1,5$ pour pouvoir calculer ensuite $\Phi(u_{(i)})$ à l'aide de la f.r. Φ de la loi $N(0, 1)$. Les valeurs permettant le calcul de la distance entre loi empirique et loi théorique figurent dans le tableau suivant :

$u_{(i)}$	$\Phi(u_{(i)})$	$\frac{i}{n} - \Phi(u_{(i)})$	$\Phi(u_{(i)}) - \frac{i-1}{n}$
-1,5	0,0668	-0,0001	0,0668
-1,25	0,1056	0,0277	-0,1611
-1,09	0,1379	0,0621	0,004
-0,79	0,2148	0,0519	0,0148
-0,72	0,2358	0,0976	-0,0309
-0,49	0,3121	0,0879	-0,0212
-0,14	0,4443	0,0224	0,0443
0,05	0,4801	0,0532	0,0134
0,25	0,5987	0,001	0,0654
0,45	0,6736	-0,007	0,0736
0,56	0,7123	0,0210	0,0456
0,82	0,7939	0,006	0,0606
1,01	0,8438	0,0229	0,0438
1,25	0,8944	0,0389	0,0277
1,45	0,9265	0,0735	-0,007

On lit sur les deux dernières colonnes du tableau les valeurs maximales $d^+(F_n, F) = 0,0976$ et $d^+(F, F_n) = 0,0736$ d'où on déduit $d(F_n, F) = 0,0976$. Pour un risque $\alpha = 0,10$ on lit dans la table p. 154, pour $n = 15$ le seuil critique $C = 0,304$ donc on accepte l'hypothèse d'une loi parente $N(1, 5; 1)$, la distance obtenue étant beaucoup plus faible que le seuil critique.

Comportement asymptotique

Introduction

La taille d'un échantillon et la quantité d'information apportée par celui-ci sont bien sûr liées, de telle sorte que, si cette taille augmente, l'information sur la loi parente P de X augmente aussi, ce qui doit se traduire par une plus grande « proximité » entre la loi empirique P_n et la loi théorique P . Cette notion intuitive de plus grande proximité d'un terme d'une suite aléatoire, avec un terme fixe, aléatoire ou non, demande à être traduite et définie de façon rigoureuse.

Nous allons donc étudier dans ce chapitre le comportement asymptotique d'une suite de v.a. X_1, \dots, X_n quand n devient infini. Nous définirons la convergence en probabilité et la convergence en loi. Nous évoquerons aussi la convergence en moyenne quadratique car elle implique la convergence en probabilité et est généralement plus facile à établir que cette dernière. À ces deux types principaux de convergence seront associés les deux théorèmes fondamentaux de la statistique asymptotique, la loi des grands nombres et le théorème central limite.

Objectifs

Présenter les deux théorèmes fondamentaux de la statistique asymptotique, la loi des grands nombres et le théorème central limite.

Étudier la convergence des lois usuelles.

Concepts clés

Convergence en probabilité et en moyenne quadratique

Loi des grands nombres

Convergence en loi

Théorème central limite

1 Convergence en probabilité

La définition de la convergence en probabilité fait intervenir une suite numérique de probabilités dont la convergence sera souvent établie grâce à l'inégalité de Bienaymé-Tchebychev qui lie une probabilité et une variance. Avant de donner la définition de cette convergence stochastique et d'étudier les propriétés, nous allons donc au préalable présenter quelques inégalités qui seront utiles par la suite dans l'établissement de certaines de ces propriétés.

1.1 Inégalité de Markov

Si X est une v.a. positive dont l'espérance existe, l'**inégalité de Markov** établit que pour tout $\lambda > 0$:

$$P\{X \geq \lambda E(X)\} \leq \frac{1}{\lambda} \quad \text{ou} \quad P(X \geq \lambda) \leq \frac{E(X)}{\lambda}$$

On peut remarquer que sous sa première forme cette inégalité est sans intérêt pour $\lambda \leq 1$ puisqu'alors le majorant de la probabilité est supérieur à un. Cette inégalité présente bien sûr essentiellement un intérêt théorique, utile pour certaines démonstrations ; sa portée pratique est limitée par sa généralité, le majorant étant indépendant de la loi, sa valeur numérique sera souvent très supérieure à la valeur exacte de la probabilité.

Exemple 6.1

Pour $\lambda = 2$ l'inégalité de Markov s'écrit $P(X \geq 2E(X)) \leq 0,5$; dans le cas particulier de la loi exponentielle $\mathcal{E}(1)$ on obtient $P(X \geq 2) = e^{-2} = 0,135$, valeur très sensiblement inférieure au majorant.

Dans le cas d'une v.a. de signe quelconque, on adapte la seconde forme de l'inégalité de Markov en l'appliquant à $|X|^k$, pour tout k tel que $E|X|^k$ existe :

$$P(|X|^k \geq \lambda) \leq \frac{E|X|^k}{\lambda}$$

On introduit alors un nombre $\varepsilon > 0$ tel que $\varepsilon^k = \lambda$ et on en déduit pour tout $\varepsilon > 0$:

$$P(|X| \geq \varepsilon) \leq \frac{E|X|^k}{\varepsilon^k}$$

1.2 Inégalité de Bienaymé-Tchebychev

On obtient l'**inégalité de Bienaymé-Tchebychev** en appliquant l'inégalité de Markov sous sa dernière forme, à la v.a. $X - E(X)$ pour $k = 2$, donc pour une variable dont la variance existe, soit pour tout $\varepsilon > 0$ fixé :

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

REMARQUE Cette inégalité relie la probabilité pour X de s'écarter de sa moyenne $E(X)$, à sa variance qui est justement un indicateur de dispersion autour de la moyenne de la loi. En choisissant $\varepsilon = \sigma(X) = \sqrt{V(X)}$, on obtient $P\{|X - E(X)| \geq \sigma(X)\} \leq 1$ et cette inégalité est donc sans intérêt pour $\varepsilon \leq \sigma(X)$. On peut écrire cette inégalité pour la variable centrée-réduite en prenant cette fois $\varepsilon = a\sigma(X)$ avec $a > 1$:

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

Exemple 6.2

Dans le cas particulier de la loi $N(m, \sigma)$, la valeur exacte de la probabilité se calcule par $P(|X - m| \geq a\sigma) = 1 - P(|U| < a) = 1 - [\Phi(a) - \Phi(-a)] = 2[1 - \Phi(a)]$ où U est une v.a. de loi $N(0, 1)$ et de f.r. Φ . Pour $a = 1,5$ on obtient comme valeur $2 \times 0,0668 = 0,1336$ alors que le majorant a pour valeur $4/9 = 0,444$, valeur bien supérieure. Cet exemple montre à nouveau que la portée pratique d'une telle inégalité est faible, mais qu'elle sera très utile pour la démonstration de certaines propriétés, étant établie pour un très grand ensemble de lois de probabilité.

1.3 Inégalité de Jensen

Si g est une fonction réelle convexe sur un intervalle I de \mathbb{R} (i.e. $g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y)$ pour tous x et y de I et tout $\alpha \in [0, 1]$, ou $g''(x) \geq 0$ si g est deux fois dérivable) qui contient $X(\Omega)$, ensemble des valeurs possibles pour la v.a. X , telle que $E(X)$ et $E[g(X)]$ existent, alors :

$$g[E(X)] \leq E[g(X)]$$

L'ordonnée (au sens de la fonction g) de la moyenne (au sens d'espérance) est plus petite que la moyenne des ordonnées.

Exemple 6.3

Si on applique cette inégalité à $g(t) = t^2$ on obtient $E^2(X) \leq E(X^2)$, résultat bien connu par ailleurs puisqu'il traduit que la variance est positive : $V(X) = E(X^2) - E^2(X) \geq 0$.

1.4 Convergence en probabilité

Si (X_n) est une suite de v.a. qui converge vers une v.a. X , cela signifie que X_n se « rapproche » de X quand n augmente. On mesure la distance entre X_n et X par $|X_n - X|$ qui sera d'autant plus petite que n sera grand ; mais, s'agissant de v.a., il faut considérer l'événement $|X_n - X| < \varepsilon$ qui sera réalisé avec une probabilité d'autant plus élevée que n sera grand. On va donc associer à la suite aléatoire (X_n) la suite numérique des probabilités de ces événements, qui devra converger vers un. Ceci conduit à la définition suivante.

a) Définition

On dit que la suite de v.a. (X_n) **converge en probabilité** vers une v.a. X si, pour tout $\varepsilon > 0$:

$$P(|X_n - X| < \varepsilon) \rightarrow 1 \quad \text{quand } n \rightarrow \infty$$

ou, de façon équivalente :

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{quand } n \rightarrow \infty$$

On écrit :

$$X_n \xrightarrow{p} X$$

Exemple 6.4

Soit (U_n) une suite de v.a. indépendantes, de même loi uniforme sur $[0,1]$; on lui associe la suite (X_n) définie pour tout entier n par $X_n = \min\{U_1, \dots, U_n\}$. Nous allons établir que :

$$X_n \xrightarrow{p} 0 \quad \text{quand } n \rightarrow \infty$$

Pour cela, nous allons calculer, pour un $\varepsilon > 0$ fixé, la probabilité $P(|X_n| > \varepsilon)$ qui est égale à $P(X_n > \varepsilon)$ puisque X_n est une variable positive. Cette dernière probabilité est nulle pour $\varepsilon \geq 1$, donc il suffit de considérer le cas $\varepsilon < 1$. Toutes les v.a. U_n étant indépendantes, de même loi que U avec $P(U < x) = x$ pour $0 \leq x \leq 1$, on obtient :

$$\begin{aligned} P(|X_n| > \varepsilon) &= P\left\{ \bigcap_{i=1}^n (U_i > \varepsilon) \right\} = \prod_{i=1}^n P(U_i > \varepsilon) = [1 - P(U < \varepsilon)]^n \\ &= (1 - \varepsilon)^n \rightarrow 0 \end{aligned}$$

qui exprime bien que (X_n) converge en probabilité vers 0.

b) Conditions suffisantes de convergence en probabilité

Dans l'exemple 6.4 précédent, on a pu établir sans trop de difficultés la convergence de la suite (X_n) , à partir de la définition. Cependant, pour des lois plus complexes il peut être difficile d'obtenir l'expression de la probabilité permettant d'établir cette convergence. Nous allons maintenant fournir un moyen plus simple de démontrer cette convergence, à partir des deux premiers moments de X_n .

Théorème 6.1

Si (X_n) est une suite de v.a. telle que :

$$\begin{cases} E(X_n) \rightarrow a \\ V(X_n) \rightarrow 0 \end{cases}$$

quand $n \rightarrow \infty$, alors :

$$X_n \xrightarrow{p} a$$

En fait, on peut énoncer un résultat plus général que celui qui s'applique à la convergence vers un nombre réel a en l'appliquant à la suite (Y_n) définie par $Y_n = X_n - X$, avec (X_n) qui converge vers la v.a. X , puisque $Y_n \xrightarrow{p} 0 \Leftrightarrow X_n \xrightarrow{p} X$. Les conditions suffisantes pour la suite (X_n) deviennent alors :

$$\begin{cases} E(X_n - X) \rightarrow 0 \\ V(X_n - X) \rightarrow 0 \end{cases}$$

quand $n \rightarrow \infty$, et impliquent :

$$X_n \xrightarrow{p} X$$

Exemple 6.5

Nous avons vu dans le chapitre précédent que pour tout x réel fixé on avait $E[F_n(x)] = F(x)$ et $V[F_n(x)] = F(x)[1 - F(x)]/n \rightarrow 0$, donc ces conditions suffisantes permettent d'établir la convergence ponctuelle de la fonction de répartition empirique vers la fonction de répartition théorique :

$$F_n(x) \xrightarrow{p} F(x).$$

En réalité, on peut établir qu'il existe une constante universelle C telle que :

$$P \left\{ \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| > \varepsilon \right\} \leq C e^{-2n\varepsilon^2}$$

Cette probabilité converge donc vers 0 pour tout $\varepsilon > 0$, mais en plus très rapidement puisqu'elle est majorée par le terme général d'une série convergente. Il y a donc convergence uniforme de la f.r. empirique vers la f.r. théorique, selon un mode plus fort qu'en probabilité. On peut énoncer la forme affaiblie du théorème de Glivenko-Cantelli :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p} 0$$

De cette convergence forte de la loi empirique vers la loi théorique découlent toutes les propriétés de convergence des moments empiriques associés à cette loi.

c) Convergence en moyenne d'ordre p

Les conditions suffisantes énoncées précédemment permettent en fait d'établir une convergence plus forte que la convergence en probabilité, qui est la convergence en moyenne quadratique, cas particulier de la convergence en moyenne d'ordre p que nous définissons pour cette raison.

DÉFINITION

On dit que la suite de v.a. (X_n) converge en moyenne d'ordre p , avec $0 < p < \infty$, vers la v.a. X si :

$$E|X_n - X|^p \rightarrow 0 \quad \text{quand } n \rightarrow \infty$$

On écrit :

$$X_n \xrightarrow{\mathcal{M}_p} X$$

Dans le cas particulier $p = 2$, la convergence en moyenne d'ordre 2 s'appelle **convergence en moyenne quadratique**. En écrivant $E(X_n - X)^2 = V(X_n - X) + E^2(X_n - X)$ qui est la somme de deux termes positifs, on retrouve les conditions suffisantes de convergence en probabilité comme conditions nécessaires de convergence en moyenne quadratique :

$$X_n \xrightarrow{m.q.} X \Leftrightarrow \begin{cases} E(X_n - X) \rightarrow 0 \\ V(X_n - X) \rightarrow 0 \end{cases}$$

Pour toute valeur de $p > 0$, la convergence en moyenne d'ordre p est plus forte que la convergence en probabilité, au sens où :

$$X_n \xrightarrow{\mathcal{M}_p} X \Rightarrow X_n \xrightarrow{p} X$$

et donc en particulier :

$$X_n \xrightarrow{m.q.} X \Rightarrow X_n \xrightarrow{p} X$$

Très souvent, on établit donc la convergence en moyenne quadratique pour obtenir la convergence en probabilité, car cela se ramène à l'étude de la convergence de deux suites numériques, généralement simple à effectuer.

d) Théorème de Slutsky

Théorème 6.2

Si f est une application réelle continue, alors :

$$X_n \xrightarrow{p} X \Rightarrow f(X_n) \xrightarrow{p} f(X)$$

Ce théorème exprime donc qu'une application continue conserve la convergence en probabilité, au sens où la limite de la suite des images est l'image de la limite de la suite. Ce résultat se généralise en deux dimensions sous la forme du théorème suivant.

Théorème 6.3

Si f est une application de \mathbb{R}^2 dans \mathbb{R} uniformément continue et si (X_n) et (Y_n) sont deux suites de v.a. qui convergent en probabilité respectivement vers les v.a. X et Y , alors :

$$f(X_n, Y_n) \xrightarrow{p} f(X, Y)$$

Si on applique ce théorème aux fonctions f définies respectivement par $f(u, v) = u + v$, $f(u, v) = uv$ et $f(u, v) = \frac{u}{v}$, de $X_n \xrightarrow{p} X$ et $Y_n \xrightarrow{p} Y$ on déduit respectivement $X_n + Y_n \xrightarrow{p} X + Y$, $X_n Y_n \xrightarrow{p} XY$ et $\frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{Y}$ à condition que $P(Y = 0) = 0$.

1.5 Loi des grands nombres

Dans le chapitre précédent, nous avons vu que la moyenne empirique d'un échantillon d'une v.a. X avait même moyenne, au sens d'espérance mathématique, m que X , mais avec une dispersion, mesurée par la variance, beaucoup plus faible, de valeur σ^2/n , c'est-à-dire que la variance de X était divisée par la taille de l'échantillon. Ainsi, quand n devient infini, cette variance tend vers zéro, assurant la convergence de la moyenne empirique vers la moyenne théorique $E(X) = m$. Ce résultat s'obtient à partir de l'inégalité de Bienaymé-Tchebychev, puisque pour tout $\varepsilon > 0$:

$$P\left(\left|\bar{X}_n - m\right| \geq \varepsilon\right) \leq \frac{V(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

ce qui permet d'énoncer le premier théorème fondamental de la statistique asymptotique, c'est-à-dire quand la taille de l'échantillon devient infinie.

Théorème 6.4

Si (X_n) est une suite de v.a. mutuellement indépendantes qui admettent les mêmes moments d'ordres un et deux, c'est-à-dire avec pour tout entier n , $E(X_n) = m$ et $V(X_n) = \sigma^2$, alors quand $n \rightarrow \infty$:

$$\bar{X}_n \xrightarrow{p} m$$

Ce théorème est obtenu sous la condition que les v.a. X_n aient mêmes moments d'ordres un et deux, mais pas forcément même loi, et s'appelle parfois loi faible des grands nombres. Il existe en effet une loi forte des grands nombres, relative à une convergence plus forte que nous n'étudierons pas ici (voir compléments), qui peut s'obtenir

plus logiquement sous la seule condition d'existence du moment d'ordre un, puisqu'elle concerne seulement le moment d'ordre un, mais en renforçant par ailleurs les hypothèses, en supposant alors que les v.a. X_n ont la même loi de probabilité.

Exemple 6.6

Théorème de Bernoulli

On effectue n expériences successives indépendantes où on s'intéresse à chaque fois à la réalisation d'un certain événement A . On associe donc à chaque expérience i , $1 \leq i \leq n$, une variable de Bernoulli :

$$X_i = \begin{cases} 1 & p \\ 0 & q = 1 - p \end{cases}$$

La fréquence empirique, c'est-à-dire le pourcentage de réalisations de A est :

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

avec $E(f_n) = p$ et $V(f_n) = pq/n$, d'où en appliquant l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$:

$$P(|f_n - p| > \varepsilon) \leq \frac{pq}{n\varepsilon^2} \rightarrow 0$$

ce qui établit que pour $n \rightarrow \infty$:

$$f_n \xrightarrow[p]{} p = P(A)$$

c'est-à-dire que la fréquence empirique (observée) d'un événement converge, en probabilité, vers la fréquence théorique ou probabilité de réalisation de cet événement quand le nombre d'expériences augmente indéfiniment. Ce résultat, connu sous le nom de théorème de Bernoulli, établit donc le lien entre l'expérience que l'on peut avoir des chances de réalisation d'un événement et la définition théorique que l'on a donnée de sa probabilité. Nous verrons dans le chapitre 7 suivant que la fréquence empirique fournit une estimation de la probabilité d'un événement.

2 Convergence en loi

2.1 Définition

DÉFINITION

On dit que la suite de v.a. (X_n) , de f.r. F_n , converge en loi vers une v.a. X de f.r. F si la suite $\{F_n(x)\}$ converge vers $F(x)$ en tout point x où F est continue ; on écrit alors :

$$X_n \xrightarrow[loi]{} X$$

Les v.a. X_n ont toutes des lois différentes, de f.r. notées F_n , à ne pas confondre avec la f.r. empirique associée à un échantillon de v.a. qui ont toutes la même loi, donc la même f.r.

2.2 Lien avec la convergence en probabilité

Le lien entre les deux convergences que nous venons de définir est précisé par le résultat suivant.

Théorème 6.5

La convergence en probabilité d'une suite (X_n) implique sa convergence en loi :

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{\text{loi}} X$$

Ce résultat se conçoit intuitivement car $F_n(x)$ représente la probabilité, associée à la loi de X_n , de l'événement $\{X_n < x\}$, et elle va converger vers la probabilité correspondante associée à la loi de X . La réciproque de ce résultat est fautive en général, mais est vraie dans le cas particulier où la limite est une v.a. certaine $X = a$, c'est-à-dire quand la loi limite est une loi de Dirac, loi dégénérée concentrée en un seul point.

2.3 Propriété

Si (X_n) et (Y_n) sont deux suites de v.a. telles que pour $n \rightarrow \infty$:

$$X_n \xrightarrow{\text{loi}} X \text{ et } Y_n \xrightarrow{p} a$$

où a est un nombre réel, alors :

$$X_n + Y_n \xrightarrow{\text{loi}} X + a$$

$$X_n Y_n \xrightarrow{\text{loi}} aX$$

$$\frac{X_n}{Y_n} \xrightarrow{\text{loi}} \frac{X}{a} \quad \text{si } a \neq 0$$

2.4 Théorème de Slutsky

Théorème 6.6

Si g est une application réelle continue, alors :

$$X_n \xrightarrow{\text{loi}} X \Rightarrow g(X_n) \xrightarrow{\text{loi}} g(X)$$

2.5 Conditions suffisantes de convergence en loi

Dans le cas d'une loi discrète ou d'une loi absolument continue, on peut énoncer des conditions suffisantes de convergence en loi relatives aux probabilités individuelles ou à la densité qui définissent ces lois.

Dans le cas discret où $X_n(\Omega) = X(\Omega) = \{a_i/i \in I\}$:

$$\forall i \in I, P(X_n = a_i) \rightarrow P(X = a_i) \Rightarrow X_n \xrightarrow[loi]{} X, \quad n \rightarrow \infty$$

Dans le cas où X_n et X admettent pour densités respectives f_n et f :

$$\forall x \in \mathbb{R}, f_n(x) \rightarrow f(x) \Rightarrow X_n \xrightarrow[loi]{} X, \quad n \rightarrow \infty$$

2.6 Théorème central limite

La loi des grands nombres énonce la convergence de la suite \bar{X}_n des moyennes empiriques vers la moyenne théorique m , c'est-à-dire une v.a. certaine ou loi de Dirac, loi dégénérée dont toute la masse est concentrée en un point. Si on équilibre la suite des v.a. centrées et réduites $(\bar{X}_n - m)/\sigma$, qui converge vers zéro, par \sqrt{n} qui tend vers l'infini, la forme indéterminée obtenue aura une limite qui sera celle d'une loi non dégénérée cette fois, la loi normale. C'est ce qu'énonce le **théorème central limite**, dont nous donnons ici l'une des multiples versions.

Théorème 6.7

Si (X_n) est une suite de v.a. indépendantes et de même loi, admettant des moments d'ordres un et deux notés $m = E(X_n)$ et $\sigma^2 = V(X_n)$, alors :

$$\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \xrightarrow[loi]{} N(0, 1)$$

C'est en raison de ce théorème que la loi normale est la loi de probabilité la plus utilisée en statistique. En effet, beaucoup de phénomènes individuels ne sont pas symétriques et donc ne peuvent pas être modélisés par la loi normale. Mais, dans beaucoup de situations, le phénomène étudié est la résultante d'un grand nombre de composantes indépendantes et le théorème central limite nous assure alors que la loi normale est tout à fait adéquate.

2.7 Limite d'une suite image

Nous savons qu'une application continue transporte la convergence en loi (théorème de Slutsky, § 2.4) ; nous allons préciser ce résultat dans le cas d'une convergence vers la loi normale, comme celle obtenue par le théorème central limite. Si une suite de v.a. (X_n) converge vers un nombre réel fixé m , avec une suite numérique (a_n) convenablement

choisie qui tend vers l'infini avec n on peut obtenir la convergence en loi de la suite $a_n(X_n - m)$ vers une loi normale centrée. La suite $(g(X_n))$ sera alors centrée sur $g(m)$ et équilibrée aussi par a_n pour converger vers une loi normale centrée dont l'écart type sera modifié. Voici l'énoncé précis de cette propriété permettant de transporter un résultat de convergence en loi.

Théorème 6.8

Si la suite (X_n) est telle que :

$$a_n (X_n - m) \xrightarrow{\text{loi}} N(0, \sigma), \quad n \rightarrow \infty$$

avec $a_n \rightarrow \infty$ et $\sigma > 0$, alors si g est une application réelle dérivable, la suite $(g(X_n))$ converge aussi en loi avec :

$$a_n [g(X_n) - g(m)] \xrightarrow{\text{loi}} N(0, \sigma |g'(m)|), \quad n \rightarrow \infty$$

2.8 Convergence des moments empiriques

a) Moyenne empirique

Les deux théorèmes fondamentaux de la statistique asymptotique établissent la convergence en probabilité et en loi de la moyenne empirique.

b) Variance empirique

Au chapitre précédent, nous avons obtenu comme expression de la variance empirique :

$$S_n'^2 = \frac{1}{n} \sum_{i=1}^n U_i - (\bar{X}_n - m)^2$$

en ayant posé $U_i = (X_i - m)^2$. De la loi des grands nombres on déduit $\bar{U}_n \xrightarrow{p} E(U) = \sigma^2$

et $(\bar{X}_n - m)^2 \xrightarrow{p} 0$ et par conséquent :

$$S_n'^2 \xrightarrow{p} \sigma^2, \quad n \rightarrow \infty$$

On obtient bien sûr le même résultat pour $S_n^2 = nS_n'^2/(n-1)$ puisque $n/(n-1) \rightarrow 1$.

En appliquant la loi des grands nombres et le théorème central limite on obtient :

$$\bar{X}_n - m \xrightarrow{p} 0 \quad \text{et} \quad \sqrt{n} (\bar{X}_n - m) \xrightarrow{\text{loi}} N(0, \sigma)$$

ce qui implique d'après la propriété § 2.3 que $\sqrt{n} \left(\overline{X}_n - m \right) \xrightarrow[loi]{} 0$. Ainsi, $\sqrt{n} \left(S_n'^2 - \sigma^2 \right)$ a même loi limite que $\sqrt{n} \left(\overline{U}_n - \sigma^2 \right)$, qui est obtenue par le théorème central limite avec $E(U) = \sigma^2 = \mu_2$ et $V(U) = \mu_4 - \mu_2^2$:

$$\sqrt{n} \left(S_n'^2 - \sigma^2 \right) \xrightarrow[loi]{} N \left(0, \sqrt{\mu_4 - \mu_2^2} \right), \quad n \rightarrow \infty$$

On a le même résultat pour S_n^2 .

c) Moments empiriques

Par application de la loi des grands nombres :

$$m_{kn} = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow[p]{} E(X^k) = m_k$$

On obtient par ailleurs $V(X^k) = m_{2k} - m_k^2$ donc par application du théorème central limite à la suite (X_n^k) :

$$\sqrt{n} \frac{m_{kn} - m_k}{\sqrt{m_{2k} - m_k^2}} \xrightarrow[loi]{} N(0, 1)$$

2.9 Convergence des lois usuelles

a) Loi binômiale

Soit X une v.a. de loi $\mathcal{B}(n, p)$ et supposons que $n \rightarrow \infty$. Nous allons considérer deux cas suivant que le second paramètre p reste fixe ou tend vers 0 avec $1/n$.

- Si $p \rightarrow 0$ quand $n \rightarrow \infty$, avec $np \rightarrow \lambda$ fini, alors en utilisant la formule de Stirling $n! = n^n e^{-n} \sqrt{2\pi n} (1 + \varepsilon_n)$, on montre que pour tout entier k fixé, $0 \leq k \leq n$:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

ce qui, d'après la condition suffisante (cf. § 2.5) de convergence en loi, établit que la loi binômiale $\mathcal{B}(n, p)$ converge vers la loi de Poisson $\mathcal{P}(\lambda)$. C'est pourquoi on appelle parfois la loi de Poisson **loi des événements rares**, puisqu'elle correspond à des événements de probabilité très faible, mais que l'on observe quand même en raison d'un très grand nombre d'épreuves, comme par exemple le nombre d'erreurs dans un livre ou le nombre d'accidents à un carrefour. On considère que l'approximation est valable pour $n \geq 30$ et $np < 5$.

Exemple 6.7

Pour $p = 0,1$ et $n = 30$ on lit dans la table 3 de la loi binômiale $P(X = 1) = 0,1413$. Si on utilise la variable Y de loi de Poisson de paramètre $np = 3$ on lit $P(Y = 1) = 0,1494$ dans la table 4, donc valeur très proche.

- Si p reste fixe quand $n \rightarrow \infty$, on utilise l'approximation normale en écrivant la loi binômiale comme somme de v.a. X_i indépendantes et de même loi de Bernoulli de paramètre p : $X = \sum_{i=1}^n X_i$. Comme $E(X_i) = p$ et $V(X_i) = pq$ on déduit du théorème central limite :

$$\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{pq}} \xrightarrow{\text{loi}} N(0, 1)$$

Avec $\bar{X}_n = X/n$, ceci peut encore s'écrire :

$$\sqrt{n} \frac{X/n - p}{\sqrt{pq}} = n \frac{X/n - p}{\sqrt{npq}} = \frac{X - np}{\sqrt{npq}} \xrightarrow{\text{loi}} N(0, 1)$$

On peut donc approximer la loi binômiale $B(n, p)$ par la loi normale $N(np, \sqrt{npq})$. On considère que l'approximation est valable pour $n \geq 30$, $np \geq 5$ et $nq \geq 5$.

Exemple 6.8

On lance 400 fois une pièce de monnaie et on veut calculer la probabilité que la fréquence d'apparition de pile soit comprise entre 0,45 et 0,55. On associe à chaque lancer i , $1 \leq i \leq n$, la v.a. de Bernoulli :

$$X_i = \begin{cases} 1 & \text{pile}(1/2) \\ 0 & \text{face}(1/2) \end{cases}$$

La variable $S = \sum_{i=1}^{400} X_i$ suit une loi binômiale $B(400, 1/2)$ avec $E(S) = 200$, $V(S) = 100$ et $\sigma(S) = 10$. La probabilité demandée s'écrit :

$$\begin{aligned} P\left(0,45 \leq \frac{S}{400} \leq 0,55\right) &= P(180 \leq S \leq 220) \\ &= P\left(-\frac{20}{10} \leq \frac{S - 200}{10} \leq \frac{20}{10}\right) \end{aligned}$$

et peut être approximée à l'aide de la f.r. Φ de la loi $N(0, 1)$ par $\Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0,9544$. Pour $n = 900$ on obtient $E(S) = 450$, $V(S) = 225$ et $\sigma(S) = 15$ et la probabilité demandée devient :

$$\begin{aligned} P\left(0,45 \leq \frac{S}{900} \leq 0,55\right) &= P(405 \leq S \leq 495) \\ &= P\left(-\frac{45}{15} \leq \frac{S - 450}{15} \leq \frac{45}{15}\right) \end{aligned}$$

qui est approximée cette fois par $2\Phi(3) - 1 = 0,9973$.

b) Loi hypergéométrique

Si X suit une loi hypergéométrique de paramètres N , n et N_A avec $N \rightarrow \infty$ et $N_A/N = p$ qui reste fixe, on obtient par la formule de Stirling, pour tout k fixé tel que $0 \leq k \leq n$:

$$P(X = k) = \frac{\binom{N_A}{k} \binom{N - N_A}{n - k}}{\binom{N}{n}} \rightarrow \binom{n}{k} p^k (1 - p)^{n - k}$$

Il y a donc convergence de cette loi hypergéométrique vers la loi binômiale $\mathcal{B}(n, p)$. L'approximation est considérée comme justifiée lorsque le rapport n/N , appelé parfois taux de sondage, est inférieur à 10 %.

c) Loi de Poisson

Si X suit une loi de Poisson dont le paramètre λ tend vers l'infini, alors :

$$\frac{X - \lambda}{\sqrt{\lambda}} \xrightarrow{\text{loi}} N(0, 1)$$

Exemple 6.9

Pour $\lambda = 16$ on lit dans la table de la loi de Poisson $P(X \leq 21) = 0,9107$. On approxime $P(X \leq 21) = P\left(\frac{X - 16}{\sqrt{16}} \leq \frac{5}{4}\right)$ par $\Phi(1,25) = 0,8944$, où Φ est la f.r. de la loi $N(0, 1)$, soit une assez bonne approximation malgré une valeur faible de λ .

Correction de continuité Nous avons ici un second exemple d'une loi discrète qui est approximée par une loi continue. Si X est une variable à valeurs entières dont la loi est approximée par celle de la variable continue U , on ne peut pas approximer $p_k = P(X = k)$ par $P(U = k) = 0$! On remplace donc cette probabilité par la valeur approchée $P(k - 1/2 < U < k + 1/2)$ qui est la probabilité de l'intervalle de longueur un qui contient l'entier k . Ainsi, $P(X \leq k) = \sum_{j=0}^k P(X = j)$ va être approximé par :

$$P(U < k + 1/2) = P(U - 1/2 < k)$$

L'approximation de la loi discrète de X par celle continue de U impose donc de remplacer X par $U - 1/2$ dans les événements exprimés à partir de X , ce qu'on appelle faire la **correction de continuité**. Dans l'exemple 6.9 précédent, $P(X \leq 21)$ doit donc être approximé par $P(U - 1/2 \leq 21) = \Phi(5,5/4) = 0,9154$ soit une meilleure approximation que sans la correction de continuité.

d) Loi gamma

Si X suit une loi $\gamma(p)$ avec $p \rightarrow \infty$, alors :

$$\frac{X - p}{\sqrt{p}} \xrightarrow{\text{loi}} N(0, 1)$$

e) Loi du khi-deux

Si X suit une loi du khi-deux dont le nombre de degrés de liberté $\nu \rightarrow \infty$, alors :

$$\frac{X - \nu}{\sqrt{2\nu}} \xrightarrow{\text{loi}} N(0, 1)$$

Cependant, pour les petites valeurs de ν , on obtient une meilleure approximation en utilisant comme résultat la convergence :

$$\sqrt{2X} - \sqrt{2\nu - 1} \xrightarrow{\text{loi}} N(0, 1)$$

Exemple 6.10

Le fractile d'ordre 0,1 de la loi du khi-deux à 50 degrés de liberté a pour valeur $u = 37,7$ (cf. table 5). Avec la première approximation, on écrit :

$$0,1 = P\left(\frac{X - \nu}{\sqrt{2\nu}} < \frac{u - \nu}{\sqrt{2\nu}}\right)$$

Le fractile d'ordre 0,1 de la loi $N(0, 1)$ a pour valeur $f = -1,2816$ d'où on déduit comme valeur approchée de u : $10f + \nu = 50 - 12,8 = 37,2$. Pour utiliser la seconde approximation, on écrit :

$$0,1 = P\left(\sqrt{2X} - \sqrt{2\nu - 1} < \sqrt{2u} - \sqrt{2\nu - 1}\right)$$

et on obtient cette fois $\sqrt{2u} - \sqrt{2\nu - 1} \simeq f$ soit $u \simeq 37,6$ valeur plus proche de la valeur exacte.

f) Loi de Student

Si X suit une loi de Student dont le nombre de degrés de liberté n devient infini, alors :

$$X \xrightarrow{\text{loi}} N(0, 1)$$

Exemple 6.11

Pour se faire une idée de l'approximation du fractile d'ordre 0,95 par celui de la loi normale qui vaut 1,645 le tableau ci-après donne les fractiles exacts de la loi de Student en fonction de n .

n	30	40	100	200	500
	1,697	1,684	1,660	1,653	1,648

3 Compléments

Nous allons présenter deux autres modes stochastiques de convergence, plus forts que la convergence en probabilité, et nous énoncerons notamment la loi forte des grands nombres.

3.1 Convergence presque sûre

On dit que la suite (X_n) converge **presque sûrement** vers la v.a. X si :

$$P \left\{ \omega \in \Omega / \lim_n X_n(\omega) = X(\omega) \right\} = 1$$

et on écrit :

$$X_n \xrightarrow[p.s.]{} X, \quad n \rightarrow \infty$$

Cette terminologie, qui peut paraître un peu étrange, traduit le fait que la suite numérique $X_n(\omega)$ converge **presque partout** vers $X(\omega)$, au sens où elle converge pour tout ω de Ω , à l'exception d'un sous-ensemble de Ω dont la probabilité est égale à zéro. Les quatre propriétés suivantes, qui sont équivalentes, peuvent permettre de mieux comprendre le sens de cette convergence :

$$\begin{aligned}
 & X_n \xrightarrow[p.s.]{} X \\
 & \Leftrightarrow \\
 & \forall \varepsilon > 0, \quad \lim_n P \left\{ \sup_{m \geq n} |X_m - X| > \varepsilon \right\} = 0 \\
 & \Leftrightarrow \\
 & \forall \varepsilon > 0, \quad \lim_n P \left\{ \bigcup_{m \geq n} (|X_m - X| > \varepsilon) \right\} = 0 \\
 & \Leftrightarrow \\
 & \forall \varepsilon > 0, \quad \lim_n P \left\{ \bigcap_{m \geq n} (|X_m - X| < \varepsilon) \right\} = 1
 \end{aligned}$$

Il est facile de voir que les événements suivants sont équivalents :

$$\bigcup_{m \geq n} (|X_m - X| > \varepsilon) = \left\{ \sup_{m \geq n} |X_m - X| > \varepsilon \right\}$$

et ont comme complémentaire :

$$\bigcap_{m \geq n} (|X_m - X| < \varepsilon)$$

ce qui montre l'équivalence des trois conditions. La dernière condition exprime que, à partir d'un certain rang $N = N(\varepsilon)$, tous les événements $(|X_n - X| < \varepsilon)$ pour $n > N$ sont réalisés avec une probabilité qui tend vers un. La première condition signifie que :

$$\sup_{m \geq n} |X_m - X| \xrightarrow{p} 0, \quad n \rightarrow \infty$$

et implique en particulier que $|X_n - X| \xrightarrow{p} 0$, donc la convergence presque sûre est plus forte que la convergence en probabilité :

$$X_n \xrightarrow{p.s.} X \Rightarrow X_n \xrightarrow{p} X$$

à titre d'illustration de cette convergence, nous pouvons énoncer le théorème suivant.

Théorème 6.9

Loi forte des grands nombres

Si (X_n) est une suite de v.a. indépendantes et de même loi, admettant une espérance notée m , alors :

$$\overline{X}_n \xrightarrow{p.s.} m, \quad n \rightarrow \infty$$

3.2 Convergence presque complète

La convergence précédente peut être difficile à établir ; cependant, si on applique l'inégalité de Boole, on observe que :

$$P \left\{ \bigcup_{m \geq n} (|X_m - X| > \varepsilon) \right\} \leq \sum_{m=n}^{\infty} P \{ |X_m - X| > \varepsilon \}$$

et donc la convergence presque sûre sera obtenue si la série de terme général $P\{|X_n - X| > \varepsilon\}$ est convergente, car le majorant est le reste de cette série. Cette condition suffisante de convergence p.s., plus facile à établir que les conditions équivalentes de la définition, sert de définition à une convergence plus forte.

On dit que la suite (X_n) converge **presque complètement sûrement** vers la v.a. X si :

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{\infty} P \{ |X_n - X| > \varepsilon \} < +\infty$$

et on écrit :

$$X_n \xrightarrow[p.co.]{n \rightarrow \infty} X$$

À titre d'illustration de cette convergence, nous pouvons énoncer le théorème suivant, énoncé sous sa forme faible dans l'exemple 6.5.

Théorème 6.10

Théorème de Glivenko-Cantelli

Si F_n est la f.r. empirique associée à un échantillon d'une loi de f.r. F , alors :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[p.co.]{n \rightarrow \infty} 0$$

Exemple 6.12

Considérons la suite (X_n) définie à partir d'une v.a. U de loi uniforme sur $[0, 1]$ par :

$$X_n = \begin{cases} \sqrt{n} & \text{si } U \leq \frac{1}{n} \\ 0 & \text{sinon} \end{cases}, \quad n \geq 1$$

et étudions sa convergence vers zéro. Pour tout $\varepsilon > 0$ fixé, l'événement $(|X_m| > \varepsilon)$ est équivalent à l'événement $U \in [0, \frac{1}{m}]$, donc :

$$\bigcup_{m=n}^{\infty} (|X_m| > \varepsilon) = \bigcup_{m=n}^{\infty} \left(U \in \left[0, \frac{1}{m} \right] \right) = \left(U \in \left[0, \frac{1}{n} \right] \right)$$

et par conséquent :

$$P \left\{ \bigcup_{m=n}^{\infty} (|X_m| > \varepsilon) \right\} = P \left(U \in \left[0, \frac{1}{n} \right] \right) = \frac{1}{n} \rightarrow 0$$

ce qui traduit le fait que :

$$X_n \xrightarrow[p.s.]{n \rightarrow \infty} 0$$

Cependant, $P(|X_n| > \varepsilon) = P\left(U \leq \frac{1}{n}\right) = \frac{1}{n}$ est le terme général d'une série divergente et donc il n'y a pas convergence presque complète de (X_n) vers zéro. Cet exemple montre bien que ces deux convergences ne sont pas équivalentes, même si dans beaucoup de cas la convergence p.s. se démontre à partir de la convergence p.co.

- 1 Il est bien sûr essentiel de retenir les énoncés des deux théorèmes fondamentaux de la statistique asymptotique : la loi des grands nombres et le théorème central limite. Ceci suppose au préalable de bien connaître également les définitions des deux convergences principales associées à ces théorèmes : la convergence en probabilité et la convergence en loi.
- 2 La convergence en probabilité est plus forte que la convergence en loi et se démontre généralement à partir des conditions suffisantes qui font intervenir l'espérance et la variance des termes de la suite. Ces conditions définissent en fait la convergence en moyenne quadratique, qui à son tour implique la convergence en probabilité.
- 3 L'inégalité de Bienaymé-Tchebychev est utile pour établir certains résultats théoriques.



Entraînez-vous

- 6.1** Étudier la convergence en probabilité, en moyenne, puis en moyenne quadratique de la suite de v.a. (X_n) dont la loi de probabilité est définie pour $n \in \mathbb{N}^*$ par :

$$P(X_n = -n) = P(X_n = n) = \frac{1}{2n^2} \text{ et } P(X_n = 0) = 1 - \frac{1}{n^2}$$

- 6.2** Soit (X_n) une suite de v.a. dont la loi est définie pour $n \in \mathbb{N}^*$ par :

$$P(X_n = 0) = 1 - \frac{1}{n} \text{ et } P(X_n = n) = \frac{1}{n}$$

- 1.** Montrer que (X_n) converge en probabilité, mais pas en moyenne quadratique, vers zéro quand n tend vers l'infini.
- 2.** Soit (Y_n) une suite de v.a. de même loi $N(0, 1)$ et indépendantes des v.a. (X_n) . Étudier la convergence en loi de $Z_n = X_n + Y_n$ et la limite de $V(Z_n)$ quand n tend vers l'infini.
- 6.3** Peut-on appliquer la loi des grands nombres à la suite (X_n) de v.a. mutuellement indépendantes dont la loi de probabilité est définie pour $n \in \mathbb{N}^*$ par :

$$P(X_n = -\sqrt{n}) = P(X_n = \sqrt{n}) = \frac{1}{2} ?$$

- 6.4** Étudier la convergence en loi de la suite de v.a. certaines $X_n = n$.
- 6.5** Soit (U_n) des v.a. indépendantes et de même loi définie par $P(U_n = 1) = p$ et $P(U_n = -1) = q = 1 - p$ avec $0 < p < 1$. Déterminer la loi exacte, puis la loi limite, de la suite (V_n) définie par :

$$V_n = \prod_{i=1}^n U_i$$

- 6.6** Soit (X_n) une suite de v.a. de loi géométrique de paramètre $p_n = \frac{\theta}{n}$, où θ est un nombre strictement positif. Montrer que la suite $\frac{X_n}{n}$ converge vers la loi exponentielle de paramètre θ .
- 6.7** Soit (X_n) une suite de v.a. de loi uniforme discrète sur l'ensemble $X_n(\Omega) = \left\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\right\}$. Montrer que cette suite converge vers la loi uniforme sur $[0, 1]$.
- 6.8** Soit (X_n) une suite de v. a. dont la loi est définie pour $n \in \mathbb{N}^*$ par :

$$P\left(X_n = 1 - \frac{1}{n}\right) = P\left(X_n = 1 + \frac{1}{n}\right) = \frac{1}{2}$$

Étudier sa convergence en loi.

- 6.9** Calculer le fractile d'ordre 0,95 de la loi χ_{40}^2 en utilisant l'approximation de la loi de $\sqrt{2\chi_v^2} - \sqrt{2v-1}$ par la loi normale standard, avec ici $v = 40$.
- 6.10** Soit X_n une v.a. positive de densité $f_n(x) = ne^{-nx}$ pour $x > 0$. Montrer que la suite (X_n) converge en moyenne quadratique vers zéro.
- 6.11** Soit (X_n) une suite de v.a. indépendantes et de même loi, de densité f définie par :

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta \\ 0 & \text{si } x \leq \theta \end{cases}$$

où θ est un nombre positif fixé. Montrer que $m_n = \min\{X_1, \dots, X_n\}$ converge en moyenne quadratique vers θ . Étudier la convergence en loi de $Y_n = n(m_n - \theta)$.

- 6.12** Soit (X_n) une suite de v. a. indépendantes et de même loi normale centrée et de variance σ^2 . Étudier la convergence en loi des v. a. :

$$D_n = \frac{1}{n} \sum_{i=1}^n |X_i| \quad \text{et} \quad S_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

- 6.13** Soit (X_n) une suite de v.a. mutuellement indépendantes et de même loi de Gumbel de densité $f(x) = \exp(x - e^x)$. Déterminer la loi de probabilité de la variable aléatoire $Z_n = e^{X_n}$, puis étudier la convergence en loi de la suite (Y_n) , avec $Y_n = -\ln\left(\frac{1}{n} \sum_{i=1}^n e^{X_i}\right)$.

Solutions



6.1 Pour tout $\varepsilon > 0$, on obtient $P(|X_n| < \varepsilon) = P(X_n = 0) = 1 - \frac{1}{n^2} \rightarrow 1$ quand $n \rightarrow \infty$, donc par définition $X_n \xrightarrow{p} 0$. D'autre part, $E(|X_n|) = \frac{1}{n} \rightarrow 0$ ce qui exprime que X_n converge en moyenne (d'ordre 1) vers zéro, ce qui implique d'ailleurs sa convergence en probabilité. Enfin, $E(X_n^2) = 1$ donc X_n ne converge pas en moyenne quadratique vers zéro.

6.2 1. Pour tout $\varepsilon > 0$, $P(|X_n| > \varepsilon) = P(X_n = n) = \frac{1}{n} \rightarrow 0$ quand $n \rightarrow \infty$, donc par définition de la convergence en probabilité :

$$X_n \xrightarrow{p} 0$$

Par ailleurs, $E(X_n^k) = n^{k-1}$ donc $E(X_n) = 1$ et $E(X_n^k) \rightarrow \infty$ pour tout entier $k > 1$. En particulier $E(X_n^2) \rightarrow \infty$ et donc (X_n) ne converge pas en moyenne quadratique vers zéro. Aucun moment de X_n ne converge d'ailleurs vers le moment correspondant de sa limite qui est toujours égal à 0.

2. Toutes les variables Y_n ont la même loi $N(0, 1)$ qu'une certaine variable Y , donc en appliquant la propriété présentée § 2.3 on en déduit que la suite $(X_n + Y_n)$ converge en loi vers Y . Cependant, $V(Z_n) = V(X_n) + V(Y_n) = n - 1 + 1 = n \rightarrow \infty$, alors que Z_n converge en loi vers la loi normale de variance 1.

6.3 On ne peut pas appliquer la loi forte des grands nombres aux variables X_n car elles n'ont pas la même loi. D'autre part, on obtient $E(X_n) = 0$ et $V(X_n) = n$ donc on ne peut pas non plus appliquer la loi faible des grands nombres car ces variables n'ont pas la même variance.

6.4 Les variables X_n ont pour f.r. :

$$F_n(x) = \begin{cases} 0 & \text{si } x \leq n \\ 1 & \text{si } n < x \end{cases}$$

La suite $F_n(x)$ converge pour tout x réel vers $F(x) = 0$; cependant il n'y a pas convergence en loi car F n'est pas une fonction de répartition.

6.5 La v.a. V_n étant le produit de termes qui valent 1 ou -1 ne prend aussi que les valeurs 1 et -1 . Si nous posons $p_n = P(V_n = 1)$ on obtient $E(V_n) = p_n - (1 - p_n) = 2p_n - 1$. D'autre part, les v.a. U_n étant indépendantes :

$$E(V_n) = \prod_{i=1}^n E(U_i) = \prod_{i=1}^n (2p - 1) = (2p - 1)^n$$

et par conséquent $p_n = \frac{1}{2} [1 + 2(p - 1)^n]$. Puisque $0 < p < 1$, on a aussi $|2p - 1| < 1$ donc $p_n \rightarrow \frac{1}{2}$ quand $n \rightarrow \infty$. Ainsi la suite (V_n) converge en loi vers la v.a. V dont la loi est définie par $P(V = 1) = P(V = -1) = \frac{1}{2}$.

6.6 La fonction de répartition de $Y_n = \frac{X_n}{n}$ est définie pour tout $x > 0$ par :

$$F_n(x) = P(Y_n < x) = P(X_n < nx) = \sum_{k=1}^{[nx]} p_n (1 - p_n)^{k-1} = 1 - (1 - p_n)^{[nx]}$$

si $nx \notin \mathbb{N}$, avec dans ce cas $nx = [nx] + \alpha$ où $0 < \alpha < 1$; on a alors :

$$\ln [1 - F_n(x)] = [nx] \ln \left(1 - \frac{\theta}{n}\right) = (nx - \alpha) \left(-\frac{\theta}{n} + \frac{\varepsilon_n}{n}\right) \rightarrow -\theta x$$

Si $nx \in \mathbb{N}$, alors :

$$F_n(x) = P(Y_n < x) = P(X_n < nx) = \sum_{k=1}^{nx-1} p_n (1 - p_n)^{k-1} = 1 - (1 - p_n)^{nx-1}$$

et :

$$\ln [1 - F_n(x)] = (nx - 1) \ln \left(1 - \frac{\theta}{n}\right) = (nx - 1) \left(-\frac{\theta}{n} + \frac{\varepsilon_n}{n}\right) \rightarrow -\theta x$$

Ainsi, dans tous les cas, pour $x > 0$ et quand $n \rightarrow \infty$:

$$F_n(x) \rightarrow 1 - e^{-\theta x}$$

la limite est la fonction de répartition sur \mathbb{R}_+ de la loi exponentielle de paramètre θ , qui est donc la loi limite de Y_n .

6.7 La fonction de répartition de X_n est définie par :

$$F_n(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{k}{n+1} & \text{si } \frac{k-1}{n} < x \leq \frac{k}{n}, 1 \leq k \leq n \\ 1 & \text{si } 1 < x \end{cases}$$

Pour $x \leq 0$ ou $x > 1$ on a $F_n(x) = F(x)$ où F est la fonction de répartition de la loi uniforme sur $[0, 1]$. Examinons maintenant le cas où $x \in [0, 1]$ et soit k l'entier tel que $\frac{k-1}{n} < x \leq \frac{k}{n}$. La fonction F étant strictement croissante :

$$F\left(\frac{k-1}{n}\right) < F(x) \leq F\left(\frac{k}{n}\right) \text{ soit } \frac{k-1}{n} < F(x) \leq \frac{k}{n}.$$

On obtient ainsi :

$$\frac{k}{n+1} - \frac{k}{n} \leq F_n(x) - F(x) < \frac{k}{n+1} - \frac{k-1}{n}$$

ou :

$$-\frac{k}{n(n+1)} \leq F_n(x) - F(x) < \frac{n+1-k}{n(n+1)}$$

Donc, pour $1 \leq k \leq n$:

$$-\frac{n}{n(n+1)} \leq F_n(x) - F(x) < \frac{n}{n(n+1)}$$

ce qui établit que $F_n(x) - F(x) \rightarrow 0$ quand $n \rightarrow \infty$ et donc que X_n converge vers la loi uniforme sur $[0, 1]$.

- 6.8** Pour tout $\varepsilon > 0$ il existe un entier $N = N(\varepsilon)$ tel que pour $n > N$ on ait $1/n < \varepsilon$ et alors :

$$P(|X_n - 1| < \varepsilon) = \frac{1}{2} + \frac{1}{2} = 1$$

Donc $P(|X_n - 1| < \varepsilon) \rightarrow 1$ quand $n \rightarrow \infty$ ce qui montre que la suite (X_n) converge en probabilité vers la v. a. certaine $X = 1$, et donc aussi en loi puisque la limite est une constante.

Cependant, $P(X_n = 1) = 0$ pour tout entier n et donc $P(X_n = 1) \rightarrow 0$ quand $n \rightarrow \infty$ alors que $P(X = 1) = 1$. Il faut donc faire attention à l'utilisation des conditions suffisantes de convergence en loi vues au § 2.5.

- 6.9** La lecture de la table 5 nous donne la valeur exacte du fractile d'ordre 0,95 de la loi χ_{40}^2 , soit $x = 55,76$. Cependant, beaucoup de tables de fractiles de la loi du χ_v^2 s'arrêtent à $v = 30$ et donc, pour un nombre plus élevé de degrés de liberté, il faut alors utiliser l'approximation proposée dans ces tables, soit :

$$\begin{aligned} P(X < x) = 0,95 &= P\left(\sqrt{2X} - \sqrt{2v-1} < \sqrt{2x} - \sqrt{2v-1}\right) \\ &\simeq \Phi\left(\sqrt{2x} - \sqrt{2v-1}\right) \end{aligned}$$

où Φ est la f.r. de la loi $N(0, 1)$. Par lecture de la table 2 des fractiles on obtient $\sqrt{2x} - \sqrt{2v-1} \simeq 1,6449$, soit avec ici $v = 40$: $x \simeq 55,47$, valeur assez proche de la valeur exacte.

- 6.10** Pour montrer que (X_n) converge en moyenne quadratique vers zéro, il faut étudier la limite de :

$$\begin{aligned} E(X_n^2) &= \int_0^{+\infty} x^2 n e^{-nx} dx = [-x^2 e^{-nx}]_0^{+\infty} + 2 \int_0^{+\infty} x e^{-nx} dx \\ &= \left[-2x \frac{e^{-nx}}{n}\right]_0^{+\infty} + \frac{2}{n} \int_0^{+\infty} e^{-nx} dx = \frac{2}{n^2} \rightarrow 0 \end{aligned}$$

ce qui exprime par définition que :

$$X_n \xrightarrow{m.q.} 0$$

- 6.11** Il faut d'abord déterminer la loi de probabilité de m_n :

$$P(m_n > x) = P\left\{\bigcap_{i=1}^n (X_i > x)\right\} = \prod_{i=1}^n P(X_i > x) = [1 - F(x)]^n$$

où F est la f.r. commune des variables X_i ; ainsi, m_n admet pour f.r. :

$$G_n(x) = 1 - [1 - F(x)]^n$$

et pour densité :

$$g_n(x) = n[1 - F(x)]^{n-1} f(x) = n e^{-n(x-\theta)}$$

pour $x > \theta$. On calcule alors :

$$\begin{aligned} E(m_n - \theta)^2 &= \int_{\theta}^{+\infty} (x - \theta)^2 n e^{-n(x-\theta)} dx = \frac{1}{n^2} \int_0^{+\infty} y^2 e^{-y} dy \\ &= \frac{\Gamma(3)}{n^2} = \frac{2}{n^2} \rightarrow 0 \end{aligned}$$

et par conséquent on en conclut que pour $n \rightarrow \infty$:

$$m_n \xrightarrow{m.q.} \theta$$

La f.r. de Y_n est définie par :

$$H_n(y) = P(Y_n < y) = P(m_n - \theta < \frac{y}{n}) = G_n\left(\theta + \frac{y}{n}\right)$$

et sa densité est donc :

$$h_n(y) = \frac{1}{n} g_n\left(\theta + \frac{y}{n}\right) = e^{-y}$$

pour $y > 0$, c'est-à-dire que Y_n suit la loi exponentielle de paramètre 1, indépendante de n , et qui est donc aussi la loi limite de Y_n .

- 6.12** Dans les deux cas, il s'agit d'une moyenne de v. a. qui vérifie les hypothèses du théorème central limite. Il nous suffit donc de calculer les deux premiers moments de ces v. a. On a d'abord :

$$E(|X|) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x| e^{-x^2/2\sigma^2} dx = \sigma\sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-u} du = \sigma\sqrt{\frac{2}{\pi}}$$

Et ensuite :

$$V(|X|) = E(X^2) - E^2(|X|) = \left(1 - \frac{2}{\pi}\right) \sigma^2$$

L'application du théorème central limite permet donc d'obtenir :

$$\sqrt{n} \frac{D_n - \sigma\sqrt{\frac{2}{\pi}}}{\sigma\sqrt{1 - \frac{2}{\pi}}} \xrightarrow{loi} N(0, 1)$$

On a ensuite $E(X^2) = \sigma^2$ et comme X^2/σ^2 suit une loi χ_1^2 on a aussi $V(X^2) = 2\sigma^4$. Ainsi :

$$\sqrt{n} \frac{S_n^2 - \sigma^2}{\sigma^2\sqrt{2}} \xrightarrow{loi} N(0, 1)$$

- 6.13** On vérifie d'abord que les v.a. $Z_n = e^{X_n}$ suivent une loi $\gamma(1)$. Nous appliquons ensuite le théorème central-limite à ces variables dont espérance et variance valent 1 :

$$\sqrt{n} (\bar{Z}_n - 1) \xrightarrow{loi} N(0, 1)$$

La suite étudiée est définie par $Y_n = -\ln \bar{Z}_n$ et on utilise alors la propriété 2.7 avec $a_n = \sqrt{n}$ et $g(t) = -\ln t$ qui nous permet d'obtenir :

$$\sqrt{n} Y_n \xrightarrow{loi} N(0, 1)$$

Introduction

Le problème qui va se poser au statisticien peut s'énoncer de la façon suivante : disposant d'observations x_1, \dots, x_n d'une certaine v.a. X associée au phénomène étudié, obtenues à partir d'expériences aléatoires identiques et indépendantes, quelle loi théorique P inconnue peut-on retenir comme loi parente, c'est-à-dire, quelle est la loi de probabilité associée à ce phénomène de production des données ? Ce qu'on peut symboliser sous la forme suivante :

$$x_1, \dots, x_n \rightarrow P ?$$

Compte tenu d'informations *a priori* dont il dispose, le statisticien va restreindre son choix à une famille donnée $(P_\theta; \theta \in \Theta)$ indexée par θ parcourant un ensemble Θ bien déterminé. On suppose donc ici que la loi de probabilité est parfaitement déterminée par la donnée d'un ou plusieurs nombres ($\Theta \subset \mathbb{R}$ ou $\Theta \subset \mathbb{R}^k$) représenté(s) par θ que nous appellerons **paramètre** de la distribution. Nous ferons l'hypothèse que pour $\theta \neq \theta'$ on a $P_\theta \neq P_{\theta'}$.

Le problème initial du statisticien peut alors se reformuler sous la forme symbolique suivante :

$$x_1, \dots, x_n \rightarrow \theta ?$$

et se nomme alors un problème d'**estimation**. Choisir une seule valeur du paramètre θ est un problème d'estimation ponctuelle, choisir un sous-ensemble de Θ , dénommé région de confiance, est un problème d'estimation ensembliste, d'estimation par intervalle dans le cas particulier, assez fréquent, où cette région est un intervalle de \mathbb{R} .

Objectifs

Montrer comment on peut, à partir d'observations indépendantes, attribuer une valeur, ou un ensemble de valeurs, à un (ou plusieurs) paramètres.

Présenter les principales propriétés d'un estimateur.

Concepts clés

- Biais d'un estimateur
- Estimateur convergent
- Erreur quadratique moyenne
- Efficacité
- Méthode du maximum de vraisemblance et méthode des moments

1 Définition d'un estimateur

Nous allons introduire la notion d'estimateur à partir d'un exemple familier issu des sondages qui, régulièrement, cherchent à connaître l'état de l'opinion publique vis-à-vis du président de la République. Le modèle statistique que l'on peut construire consiste à retenir comme ensemble Ω l'ensemble des électeurs français et à considérer le sous-ensemble A de ceux qui sont favorables au président de la République. La v.a. X associée sera alors la variable indicatrice de A , définie par :

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

La loi de probabilité de X est donc connue, c'est la loi de Bernoulli $B(1, p)$, et le paramètre qui définit complètement le modèle est ici $\theta = p = P(X = 1) = P(A)$. La famille de lois de probabilité est donc $\{B(1, p); p \in [0, 1]\}$ indexée par $\Theta = [0, 1]$. Pour se faire une « idée » de la vraie valeur de ce paramètre, on effectue un sondage en interrogeant n personnes tirées au hasard dans Ω et on associe à chacune d'elles une variable de Bernoulli $X_i, 1 \leq i \leq n$, de même loi que X .

Sans être statisticien, si on constate que 48 % des personnes interrogées sont favorables au président de la République, on en déduira qu'environ un Français sur deux y est favorable. En langage courant, on dira que l'on « estime » la valeur de p à 48 %. Cette valeur, calculée sur l'échantillon, est une **estimation**, cette fois en terme statistique, du paramètre p , obtenue à partir de la fréquence empirique $f_n = \frac{1}{n} \sum_{i=1}^n X_i$ qu'on appellera donc un **estimateur de p** .

Prenons un autre exemple. Supposons qu'avant de choisir un véhicule automobile on se fixe un critère de choix basé sur le nombre moyen N de pannes par an que l'on est susceptible d'avoir avec un modèle donné. Ayant la possibilité de faire une étude statistique chez un concessionnaire, on prélève au hasard n dossiers de véhicules et on note pour chacun le nombre $N_i, 1 \leq i \leq n$, de pannes subies la dernière année de mise en circulation. La loi de Poisson étant adaptée pour modéliser le nombre de pannes, on retient ici la famille de lois $\{\mathcal{P}(\theta); \theta \in \mathbb{R}_+\}$. Le paramètre est ici la moyenne de la loi, i.e. $\theta = E(N)$, donc on estime sa valeur par la moyenne des valeurs observées sur l'échantillon, soit :

$$\overline{N}_n = \frac{1}{n} \sum_{i=1}^n N_i$$

Dans ces deux exemples, on a donc construit un modèle statistique où la v.a. X suit une loi P_θ et pour se faire une idée de la valeur inconnue du paramètre θ qui détermine la loi de probabilité, on utilise un échantillon de cette loi.

À partir des valeurs observées x_1, \dots, x_n on calcule ensuite une certaine valeur numérique que l'on considérera comme une valeur approchée de θ et qu'on appellera une

estimation de θ . La règle qui permettra d'effectuer ce calcul est un **estimateur**, dont la définition précise est la suivante.

DÉFINITION

Un estimateur de θ est une application T_n de E^n dans F qui à un échantillon (X_1, \dots, X_n) de la loi P_θ associe une variable aléatoire réelle (ou plusieurs dans le cas d'un paramètre multidimensionnel) dont on peut déterminer la loi de probabilité.

Il s'agit d'un modèle paramétrique où $E = X(\Omega)$ est l'ensemble des valeurs possibles pour la v.a. X , le paramètre θ étant à valeur(s) réelle(s) ($F \subset \mathbb{R}$ ou $F \subset \mathbb{R}^k$).

La loi de la v.a. $T_n(X_1, \dots, X_n)$ dépend de celle de X , et donc de θ , et chaque réalisation $T_n(x_1, \dots, x_n)$ sera une estimation de θ . Cette définition est extrêmement générale et ne donne pas de méthode de construction d'un estimateur. Cependant, comme nous l'avons vu sur les exemples introductifs, l'expression de l'estimateur découle très naturellement de l'interprétation que l'on peut très souvent donner du paramètre. Nous avons dans ces exemples $\theta = E(X)$, c'est-à-dire la moyenne théorique de la loi, et on retient donc très logiquement comme estimateur du paramètre θ la moyenne empirique ou moyenne de l'échantillon :

$$T_n(X_1, \dots, X_n) = \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Ce n'est que dans certains cas, assez rares, où le paramètre n'admet pas d'interprétation évidente, que l'on est amené à utiliser une méthode de construction d'un estimateur, comme la méthode des moments, généralisant cette méthode intuitive, ou d'autres méthodes aux propriétés voisines, comme la méthode du maximum de vraisemblance et la méthode des moindres carrés.

Un estimateur est une statistique particulière, en tant que fonction de l'échantillon, qui permet d'attribuer une valeur au paramètre θ à estimer. On pourrait donc le définir comme une statistique à valeurs dans Θ . Comme nous n'avons ici imposé aucune condition à l'ensemble F , nous allons définir une propriété que l'on pourrait qualifier de minimale pour un estimateur, c'est-à-dire de prendre ses valeurs dans le même ensemble que le paramètre. On dira qu'un estimateur est *strict* si $T_n(E^n) \subset \Theta$.

Exemple 7.1

Si X suit une loi de Bernoulli $B(1, \theta)$, T_n sera un estimateur strict si $0 \leq T_n(x_1, \dots, x_n) \leq 1$ pour tout échantillon observé (x_1, \dots, x_n) .

Nous allons maintenant définir les propriétés essentielles que doit vérifier un estimateur.

2 Propriétés d'un estimateur

Ayant fait choix d'un estimateur, on se pose la question de savoir s'il s'agit d'un « bon » estimateur. Il va donc falloir se donner des critères de qualité. Par ailleurs, il se peut que pour certains problèmes on ait le choix entre plusieurs estimateurs et il va donc falloir aussi définir un critère de comparaison pour déterminer quel est le meilleur. Enfin, on peut se poser la question de savoir s'il n'existerait pas un estimateur qui, pour un problème donné, ne serait pas meilleur que tous les autres : c'est le problème de l'optimalité. Toutes ces questions vont trouver des réponses dans l'étude des propriétés ci-après où nous supposerons pour simplifier que θ est un paramètre réel, c'est-à-dire que $\Theta \subset \mathbb{R}$. Ces propriétés peuvent bien sûr s'étendre au cas d'un paramètre multidimensionnel.

2.1 Biais d'un estimateur

Bien entendu, pour pouvoir considérer $T_n(x_1, \dots, x_n)$ comme une valeur approchée de θ , il faut que les valeurs prises par la v.a. T_n ne s'écartent pas trop de la valeur, fixe, de θ . Comme T_n est une v.a., on ne peut imposer une condition qu'à sa valeur moyenne, ce qui nous amène à définir le *biais* d'un estimateur comme l'écart entre sa moyenne et la vraie valeur du paramètre :

$$b_n(\theta) = E_\theta(T_n) - \theta$$

D'où une propriété relative au biais d'un estimateur.

DÉFINITION

Un estimateur T_n de θ est dit *sans biais* si pour tout θ de Θ et tout entier positif n :

$$E_\theta(T_n) = \theta$$

Exemple 7.2

Si le paramètre à estimer est la moyenne théorique de la loi, i.e. $\theta = E(X)$, l'estimateur naturel est la moyenne empirique $T_n = \bar{X}_n$. Nous avons vu dans le chapitre 5 que $E_\theta(T_n) = E(X) = \theta$, donc nous en déduisons le résultat très général que la moyenne empirique est toujours un estimateur sans biais de la moyenne théorique (espérance mathématique), quelle que soit la loi de probabilité de la variable X .

Cependant, cette propriété peut ne pas être strictement vérifiée, le biais diminuant seulement quand la taille d'échantillon augmente. Ceci correspond à la définition suivante.

DÉFINITION

Un estimateur T_n de θ est dit **asymptotiquement sans biais** si pour tout θ de Θ :

$$E_\theta(T_n) \rightarrow \theta \quad \text{quand } n \rightarrow \infty$$

Exemple 7.3

Si le paramètre à estimer est la variance théorique de la loi, i.e. $\theta = V(X)$, l'estimateur naturel est la variance empirique $T_n = S_n'^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ et nous avons vu dans le chapitre 5 que $E_\theta(T_n) = \frac{n-1}{n}\theta$. Donc $S_n'^2$ est un estimateur asymptotiquement sans biais de θ . Nous avons vu aussi que $E_\theta(S_n^2) = V(X) = \theta$, où $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est la variance empirique modifiée qui est donc toujours un estimateur sans biais de la variance théorique, quelle que soit la loi de probabilité de la variable X .

2.2 Convergence d'un estimateur

Intuitivement, on est amené à penser que si la taille de l'échantillon augmente, l'information sur le paramètre θ va augmenter et donc l'estimateur devrait d'une certaine manière se « rapprocher » de la valeur de θ . Cet estimateur étant une v.a., la formulation mathématique de cette notion intuitive va faire appel à la convergence en probabilité d'une suite de v.a. : T_n prendra des valeurs proches de θ avec une probabilité d'autant plus proche de un que la taille d'échantillon n sera grande. Ceci conduit à la définition suivante.

DÉFINITION

Un estimateur T_n est *convergent* si la suite de v.a. (T_n) converge en probabilité vers la valeur du paramètre, c'est-à-dire pour tout θ de Θ :

$$T_n \xrightarrow{p} \theta \Leftrightarrow P_\theta(|T_n - \theta| < \epsilon) \rightarrow 1, \quad \forall \epsilon > 0, n \rightarrow \infty$$

$$\Leftrightarrow P_\theta(|T_n - \theta| > \epsilon) \rightarrow 0$$

Il existe un moyen en général simple de vérifier cette propriété de convergence, à partir de conditions suffisantes faisant intervenir les deux premiers moments de l'estimateur, et que nous énonçons dans le théorème ci-après.

Théorème 7.1

Tout estimateur sans biais dont la variance tend vers zéro est convergent :

$$\left. \begin{array}{l} E_\theta(T_n) = \theta \\ V_\theta(T_n) \rightarrow 0 \end{array} \right\} \Rightarrow T_n \xrightarrow{p} \theta, \quad n \rightarrow \infty$$

Ce résultat se déduit directement de l'inégalité de Bienaymé-Tchebychev :

$$P_\theta(|T_n - \theta| > \epsilon) \leq \frac{V_\theta(T_n)}{\epsilon^2} \rightarrow 0, \quad \forall \epsilon > 0, n \rightarrow \infty$$

En fait, sous les hypothèses du théorème, on obtient une convergence plus forte qui est la convergence en moyenne quadratique puisqu'ici

$$V_{\theta}(T_n) = E_{\theta}(T_n - \theta)^2 \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Exemple 7.4

Si $\theta = E(X)$, ce paramètre est estimé sans biais par $T_n = \overline{X}_n$, et on sait de plus que $V_{\theta}(T_n) = V(X)/n$ donc $V_{\theta}(T_n) \rightarrow 0$ quand $n \rightarrow \infty$ et T_n est convergent (résultat que l'on aurait pu déduire directement de la loi des grands nombres). Ainsi, la moyenne empirique est un estimateur sans biais et convergent de la moyenne théorique $E(X)$, quelle que soit la loi de X .

Le résultat précédent peut être obtenu sous des conditions un peu plus générales énoncées ci-après.

Théorème 7.2

Tout estimateur asymptotiquement sans biais dont la variance tend vers zéro est convergent :

$$\left. \begin{array}{l} E_{\theta}(T_n) \rightarrow \theta \\ V_{\theta}(T_n) \rightarrow 0 \end{array} \right\} \Rightarrow T_n \xrightarrow{p} \theta, \quad n \rightarrow \infty$$

Là encore, ces conditions suffisantes de convergence en probabilité de T_n établissent en fait que cette suite converge en moyenne quadratique.

2.3 Estimateur optimal

a) Qualité d'un estimateur

La qualité d'un estimateur va se mesurer à l'aide d'une distance au paramètre qui peut être par exemple $|T_n - \theta|$ ou $(T_n - \theta)^2$. Pour obtenir un indicateur numérique on peut alors déterminer la valeur moyenne de cette distance. L'indicateur généralement retenu, car il se prête facilement aux calculs, est l'**erreur quadratique moyenne** définie pour tout θ par :

$$EQ(T_n) = E_{\theta}(T_n - \theta)^2 = V_{\theta}(T_n) + b_n^2(\theta)$$

Dans le cas particulier d'un estimateur sans biais, cette erreur quadratique se confond avec la variance de l'estimateur. Si dans l'erreur totale d'estimation on privilégie l'erreur structurelle, mesurée par $b_n^2(\theta)$, on fera choix d'un estimateur sans biais et l'erreur d'estimation se réduira à l'erreur statistique mesurée par la variance de l'estimateur. Si on se place donc dorénavant dans la classe des estimateurs sans biais, on pourra comparer deux estimateurs T_n et T'_n de cette classe par leur variance qui mesure alors leur dispersion par rapport au paramètre qui est leur espérance commune. Nous dirons que

l'estimateur T_n est plus **efficace** que T'_n si pour tout θ de Θ et pour une taille d'échantillon $n > N$:

$$V_{\theta}(T_n) \leq V_{\theta}(T'_n)$$

La question se pose alors de savoir si on pourrait trouver un troisième estimateur qui serait à son tour meilleur que T_n . En cas de réponse positive, il faudrait poursuivre la recherche, ce qui nous conduirait à essayer d'améliorer indéfiniment un estimateur. Le problème n'admettrait une fin que si l'on savait que l'estimateur obtenu est le meilleur. Le paragraphe suivant va fournir des éléments de réponse.

b) Inégalité de Fréchet-Darmois-Cramer-Rao

Nous allons voir que dans certaines conditions il existe une borne inférieure pour l'ensemble des variances des estimateurs sans biais, ce qui va constituer un butoir ne permettant pas d'améliorer sans cesse les estimateurs. D'autre part, si cette borne est atteinte par un estimateur, il deviendra le meilleur et sera qualifié d'optimal dans la classe des estimateurs sans biais. Pour énoncer ce résultat, nous avons besoin d'introduire la définition suivante.

DÉFINITION

On appelle **vraisemblance** (*likelihood*) de l'échantillon (X_1, \dots, X_n) la loi de probabilité de ce n -uplet, notée $L(x_1, \dots, x_n; \theta)$, et définie par :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n P(X_i = x_i | \theta)$$

si X est une v.a. discrète, et par :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

si X est une v.a. continue de densité $f(x; \theta)$.

Le théorème suivant va préciser la borne inférieure pour la variance des estimateurs sans biais, sous certaines hypothèses relatives à la loi de probabilité de X et que nous appellerons hypothèses de Cramer-Rao. Nous ne donnerons pas le détail de ces hypothèses (voir compléments) qui sont essentiellement des conditions techniques portant sur l'existence de dérivées de la densité f de X et la possibilité d'intervertir les opérations de dérivation et d'intégration. Cependant, nous ferons figurer dans l'énoncé du théorème la condition indispensable pour que cette inégalité soit vraie et qui n'est pas satisfaite pour certaines lois usuelles, alors que les autres hypothèses sont généralement vérifiées. Enfin, cet énoncé fait intervenir la notion de **quantité d'information de Fisher** qui est définie par :

$$I_n(\theta) = E_{\theta} \left(\frac{\partial \ln L}{\partial \theta} \right)^2$$

Théorème 7.3

Sous les hypothèses de Cramer-Rao, en particulier si $E = X(\Omega)$ est indépendant du paramètre à estimer θ , pour tout estimateur sans biais T_n de θ on a :

$$V_\theta(T_n) \geq \frac{1}{I_n(\theta)} = B_F(\theta)$$

La quantité $B_F(\theta)$ est la borne inférieure de Fréchet-Darmois-Cramer-Rao (FDCR en abrégé). Notons que dans les conditions d'application de ce théorème, en particulier si $E = X(\Omega)$ est indépendant du paramètre à estimer θ , on obtient une expression équivalente de la quantité d'information de Fisher qui est généralement plus simple à calculer :

$$I_n(\theta) = E_\theta \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right)$$

Exemple 7.5

Soit X une v.a. de loi exponentielle de paramètre $1/\theta$, ou loi $\gamma(1, 1/\theta)$, avec $\theta > 0$, de densité pour $x > 0$:

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}$$

La vraisemblance admet ici pour expression :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \frac{1}{\theta^n} \exp \left(-\frac{1}{\theta} \sum_{i=1}^n x_i \right)$$

Pour calculer la quantité d'information de Fisher nous écrivons la log-vraisemblance :

$$\ln L(x_1, \dots, x_n; \theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n x_i$$

Nous dérivons par rapport au paramètre :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i$$

Soit en élevant au carré :

$$\left(\frac{\partial \ln L}{\partial \theta} \right)^2 = \frac{1}{\theta^2} \left[n^2 - 2\frac{n}{\theta} \sum_{i=1}^n x_i + \frac{1}{\theta^2} \left(\sum_{i=1}^n x_i \right)^2 \right]$$

Si on pose $S_n = \sum_{i=1}^n X_i$, on obtient :

$$I_n(\theta) = \frac{1}{\theta^2} \left[n^2 - 2\frac{n}{\theta} E_\theta(S_n) + \frac{1}{\theta^2} E_\theta(S_n^2) \right]$$

Avec $E_\theta(S_n) = nE_\theta(X) = n\theta$, $E_\theta(S_n^2) = V_\theta(S_n) + E_\theta^2(S_n) = nV_\theta(X) + n^2\theta^2 = n(n+1)\theta^2$ on obtient :

$$I_n(\theta) = \frac{n}{\theta^2}$$

Comme $X(\Omega) = \mathbb{R}_+$ est indépendant de θ , on peut utiliser la seconde expression de la quantité d'information de Fisher, calculée à partir de :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} S_n$$

ce qui permet d'obtenir plus facilement :

$$I_n(\theta) = E_\theta\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right) = -\frac{n}{\theta^2} + 2\frac{n\theta}{\theta^3} = \frac{n}{\theta^2}$$

Exemple 7.6

Si nous prenons maintenant l'exemple de la loi exponentielle sur $[\theta, +\infty[$, de densité :

$$f(x; \theta) = \begin{cases} e^{-(x-\theta)} & \text{si } x \geq \theta \\ 0 & \text{sinon} \end{cases}$$

La vraisemblance s'écrit :

$$L(x_1, \dots, x_n; \theta) = \exp\left\{-\sum_{i=1}^n (x_i - \theta)\right\}$$

si tous les x_i sont plus grands que θ , c'est-à-dire si $\min\{x_i/1 \leq i \leq n\} \geq \theta$.

On a alors :

$$\ln L(x_1, \dots, x_n; \theta) = -\sum_{i=1}^n x_i + n\theta$$

d'où $\frac{\partial \ln L}{\partial \theta} = n$ et :

$$I_n(\theta) = E_\theta\left(\frac{\partial \ln L}{\partial \theta}\right)^2 = n^2$$

On peut remarquer ici que $\frac{\partial^2 \ln L}{\partial \theta^2} = 0$ et par conséquent $E_\theta\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right) = 0$ ne coïncide pas avec la valeur de la quantité d'information de Fisher, ce qui s'explique par le fait qu'ici $X(\Omega) = [\theta, +\infty[$ dépend du paramètre à estimer θ .

Estimateur efficace

Le théorème précédent fournit une borne inférieure pour la variance des estimateurs sans biais, qui peut ou non être atteinte. Si cette borne est effectivement atteinte par un

estimateur, il sera donc le meilleur, selon ce critère, dans la classe des estimateurs sans biais. Cette optimalité se traduit par la définition suivante.

DÉFINITION

Un estimateur sans biais T_n est dit **efficace** si sa variance est égale à la borne inférieure de FDCR :

$$V_{\theta}(T_n) = \frac{1}{I_n(\theta)}$$

Exemple 7.7

Si nous reprenons l'exemple de la loi exponentielle de paramètre $1/\theta$, comme $E_{\theta}(X) = \theta$, on sait que $T_n = \bar{X}_n$ est un estimateur sans biais et convergent. De plus :

$$V_{\theta}(T_n) = V_{\theta}(\bar{X}_n) = \frac{V_{\theta}(X)}{n} = \frac{\theta^2}{n} = \frac{1}{I_n(\theta)}$$

donc cet estimateur est aussi efficace.

REMARQUE Un estimateur efficace est bien sûr optimal, mais dans la classe des estimateurs sans biais. Si on utilise comme critère l'erreur quadratique, qui est une mesure de l'erreur totale où les erreurs structurelle et statistique jouent le même rôle, on peut trouver un estimateur qui soit meilleur qu'un estimateur efficace. Par exemple, dans le cas d'un échantillon d'une loi $N(0, \sigma)$, on sait que $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ est un estimateur efficace de σ^2 , avec $EQ(\sigma_n^2) = V(\sigma_n^2) = \frac{2\sigma^4}{n}$. Mais, si on retient l'estimateur avec biais $T_n = \frac{1}{n+2} \sum_{i=1}^n X_i^2$, on obtient un estimateur meilleur, c'est-à-dire d'erreur totale plus faible puisque $EQ(T_n) = \frac{2\sigma^4}{n+2} < EQ(\sigma_n^2)$.

3 Méthodes de construction d'un estimateur

Dans les situations où il n'y a pas d'estimateur évident, on est amené à recourir à une méthode de construction d'un estimateur, les deux méthodes que nous présenterons ici étant celles du **maximum de vraisemblance** et des **moments**.

3.1 Méthode du maximum de vraisemblance

La vraisemblance $L(x_1, \dots, x_n; \theta)$ représente la probabilité d'observer le n -uple (x_1, \dots, x_n) pour une valeur fixée de θ . Dans la situation inverse ici où on a observé (x_1, \dots, x_n) sans connaître la valeur de θ , on va attribuer à θ la valeur qui paraît la plus vraisemblable, compte tenu de l'observation dont on dispose, c'est-à-dire celle qui va lui attribuer la plus forte probabilité. On se fixe donc la règle suivante : à (x_1, \dots, x_n) fixé, on considère la vraisemblance L comme une fonction de θ et on attribue à θ la valeur qui maximise cette fonction. D'où la définition suivante.

DÉFINITION

On appelle **estimateur du maximum de vraisemblance** (emv) toute fonction $\hat{\theta}_n$ de (x_1, \dots, x_n) qui vérifie :

$$L(x_1, \dots, x_n; \hat{\theta}_n) = \max_{\theta \in \Theta} L(x_1, \dots, x_n; \theta)$$

Cette définition ne renseigne en aucune façon, ni sur l'existence, ni sur l'unicité, d'un tel estimateur. La recherche de l'emv peut se faire directement par recherche du maximum de L , ou dans le cas particulier où la fonction L est deux fois dérivable par rapport à θ , comme solution de l'équation $\frac{\partial L}{\partial \theta} = 0$ qui vérifie aussi $\frac{\partial^2 L}{\partial \theta^2} < 0$. Cependant, la vraisemblance se calculant à partir d'un produit, on préfère remplacer ce dernier problème par le problème équivalent pour la log-vraisemblance, puisque la fonction \ln est strictement croissante, $\frac{\partial \ln L}{\partial \theta} = 0$ avec $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$ et qui aura une expression généralement simplifiée. Remarquons enfin que si $\hat{\theta}_n$ est un emv du paramètre θ , alors $g(\hat{\theta}_n)$ est un emv du paramètre $g(\theta)$ pour toute fonction g . Si par exemple la variance empirique modifiée S_n^2 , qui est un estimateur sans biais de $\theta = V_\theta(X)$, est un emv pour un modèle statistique donné, alors S_n est un emv du paramètre $g(\theta) = \sigma_\theta(X) = \sqrt{V_\theta(X)} = \sqrt{\theta}$. Notons cependant que S_n ne peut pas être aussi un estimateur sans biais car on aurait alors $V_\theta(S_n) = E_\theta(S_n^2) - E_\theta^2(S_n) = \theta - \theta = 0$.

Exemple 7.8

Cherchons l'emv pour la famille de lois exponentielles de paramètre $1/\theta$. La log-vraisemblance est indéfiniment dérivable pour $\theta > 0$ et nous avons obtenu dans l'exemple 7.5 :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i$$

qui s'annule en changeant de signe pour $\theta = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$, avec :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n x_i = \frac{n}{\theta^3} (\theta - 2\bar{x}_n)$$

soit pour $\theta = \bar{x}_n$:

$$\left(\frac{\partial^2 \ln L}{\partial \theta^2} \right)_{\theta=\bar{x}_n} = -\frac{n}{\bar{x}_n^2} < 0$$

donc l'emv est $\hat{\theta}_n = \bar{X}_n$.

Exemple 7.9

Dans le cas de la loi exponentielle sur $[\theta, +\infty[$, la vraisemblance avait pour expression :

$$L(x_1, \dots, x_n; \theta) = \begin{cases} e^{n\theta} e^{-n\bar{x}_n} & \theta \leq \min \{x_i/1 \leq i \leq n\} \\ 0 & \min \{x_i/1 \leq i \leq n\} < \theta \end{cases}$$

Pour $\theta \leq \min \{x_i/1 \leq i \leq n\}$ la vraisemblance est une fonction croissante de θ et ensuite elle s'annule ; l'allure générale du graphe de L (cf. figure 7.1) montre bien que L est maximum pour $\theta = \min \{x_i/1 \leq i \leq n\}$, ce qui correspond à l'emv :

$$\hat{\theta}_n = \min \{X_i/1 \leq i \leq n\}$$

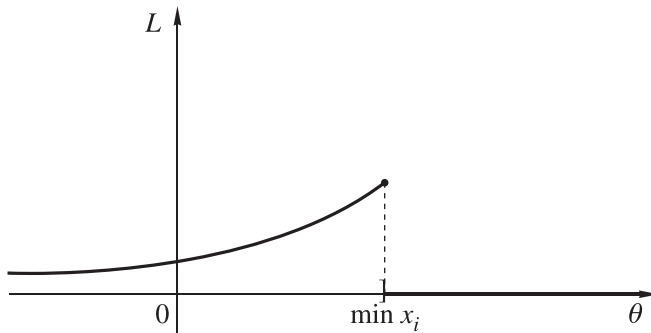


Figure 7.1

REMARQUE Dans l'exemple 7.9, nous avons $\ln L = n\theta - n\bar{x}_n$ et $\frac{\partial \ln L}{\partial \theta} = n$ ne s'annule pas ; la recherche du maximum de L (ou de $\ln L$) s'est effectuée directement, sans avoir recours aux dérivées. Rappelons à ce propos que dans la recherche des extremums d'une fonction réelle g , la condition $g'(x_0) = 0$ n'est ni nécessaire, ni suffisante, pour qu'il existe un extremum en x_0 . Cette recherche d'extremum doit se faire directement, à partir de la condition qui le définit, c'est-à-dire $g(x) - g(x_0)$ a un signe constant au voisinage de x_0 . Ce n'est que dans le cas d'une fonction deux fois dérivable sur un intervalle ouvert, que le maximum est obtenu à partir des conditions $g'(x_0) = 0$ et $g''(x_0) < 0$.

3.2 Méthode des moments

Dans le cas où le paramètre à estimer est $\theta = E_\theta(X)$, moyenne théorique de la loi, nous avons vu que l'estimateur naturel était la moyenne empirique, ou moyenne de l'échantillon, \bar{X}_n . De même, pour estimer le paramètre $\theta = V_\theta(X)$, variance de la loi, nous retenons logiquement comme estimateur la variance empirique $S_n'^2$. Plus généralement, si l'un des moments d'ordre $k \in \mathbb{N}^*$, non centré $m_k = E_\theta(X^k) = m_k(\theta)$, ou centré $\mu_k = E_\theta(X - m_1)^k = \mu_k(\theta)$, dépend de θ , nous allons chercher un estimateur par résolution de l'équation en θ obtenue en égalant moment théorique et moment empirique correspondant, soit :

$$m_{kn} = \frac{1}{n} \sum_{i=1}^n X_i^k = m_k(\theta) \quad \text{ou} \quad \mu_{kn} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k = \mu_k(\theta)$$

La solution de l'équation, si elle existe et est unique, sera appelée estimateur obtenu par la **méthode des moments**. Dans les exemples introductifs où $\theta = E_\theta(X)$ et $\theta = V_\theta(X)$, les équations à résoudre s'écrivent sous forme résolue $\theta = \bar{X}_n$ et $\theta = S_n'^2$.

Exemple 7.10

Si X suit une loi exponentielle de paramètre θ , on sait que $E_\theta(X) = 1/\theta$ et l'équation à résoudre s'écrit $\bar{X}_n = 1/\theta$, de solution immédiate $\theta = 1/\bar{X}_n$, qui correspond à l'estimateur obtenu par la méthode des moments :

$$\hat{\theta}_n = 1/\bar{X}_n$$

Bien entendu, on pourrait utiliser cette méthode avec des moments d'ordres plus élevés et obtenir ainsi d'autres estimateurs. En utilisant par exemple la variance $V_\theta(X) = 1/\theta^2$ on obtient le nouvel estimateur $\hat{\theta}_n = 1/S_n'^2$.

Cette méthode intuitive se justifie par les propriétés de convergence des moments empiriques vers les moments théoriques correspondants (cf. chap. 6, 2.8).

4 Estimation par intervalle de confiance

4.1 Exemple introductif

Un industriel commande un lot de tiges métalliques qu'il ne peut utiliser que si leur longueur est comprise entre 23,60 mm et 23,70 mm. Ces tiges ont été fabriquées par une machine qui, lorsqu'elle est réglée à la valeur m , produit des tiges dont la longueur peut être considérée comme une v.a. X de loi normale $N(m, \sigma)$, où l'écart type σ est une caractéristique de la machine, de valeur connue, ici $\sigma = 0,02$ mm. Compte tenu de la symétrie de la distribution normale, la proportion de tiges utilisables par l'industriel sera maximale si le réglage a été effectué à $m_0 = 23,65$ mm. Ne connaissant pas cette valeur, à la réception d'un lot de tiges l'industriel prélève au hasard n tiges dont il mesure les longueurs X_1, \dots, X_n pour se faire une idée de la valeur du paramètre de réglage m . Il calcule la moyenne des longueurs observées et, ayant obtenu la valeur $\bar{x}_n = 23,63$, il en conclut que, s'il est peu réaliste de croire que la valeur de m est exactement 23,63 mm, elle doit malgré tout être très proche de cette valeur moyenne observée sur l'échantillon. Il lui paraît raisonnable d'aboutir à une conclusion de la forme « il y a 95 chances sur 100 que la valeur de m soit comprise entre $23,63 - a$ et $23,63 + b$ ». Le problème consiste alors à fixer des valeurs précises pour a et b et on conçoit bien qu'elles doivent dépendre des « chances » que l'on a attribué à cet intervalle de contenir effectivement la vraie valeur de m . L'intervalle ainsi obtenu s'appellera **intervalle de confiance** et sa probabilité qui a permis de le déterminer, **niveau de confiance**. La longueur de cet

intervalle sera bien sûr proportionnelle à ce niveau de confiance. On peut par exemple toujours fournir un intervalle qui contient avec certitude le paramètre en le choisissant suffisamment large ; mais dans ce cas, cet intervalle ne nous renseigne en aucune façon sur la vraie valeur du paramètre. Il faut donc arriver à un compromis entre un intervalle pas trop grand et une probabilité assez élevée de contenir le paramètre.

Pour une famille quelconque de lois de probabilité $(P_\theta; \theta \in \Theta)$ on peut donner la définition suivante.

DÉFINITION

Un **intervalle de confiance** pour le paramètre θ , de **niveau de confiance** $1 - \alpha \in]0, 1[$, est un intervalle qui a la probabilité $1 - \alpha$ de contenir la vraie valeur du paramètre θ .

4.2 Principe de construction

Dans l'exemple précédent, nous avons abouti à un intervalle de la forme $\bar{x}_n - a < m < \bar{x}_n + b$ qui correspond à la réalisation d'un événement devant se produire avec une probabilité fixée $1 - \alpha$. La détermination des valeurs de a et b va donc se faire à partir de la valeur $1 - \alpha$ de la probabilité, fixée par le statisticien, à partir de la condition qui s'écrit ici :

$$1 - \alpha = P \left\{ \bar{X}_n - a < m < \bar{X}_n + b \right\}$$

qui est équivalente à :

$$1 - \alpha = P \left\{ -b < \bar{X}_n - m < a \right\}.$$

Il n'y a donc qu'une seule condition pour déterminer ces deux valeurs ; cependant, la loi de la v.a. $\bar{X}_n - m$ qui sert à construire cet intervalle étant symétrique, on choisit $b = a$ et on utilise la variable centrée et réduite pour déterminer la valeur de a qui vérifie la condition :

$$1 - \alpha = P \left\{ -\frac{a}{\sigma/\sqrt{n}} < \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} < \frac{a}{\sigma/\sqrt{n}} \right\}$$

Si Φ est la f.r. de la loi $N(0, 1)$, alors a est solution de :

$$1 - \alpha = \Phi \left(a\sqrt{n}/\sigma \right) - \Phi \left(-a\sqrt{n}/\sigma \right) = 2\Phi \left(a\sqrt{n}/\sigma \right) - 1$$

ou $1 - \alpha/2 = \Phi \left(a\sqrt{n}/\sigma \right)$, soit $a\sqrt{n}/\sigma = \Phi^{-1} (1 - \alpha/2)$. Pour un niveau de confiance de 0,95, soit $\alpha = 0,05$, et pour une taille d'échantillon $n = 100$, le fractile d'ordre 0,975 de la loi $N(0, 1)$ a pour valeur 1,96 et on en déduit $a = 0,004$, d'où l'intervalle :

$$23,626 < m < 23,634$$

obtenu pour cet échantillon particulier.

À partir d'un intervalle que l'on souhaitait obtenir pour le paramètre et dont les bornes s'exprimaient en fonction d'un estimateur de ce paramètre, on est arrivé à un intervalle pour l'estimateur, qui pouvait être déterminé avec précision puisqu'on connaissait sa loi. Ceci va nous fournir le principe de construction d'un intervalle de confiance en effectuant la démarche en sens contraire.

Le point de départ est fourni par un estimateur T_n du paramètre θ , construit à partir d'un échantillon (X_1, \dots, X_n) et dont on connaît la loi en fonction de θ , ce qui va permettre de déterminer les valeurs $t_1 = t_1(\theta)$ et $t_2 = t_2(\theta)$ telles que :

$$P_{\theta}\{t_1(\theta) \leq T_n \leq t_2(\theta)\} = 1 - \alpha$$

Il faut ensuite inverser cet intervalle pour T_n , dont les bornes dépendent du paramètre θ , pour obtenir un intervalle pour θ , dont les bornes vont dépendre de l'estimateur T_n , c'est-à-dire déterminer les valeurs $a = a(T_n)$ et $b = b(T_n)$ telles que :

$$P_{\theta}\{a(T_n) \leq \theta \leq b(T_n)\} = 1 - \alpha$$

On peut écrire également $P_{\theta}\{\theta \in [a(T_n), b(T_n)]\} = 1 - \alpha$ et on voit ainsi que $[a(T_n), b(T_n)]$ est un intervalle de confiance de niveau $1 - \alpha$ pour le paramètre θ . Il s'agit d'un intervalle aléatoire, au sens où ses bornes sont des v.a., dont chaque échantillon fournira une réalisation. Pour obtenir cet intervalle, il faut donc trouver l'ensemble des valeurs de θ pour lesquelles on a simultanément, pour T_n fixé :

$$t_1(\theta) \leq T_n \text{ et } T_n \leq t_2(\theta)$$

Cet ensemble sera plus ou moins facile à obtenir selon le comportement des fonctions t_1 et t_2 . Si par exemple ces deux fonctions sont croissantes, on voit sur la figure 7.2 que :

$$t_1(\theta) \leq T_n \Leftrightarrow \theta \leq t_1^{-1}(T_n) \quad \text{et} \quad T_n \leq t_2(\theta) \Leftrightarrow \theta \geq t_2^{-1}(T_n)$$

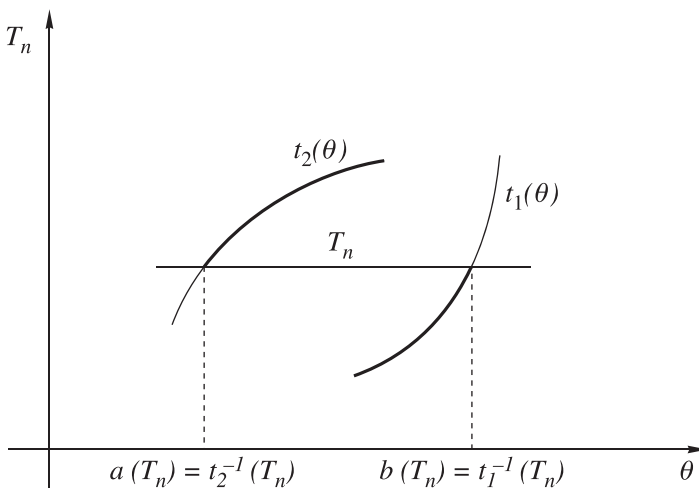


Figure 7.2

L'intervalle est facile à obtenir ici car les fonctions t_1 et t_2 sont inversibles, et on obtient alors :

$$t_1(\theta) \leq T_n \leq t_2(\theta) \Leftrightarrow a(T_n) = t_2^{-1}(T_n) \leq \theta \leq t_1^{-1}(T_n) = b(T_n)$$

Il existe cependant une part d'arbitraire dans le choix simultané de t_1 et t_2 puisqu'il n'y a qu'une seule condition pour déterminer ces deux valeurs, qui peut d'ailleurs aussi s'écrire $P_\theta(T_n < t_1) + P_\theta(T_n > t_2) = \alpha$, le risque total α pouvant être *a priori* réparti de multiples façons. Posons $\alpha_1 = P_\theta\{\theta > b(T_n)\}$ et $\alpha_2 = P_\theta\{\theta < a(T_n)\}$; les différents choix possibles sont les suivants.

a) Intervalle bilatéral ($\alpha_1 \alpha_2 > 0$)

• **Symétrique** : $\alpha_1 = \alpha_2 = \alpha/2$

C'est le choix que l'on fait si la loi de T_n est symétrique, ou si on n'a aucune information particulière, ce choix étant le moins arbitraire.

• **Dissymétrique** : $\alpha_1 \neq \alpha_2$

Seules des raisons très particulières peuvent permettre de fixer les valeurs de α_1 et α_2 telles que $\alpha_1 + \alpha_2 = \alpha$.

b) Intervalle unilatéral ($\alpha_1 \alpha_2 = 0$)

• **À droite** : $\alpha_1 = 0, \alpha_2 = \alpha$

C'est l'interprétation donnée au paramètre θ , comme par exemple la résistance d'un matériau qui doit être supérieure à un seuil minimum, qui conduit à un intervalle de la forme $\theta > a(T_n)$.

• **À gauche** : $\alpha_1 = \alpha, \alpha_2 = 0$

Si par exemple le paramètre est une proportion de défectueux dans un lot, ou un taux d'impuretés, on souhaite qu'il soit inférieur à un seuil maximum, d'où un intervalle de la forme $\theta < b(T_n)$.

4.3 Intervalle pour une proportion

Nous allons voir sur cet exemple que l'inversion de l'intervalle pour la statistique n'est pas toujours aisée, surtout lorsque sa loi est discrète. Supposons que l'on ait effectué un sondage pour déterminer les intentions de vote pour un certain candidat à une élection présidentielle. À chaque individu interrogé, on associe une variable de Bernoulli qui prend la valeur 1 si cet individu i déclare vouloir voter en faveur de ce candidat :

$$X_i = \begin{cases} 1 & p \\ 0 & q = 1 - p \end{cases} \quad 1 \leq i \leq n$$

Le paramètre p de cette loi de Bernoulli représente la proportion, dans la population, d'électeurs favorables à ce candidat et est estimé par la proportion observée dans l'échantillon :

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Nous savons que, en tant que moyenne empirique estimant une moyenne théorique, cet estimateur naturel de p est sans biais et convergent. Nous allons établir maintenant que c'est aussi l'estimateur du maximum de vraisemblance et qu'il est efficace. Pour cela, on suppose que la taille de la population est suffisamment grande pour que les variables X_i puissent être considérées comme indépendantes, c'est-à-dire que l'on se place dans le cadre d'un schéma binomial avec remise. La vraisemblance s'écrit alors :

$$L(x_1, \dots, x_n; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

avec $x_i \in \{0, 1\}$. On obtient donc comme expression de la log-vraisemblance :

$$\ln L(x_1, \dots, x_n; p) = \left(\sum_{i=1}^n x_i \right) \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln(1-p)$$

d'où en dérivant :

$$\frac{\partial \ln L}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = \frac{(1-p)s_n - (n-s_n)p}{p(1-p)}$$

en ayant posé $s_n = \sum_{i=1}^n x_i$; cette dérivée s'annule pour $s_n - np = 0$, soit $p = s_n/n$, avec comme dérivée seconde :

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{s_n}{p^2} - \frac{n-s_n}{(1-p)^2} < 0$$

car $0 \leq s_n \leq n$; $\hat{p}_n = s_n/n$ est donc l'emv de p . La variable S_n suit une loi $\mathcal{B}(n, p)$ et comme nous sommes dans les conditions d'application de l'inégalité FDCR, nous pouvons calculer la quantité d'information de Fisher par :

$$I_n(p) = E\left(-\frac{\partial^2 \ln L}{\partial p^2}\right) = \frac{E(S_n)}{p^2} + \frac{n - E(S_n)}{(1-p)^2} = \frac{n}{p(1-p)}$$

Comme $E(\hat{p}_n) = p$ et $V(\hat{p}_n) = p(1-p)/n = 1/I_n(p)$, on en conclut que \hat{p}_n est efficace.

Pour construire un intervalle de confiance pour p de niveau 0,95 il faut déterminer les valeurs de $t_1 = t_1(p)$ et $t_2 = t_2(p)$ telles que $P(t_1 \leq \hat{p}_n \leq t_2) = 0,95$. Nous pouvons pour cela utiliser la loi exacte de $n\hat{p}_n = S_n$ et retenir un intervalle à risques symétriques, c'est-à-dire retenir les conditions $P(n\hat{p}_n \leq nt_1) = P(n\hat{p}_n \geq nt_2) = 0,025$. Cependant, une première difficulté se présente puisque la loi est discrète et que ces égalités n'admettent pas forcément de solution, au sens où il n'y a pas de fractiles d'ordre 0,025 ou 0,975. Nous allons donc construire un intervalle qui sera de niveau au moins égal à 0,95 en retenant plutôt les inégalités $P(n\hat{p}_n \leq nt_1) \leq 0,025$ et $P(n\hat{p}_n \geq nt_2) \leq 0,025$. On doit donc déterminer le plus grand entier n_1 tel que :

$$\sum_{k=0}^{n_1} \binom{n}{k} p^k (1-p)^{n-k} \leq 0,025$$

et le plus petit entier n_2 tel que :

$$\sum_{k=0}^{n_2} \binom{n}{k} p^k (1-p)^{n-k} \geq 0,975$$

Pour tout entier fixé n , et pour chaque valeur de p , on peut déterminer les solutions $t_1 = n_1/n$ et $t_2 = n_2/n$ de ces inéquations et tracer ainsi point par point le graphe de ces deux fonctions de p . Pour chaque verticale, c'est-à-dire pour p fixé, ces deux courbes déterminent un intervalle $[t_1(p), t_2(p)]$ tel que $P(t_1 \leq \hat{p}_n \leq t_2) \geq 0,95$. Pour une valeur observée \hat{p}_n l'intersection de l'horizontale correspondante avec ces deux courbes permet de lire sur l'axe des abscisses l'intervalle des valeurs de p pour lesquelles on a simultanément $t_1(p) \leq \hat{p}_n$ et $t_2(p) \geq \hat{p}_n$ (cf. figure 7.3). Ainsi, l'intervalle de confiance s'obtient par simple lecture de l'**abaque** correspondant à un niveau de confiance donné. On a un couple de courbes associées à chaque valeur de n .

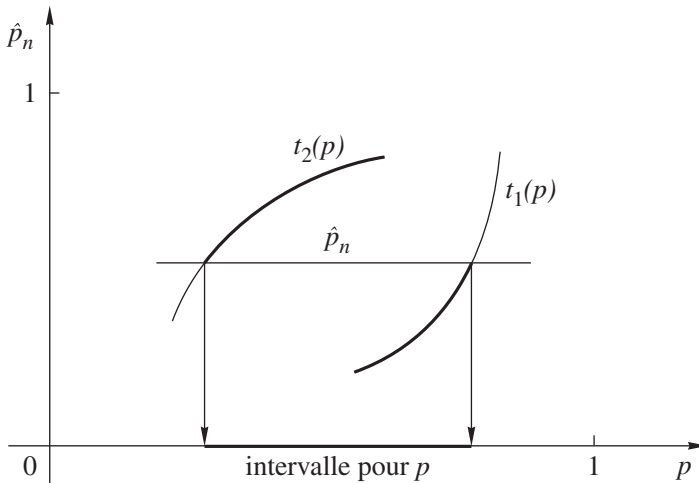


Figure 7.3

Dans l'exemple proposé, si pour $n = 100$ on a observé l'estimation $\hat{p}_{100} = 0,44$ la lecture de l'abaque 2 (où $\hat{p}_n = \bar{x}_n$ est en abscisse et p en ordonnée) conduit à l'intervalle de confiance bilatéral :

$$0,341 < p < 0,538$$

Si on ne dispose pas d'abaque permettant d'obtenir sans calculs l'intervalle exact, et si la valeur de n est suffisamment grande, on obtiendra un intervalle approché (i.e. de niveau voisin de 0,95) en utilisant la loi asymptotique de \hat{p}_n , déduite du théorème central limite :

$$\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{pq}} \xrightarrow{\text{loi}} N(0, 1)$$

On retient alors un intervalle symétrique, à partir de la valeur de a lue dans la table 2 de la loi normale, telle que :

$$P\left(-a < \sqrt{n} \frac{\hat{p}_n - p}{\sqrt{pq}} < a\right) = 1 - \alpha$$

Soit pour $\alpha = 0,05$: $a = \Phi^{-1}(0,975) = 1,96$ d'où :

$$P\left(\hat{p}_n - 1,96\sqrt{\frac{pq}{n}} < p < \hat{p}_n + 1,96\sqrt{\frac{pq}{n}}\right) = 0,95$$

Pour en déduire un intervalle pour p , il faudrait résoudre ces deux inéquations du second degré en p . Il est plus simple de faire une seconde approximation en remplaçant p par \hat{p}_n dans les bornes de l'intervalle, ce qui conduit à l'intervalle de confiance pour p :

$$\hat{p}_n - a\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} < p < \hat{p}_n + a\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

où $a = \Phi^{-1}(1 - \alpha/2)$. On considère que cette approximation est acceptable pour $np(1-p) \geq 3$. Avec les données de l'exemple on obtient l'intervalle :

$$0,343 < p < 0,537$$

On aurait pu également remplacer $p(1-p)$ par sa borne supérieure, qui vaut $1/4$, et obtenir ainsi l'intervalle approché le plus grand :

$$\hat{p}_n - \frac{a}{2\sqrt{n}} < p < \hat{p}_n + \frac{a}{2\sqrt{n}}$$

Dans cet exemple on obtient :

$$0,342 < p < 0,538$$

intervalle qui est donc bien un peu plus grand que le précédent ; seul le premier intervalle est exact, mais les approximations utilisées ici avec $n = 100$ donnent des résultats très voisins.

4.4 Intervalles associés aux paramètres de la loi normale

On dispose d'un échantillon (X_1, \dots, X_n) d'une v.a. X qui suit une loi $N(m, \sigma)$ pour construire un intervalle de confiance relatif à l'un des deux paramètres de cette loi, l'autre paramètre étant connu ou inconnu. Ceci correspond aux différentes situations que nous allons étudier maintenant.

Intervalle pour la moyenne d'une loi normale d'écart type connu

L'exemple introductif se situait dans ce cadre-là et nous avons obtenu un intervalle de confiance, de niveau $1 - \alpha$, pour le paramètre m , centré sur l'estimateur \overline{X}_n

$$\overline{X}_n - u \frac{\sigma}{\sqrt{n}} < m < \overline{X}_n + u \frac{\sigma}{\sqrt{n}}$$

où u est le fractile d'ordre $1 - \alpha/2$ de la loi normale standard, soit $u = \Phi^{-1}(1 - \alpha/2)$. Cet intervalle est de longueur non aléatoire $l = 2u\sigma/\sqrt{n}$, proportionnelle au niveau de confiance par l'intermédiaire de u et inversement proportionnelle à la taille de l'échantillon par le terme $1/\sqrt{n}$. Pour un niveau de confiance fixé, c'est l'intervalle symétrique de longueur minimale puisque la loi de l'estimateur utilisé est symétrique.

Intervalle pour la moyenne d'une loi normale d'écart type inconnu

La statistique utilisée dans la situation précédente, et dont la loi était connue, était la variable normale centrée-réduite :

$$\frac{\overline{X}_n - m}{\sigma/\sqrt{n}} = \sqrt{n} \frac{\overline{X}_n - m}{\sigma}$$

Elle ne peut plus convenir ici puisque le paramètre σ est inconnu et va donc devoir être remplacé par un estimateur, basé sur la variance empirique modifiée qui est un estimateur sans biais de la variance théorique σ^2 :

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$$

On utilise donc comme nouvelle statistique :

$$\sqrt{n} \frac{\overline{X}_n - m}{S_n}$$

dont la loi est connue ; en effet, si nous divisons numérateur et dénominateur par σ , elle peut aussi s'écrire :

$$\frac{(\overline{X}_n - m) / (\sigma/\sqrt{n})}{\sqrt{(n-1) \frac{S_n^2}{\sigma^2} / (n-1)}}$$

où on voit que le numérateur suit une loi normale centrée-réduite et que le dénominateur est la racine carrée d'une v.a. de loi du khi-deux réduite (divisée) par son nombre de degrés de liberté, car nous avons vu au chapitre 5 que $(n-1) \frac{S_n^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$. Comme d'après le théorème de Fisher numérateur et dénominateur sont indépendants, on en conclut que la statistique $\sqrt{n} \frac{\overline{X}_n - m}{S_n}$, utilisée ici, suit une loi de Student à $n-1$ degrés de liberté. Par

lecture de la table 6, on peut donc déterminer la valeur de t telle que :

$$P\left(-t < \sqrt{n} \frac{\bar{X}_n - m}{S_n} < t\right) = 1 - \alpha$$

La valeur de t est donc le fractile d'ordre $1 - \alpha/2$ de la loi de Student T_{n-1} . L'intervalle a bien sûr été choisi symétrique puisque la loi utilisée est symétrique. Par inversion de cet intervalle, on obtient :

$$P\left(\bar{X}_n - t \frac{S_n}{\sqrt{n}} < m < \bar{X}_n + t \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha$$

ce qui fournit l'intervalle de confiance pour m de niveau $1 - \alpha$, centré sur \bar{X}_n , mais ici de longueur aléatoire $L_n = 2tS_n/\sqrt{n}$.

Exemple 7.11

Sur un échantillon de $n = 30$ durées de vie d'un certain modèle de lampe on a obtenu comme moments empiriques $\bar{x}_{30} = 2000h$ et $s_{30} = 300h$. L'intervalle de confiance de niveau 0,95 pour la durée de vie moyenne m est donc :

$$\bar{x}_{30} - t \frac{s_{30}}{\sqrt{30}} < m < \bar{x}_{30} + t \frac{s_{30}}{\sqrt{30}}$$

où t est défini par $P(-t < T_{29} < t) = 0,95$ ou $P(T_{29} < t) = 0,975$ soit $t = 2,045$ d'où l'intervalle :

$$1888 < m < 2112$$

de longueur $l = 224 h$ observée sur cet échantillon. Si σ avait été connu, de même valeur que celle observée sur l'échantillon, soit $\sigma = 300$, l'intervalle correspondant aurait été :

$$\bar{x}_{30} - u \frac{s_{30}}{\sqrt{30}} < m < \bar{x}_{30} + u \frac{s_{30}}{\sqrt{30}}$$

avec $u = \Phi^{-1}(0,975) = 1,96$ soit l'intervalle $1893 < m < 2107$, de longueur $l = 214h$, inférieure à la précédente. Ainsi, même en cas d'estimation parfaite, c'est-à-dire telle que $s_n = \sigma$, l'intervalle obtenu est plus grand ; la connaissance du paramètre σ conduit logiquement à un intervalle plus précis.

Intervalle pour la variance d'une loi normale d'espérance connue

Pour estimer la précision d'un thermomètre, on réalise n mesures indépendantes de la température d'un liquide qui est maintenu à température constante, égale à 20 degrés Celsius. Compte tenu des erreurs de mesure, la valeur indiquée par le thermomètre peut être considérée comme une v.a. normale dont la moyenne m est la valeur exacte de la température, soit ici $m = 20$, et dont l'écart type σ est inconnu et caractérise la précision

du thermomètre. L'estimateur, basé sur l'échantillon (X_1, \dots, X_n) de cette loi $N(m, \sigma)$, est ici :

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

estimateur sans biais, convergent et efficace, de loi connue : $n \widehat{\sigma}_n^2 / \sigma^2 \rightsquigarrow \chi_n^2$. On peut donc déterminer les valeurs de a et b telles que :

$$P \left(a < n \frac{\widehat{\sigma}_n^2}{\sigma^2} < b \right) = 1 - \alpha$$

ce qui conduit à l'intervalle de confiance défini par :

$$P \left(n \frac{\widehat{\sigma}_n^2}{b} < \sigma^2 < n \frac{\widehat{\sigma}_n^2}{a} \right) = 1 - \alpha$$

Cependant, il n'y a qu'une seule condition pour déterminer les deux valeurs a et b et il reste un degré d'incertitude puisque la loi utilisée n'est pas symétrique. Si on pose $\alpha_1 = P(\chi_n^2 < a)$ et $\alpha_2 = P(\chi_n^2 > b)$, la seule contrainte dans le choix de α_1 et α_2 est $\alpha_1 + \alpha_2 = \alpha$. Dans l'exemple retenu, si on a observé sur un échantillon de taille 15 la valeur $\widehat{\sigma}_{15}^2 = 18$ et qu'on retient un intervalle à erreurs symétriques (choix le moins arbitraire), pour un niveau de confiance $1 - \alpha = 0,99$ on lit dans la table 5 les valeurs $a = 4,60$ et $b = 32,8$ d'où l'intervalle :

$$8,23 < \sigma^2 < 58,70$$

Mais, compte tenu de l'interprétation du paramètre qui mesure ici un degré d'imprécision, on souhaite qu'il soit le plus faible possible et on retient plus logiquement un intervalle unilatéral à gauche, de la forme $\sigma^2 < \text{constante}$, ce qui correspond au choix $\alpha_1 = \alpha = 0,01$ et $\alpha_2 = 0$, soit $a = 5,23$ et l'intervalle :

$$\sigma^2 < 51,63$$

Intervalle pour la variance d'une loi normale d'espérance inconnue

Quand le second paramètre m de la loi normale est inconnu, l'estimateur sans biais et convergent de σ^2 qu'il faut retenir est :

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$$

Sa loi est connue, $(n-1) S_n^2 / \sigma^2 \rightsquigarrow \chi_{n-1}^2$, et on peut donc déterminer les valeurs de a et b telles que :

$$P \left\{ a < (n-1) \frac{S_n^2}{\sigma^2} < b \right\} = 1 - \alpha$$

ce qui permet d'en déduire l'intervalle de confiance défini par :

$$P \left\{ (n-1) \frac{S_n^2}{b} < \sigma^2 < (n-1) \frac{S_n^2}{a} \right\} = 1 - \alpha$$

Là encore, il n'y a qu'une seule contrainte pour déterminer les valeurs de a et b ; si nous posons $\alpha_1 = P(\chi_{n-1}^2 < a)$ et $\alpha_2 = P(\chi_{n-1}^2 > a)$ la contrainte est $\alpha_1 + \alpha_2 = \alpha$.

Exemple 7.12

Sur un échantillon de seize chiffres d'affaires de magasins d'une chaîne de grandes surfaces on a observé $S_{16}^2 = 72, 53$. L'intervalle de niveau 0,95 à risques symétriques est défini à partir de $\alpha_1 = \alpha_2 = 0, 025$ et on lit dans la table 5, $a = 6, 26$ et $b = 27, 49$ d'où l'intervalle $39, 56 < \sigma^2 < 173, 79$. Si on fait le choix d'un intervalle unilatéral à gauche, soit $\alpha_1 = \alpha = 0, 05$ et $\alpha_2 = 0$ on obtient $a = 7, 26$ et l'intervalle $\sigma^2 < 149, 86$ qui est de longueur plus grande que le précédent.

Tous les intervalles précédents ont été construits à partir d'une statistique dont la loi était indépendante du paramètre θ à estimer et que lon appelle une **fonction pivotale** pour θ . Cependant, dans certains problèmes on peut seulement trouver une fonction **quasi pivotale** pour θ , c'est-à-dire une statistique dont seule la loi asymptotique est indépendante de θ .

Intervalle pour l'écart type d'une loi normale d'espérance connue

Supposons pour simplifier que la loi normale soit centrée ; l'estimateur naturel de la variance σ^2 , c'est-à-dire l'estimateur de la méthode des moments, sans biais et convergent, est :

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

et sa loi est connue : $n\sigma_n^2/\sigma^2 \rightsquigarrow \chi_n^2$. Si on retient σ_n comme estimateur de σ , cet estimateur ne pourra pas être sans biais, car dans ce cas on aurait $V(\sigma_n) = E(\sigma_n^2) - E^2(\sigma_n) = \sigma^2 - \sigma^2 = 0$. On démontre que cet estimateur est asymptotiquement sans biais, avec $E(\sigma_n) = \sigma a_n$ où $a_n^2 = 1 - \frac{1}{2n} + \frac{\epsilon_n}{n}$ avec $\epsilon_n \rightarrow 0$ quand $n \rightarrow \infty$. Nous allons écrire la vraisemblance pour établir que σ_n est l'emv de σ :

$$L(x_1, \dots, x_n; \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2$$

et la log-vraisemblance s'écrit :

$$\ln L(x_1, \dots, x_n; \sigma) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2$$

d'où en dérivant :

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n x_i^2$$

cette dérivée s'annule pour $\sigma = \sigma_n$ et la dérivée seconde :

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n x_i^2$$

a pour valeur en ce point ; $\frac{n}{\sigma_n^2} - \frac{3n\sigma_n^2}{\sigma_n^4} = -\frac{2n}{\sigma_n^2} < 0$; cette valeur correspond bien à un maximum. La quantité d'information de Fisher peut se calculer ici par :

$$I_n(\sigma) = E\left(-\frac{\partial^2 \ln L}{\partial \sigma^2}\right) = -\frac{n}{\sigma^2} + \frac{3nE(X^2)}{\sigma^4} = \frac{2n}{\sigma^2}$$

Pour obtenir un estimateur qui puisse être efficace, il faut d'abord qu'il soit sans biais et c'est le cas ici de l'estimateur $\hat{\sigma}_n = \sigma_n/a_n$. Sa variance vaut :

$$V(\hat{\sigma}_n) = \frac{V(\sigma_n)}{a_n^2} = \frac{\sigma^2 - a_n^2 \sigma^2}{a_n^2} = \frac{\sigma^2}{2n} (1 + \epsilon_n)$$

et donc $\hat{\sigma}_n$ est asymptotiquement efficace. Bien que la loi exacte de $\hat{\sigma}_n$ ne soit pas connue, on va pouvoir construire un intervalle de confiance à partir de la fonction pivotale $n\sigma_n^2/\sigma^2$. On peut en effet déterminer les valeurs de a et b telles que :

$$P\left(a < n \frac{\sigma_n^2}{\sigma^2} < b\right) = 1 - \alpha$$

et, comme la fonction racine carrée est croissante, on en déduit l'intervalle de confiance pour σ défini par :

$$P\left(\sigma_n \sqrt{\frac{n}{b}} < \sigma < \sigma_n \sqrt{\frac{n}{a}}\right) = 1 - \alpha$$

Exemple 7.13

Sur un échantillon de taille $n = 100$ on veut construire un intervalle de confiance de niveau 0,95 pour σ à partir de l'observation $\sigma_{100}^2 = 1,945$. Si on retient un intervalle à risques symétriques, on lit dans la table 5 les fractiles $a = 74,22$ et $b = 129,56$ d'où l'intervalle $1,23 < \sigma < 1,62$ de longueur 0,39.

Nous allons comparer l'estimateur précédent à un estimateur basé sur l'écart absolu moyen :

$$D_n = \frac{1}{n} \sum_{i=1}^n |X_i|$$

Son espérance est égale à celle de :

$$E(|X|) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x| e^{-x^2/2\sigma^2} dx$$

soit, avec le changement de variable $x^2 = 2\sigma^2u$:

$$E(|X|) = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-u} \sigma du = \sigma \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-u} du = \sigma \sqrt{\frac{2}{\pi}}$$

L'estimateur sans biais que lon retient est donc :

$$T_n = \sqrt{\frac{\pi}{2}} D_n$$

D'après la loi des grands nombres :

$$D_n \xrightarrow[p]{} E(|X|)$$

donc T_n est convergent. Sa variance est :

$$V(T_n) = \frac{\pi}{2} V(D_n) = \frac{\pi}{2n} V(|X|) = \frac{\pi}{2n} \left(\sigma^2 - \frac{2\sigma^2}{\pi} \right) = \left(\frac{\pi}{2} - 1 \right) \frac{\sigma^2}{n}$$

Si on compare les deux estimateurs sans biais $\hat{\sigma}_n$ et T_n :

$$\frac{V(T_n)}{V(\hat{\sigma}_n)} = \frac{(\pi - 2) a_n^2}{2n(1 - a_n^2)} \rightarrow \pi - 2 > 1$$

donc $\hat{\sigma}_n$ est préférable à T_n , ce qui était prévisible car nous avons vu qu'il était asymptotiquement efficace.

Pour construire un intervalle de confiance, nous allons utiliser la loi asymptotique de l'estimateur T_n , déduite du théorème central limite :

$$\sqrt{n} \frac{D_n - E(|X|)}{\sigma(|X|)} \xrightarrow[loi]{} N(0, 1)$$

d'où :

$$\sqrt{n} \frac{T_n - \sigma}{\sigma \sqrt{\pi/2 - 1}} \xrightarrow[loi]{} N(0, 1)$$

On peut donc déterminer une valeur approchée de u telle que :

$$P \left\{ -u < \sqrt{n} \frac{T_n - \sigma}{\sigma \sqrt{\pi/2 - 1}} < u \right\} = 1 - \alpha$$

ce qui conduit à l'intervalle de confiance :

$$\frac{T_n}{1 + u \sqrt{\pi/2 - 1} / \sqrt{n}} < \sigma < \frac{T_n}{1 - u \sqrt{\pi/2 - 1} / \sqrt{n}}$$

Exemple 7.14

Sur le même échantillon que dans l'exemple 7.13, de taille $n = 100$, on a observé $d_n = 1,084$. L'intervalle de niveau voisin de 0,95 est alors $1,16 < \sigma < 1,56$ de longueur 0,40.

5 Compléments

5.1 Inégalité de Fréchet-Darmonis-Cramer-Rao

Nous allons préciser les hypothèses de Cramer-Rao évoquées dans l'énoncé du théorème 7.3. L'ensemble Θ est un ouvert sur lequel la densité $f(x; \theta)$ ne s'annule en aucun point x et est dérivable par rapport à θ . On suppose également que l'on peut intervertir dérivation par rapport à θ et intégration, et que la quantité d'information de Fisher est strictement positive. La variance d'un estimateur sans biais est donc minorée par un nombre positif, d'autant plus petit que l'information de Fisher est grande. On ajoute souvent l'hypothèse que la densité est deux fois dérivable par rapport à θ et que l'on peut toujours intervertir dérivation par rapport à θ et intégration. On obtient alors une autre expression de l'information de Fisher, vue au § 2.3, et généralement plus simple à calculer.

L'inégalité FDCR se généralise au cas où le paramètre à estimer est $g(\theta)$, g étant une fonction réelle. Si T_n est un estimateur sans biais de $g(\theta)$, si on peut encore échanger dérivation par rapport à θ et intégration de T_n , alors :

$$V_\theta(T_n) \geq \frac{[g'(\theta)]^2}{I_n(\theta)}$$

Si l'estimateur présente un biais $b_n(\theta)$, alors :

$$V_\theta(T_n) \geq b_n^2(\theta) + \frac{[b'_n(\theta) + g'(\theta)]^2}{I_n(\theta)}$$

L'inégalité FDCR se généralise également au cas multidimensionnel, Θ étant un ouvert de \mathbb{R}^p , l'information de Fisher devenant la matrice dont l'élément ligne i , colonne j $1 \leq i, j \leq p$, est :

$$E_\theta \left(\frac{\partial \ln L}{\partial \theta_i} \times \frac{\partial \ln L}{\partial \theta_j} \right)$$

ou, dans le cas où on peut dériver deux fois et échanger avec l'intégration :

$$E_\theta \left(-\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right)$$

Si T_n est un estimateur sans biais de $g(\theta)$, g étant une application de \mathbb{R}^p dans \mathbb{R}^q , alors l'inégalité FDCR se traduit par l'affirmation que la matrice :

$$V_\theta(T_n) - \left(\frac{dg}{d\theta}\right) I_n^{-1}(\theta) \left(\frac{dg}{d\theta}\right)$$

est semi-définie positive, $\frac{dg}{d\theta}$ étant la matrice dont l'élément ligne i , $1 \leq i \leq q$, colonne j , $1 \leq j \leq p$, est $\frac{\partial g_i}{\partial \theta_j}$, avec $g = (g_1, \dots, g_q)$.

5.2 Statistique exhaustive

Dans l'exercice 3, pour une famille de lois géométriques, on établit que l'information apportée par la somme des observations est égale à celle apportée par les données individuelles. Ainsi, la connaissance de cette seule statistique n'a pas diminuée l'information apportée par toutes les données. Nous allons voir sur un autre exemple comment peut se traduire cette propriété intéressante pour une statistique et qui conduira à la notion d'exhaustivité.

Exemple 7.15

Considérons un contrôle industriel de pièces défectueuses, effectué en tirant avec remise n pièces dans un lot, pour connaître la proportion θ de défectueuses. À chaque pièce tirée on associe une v.a. de Bernoulli :

$$X_i = \begin{cases} 1 & \theta \\ 0 & 1 - \theta \end{cases} \quad 1 \leq i \leq n$$

Le rang de tirage des pièces défectueuses n'apporte évidemment aucune information sur le paramètre θ et toute l'information apportée par l'échantillon (X_1, \dots, X_n) est contenue dans la statistique $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ qui suit une loi $\mathcal{B}(n, \theta)$. Pour préciser cette évidence, nous allons déterminer la loi d'un échantillon lorsque cette statistique a une valeur fixée t , soit :

$$P_\theta(X_1 = x_1, \dots, X_n = x_n | T = t) = \frac{P_\theta(X_1 = x_1, \dots, X_n = x_n, T = t)}{P_\theta(T = t)}$$

On a $P_\theta(X_i = x_i) = \theta^{x_i}(1 - \theta)^{1-x_i}$, $x_i \in \{0, 1\}$ et donc :

$$\begin{aligned} P_\theta(X_1 = x_1, \dots, X_n = x_n, T = t) &= P_\theta\left(X_1 = x_1, \dots, X_n = t - \sum_{i=1}^{n-1} x_i\right) \\ &= \theta^t (1 - \theta)^{n-t} \end{aligned}$$

Comme :

$$P_\theta(T = t) = \binom{n}{t} \theta^t (1 - \theta)^{n-t}$$

on en conclut :

$$P_\theta(X_1 = x_1, \dots, X_n = x_n | T = t) = \frac{1}{\binom{n}{t}}$$

résultat prévisible en remarquant que $\binom{n}{t}$ représente le nombre de choix des t indices i pour lesquels $x_i = 1$. On constate donc que cette probabilité est indépendante du paramètre inconnu θ .

Cet exemple nous montre que si la loi de l'échantillon, conditionnellement à une valeur fixée de la statistique, est indépendante de la loi de X , c'est-à-dire du paramètre θ , cela traduit le fait que les valeurs individuelles n'apportent pas plus d'information sur θ que la seule valeur de T . Nous dirons qu'une statistique T_n est **exhaustive** (*sufficient*) si la loi conditionnelle de (X_1, \dots, X_n) pour $T = t$ fixé est indépendante de θ . Cette propriété d'exhaustivité pourra se démontrer facilement à l'aide du résultat suivant.

Théorème 7.4

Théorème de factorisation (Neyman-Fisher)

Une statistique T_n est exhaustive s'il existe deux applications mesurables positives g et h telles que la densité L de l'échantillon puisse se factoriser sous la forme :

$$L(x_1, \dots, x_n; \theta) = g(t; \theta)h(x_1, \dots, x_n)$$

Cette factorisation n'est bien sûr pas unique, mais la quantité $h(x_1, \dots, x_n) = \psi(x_1, \dots, x_n; t(x_1, \dots, x_n))$ peut représenter la densité de $(X_1, \dots, X_n) | T_n = t$ et dans ce cas $g(t; \theta)$ est la densité de T_n .

Exemple 7.16

Soit X une v.a. de loi $\mathcal{E}(1/\theta) \equiv \gamma(1, 1/\theta)$ et considérons la statistique $T_n = \sum_{i=1}^n X_i$ qui suit la loi $\gamma(n, 1/\theta)$. La vraisemblance s'écrit :

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \left(\prod_{i=1}^n \mathbf{1}_{\mathbb{R}_+}(x_i) \right) \frac{1}{\theta^n} \exp -\frac{1}{\theta} \sum_{i=1}^n x_i \\ &= g(t; \theta) h(x_1, \dots, x_n) \end{aligned}$$

en ayant posé $g(t; \theta) = 1/\theta^n e^{-t/\theta}$ et $h(x_1, \dots, x_n) = \mathbf{1}_{\mathbb{R}_+^n}(x_1, \dots, x_n)$, ce qui montre que T_n est exhaustive. Mais on peut également prendre :

$$g(t; \theta) = \frac{1}{\theta^n \Gamma(n)} e^{-t/\theta} t^{n-1} \mathbf{1}_{\mathbb{R}_+}(t)$$

qui est la densité de T_n , et dans ce cas :

$$h(x_1, \dots, x_n) = \frac{\Gamma(n)}{t^{n-1}} \mathbf{1}_{\mathbb{R}_+^n}(x_1, \dots, x_n) = \frac{(n-1)!}{\left(\sum_{i=1}^n x_i \right)^{n-1}} \mathbf{1}_{\mathbb{R}_+^n}(x_1, \dots, x_n)$$

est la densité de l'échantillon conditionnellement à $T_n = t$.

Quand T prend ses valeurs dans l'espace des paramètres Θ , on dit que T est un **résumé exhaustif** pour θ . Quand T est à valeurs dans \mathbb{R}^s , on dit que T est une statistique exhaustive d'**ordre** s . Une statistique exhaustive pour un échantillon est bien entendu exhaustive pour toute fonction de cet échantillon.

Dans l'exemple précédent, nous aurions pu choisir d'autres statistiques exhaustives comme par exemple $(X_1 + X_2, X_3, \dots, X_n)$, $(X_1 + X_2 + X_3, X_4, \dots, X_n)$... ou l'échantillon lui-même (X_1, X_2, \dots, X_n) puisque l'application identique est toujours une statistique exhaustive. Notons aussi que T_n est une fonction de toutes ces statistiques. Il est donc souhaitable de réduire au maximum l'espace des observations sans perdre d'information sur θ . C'est ce que réalise une **statistique exhaustive minimale** (*minimal sufficient*, ou *necessary and sufficient*) S qui est une fonction de toute autre statistique exhaustive T . De façon précise, si T est à valeurs dans G et S à valeurs dans F , alors il existe une application mesurable $g : G \rightarrow F$ telle que $S = g(T)$.

Nous allons introduire une troisième notion et faire le lien entre elles. Une statistique T est dite **complète** (ou totale) si toute v.a. h telle que $h(T)$ est d'intégrale nulle pour P_θ est nulle P_θ -presque partout. Avec cette nouvelle définition, on peut énoncer le résultat suivant.

Théorème 7.5

Toute statistique exhaustive et complète est minimale.

Il faut cependant faire attention : une statistique exhaustive minimale n'est pas forcément complète.

5.3 Famille exponentielle

La famille des lois exponentielles joue un rôle très important en statistique car elle possède un certain nombre de propriétés intéressantes. Il s'agit des lois dont la densité peut s'écrire sous la forme :

$$f(x; \theta) = a(\theta) b(x) \exp \left[\sum_{j=1}^k \alpha_j(\theta) T_j(x) \right]$$

Exemple 7.17

Loi binômiale :

$$f(x; p) = (1-p) \binom{n}{x} \exp \left(x \ln \frac{p}{1-p} \right)$$

$$\alpha(p) = \ln \frac{p}{1-p} \text{ et } T(x) = x$$

Exemple 7.18

Loi de Poisson :

$$f(x; \theta) = e^{-\theta} \frac{1}{x!} \exp(x \ln \theta)$$

$$\alpha(\theta) = \ln \theta \text{ et } T(x) = x$$

Exemple 7.19

Famille des lois $\gamma(p; \theta)$:

$$f(x; p, \theta) = \frac{\theta^p}{\Gamma(p)} \mathbf{1}_{\mathbb{R}_+}(x) \exp[-\theta x + (p - 1) \ln x]$$

$$\alpha_1 = -\theta, \alpha_2 = p - 1 \text{ et } T_1(x) = x, T_2(x) = \ln x$$

Le théorème de factorisation permet de conclure que la statistique :

$$T(X_1, \dots, X_n) = \left(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

est exhaustive. En effet, la vraisemblance s'écrit dans ce cas :

$$L(x_1, \dots, x_n; \theta) = a^n(\theta) \left(\prod_{i=1}^n b(x_i) \right) \exp \left[\sum_{j=1}^k \alpha_j(\theta) T_j \right]$$

ayant posé $T_j = \sum_{i=1}^n T_j(x_i)$. Dans le cas où les fonctions α_j sont linéairement indépendantes, cette statistique est de plus minimale. On fait alors le changement de paramètres $\theta_j = \alpha_j(\theta)$, $1 \leq j \leq k$, dans la famille exponentielle, ce qui correspond à une nouvelle famille P_θ où $\theta = (\theta_1, \dots, \theta_k)$ est appelé paramètre **naturel** de la famille exponentielle. On peut énoncer le résultat suivant.

Théorème 7.6

Si la densité de la famille exponentielle s'écrit sous la forme :

$$f(x; \theta) = a(\theta) b(x) \exp \left[\sum_{j=1}^k \theta_j T_j(x) \right]$$

où $\theta = (\theta_1, \dots, \theta_k)$ appartient à Θ qui contient un pavé de \mathbb{R}^k (ou qui est d'intérieur non vide), alors la statistique $T = (T_1, \dots, T_k)$ est exhaustive, complète et minimale.

Exemple 7.20

Loi binômiale : $T(x) = x$, donc $\sum_{i=1}^n X_i$, ou \bar{X}_n , est une statistique exhaustive et complète pour le paramètre p .

Exemple 7.21

Loi de Poisson : \bar{X}_n est une statistique exhaustive et complète pour le paramètre θ .

Exemple 7.22

Famille des lois $\gamma(p; \theta)$: la statistique $(T_1 = \sum_{i=1}^n X_i, T_2 = \sum_{i=1}^n \ln X_i)$ est exhaustive pour le paramètre (θ, p) .

Exemple 7.23

Loi normale :

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}} \frac{e^{-m^2/2\sigma^2}}{\sigma} \exp\left(\frac{m}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right)$$

et $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ est exhaustive complète pour le paramètre $\theta = (m, \sigma^2)$.

Dans le cas où l'ensemble $E = X(\Omega)$ des valeurs possibles pour X est indépendant du paramètre θ , le théorème de Darmois-Koopman énonce que l'appartenance à la famille exponentielle est une condition nécessaire et suffisante d'existence d'un résumé exhaustif d'ordre k pour un paramètre vectoriel.

Exemple 7.24

Loi de Cauchy :

$$f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}$$

cette densité n'appartient pas à la famille exponentielle, donc il n'existe pas de résumé exhaustif pour θ .

Exemple 7.25

Loi uniforme sur $[0, \theta]$:

$$f(x; \theta) = \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(x)$$

cette densité n'appartient pas à la famille exponentielle et cependant on peut établir grâce au théorème de factorisation que $X_{(n)} = \max\{X_1, \dots, X_n\}$ est une statistique exhaustive pour θ . Ceci provient du fait que le théorème de Darmois-Koopman ne s'applique pas, l'ensemble $E = X(\Omega) = [0, \theta]$ dépendant du paramètre θ .

La famille exponentielle permet d'obtenir un estimateur efficace et comme le théorème suivant le précise, c'est une condition nécessaire.

Théorème 7.7

Théorème de Koopman

Sous les hypothèses de Cramer-Rao, la dérivée de f par rapport à θ étant continue en θ , si T_n est un estimateur sans biais du paramètre réel $g(\theta)$, alors T_n est un estimateur efficace si, et seulement si, il existe des fonctions réelles α, β et γ telles que :

$$\ln f(x; \theta) = \beta(\theta) + \gamma(x) + \alpha(\theta) T(x)$$

La statistique $T_n = \frac{1}{n} \sum_{i=1}^n T(X_i)$ est exhaustive, complète et minimale et constitue un résumé exhaustif d'ordre un ; c'est aussi un estimateur efficace du paramètre $g(\theta) = -\beta'(\theta)/\alpha'(\theta)$ qui est le seul paramètre que l'on peut estimer efficacement (ou une fonction affine de celui-ci) et qui n'est pas forcément le paramètre d'intérêt.

Exemple 7.26

Loi de Poisson :

$$\begin{aligned} \ln f(x; \theta) &= -\theta - \ln x! + x \ln \theta \\ \beta(\theta) &= -\theta, \alpha(\theta) = \ln \theta, T(x) = x \end{aligned}$$

L'estimateur $T_n = \frac{1}{n} \sum_{i=1}^n T(X_i)$ est efficace pour $g(\theta) = \theta$.

Exemple 7.27

Famille de lois $\gamma(p)$:

$$\begin{aligned} \ln f(x; p) &= -\ln \Gamma(p) - x + (p-1) \ln x \\ \beta(p) &= -\ln \Gamma(p), \alpha(p) = p-1, T(x) = \ln x \end{aligned}$$

L'estimateur $T_n = \frac{1}{n} \sum_{i=1}^n \ln X_i$ est efficace pour $g(p) = \Gamma'(p)/\Gamma(p)$ qui est un paramètre de peu d'intérêt.

5.4 Amélioration d'un estimateur

L'existence d'une statistique exhaustive pour le paramètre permet d'améliorer, en utilisant le critère de la variance, un estimateur sans biais.

Théorème 7.8

Théorème de Rao-Blackwell

Si T est une statistique exhaustive pour la famille P_θ et S un estimateur sans biais de $g(\theta)$, l'estimateur $E_\theta(S|T)$ est sans biais et préférable à S .

L'espérance conditionnelle $E_\theta(S|T)$ permet bien de construire un estimateur, car elle est indépendante de θ en raison de l'exhaustivité de T .

Avec le critère de la variance, on peut espérer trouver un estimateur optimal dans la classe des estimateurs sans biais de variance finie. Cet estimateur n'est pas nécessairement efficace, mais bien sûr s'il existe un estimateur efficace il est optimal. Dans le cas d'une statistique exhaustive complète, l'estimateur amélioré de Rao-Blackwell est optimal.

Théorème 7.9

Théorème de Lehmann-Scheffé

Si T est une statistique exhaustive complète pour la famille P_θ et S un estimateur sans biais de $g(\theta)$, l'estimateur $E_\theta(S|T)$ est optimal dans la classe des estimateurs sans biais.

Exemple 7.28

Pour la loi uniforme sur $[0, \theta]$, $S = 2\overline{X}_n$ est un estimateur sans biais de θ . On peut établir que $X_{(n)} = \max\{X_1, \dots, X_n\}$ est une statistique exhaustive et complète pour θ . Donc l'estimateur $E_\theta(2\overline{X}_n | X_{(n)})$ est sans biais et optimal. En écrivant :

$$E_\theta(\overline{X}_n | X_{(n)}) = \frac{1}{n} X_{(n)} + \frac{n-1}{n} E_\theta \left[\frac{1}{n-1} \sum_{i=1}^{n-1} X_{(i)} | X_{(n)} \right]$$

on établit que $E_\theta(2\overline{X}_n | X_{(n)}) = \frac{n+1}{n} X_{(n)}$.

- 1 Un estimateur est une statistique, c'est-à-dire une variable aléatoire fonction d'un échantillon, dont les réalisations sont censées être proches de la valeur inconnue du paramètre à estimer. Pour qu'il en soit ainsi, on demande à cet estimateur de posséder un certain nombre de propriétés, comme par exemple d'être sans biais, i.e. d'avoir une valeur moyenne (au sens d'espérance) égale au paramètre à estimer. On souhaite ensuite qu'il soit le plus efficace possible, i.e. qu'il ait une dispersion, mesurée par la variance, la plus petite possible. Dans certaines conditions, notamment quand l'ensemble des valeurs possibles pour la variable ne dépend pas du paramètre à estimer, on peut trouver un estimateur optimal, qu'on appelle efficace, et qui est le meilleur estimateur sans biais que l'on puisse obtenir.
- 2 Si un estimateur est seulement asymptotiquement sans biais, i.e. que son espérance tend vers le paramètre quand la taille de l'échantillon devient infinie, sa qualité est mesurée par l'erreur quadratique moyenne.
- 3 Un estimateur est dit convergent s'il converge en probabilité vers le paramètre à estimer. Tout estimateur sans biais, ou asymptotiquement sans biais, dont la variance tend vers zéro est convergent.
- 4 Si le paramètre à estimer est la moyenne (espérance) de la loi, la moyenne empirique est un estimateur sans biais et convergent.
- 5 Si le paramètre à estimer est la variance de la loi, la variance empirique modifiée (divisée par $n - 1$ au lieu de n) est un estimateur sans biais et convergent.
- 6 Dans les autres cas, on construit un estimateur par la méthode des moments ou la méthode du maximum de vraisemblance.



- 7.1** À partir d'observations indépendantes (X_1, \dots, X_n) d'une certaine grandeur économique X , on retient le modèle suivant :

$$X_t = a(1 + \varepsilon_t), \quad 1 \leq t \leq n$$

où les v.a. ε_t sont indépendantes et de même loi normale standard.

1. Déterminer l'estimateur du maximum de vraisemblance \hat{a}_n de a .
 2. Étudier les propriétés de l'estimateur de a obtenu par la méthode des moments.
 3. Proposer un estimateur de la variance du modèle.
- 7.2** Afin d'organiser au mieux l'accueil des groupes de visiteurs d'un parc d'attractions, on note la durée X séparant l'arrivée de deux groupes successifs. À partir des observations (X_1, \dots, X_n) recueillies dans une journée, on retient comme modèle la loi uniforme sur $[0, \theta]$.
1. Déterminer par la méthode des moments un estimateur sans biais du paramètre $\theta > 0$ et étudier ses propriétés.
 2. Déterminer par la méthode du maximum de vraisemblance un estimateur sans biais du paramètre θ et étudier ses propriétés. Comparer les deux estimateurs.
- 7.3** Une v.a. X suit une loi uniforme discrète sur l'ensemble des entiers $\{1, 2, \dots, \theta\}$ où θ est un entier positif inconnu. Déterminer, par la méthode des moments, un estimateur de θ construit à partir d'un échantillon (X_1, \dots, X_n) de X et étudier ses propriétés. Est-il efficace ?
- 7.4** Une urne contient un nombre de boules inconnu $\theta \geq 2$, une seule d'entre elles étant blanche. On effectue dans cette urne des tirages successifs avec remise, jusqu'à ce qu'on obtienne une boule blanche et on note X la variable aléatoire qui représente le nombre de tirages effectués. À partir d'un échantillon (X_1, \dots, X_n) de X , déterminer un estimateur T_n de θ par la méthode des moments. Étudier ses propriétés. Est-il efficace ?
- 7.5** Soit X une variable aléatoire dont la densité a pour expression, pour $x > 1$:

$$f(x) = \frac{1}{\theta} x^{-1/\theta-1} \quad \text{avec } \theta > 0$$

et nulle sinon.

1. Calculer $E(X)$ et en déduire un estimateur T_n de θ par la méthode des moments, construit à partir d'un échantillon (X_1, \dots, X_n) de X .
 2. Déterminer la loi de probabilité de la variable aléatoire $Y = \ln X$.
 3. Déterminer l'estimateur $\hat{\theta}_n$ du paramètre θ par la méthode du maximum de vraisemblance et étudier ses propriétés. Est-il efficace ?
 4. On pose $\Sigma_n = \sum_{i=1}^n \ln X_i$. Déterminer la loi de probabilité de $2\Sigma_n/\theta$ et en déduire un intervalle de confiance bilatéral pour θ de niveau $1 - \alpha$.
- 7.6** Le total des ventes hebdomadaires d'un produit alimentaire dans un magasin i , $1 \leq i \leq n$, est une v.a. X_i de loi normale $N(m_i, \sigma)$ où les valeurs m_i et σ sont supposées

connues. Une campagne publicitaire de ce produit a pour conséquence d'augmenter les ventes, de telle sorte que chaque moyenne m_i est augmentée d'une même quantité a .

1. Déterminer un estimateur de a construit à partir d'observations indépendantes (X_1, \dots, X_n) des ventes après cette campagne et étudier ses propriétés, puis construire un intervalle de confiance de niveau 0,95.
2. Déterminer un estimateur du paramètre b dans le cas où chaque moyenne m_i est cette fois multipliée par b et étudier ses propriétés.
3. Application aux données suivantes dans le cas où $\sigma = 3$.

m_i	98	101	104	99	100	102	95	97	105	103
x_i	109	105	110	106	110	114	108	104	115	118

- 7.7** La durée de vie d'un certain matériel est représentée par une v.a. positive X de densité :

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

où θ est un paramètre inconnu strictement positif.

Étudier les propriétés de l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ construit à partir d'un échantillon (X_1, \dots, X_n) de la v.a. X .

Construire un intervalle de confiance pour θ de niveau 0,95 dans le cas où les observations ont conduit à $\sum_{i=1}^{10} x_i = 11,5$.

- 7.8** Soit (X_1, \dots, X_n) un échantillon d'une v.a. X de loi log-normale de paramètres m et $\sigma > 0$.

Étudier les propriétés de l'estimateur du maximum de vraisemblance de m .

Construire un intervalle de confiance pour m de niveau 0,95 dans le cas où $\sigma = 1$ et où on a observé $\sum_{i=1}^{25} \ln x_i = 54,94$.

- 7.9** Soit X une v.a. de densité :

$$f(x; \theta) = \begin{cases} \frac{2x}{\theta^2} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}$$

où θ est un paramètre strictement positif.

1. Déterminer un estimateur de θ , par la méthode des moments, construit à partir d'un échantillon (X_1, \dots, X_n) de X , et étudier ses propriétés.

2. Déterminer un estimateur sans biais $\hat{\theta}_n$ de θ , construit à partir de l'estimateur du maximum de vraisemblance, et étudier ses propriétés.

3. Construire un intervalle de confiance de niveau 0,95 pour θ dans le cas où on a observé $\max \{x_1, \dots, x_{20}\} = 5$.

- 7.10** Deux ateliers de fabrication produisent des doubles-vitrages dont l'épaisseur peut être considérée comme une v.a. de loi normale d'espérance $m = 6$ mm, soit pour chaque

atelier les v.a. X et Y de lois respectives $N(m, \sigma_1)$ et $N(m, \sigma_2)$. Pour comparer les caractéristiques de fabrication de chacun de ces ateliers, on prélève respectivement n_1 et n_2 vitrages d'épaisseurs notées X_1, \dots, X_{n_1} et Y_1, \dots, Y_{n_2} . Construire un intervalle de confiance de niveau $1 - \alpha$ pour le rapport σ_1^2 / σ_2^2 .

7.11 Soit (X_1, \dots, X_n) un échantillon d'une v.a. X de loi normale d'espérance et de variance égales à un paramètre inconnu $\theta > 0$.

1. Déterminer deux estimateurs de θ par la méthode des moments, étudier leurs propriétés et les comparer entre eux.

2. Construire un intervalle de confiance pour θ de niveau 0,95 ayant observé :

$$\sum_{i=1}^{25} x_i = 50,23 \quad \text{et} \quad \sum_{i=1}^{25} (x_i - \bar{x})^2 = 48,12$$

7.12 Soit X une variable aléatoire dont la densité a pour expression :

$$f(x) = \frac{\theta}{2} e^{-\theta|x|}$$

où θ est un paramètre réel strictement positif.

1. Déterminer l'estimateur $\hat{\theta}_n$ du paramètre θ par la méthode du maximum de vraisemblance, construit à partir d'un échantillon (X_1, \dots, X_n) de X et étudier ses propriétés. Est-il efficace ?

2. On pose $\sum_n = \sum_{i=1}^n |X_i|$. Déterminer la loi de $2\theta \sum_n$ et en déduire un intervalle de confiance bilatéral pour θ de niveau $1 - \alpha$.

3. Déterminer la loi limite de $\hat{\theta}_n$ et en déduire un intervalle de confiance bilatéral pour θ de niveau voisin de $1 - \alpha$.

7.13 Soit X une v.a. de densité :

$$f(x; \theta) = \begin{cases} \frac{1}{2\theta\sqrt{x}} e^{-\sqrt{x}/\theta} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

où θ est un paramètre strictement positif que l'on se propose d'estimer à partir d'un échantillon (X_1, \dots, X_n) de X .

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ et étudier ses propriétés.

2. Construire un intervalle de confiance de niveau 0,90 pour θ dans le cas où on a $\sum_{i=1}^{20} \sqrt{x_i} = 47,4$.

7.14 Soit (X, Y) un couple normal dont la densité est définie par :

$$f(x, y) = \frac{1}{2\pi\sqrt{\theta}} \exp - \frac{1}{2\theta} [(1 + \theta)x^2 + 2(1 + 2\theta)xy + (1 + 4\theta)y^2]$$

1. Déterminer la loi de $X + Y$ et en déduire un estimateur T_n de θ par la méthode des moments construit à partir d'un échantillon de taille n du couple (X, Y) .

2. Étudier les propriétés de T_n . Est-il efficace ? Déterminer la loi de T_n et en déduire un intervalle de confiance bilatéral pour θ de niveau $1 - \alpha$.

3. Déterminer la loi limite de T_n et en déduire un intervalle de confiance bilatéral pour θ de niveau voisin de $1 - \alpha$.

7.15 Soit X une v.a. de densité :

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} \exp\left(-\frac{x-\theta}{\theta}\right) & \text{si } \theta \leq x \\ 0 & \text{sinon} \end{cases}$$

où θ est un réel strictement positif.

1. Calculer $E(X)$

2. Soit X_1, \dots, X_n des v.a. indépendantes, de même loi que X . Déterminer un estimateur T_n de θ par la méthode des moments. Étudier ses propriétés. Est-il efficace ?

3. Déterminer la loi limite de T_n quand n devient infini et en déduire un intervalle de confiance bilatéral symétrique de niveau voisin de $1 - \alpha = 0,95$ pour θ , dans le cas où on a observé la valeur $\sum_{i=1}^{100} x_i = 660$ sur un échantillon de taille $n = 100$.



7.1 1. Les v.a. X_i suivent la même loi normale $N(a, a)$, donc la log-vraisemblance a pour expression :

$$\ln L(x_1, \dots, x_n; a) = -\frac{n}{2} \ln(2\pi) - n \ln a - \frac{1}{2a^2} \sum_{i=1}^n (x_i - a)^2$$

de dérivées :

$$\begin{aligned} \frac{\partial \ln L}{\partial a} &= -\frac{n}{a} + \frac{1}{a^3} \sum_{i=1}^n (x_i - a)^2 + \frac{1}{a^2} \sum_{i=1}^n (x_i - a) \\ \frac{\partial^2 \ln L}{\partial a^2} &= \frac{2n}{a^2} - \frac{3}{a^4} \sum_{i=1}^n (x_i - a)^2 - \frac{4}{a^3} \sum_{i=1}^n (x_i - a) \end{aligned}$$

Nous allons poser :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad s'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

On peut écrire :

$$\frac{1}{n} \sum_{i=1}^n (x_i - a)^2 = s'^2 + (\bar{x} - a)^2$$

et on obtient alors :

$$\begin{aligned} \frac{\partial \ln L}{\partial a} &= \frac{n}{a^3} (s'^2 + \bar{x}^2 - a\bar{x} - a^2) \\ \frac{\partial^2 \ln L}{\partial a^2} &= \frac{n}{a^2} + \frac{2n\bar{x}}{a^3} - \frac{3n}{a^4} (\bar{x}^2 + s'^2) \end{aligned}$$

La dérivée première admet comme racine positive $a = \frac{1}{2} (\sqrt{5\bar{x}^2 + s'^2} - \bar{x})$ qui vérifie $s'^2 + \bar{x}^2 = a^2 + a\bar{x}$ et donc la dérivée seconde admet comme valeur en ce point :

$$-\frac{2n}{a^2} - \frac{n\bar{x}}{a^3} < 0$$

La vraisemblance est donc maximum en ce point, l'expression de l'emv étant donc :

$$\hat{a}_n = \frac{1}{2} \left(\sqrt{5\bar{X}_n^2 + S_n'^2} - \bar{X}_n \right)$$

2. L'expression précédente est beaucoup trop compliquée et il est préférable d'utiliser la méthode des moments puisque $a = E(X)$. On peut donc prendre comme estimateur sans biais et convergent le moment empirique \bar{X}_n . Calculons l'information de Fisher :

$$I_n(a) = E \left(-\frac{\partial^2 \ln L}{\partial a^2} \right) = -\frac{n}{a^2} - \frac{2na}{a^3} + \frac{3n}{a^4} \left[E(\bar{X}_n^2) + E(S_n'^2) \right]$$

avec $E(\bar{X}_n^2) = V(\bar{X}_n) = \frac{a^2}{n} + a^2$ et $E(S_n'^2) = \frac{n-1}{n} a^2$ on obtient :

$$I_n(a) = \frac{3n}{a^2}$$

Ainsi :

$$V\left(\bar{X}_n\right) = \frac{a^2}{n} > \frac{1}{I_n(a)} = \frac{a^2}{3n}$$

donc cet estimateur n'est pas efficace.

3. Le modèle a pour variance a dont un estimateur sans biais et convergent est S_n^2 . On pourrait prendre aussi comme estimateur \bar{X}_n^2 qui, d'après ce qui précède, est un estimateur asymptotiquement sans biais de a^2 .

7.2 1. Comme $E(X) = \frac{\theta}{2}$, l'estimateur est obtenu par la méthode des moments comme solution de l'équation $\bar{X}_n = \frac{\theta}{2}$, soit $T_n = 2\bar{X}_n$. Cet estimateur est sans biais et convergent, de variance :

$$V(T_n) = 4V\left(\bar{X}_n\right) = 4\frac{V(X)}{n} = \frac{\theta^2}{3n}$$

La question de l'efficacité d'un estimateur ne se pose pas ici car nous ne sommes pas dans les conditions d'application de l'inégalité FDCR, l'ensemble des valeurs possibles pour X étant $X(\Omega) = [0, \theta]$ qui dépend donc du paramètre à estimer.

2. La vraisemblance a pour expression :

$$L(x_1, \dots, x_n; \theta) = \begin{cases} \theta^{-n} & \text{si } 0 \leq \min x_i \leq \max x_i \leq \theta \\ 0 & \text{sinon} \end{cases}$$

Ainsi, L est nulle pour $\theta < \max x_i$ et ensuite est décroissante pour $\theta \geq \max x_i$, donc est maximum pour $\theta = \max x_i$, ce qui correspond à l'emv :

$$M_n = \max\{X_1, \dots, X_n\}$$

Pour étudier ses propriétés, nous devons déterminer sa loi de probabilité :

$$P(M_n < x) = P\left\{\bigcap_{i=1}^n (X_i < x)\right\} = \prod_{i=1}^n P(X_i < x) = F^n(x)$$

en raison de l'indépendance et de l'identité des lois des v.a. X_i , de f.r. F . La densité de M_n est donc :

$$g(x) = nF^{n-1}(x)f(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}$$

Par conséquent :

$$E(M_n) = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{n+1}\theta$$

et l'estimateur sans biais est donc :

$$\hat{\theta}_n = \frac{n+1}{n}M_n$$

Pour calculer sa variance, on calcule d'abord :

$$E(M_n^2) = \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx = \frac{n}{n+2} \theta^2$$

d'où on déduit :

$$V(M_n) = \frac{n}{(n+1)^2(n+2)} \theta^2$$

puis :

$$V(\hat{\theta}_n) = \frac{\theta^2}{n(n+2)}$$

ce qui montre que $\hat{\theta}_n$ est convergent. Le rapport :

$$\frac{V(\hat{\theta}_n)}{V(T_n)} = \frac{3}{n+2} \rightarrow 0$$

montre que $\hat{\theta}_n$ est infiniment plus efficace que T_n .

- 7.3** La v.a. X admet comme espérance $E(X) = \frac{\theta+1}{2}$. La méthode des moments consiste à écrire l'égalité entre moment théorique, ici l'espérance, et moment empirique correspondant, ici la moyenne empirique \bar{X}_n . On résout donc l'équation en θ :

$$\frac{\theta+1}{2} = \bar{X}_n$$

La solution donne l'estimateur $T_n = 2\bar{X}_n - 1$. Cet estimateur est sans biais :

$$E(T_n) = E(2\bar{X}_n - 1) = 2E(\bar{X}_n) - 1 = 2E(X) - 1 = \theta$$

Il est aussi convergent d'après la loi des grands nombres :

$$\bar{X}_n \xrightarrow{p} E(X) = \frac{\theta+1}{2}$$

On en déduit du théorème de Slutsky que :

$$T_n = 2\bar{X}_n - 1 \xrightarrow{p} 2E(X) - 1 = \theta$$

La question de l'efficacité ne se pose pas puisque l'ensemble des valeurs possibles pour X est $\{1, 2, \dots, \theta\}$, qui dépend du paramètre à estimer θ .

- 7.4** La v.a. X suit une loi géométrique de paramètre $p = \frac{1}{\theta}$, donc avec :

$$E(X) = \frac{1}{p} = \theta \quad V(X) = \frac{1-p}{p^2} = \theta(\theta-1)$$

Le paramètre à estimer est la moyenne théorique, donc l'estimateur T_n de θ par la méthode des moments est la moyenne empirique \bar{X}_n . Cet estimateur est sans biais et convergent d'après la loi des grands nombres. Pour savoir s'il est efficace, on détermine

d'abord l'expression de la vraisemblance :

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n p(1-p)^{x_i-1} \\ &= p^n (1-p)^{s_n-n} = \frac{(\theta-1)^{s_n-n}}{\theta^{s_n}} \end{aligned}$$

ayant posé $s_n = \sum_{i=1}^n x_i$. La log-vraisemblance est donc :

$$\ln L(x_1, \dots, x_n; \theta) = (s_n - n) \ln(\theta - 1) - s_n \ln \theta$$

Soit en dérivant :

$$\frac{\partial \ln L}{\partial \theta} = \frac{s_n - n}{\theta - 1} - \frac{s_n}{\theta}$$

On dérive une nouvelle fois :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{s_n - n}{(\theta - 1)^2} + \frac{s_n}{\theta^2}$$

On calcule alors la quantité d'information de Fisher :

$$I_n(\theta) = E\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right) = \frac{E(S_n) - n}{(\theta - 1)^2} - \frac{E(S_n)}{\theta^2} = \frac{n\theta - n}{(\theta - 1)^2} - \frac{n\theta}{\theta^2} = \frac{n}{\theta(\theta - 1)}$$

Par ailleurs :

$$V(T_n) = V(\bar{X}_n) = \frac{V(X)}{n} = \frac{\theta(\theta - 1)}{n} = \frac{1}{I_n(\theta)}$$

L'estimateur T_n est donc efficace.

7.5 1. On obtient :

$$E(X) = \frac{1}{\theta} \int_1^{+\infty} x^{-1/\theta} dx = \frac{1}{1-\theta}$$

sous réserve que cette intégrale soit convergente, c'est-à-dire que $1 - 1/\theta < 0$ soit $0 < \theta < 1$.

L'équation $\bar{X}_n = 1/(1 - \theta)$ donne comme solution l'estimateur $T_n = 1 - 1/\bar{X}_n$.

2. La v.a. Y est positive, donc $G(y) = P(Y < y) = 0$ pour $y \leq 0$. Pour $y > 0$:

$$G(y) = P(\ln X < y) = P(X < e^y) = F(e^y)$$

où F est la f.r. de X . La densité obtenue par dérivation est :

$$g(y) = e^y f(e^y) = \frac{1}{\theta} e^{-y/\theta}$$

qui est la densité de la loi exponentielle de paramètre $1/\theta$.

3. L'expression de la vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i) = \left(\frac{1}{\theta}\right)^n \left(\prod_{i=1}^n x_i\right)^{-1-1/\theta}$$

La log-vraisemblance s'écrit :

$$\ln L(x_1, \dots, x_n; \theta) = -n \ln \theta - \left(1 + \frac{1}{\theta}\right) \sum_{i=1}^n \ln x_i$$

Cette fonction est dérivable pour tout $\theta > 0$ avec :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n \ln x_i$$

La dérivée s'annule pour $\theta = \sum_{i=1}^n \ln x_i / n$; la dérivée seconde est négative pour cette valeur :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n \ln x_i$$

L'estimateur du maximum de vraisemblance est donc :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \ln X_i = \bar{Y}_n$$

Cet estimateur moyenne empirique est un estimateur sans biais de la moyenne théorique de Y et convergent d'après la loi des grands nombres.

Sa variance est :

$$V(\bar{Y}_n) = \frac{V(Y)}{n} = \frac{\theta^2}{n}$$

La quantité d'information de Fisher est :

$$I_n(\theta) = E \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right) = -\frac{n}{\theta^2} + \frac{2}{\theta^3} \sum_{i=1}^n E(Y_i) = \frac{n}{\theta^2}$$

L'estimateur est donc efficace.

4. La v.a. Σ_n est la somme de n v.a. indépendantes et de même loi $\gamma(1, 1/\theta)$ donc suit une loi $\gamma(n, 1/\theta)$. On en déduit successivement que Σ_n/θ suit une loi $\gamma(n)$ et que $2\Sigma_n/\theta$ suit une loi $\gamma(n, 1/2)$ qui est la loi χ_{2n}^2 (cf. chap. 3, 2.4 et 2.5). On en déduit un intervalle de confiance bilatéral pour θ par la condition :

$$1 - \alpha = P \left\{ a < \frac{2\Sigma_n}{\theta} < b \right\} = P \left\{ 2n \frac{\hat{\theta}_n}{b} < \theta < 2n \frac{\hat{\theta}_n}{a} \right\}$$

a et b étant les fractiles d'ordres respectifs $\alpha/2$ et $1 - \alpha/2$ de la loi χ_{2n}^2 .

7.6 1. Chaque variable X_i suit une loi $N(m_i + a, \sigma)$, $1 \leq i \leq n$; les variables $Y_i = X_i - m_i$ constituent donc un échantillon d'une v.a. Y de loi $N(a, \sigma)$. Par conséquent, l'estimateur :

$$\hat{a}_n = \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n (X_i - m_i)$$

est un estimateur sans biais et convergent du paramètre a .

L'expression de la log-vraisemblance est :

$$\ln L(x_1, \dots, x_n; a) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_i - a)^2$$

de dérivées :

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m_i - a)$$

$$\frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{\sigma^2}$$

On remarque ainsi que \hat{a}_n est un emv et qu'il est efficace car :

$$V(\hat{a}_n) = \frac{\sigma^2}{n} = \frac{1}{I_n(a)}$$

puisque :

$$I_n(a) = E \left(-\frac{\partial^2 \ln L}{\partial a^2} \right) = \frac{n}{\sigma^2}$$

On obtient un intervalle de confiance à partir de \hat{a}_n dont on connaît la loi, qui est la loi normale $N(a, \sigma/\sqrt{n})$:

$$\hat{a}_n - u \frac{\sigma}{\sqrt{n}} < a < \hat{a}_n + u \frac{\sigma}{\sqrt{n}}$$

où u est le fractile d'ordre $1 - \alpha/2$ de la loi $N(0, 1)$, soit $u = 1,96$ pour $\alpha = 0,05$.

2. Les variables $Z_i = X_i/m_i$ sont indépendantes et de loi $N(b, \frac{\sigma}{m_i})$, donc la statistique :

$$\hat{b}_n = \frac{1}{n} \sum_{i=1}^n Z_i = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{m_i}$$

est un estimateur sans biais et convergent, de loi normale d'espérance b et de variance :

$$V(\hat{b}_n) = \frac{1}{n^2} \sum_{i=1}^n \frac{\sigma^2}{m_i^2}$$

La log-vraisemblance a pour expression :

$$\ln L(x_1, \dots, x_n; b) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - bm_i)^2$$

de dérivées :

$$\frac{\partial \ln L}{\partial b} = \frac{1}{\sigma^2} \sum_{i=1}^n m_i (x_i - bm_i)$$

$$\frac{\partial^2 \ln L}{\partial b^2} = -\frac{1}{\sigma^2} \sum_{i=1}^n m_i^2$$

donc l'information de Fisher a pour expression :

$$I_n(b) = E \left(-\frac{\partial^2 \ln L}{\partial b^2} \right) = \frac{1}{\sigma^2} \sum_{i=1}^n m_i^2 = \frac{n\bar{m}}{\sigma^2}$$

en ayant posé $\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i^2$. La variance de l'estimateur peut aussi s'écrire $V(\hat{b}_n) = \frac{\sigma^2}{n\bar{m}}$, en ayant noté par h la moyenne harmonique des nombres m_i^2 , définie par

$\frac{1}{h} = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i^2}$. Comme les moyennes arithmétique et géométrique vérifient toujours la relation $h < \bar{m}$, on en conclut que :

$$V(b_n) > \frac{1}{I_n(b)}$$

c'est-à-dire que cet estimateur n'est pas efficace, sauf dans le cas particulier où tous les m_i sont égaux.

3. L'estimation $\hat{a}_{10} = 9,5$ conduit à l'intervalle :

$$7,64 < a < 11,36$$

On obtient $\hat{b}_{10} = 1,095$.

7.7 Nous avons vu dans l'exemple 7.8 que l'emv est $\hat{\theta}_n = \bar{X}_n$, qui est donc un estimateur sans biais et convergent, comme toute moyenne empirique lorsque le paramètre à estimer est la moyenne théorique, avec ici :

$$V(\hat{\theta}_n) = \frac{\theta^2}{n}$$

La quantité d'information de Fisher a pour valeur :

$$I_n(\theta) = E \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right) = -\frac{n}{\theta^2} + \frac{2nE(X)}{\theta^3} = \frac{n}{\theta^2}$$

Donc cet estimateur est aussi efficace.

Pour construire un intervalle de confiance basé sur $\hat{\theta}_n$ on a besoin de connaître sa loi de probabilité. On va pour cela utiliser le fait que la loi exponentielle appartient à la famille des lois gamma : $X \rightsquigarrow \mathcal{E}(1/\theta) \equiv \gamma(1, 1/\theta)$. Utilisant l'une des propriétés des lois gamma on en conclut que $S_n = \sum_{i=1}^n X_i \rightsquigarrow \gamma(n, 1/\theta)$ et donc $S_n/\theta \rightsquigarrow \gamma(n)$. Le résultat du chapitre 3 §2.5 permet de recourir aux tables des lois usuelles puisqu'on y a établi que $2S_n/\theta \rightsquigarrow \chi_{2n}^2$. On peut donc déterminer deux valeurs a et b telles que :

$$p \left(a < \frac{2S_n}{\theta} < b \right) = 1 - \alpha$$

et déduire de là l'intervalle de confiance de niveau $1 - \alpha$:

$$\frac{2S_n}{b} < \theta < \frac{2S_n}{a}$$

En choisissant des risques symétriques, pour $\alpha = 0,05$ on lit dans la table 5 les fractiles d'ordres respectifs 0,025 et 0,975 de la loi χ_{20}^2 : $a = 9,59$ et $b = 34,17$ d'où l'intervalle $0,67 < \theta < 2,40$.

7.8 La log-vraisemblance a pour expression (cf. chap. 3, 2.7) :

$$\ln L(x_1, \dots, x_n; m) = -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \ln x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - m)^2$$

de dérivées :

$$\frac{\partial \ln L}{\partial m} = \frac{1}{\sigma^2} \sum_{i=1}^n (\ln x_i - m)$$

$$\frac{\partial^2 \ln L}{\partial m^2} = -\frac{n}{\sigma^2} < 0$$

donc l'emv est :

$$\widehat{m}_n = \frac{1}{n} \sum_{i=1}^n \ln X_i$$

Or on sait, par définition de la loi log-normale, que $\ln X \rightsquigarrow N(m, \sigma)$ donc cet estimateur est sans biais et convergent, de variance :

$$V(\widehat{m}_n) = \frac{\sigma^2}{n}$$

donc il est aussi efficace puisque $I_n(m) = \frac{n}{\sigma^2}$. Il suit la loi $N(m, \sigma/\sqrt{n})$ d'où l'intervalle de niveau 0,95 :

$$\widehat{m}_n - 1,96 \frac{\sigma}{\sqrt{n}} < m < \widehat{m}_n + 1,96 \frac{\sigma}{\sqrt{n}}$$

soit pour $\sigma = 1$ et $n = 25$: $1,81 < m < 2,59$.

7.9 1. On calcule :

$$E(X) = \frac{2}{\theta^2} \int_0^\theta x^2 dx = \frac{2}{3}\theta$$

et l'estimateur sans biais est solution de l'équation en θ , $\bar{X}_n = \frac{2}{3}\theta$, soit :

$$T_n = \frac{3}{2}\bar{X}_n$$

qui est sans biais et convergent d'après la loi des grands nombres. On peut calculer sa variance à partir de :

$$E(X^2) = \frac{2}{\theta^2} \int_0^\theta x^3 dx = \frac{1}{2}\theta^2$$

d'où $V(T_n) = \frac{9}{4}V(\bar{X}_n) = \frac{9}{4n}V(X) = \frac{\theta^2}{8n}$. La question de l'efficacité ne se pose pas ici car les hypothèses de Cramer-Rao ne sont pas vérifiées, $X(\Omega) = [0, \theta]$ dépend du paramètre à estimer.

2. La vraisemblance a pour expression :

$$L(x_1, \dots, x_n; \theta) = \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^n x_i \quad \text{pour } 0 \leq \min x_i \leq \max x_i \leq \theta$$

elle est donc nulle pour $\theta < \max x_i$ et décroissante pour $\theta \geq \max x_i$, ce qui montre que l'emv est :

$$M_n = \max\{X_1, \dots, X_n\}$$

Pour étudier ses propriétés nous devons déterminer sa loi de probabilité :

$$P(M_n < x) = P\left\{\bigcap_{i=1}^n (X_i < x)\right\} = \prod_{i=1}^n P(X_i < x) = F^n(x; \theta)$$

en raison de l'indépendance et de l'identité des lois des v.a. X_i de f.r. F . La densité de M_n est donc :

$$g(x; \theta) = nF^{n-1}(x; \theta)f(x; \theta)$$

La f.r. de X étant définie par :

$$F(x; \theta) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{\theta} & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } \theta < x \end{cases}$$

Par conséquent :

$$g(x; \theta) = \begin{cases} 2n \frac{x^{2n-1}}{\theta^{2n}} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}$$

et :

$$E(M_n) = \frac{2n}{\theta^{2n}} \int_0^\theta x^{2n} dx = \frac{2n}{2n+1} \theta$$

L'estimateur sans biais est donc :

$$\hat{\theta}_n = \frac{2n+1}{2n} M_n$$

dont la variance se déduit de :

$$E(M_n^2) = \frac{2n}{\theta^{2n}} \int_0^\theta x^{2n+1} dx = \frac{2n}{2n+2} \theta^2$$

soit :

$$V(\hat{\theta}_n) = \left(\frac{2n+1}{2n}\right)^2 V(M_n) = \frac{\theta^2}{2n(2n+2)}$$

Cet estimateur est donc convergent et infiniment plus efficace que T_n , au sens où :

$$\frac{V(\hat{\theta}_n)}{V(T_n)} = \frac{2}{n+1} \rightarrow 0$$

3. Nous allons construire un intervalle de confiance à partir de M_n dont la loi est connue, de f.r. :

$$G(x; \theta) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^{2n}}{\theta^{2n}} & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } \theta < x \end{cases}$$

Nous cherchons a et b tels que $1 - \alpha = P(\theta a < M_n < \theta b)$ ou tels que $\alpha_1 = P(M_n < \theta a) = G(\theta a; \theta)$ et $\alpha_2 = P(M_n > \theta b) = 1 - G(\theta b; \theta)$ avec $\alpha_1 + \alpha_2 = \alpha$, ce qui va définir a et b par :

$$\alpha_1 = a^{2n}, 1 - \alpha_2 = b^{2n} \text{ soit } a = \alpha_1^{1/2n} \text{ et } b = (1 - \alpha_2)^{1/2n}$$

d'où l'intervalle de confiance :

$$\frac{M_n}{b} < \theta < \frac{M_n}{a}$$

Application : $a = 0,91, b = 0,999$ soit l'intervalle $5,00 < \theta < 5,48$.

7.10 Les estimateurs sans biais de σ_1^2 et σ_2^2 sont respectivement :

$$\hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - m)^2 \text{ et } \hat{\sigma}_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - m)^2$$

dont les lois sont données par $n_1 \hat{\sigma}_1^2 / \sigma_1^2 \rightsquigarrow \chi_{n_1}^2$ et $n_2 \hat{\sigma}_2^2 / \sigma_2^2 \rightsquigarrow \chi_{n_2}^2$. Les deux échantillons étant indépendants, on sait (cf. chap. 5, § 3.2) que $(\hat{\sigma}_1^2 / \sigma_1^2) / (\hat{\sigma}_2^2 / \sigma_2^2) \rightsquigarrow F(n_1, n_2)$ donc l'intervalle de confiance est défini à partir de :

$$1 - \alpha = P\left(a < \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} < b\right)$$

soit :

$$\frac{1}{b} \times \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{1}{a} \times \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

7.11 1. Comme $E(X) = V(X) = \theta$, on peut retenir comme estimateurs sans biais de θ les moments empiriques :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Ces deux estimateurs sont aussi convergents, de variances respectives :

$$V(\bar{X}_n) = \frac{\theta}{n} \text{ et } V(S_n^2) = \frac{2\theta^2}{n-1}$$

Le rapport $V(S_n^2) / V(\bar{X}_n)$ tend vers 2θ quand n devient infini, valeur inconnue qui ne permet pas de comparer ces deux estimateurs. Étudions leur efficacité en écrivant d'abord la log-vraisemblance :

$$\ln L(x_1, \dots, x_n; \theta) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \frac{1}{2\theta} \sum_{i=1}^n (x_i - \theta)^2$$

puis en dérivant :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 - \frac{n}{2}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3} \sum_{i=1}^n x_i^2$$

D'où la quantité d'information de Fisher :

$$I_n(\theta) = E \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right) = -\frac{n}{2\theta^2} + \frac{1}{\theta^3} nE(X^2)$$

soit, avec $E(X^2) = V(X) + E^2(X) = \theta + \theta^2$:

$$I_n(\theta) = \frac{n}{2\theta^2} + \frac{n}{\theta}$$

On voit ainsi que :

$$\frac{1}{V(\bar{X}_n)} < I_n(\theta) \quad \text{et} \quad \frac{1}{V(S_n^2)} < I_n(\theta)$$

donc aucun de ces estimateurs n'est efficace.

2. Pour construire un intervalle de confiance, nous allons utiliser le fait que (cf. chap. 5, 3.1) :

$$\sqrt{n} \frac{\bar{X}_n - \theta}{S_n} \rightsquigarrow T_{n-1}$$

ce qui permet de déterminer les valeurs de a et b telles que :

$$P \left(a < \sqrt{n} \frac{\bar{X}_n - \theta}{S_n} < b \right) = 1 - \alpha$$

d'où l'intervalle :

$$\bar{X}_n - b \frac{S_n}{\sqrt{n}} < \theta < \bar{X}_n - a \frac{S_n}{\sqrt{n}}$$

Pour $\alpha = 0,05$ on retient des risques symétriques puisque la loi est symétrique et on lit dans la table 6 les fractiles $b = -a = 2,064$ avec $\bar{x}_{25} = 2,01$ et $s_{25} = 1,42$ d'où l'intervalle :

$$1,42 < \theta < 2,59$$

7.12 1. L'expression de la vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i) = \left(\frac{\theta}{2}\right)^n \exp \left\{ -\theta \sum_{i=1}^n |x_i| \right\}$$

La log-vraisemblance s'écrit :

$$\ln L(x_1, \dots, x_n; \theta) = -n \ln 2 + n \ln \theta - \theta \sum_{i=1}^n |x_i|$$

Cette fonction est dérivable pour tout $\theta > 0$ avec :

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n |x_i|$$

La dérivée s'annule pour $\theta = n / \sum_{i=1}^n |x_i|$; la dérivée seconde est négative :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2}$$

L'estimateur du maximum de vraisemblance est donc :

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^n |X_i|}$$

Pour étudier les propriétés de cet estimateur nous allons déterminer la loi de $U = \theta |X|$.

La fonction de répartition est nulle pour $u \leq 0$ et définie pour $u > 0$ par :

$$G(u) = P(U < u) = F\left(\frac{u}{\theta}\right) - F\left(-\frac{u}{\theta}\right)$$

D'où une densité définie par :

$$g(u) = \frac{2}{\theta} f\left(\frac{u}{\theta}\right) = e^{-u}$$

Il s'agit donc de la loi exponentielle ou loi $\gamma(1)$. Par conséquent $S_n = \theta \sum_{i=1}^n |X_i|$ suit la loi $\gamma(n)$. Un calcul intégral nous permet alors d'obtenir :

$$E\left(\frac{1}{S_n}\right) = \frac{1}{n-1} \text{ et } V\left(\frac{1}{S_n}\right) = \frac{1}{(n-2)(n-1)^2}$$

L'estimateur s'écrivant sous la forme $\hat{\theta}_n = n\theta/S_n$, on en déduit :

$$E(\hat{\theta}_n) = \frac{n}{n-1}\theta \text{ et } V(\hat{\theta}_n) = \frac{n^2}{(n-2)(n-1)^2}\theta^2$$

Cet estimateur est biaisé, donc il ne peut pas être efficace. Il est asymptotiquement sans biais et de variance qui tend vers 0, donc il est convergent.

2. On a $\Sigma_n = S_n/\theta$ qui suit donc la loi $\gamma(n, \theta)$ et donc $2\theta\Sigma_n$ suit la loi $\gamma(n, 1/2)$ ou loi χ_{2n}^2 (cf. chapitre 3 § 2.4 et 2.5). On en déduit un intervalle de confiance bilatéral pour θ par la condition :

$$1 - \alpha = P\{a < 2\theta\Sigma_n < b\} = P\left\{\frac{a}{2\Sigma_n} < \theta < \frac{b}{2\Sigma_n}\right\}$$

a et b étant les fractiles d'ordres respectifs $\alpha/2$ et $1 - \alpha/2$ de la loi χ_{2n}^2 .

3. On déduit de la question 1 que $E(|X|) = 1/\theta$ et $V(|X|) = 1/\theta^2$. L'application du théorème central limite à $1/\hat{\theta}_n$ permet donc d'obtenir :

$$\sqrt{n} \frac{1/\hat{\theta}_n - 1/\theta}{1/\theta} \xrightarrow{\text{loi}} N(0, 1)$$

En appliquant la propriété 2.7 du chapitre 6 on en conclut :

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\theta} \xrightarrow{\text{loi}} N(0, 1)$$

On en déduit un intervalle de confiance bilatéral pour θ par la condition :

$$1 - \alpha = P \left\{ -u < \sqrt{n} \frac{\hat{\theta}_n - \theta}{\theta} < u \right\}$$

où u est approximé par le fractile d'ordre $\alpha/2$ de la loi normale centrée réduite.

L'intervalle est :

$$\frac{\hat{\theta}_n}{1 + u/\sqrt{n}} < \theta < \frac{\hat{\theta}_n}{1 - u/\sqrt{n}}$$

7.13 1. La vraisemblance s'écrit :

$$L(x_1, \dots, x_n; \theta) = (2\theta)^{-n} \prod_{i=1}^n x_i^{-1/2} \exp -\frac{1}{\theta} \sum_{i=1}^n \sqrt{x_i}$$

d'où la log-vraisemblance :

$$\ln L(x_1, \dots, x_n; \theta) = -n \ln 2 - n \ln \theta - \frac{1}{2} \ln \prod_{i=1}^n x_i - \frac{1}{\theta} \sum_{i=1}^n \sqrt{x_i}$$

et ses dérivées :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n \sqrt{x_i}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n \sqrt{x_i}$$

La dérivée première s'annule pour $\theta = \frac{1}{n} \sum_{i=1}^n \sqrt{x_i}$ avec une dérivée seconde en ce point de valeur $-n/\theta^2$, donc l'emv est :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \sqrt{X_i}$$

Pour étudier ses propriétés, nous allons déterminer la loi de la v.a. $Y = \sqrt{X}$:

$$G(y; \theta) = P(Y < y) = P(\sqrt{X} < y) = P(X < y^2) = F(y^2; \theta)$$

pour $y > 0$. Sa densité est alors définie par :

$$g(y; \theta) = 2yf(y^2; \theta) = \frac{1}{\theta} e^{-y/\theta}$$

donc $Y \rightsquigarrow \mathcal{E}(1/\theta)$ et l'estimateur $\hat{\theta}_n = \bar{Y}_n$ est un estimateur sans biais et convergent de $\theta = E(Y)$, de variance θ^2/n .

L'information de Fisher se calcule par :

$$I_n(\theta) = E \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right) = -\frac{n}{\theta^2} + \frac{2n}{\theta^3} E(\hat{\theta}_n) = \frac{n}{\theta^2}$$

donc $\hat{\theta}_n$ est efficace.

2. Pour construire un intervalle de confiance à partir de $\hat{\theta}_n$, on utilise le fait que $Y \rightsquigarrow \mathcal{E}(1/\theta) \equiv \gamma(1, 1/\theta)$ donc $n\hat{\theta}_n \rightsquigarrow \gamma(n, 1/\theta)$ puis $n\hat{\theta}_n/\theta \rightsquigarrow \gamma(n, 1/\theta)$ et enfin (cf. chapitre 3 § 2.4 et 2.5) $2n\hat{\theta}_n/\theta \rightsquigarrow \chi_{2n}^2$. On peut donc trouver deux valeurs a et b telles que :

$$1 - \alpha = P \left(a < \frac{2n\hat{\theta}_n}{\theta} < b \right)$$

ce qui permet d'obtenir l'intervalle :

$$\frac{2n\hat{\theta}_n}{b} < \theta < \frac{2n\hat{\theta}_n}{a}$$

Pour $\alpha = 0, 10$ et en prenant des risques symétriques, on lit dans la table 5 : $a = 26, 51$ et $b = 55, 76$; on obtient par ailleurs $\hat{\theta}_{20} = 2, 37$ d'où l'intervalle :

$$1, 70 < \theta < 3, 58$$

intervalle très grand en raison de la faible taille de l'échantillon.

7.14 1. D'après l'exemple 4.13 et la forme de la densité on sait que X et Y suivent des lois normales centrées et de variances respectives $1 + 4\theta$ et $1 + \theta$.

On a de plus $Cov(X, Y) = -(1 + 2\theta)$.

D'après la définition de la loi normale vectorielle, $X + Y$ suit une loi normale. Cette loi est centrée, avec :

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y) = \theta = E(X + Y)^2$$

L'estimateur est donc :

$$T_n = \frac{1}{n} \sum_{i=1}^n (X_i + Y_i)^2$$

2. Cet estimateur est sans biais et convergent d'après la loi des grands nombres. On sait également que nT_n/θ suit une loi χ_n^2 et donc $V(T_n) = 2\theta^2/n$. Pour savoir s'il est efficace, nous devons calculer la quantité d'information de Fisher. L'expression de la vraisemblance est :

$$\begin{aligned} L(x_1, y_1, \dots, x_n, y_n; \theta) &= \prod_{i=1}^n f(x_i, y_i) \\ &= \left(\frac{1}{2\pi\sqrt{\theta}} \right)^n \exp -\frac{1}{2\theta} \sum_{i=1}^n [(1 + \theta)x_i^2 + 2(1 + 2\theta)x_i y_i + (1 + 4\theta)y_i^2] \end{aligned}$$

La log-vraisemblance s'écrit :

$$\ln L(x_1, y_1, \dots, x_n, y_n; \theta) = -n \ln 2\pi - \frac{n}{2} \ln \theta - \frac{1}{2\theta} \sum_{i=1}^n [(1 + \theta)x_i^2 + 2(1 + 2\theta)x_i y_i + (1 + 4\theta)y_i^2]$$

Cette fonction est dérivable pour tout $\theta > 0$ avec :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n (x_i + y_i)^2$$

La dérivée seconde est :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3} \sum_{i=1}^n (x_i + y_i)^2$$

La quantité d'information de Fisher est donc :

$$I_n(\theta) = E \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right) = -\frac{n}{2\theta^2} + \frac{1}{\theta^3} \sum_{i=1}^n E(X_i + Y_i)^2 = \frac{n}{2\theta^2}$$

L'estimateur est donc efficace.

L'intervalle de confiance bilatéral pour θ est obtenu par la condition :

$$1 - \alpha = P \left\{ a < nT_n/\theta < b \right\} = P \left\{ \frac{nT_n}{b} < \theta < \frac{nT_n}{a} \right\}$$

a et b étant les fractiles d'ordres respectifs $\alpha/2$ et $1 - \alpha/2$ de la loi χ_{2n}^2 .

3. L'application du théorème central limite à T_n permet d'obtenir :

$$\sqrt{n} \frac{T_n - \theta}{\theta \sqrt{2}} \xrightarrow{\text{loi}} N(0, 1)$$

On en déduit un intervalle de confiance bilatéral pour θ par la condition :

$$1 - \alpha = P \left\{ -u < \sqrt{n} \frac{T_n - \theta}{\theta \sqrt{2}} < u \right\}$$

où u est approximé par le fractile d'ordre $\alpha/2$ de la loi normale centrée réduite.

L'intervalle est :

$$\frac{T_n}{1 + u/\sqrt{2/n}} < \theta < \frac{T_n}{1 - u/\sqrt{2/n}}$$

7.15 1. L'espérance se calcule par :

$$E(X) = \int_{-\infty}^{+\infty} x f(x; \theta) dx = \frac{1}{\theta} \int_{\theta}^{+\infty} x \exp\left(-\frac{x-\theta}{\theta}\right) dx$$

On fait alors le changement $x - \theta = \theta u$, puis on intègre par parties :

$$E(X) = \theta \int_0^{+\infty} (1 + u)e^{-u} du = \theta \left[-(1 + u)e^{-u} \right]_0^{+\infty} + \theta \int_0^{+\infty} e^{-u} du = 2\theta$$

2. L'estimateur T_n de θ obtenu par la méthode des moments est solution de l'équation en θ :

$$2\theta = \overline{X}_n$$

Il s'agit donc de $T_n = \frac{1}{2}\overline{X}_n$. Il est sans biais par construction :

$$E(T_n) = \frac{1}{2}E(\overline{X}_n) = \frac{1}{2}E(X) = \theta$$

Il est convergent d'après la loi des grands nombres :

$$\overline{X}_n \xrightarrow{p} E(X) = 2\theta$$

Donc :

$$T_n = \frac{1}{2}\overline{X}_n \xrightarrow{p} \frac{1}{2}E(X) = \theta$$

La question de l'efficacité ne se pose pas puisque X prend ses valeurs dans l'intervalle $[\theta, +\infty[$ qui dépend de θ .

3. D'après le théorème central-limite :

$$\sqrt{n} \frac{\overline{X}_n - E(X)}{\sigma(X)} \xrightarrow{\text{loi}} N(0, 1)$$

On doit calculer la variance de X :

$$E(X^2) = \frac{1}{\theta} \int_{\theta}^{+\infty} x^2 \exp\left(-\frac{x-\theta}{\theta}\right) dx$$

On fait alors le changement $x - \theta = \theta u$ puis on intègre par parties :

$$\begin{aligned} E(X^2) &= \theta^2 \int_0^{+\infty} (1+u)^2 e^{-u} du \\ &= \theta^2 \left[-(1+u)^2 e^{-u} \right]_0^{+\infty} + 2\theta^2 \int_0^{+\infty} (1+u) e^{-u} du \\ &= \theta^2 + 2\theta E(X) = 5\theta^2 \end{aligned}$$

Ainsi, $V(X) = \theta^2$ et on obtient :

$$\sqrt{n} \frac{\overline{X}_n - 2\theta}{\theta} = \sqrt{n} \frac{T_n - \theta}{\theta/2} \xrightarrow{\text{loi}} N(0, 1)$$

L'estimateur T_n suit asymptotiquement la loi normale $N\left(\theta, \frac{\theta}{2\sqrt{n}}\right)$. On construit donc un intervalle de confiance bilatéral symétrique, puisque cette loi est symétrique. On peut trouver la valeur approchée de la constante u telle que :

$$1 - \alpha = P\left(-u < \sqrt{n} \frac{T_n - \theta}{\theta/2} < u\right)$$

L'intervalle de confiance est alors défini par :

$$-\theta \frac{u}{2\sqrt{n}} < T_n - \theta < \theta \frac{u}{2\sqrt{n}}$$

Ce qui est équivalent à :

$$\frac{T_n}{1 + u/2\sqrt{n}} < \theta < \frac{T_n}{1 - u/2\sqrt{n}}$$

où u est approximé par le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale standard $N(0, 1)$. Pour $1 - \alpha = 0,95$ on retient la valeur approchée $u = 2$; pour cet échantillon $t_{100} = 3,3$ d'où l'intervalle $3 < \theta < 3,67$.

Introduction

On appelle **théorie des tests** la seconde branche de la statistique mathématique, celle qui permet de confronter deux hypothèses retenues *a priori*. Comme dans le cadre d'un problème d'estimation, on retient un modèle statistique où la v.a. X suit une loi de probabilité P_θ , qui dépend d'un paramètre θ inconnu. On dispose cependant ici d'informations supplémentaires qui font penser *a priori* que la valeur de ce paramètre est égale à une valeur fixée θ_0 et on cherche à valider (à tester) cette hypothèse, au vu d'un échantillon de la loi de X . Cette hypothèse qui est privilégiée, parce qu'elle paraît la plus vraisemblable *a priori*, est appelée hypothèse nulle et notée H_0 . Construire un test va consister à partitionner l'ensemble \mathbb{R}^n des réalisations possibles du n -échantillon en deux régions, celle où l'on décidera d'accepter H_0 , et celle où l'on décidera de la rejeter, qui se nommera région critique du test. Pour délimiter ces deux régions, on fixe *a priori* une valeur (faible) à la probabilité de l'erreur qui consiste à décider, au vu de l'échantillon, de rejeter l'hypothèse nulle alors que celle-ci est vérifiée. Cette probabilité se nomme risque de première espèce et sa valeur standard est de 5 %. Lorsque le paramètre θ ne peut prendre que deux valeurs distinctes θ_0 et θ_1 , c'est le théorème de Neyman et Pearson qui permet de déterminer la forme de la région critique, à partir du rapport des vraisemblances associées à chacune des deux valeurs possibles du paramètre. Dans le cas où on peut attribuer des probabilités *a priori* à ces deux valeurs, ainsi que des coûts d'erreur, on utilise la méthode de Bayes.

Objectifs

Montrer comment, à partir d'observations indépendantes on peut choisir entre deux hypothèses relatives à la valeur du paramètre qui caractérise la loi retenue dans le modèle.

Présenter un test d'indépendance de deux caractères

Concepts clés

Hypothèses nulle et alternative
Région critique
Risque de première espèce
Méthodes de Bayes et de Neyman et Pearson
Test du khi-deux

1 Concepts principaux en théorie des tests

Dans le chapitre précédent, nous avons retenu un modèle statistique paramétrique pour décrire de façon simplifiée, mais théorique, un certain phénomène réel. Les valeurs observées, liées à ce phénomène, sont alors considérées comme des réalisations d'une variable aléatoire dont la loi est inconnue, mais appartient à une famille donnée.

Cette loi est totalement spécifiée par la connaissance d'un nombre réel θ , appelé paramètre, et qui permet de repérer précisément l'élément de cette famille de lois. La théorie de l'estimation fournit les outils permettant de se faire une idée de la valeur numérique de ce paramètre.

Ici, la théorie des tests va être un outil d'aide à la décision. Dans le cadre du même modèle, on dispose cette fois de plus d'informations *a priori* sur le paramètre. Ces informations se traduisent par deux hypothèses seulement sur les valeurs possibles du paramètre. En fonction des observations, on devra choisir l'une de ces deux hypothèses. Nous allons montrer, à partir d'un exemple, comment on peut se fixer une règle de décision entre ces deux hypothèses.

Le ministre de l'Économie et des Finances s'interroge pour savoir s'il doit prendre des mesures de relance de l'économie. Sa décision va être fondée sur les observations de l'accroissement mensuel de l'indice de la production industrielle. Cet accroissement est mesuré par l'Insee avec une certaine incertitude, ce qui amène à le considérer comme une v.a. I de loi normale, de moyenne m , et d'écart type connu $\sigma = 0,2 \%$. Dans la période antérieure, le paramètre m avait pour valeur $m = 0,5\%$. En période de récession, on considère que ce paramètre prend la valeur $m = 0,3\%$. Pour faire un choix entre ces deux valeurs, le ministre attend de disposer des valeurs de I pour le dernier trimestre. Inquiet de l'effet de mesures de relance sur l'inflation, il se fixe *a priori* la règle de décision suivante : *si la moyenne des accroissements du trimestre est inférieure à 0,35 % alors je prends des mesures de relance*. On peut alors se poser les questions suivantes : est-il possible de mesurer les risques associés à cette règle arbitraire ? peut-on fixer à l'aide de critères objectifs un autre seuil que la valeur retenue de 0,35 % ?

Le modèle statistique va nous permettre de répondre et d'associer des évaluations numériques à cette règle de décision. La v. a. I appartient ici à la famille des lois normales, d'écart type connu $\sigma = 0,2$. L'autre paramètre de cette loi, la moyenne m , est inconnu, mais ne peut prendre ici que deux valeurs. Il s'agit donc de faire un choix entre les deux hypothèses :

$$\begin{cases} H_0 : I \rightsquigarrow N(0,3; 0,2) \\ H_1 : I \rightsquigarrow N(0,5; 0,2) \end{cases}$$

Chacune de ces hypothèses a pour conséquence une décision qui lui est associée :

- D_0 : prendre des mesures de relance de l'économie ;
- D_1 : ne rien faire.

Cette décision va être prise au vu d'un échantillon (I_1, I_2, I_3) de cette v.a. I observée au cours du dernier trimestre. La règle de décision retenue par le ministre se formalise alors de la façon suivante :

$$\begin{array}{ll} \text{si} & \frac{1}{3} (I_1 + I_2 + I_3) < k & \text{on décide } D_0 \\ \text{si} & \frac{1}{3} (I_1 + I_2 + I_3) \geq k & \text{on décide } D_1 \end{array}$$

La valeur de k , appelé seuil critique, est fixée arbitrairement ici à $k = 0,35$.

Chacune de ces décisions a pour conséquence une erreur éventuelle :

- relancer l'économie (D_0) en période d'expansion (H_1) et favoriser ainsi l'inflation ;
- ne rien faire (D_1) en période de récession (H_0) et accroître le chômage.

Le modèle statistique retenu permet alors de calculer les probabilités associées à ces deux erreurs. Par exemple :

$$\alpha = P(\text{ne rien faire} | m = 0,3) = P(D_1 | H_0) = P\left(\frac{1}{3} \sum_{j=1}^3 I_j \geq k | H_0\right)$$

Sous l'hypothèse H_0 , la loi de $\bar{I} = \frac{1}{3} \sum_{j=1}^3 I_j$ est la loi normale $N\left(0,3; \frac{0,2}{\sqrt{3}}\right)$.

On peut donc calculer la probabilité précédente en utilisant une v.a. U de loi $N(0, 1)$:

$$\begin{aligned} \alpha &= P(\bar{I} \geq 0,35 | H_0) = P\left(\frac{\bar{I} - 0,3}{0,2/\sqrt{3}} \geq \frac{0,05}{0,2} \sqrt{3} | H_0\right) \\ &= P(U \geq 0,43) = 0,33 \end{aligned}$$

De même, l'autre risque d'erreur se calcule par :

$$\begin{aligned} \beta &= P(\text{relancer} | m = 0,5) = P(D_0 | H_1) = P(\bar{I} < k | H_1) \\ &= P\left(\frac{\bar{I} - 0,5}{0,2/\sqrt{3}} < -\frac{0,15}{0,2} \sqrt{3} | H_1\right) = P(U < -1,30) = 0,097 \end{aligned}$$

Ces deux risques ne sont donc pas équivalents, le premier étant trois fois supérieur au second. Cette règle correspond donc bien à un souhait de se garantir avant tout contre l'inflation. Si on veut que le seuil ne soit pas fixé arbitrairement, c'est par le choix d'une valeur de risque que l'on en déduira alors une valeur de seuil critique. Si on souhaite plutôt se prémunir prioritairement contre le chômage, on fixe une valeur faible au risque α , par exemple $\alpha = 5\%$. Il va en découler une valeur du seuil k par la condition :

$$\alpha = 0,05 = P\left(\frac{1}{3} \sum_{j=1}^3 I_j \geq k | H_0\right) = P\left(U \geq \frac{k - 0,3}{0,2/\sqrt{3}}\right)$$

On obtient ainsi :

$$\frac{k - 0,3}{0,2/\sqrt{3}} = 1,6449 \quad \text{soit} \quad k = 0,3 + \frac{0,2}{\sqrt{3}} 1,6449 = 0,49$$

L'autre risque a alors pour valeur :

$$\beta = P(\bar{I} < k | H_1) = P\left(U < \frac{0,49 - 0,5}{0,2/\sqrt{3}}\right) = P(U < -0,09) = 0,4641$$

Le risque de relancer à tort est cette fois très élevé. Pour une décision où ce risque serait considéré comme le plus dommageable, il faudrait fixer le seuil k par la condition :

$$\beta = 0,05 = P(\text{relancer} | m = 0,5) = P\left(U < \frac{k - 0,5}{0,2/\sqrt{3}}\right)$$

On obtient alors comme valeur :

$$k = 0,5 - \frac{0,2}{\sqrt{3}} 1,6449 = 0,31$$

La règle de décision, déterminée par le seuil, est fortement dépendante du risque contre lequel on souhaite se prémunir en priorité.

Cet exemple introductif va nous permettre de formaliser un problème général de test. On considère un modèle statistique où la loi de probabilité P_θ de la v.a. X dépend d'un paramètre inconnu θ qui varie dans un sous-ensemble donné Θ de \mathbb{R} . On suppose que cet ensemble est partitionné en deux sous-ensembles donnés Θ_0 et Θ_1 , auxquels vont être associées les deux **hypothèses notées** $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$. Construire un *test* consiste à définir une **règle de décision** qui va associer une décision à un échantillon observé (X_1, \dots, X_n) de la loi de X , les deux décisions possibles étant D_0 : accepter H_0 , et D_1 : accepter H_1 . À chaque décision correspond une région de \mathbb{R}^n , qui va donc être partitionné en deux sous-ensembles W et \overline{W} , c'est-à-dire que si la réalisation de l'échantillon est un point (x_1, \dots, x_n) de W on décide D_1 , donc on rejette H_0 . Dans le cas contraire, c'est-à-dire pour un point de \overline{W} , on décide D_0 , donc on accepte H_0 .

DÉFINITION

La région W de rejet de l'hypothèse nulle H_0 se nomme **région critique** du test et la région \overline{W} **région d'acceptation**.

La construction d'un test va donc consister à déterminer cette région critique. La méthode pour l'obtenir dépendra des conséquences que l'on attribue à chacune des deux erreurs qui sont associées aux deux décisions possibles. Ces erreurs sont les suivantes.

DÉFINITION

L'**erreur de première espèce** consiste à décider D_1 alors que H_0 est vraie, soit rejeter à tort l'hypothèse nulle H_0 . L'**erreur de seconde espèce** consiste à décider D_0 alors que H_1 est vraie, soit accepter à tort l'hypothèse nulle H_0 .

Nous allons présenter deux méthodes de construction d'un test, basées sur des principes très différents. La méthode de Bayes est utilisée lorsqu'on dispose encore plus d'informations *a priori* sur les hypothèses, permettant de leur attribuer une probabilité *a priori*, et lorsque l'on peut en plus quantifier le coût de chaque décision en fonction de l'hypothèse effectivement réalisée.

2 Méthode de Bayes

On se place dans le cas où on a attribué des probabilités *a priori* p_0 et $p_1 = 1 - p_0$ à chacune des hypothèses respectives H_0 et H_1 et que l'on a également associé un coût à chaque décision, en fonction de l'hypothèse qui est effectivement réalisée. Le tableau ci-après contient ces coûts, la décision prise figurant en colonne et l'hypothèse vraie en ligne :

	D_0	D_1
$H_0(p_0)$	C_{00}	C_{01}
$H_1(p_1)$	C_{10}	C_{11}

Une bonne décision peut avoir également un coût et donc on aura généralement $C_{00} > 0$ et $C_{11} > 0$.

Après la réalisation (x_1, \dots, x_n) on peut calculer, à l'aide du théorème de Bayes, les probabilités *a posteriori* π_0 et π_1 des hypothèses H_0 et H_1 :

$$\pi_0 = \frac{p_0 L_0}{p_0 L_0 + p_1 L_1} \quad \text{et} \quad \pi_1 = \frac{p_1 L_1}{p_0 L_0 + p_1 L_1}$$

où on a noté L_0 la valeur de la vraisemblance $L(x_1, \dots, x_n; \theta)$, quand $\theta \in \Theta_0$, et L_1 , quand $\theta \in \Theta_1$. On peut alors calculer les espérances du coût de chaque décision pour cette distribution *a posteriori* :

$$E[C(D_0)] = C_{00}\pi_0 + C_{10}\pi_1 \quad \text{et} \quad E[C(D_1)] = C_{01}\pi_0 + C_{11}\pi_1$$

La **règle de décision de Bayes** consiste à associer à l'observation (x_1, \dots, x_n) la décision dont l'espérance de coût est la plus faible.

Exemple 8.1

Nous allons reprendre l'exemple introductif en supposant cette fois que les informations disponibles permettent d'attribuer la probabilité $p_0 = 0,6$ à l'hypothèse d'entrée dans une période de récession, qui se traduit par $m = 0,3$. On considère de plus qu'une bonne décision est sans coût, soit $C_{00} = C_{11} = 0$, et que le coût de relance à tort est trois fois plus élevé que celui de ne rien faire en période de récession, soit $C_{10} = 3C_{01}$. Dans ces conditions, les espérances du coût de chaque décision sont :

$$E[C(D_0)] = C_{10}\pi_1 = 3C_{01}\pi_1 \quad \text{et} \quad E[C(D_1)] = C_{01}\pi_0$$

On a donc :

$$\begin{aligned} E[C(D_0)] < E[C(D_1)] &\iff 3\pi_1 < \pi_0 \\ &\iff 3p_1 L_1 < p_0 L_0 \iff 3 \frac{p_1}{p_0} < \frac{L_0}{L_1} \iff 2 < \frac{L_0}{L_1} \end{aligned}$$

L'expression de la vraisemblance est ici :

$$L(x_1, x_2, x_3; m) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^3 \exp - \frac{1}{2\sigma^2} \sum_{i=1}^3 (x_i - m)^2$$

Le rapport des vraisemblances est donc :

$$\frac{L_0}{L_1} = \exp - \frac{1}{2\sigma^2} \sum_{i=1}^3 [(x_i - m_0)^2 - (x_i - m_1)^2]$$

La décision D_0 a donc le coût le plus faible si :

$$\begin{aligned} \ln 2 &< -\frac{1}{2\sigma^2} \sum_{i=1}^3 [(x_i - m_0)^2 - (x_i - m_1)^2] \\ \ln 2 &< \frac{m_0 - m_1}{2\sigma^2} \left[2 \sum_{i=1}^3 x_i - 3(m_0 + m_1) \right] \end{aligned}$$

La règle de Bayes consiste donc à choisir la décision D_0 dans le cas où :

$$\ln 2 < 6 - 5 \sum_{i=1}^3 x_i \iff \frac{1}{3} \sum_{i=1}^3 x_i < \frac{2}{5} - \frac{\ln 2}{15} = 0,353$$

On retrouve la même règle de décision avec le même seuil critique arrondi qui avait été fixé à la valeur $k = 0,35$.

3 Méthode de Neyman et Pearson

3.1 Principe de la règle de Neyman et Pearson

On privilégie l'une des deux hypothèses, par exemple celle que l'on considère comme la plus vraisemblable, et on la choisit comme **hypothèse nulle** H_0 . Cette hypothèse sera celle dont le rejet à tort est le plus préjudiciable. L'autre hypothèse H_1 est l'**hypothèse alternative**. Il n'y a donc pas de symétrie entre ces deux hypothèses. L'hypothèse H_0 est privilégiée et il faut des observations très éloignées de cette hypothèse pour la rejeter.

DÉFINITION

On appelle **risque de première espèce** la probabilité de rejeter à tort l'hypothèse nulle, soit :

$$\alpha = P_\theta (D_1 | H_0) = P_\theta (H_1 | H_0) = P_\theta (W | \theta \in \Theta_0)$$

On appelle **risque de seconde espèce** la probabilité d'accepter à tort l'hypothèse nulle, soit :

$$\beta = P_\theta (D_0 | H_1) = P_\theta (H_0 | H_1) = P_\theta (\bar{W} | \theta \in \Theta_1)$$

L'erreur la plus grave consistant à rejeter à tort l'hypothèse nulle, la méthode de Neyman et Pearson fixe une valeur maximum α_0 au risque de première espèce. Le test est alors déterminé par la recherche de la règle qui minimise l'autre risque, celui de seconde espèce.

DÉFINITION

On appelle *puissance* d'un test la probabilité de refuser H_0 avec raison, c'est-à-dire lorsque H_1 est vérifiée, soit :

$$\eta = P_\theta(D_1|H_1) = P_\theta(H_1|H_1) = P_\theta(W|\theta \in \Theta_1) = 1 - \beta$$

La **règle de décision de Neyman et Pearson** consiste à déterminer la région critique W pour laquelle la puissance est maximum, sous la contrainte $\alpha \leq \alpha_0$. Le choix de la valeur de α_0 peut être déterminant quant à la conclusion tirée au vu d'un échantillon. La valeur standard retenue est $\alpha_0 = 0,05$. Choisir une valeur plus faible (par exemple $\alpha_0 = 0,01$) conduit à exiger des contre-preuves très fortes pour rejeter H_0 , qui est ainsi admise *a priori*. Choisir une valeur plus forte (par exemple $\alpha_0 = 0,10$) signifie que l'on est moins convaincu *a priori* de la validité de H_0 et que l'on est prêt plus facilement à la rejeter au vu des observations.

3.2 Hypothèses simples

Une hypothèse est qualifiée de **simple** si la loi de la v.a. X est totalement spécifiée quand cette hypothèse est réalisée. Dans le cas contraire elle est dite **multiple**. Nous allons examiner le cas où le paramètre θ ne peut prendre que deux valeurs θ_0 et θ_1 , ce qui correspond au choix entre les deux hypothèses simples suivantes :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

Même si cette situation est peu fréquente dans la réalité, de nombreux autres cas peuvent être résolus à partir de ce cas élémentaire. La forme de la région critique est alors déterminée par le théorème suivant.

Théorème 8.1

Théorème de Neyman et Pearson

Pour un risque de première espèce fixé à α_0 , le test de puissance maximum entre les hypothèses simples ci-dessus est défini par la région critique :

$$W = \left\{ (x_1, \dots, x_n) / \frac{L_0(x_1, \dots, x_n)}{L_1(x_1, \dots, x_n)} \leq k \right\}$$

où la valeur de la constante k est déterminée par le risque fixé $\alpha_0 = P_\theta(W|\theta = \theta_0)$, ayant posé $L_0(x_1, \dots, x_n) = L(x_1, \dots, x_n; \theta_0)$ et $L_1(x_1, \dots, x_n) = L(x_1, \dots, x_n; \theta_1)$.

Exemple 8.2

Nous allons appliquer ce théorème au cas de la loi exponentielle de paramètre $\frac{1}{\theta}$, avec $\theta_1 > \theta_0$. La vraisemblance a pour expression :

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \exp \left[-\frac{1}{\theta} \sum_{i=1}^n x_i \right]$$

avec $x_i > 0, 1 \leq i \leq n$. Le rapport des vraisemblances est donc :

$$\frac{L_0(x_1, \dots, x_n)}{L_1(x_1, \dots, x_n)} = \left(\frac{\theta_1}{\theta_0} \right)^n \exp \left[\left(\frac{1}{\theta_1} - \frac{1}{\theta_0} \right) \sum_{i=1}^n x_i \right]$$

La région critique est donc définie par la condition :

$$\left(\frac{\theta_1}{\theta_0} \right)^n \exp \left[\left(\frac{1}{\theta_1} - \frac{1}{\theta_0} \right) \sum_{i=1}^n x_i \right] \leq k$$

Cette condition est équivalente à :

$$\exp \left[\left(\frac{1}{\theta_1} - \frac{1}{\theta_0} \right) \sum_{i=1}^n x_i \right] \leq k_1$$

En prenant le logarithme, on obtient comme nouvelle condition équivalente :

$$\left(\frac{1}{\theta_1} - \frac{1}{\theta_0} \right) \sum_{i=1}^n x_i \leq k_2$$

Puisque $\theta_1 > \theta_0$, on arrive à la condition :

$$\sum_{i=1}^n x_i \geq C$$

La valeur de la constante C , qui va totalement préciser la région critique, est déterminée par la condition :

$$\alpha_0 = P \left\{ \sum_{i=1}^n X_i \geq C \mid \theta = \theta_0 \right\}$$

Dans l'exercice 7.7 nous avons établi que $2S_n/\theta$ suivait une loi du khi-deux à $2n$ degrés de liberté, avec $S_n = \sum_{i=1}^n X_i$. La condition précédente se réécrit donc sous la forme :

$$\alpha_0 = P \left\{ 2 \frac{S_n}{\theta_0} \geq 2 \frac{C}{\theta_0} \right\}$$

La valeur de $2 \frac{C}{\theta_0}$ est donc celle du fractile d'ordre $1 - \alpha_0$ de la loi du khi-deux à $2n$ degrés de liberté. La puissance de ce test peut ensuite se calculer par :

$$\eta = P \left\{ \sum_{i=1}^n X_i \geq C \mid \theta = \theta_1 \right\} = P \left\{ 2 \frac{S_n}{\theta_1} \geq 2 \frac{C}{\theta_1} \right\}$$

3.3 Hypothèses multiples

Nous allons d'abord considérer le cas d'une hypothèse simple contre une hypothèse multiple de l'une des formes suivantes :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

On détermine au préalable, par la méthode de Neyman et Pearson, la région critique W du test suivant :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

où θ_1 est une valeur fixée quelconque, mais vérifiant l'hypothèse alternative H_1 . Si la région W obtenue pour ce test entre hypothèses simples ne dépend pas de la valeur choisie θ_1 , alors on aura obtenu un **test uniformément le plus puissant (UPP)** pour le problème de test initial. Cela signifie que pour toute autre région critique W' on aura $P_\theta(W|\theta \in \Theta_1) \geq P_\theta(W'|\theta \in \Theta_1)$ pour tout θ de Θ_1 .

Exemple 8.3

Si nous reprenons l'exemple 8.2, la région critique dépendait de la condition $\theta_1 > \theta_0$, mais pas de la valeur précise θ_1 . La région critique obtenue est donc aussi celle du test UPP de $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$. Cependant, on ne peut pas cette fois calculer la puissance de ce test puisque la valeur du paramètre n'est pas connue dans l'hypothèse alternative. On peut seulement définir une fonction puissance de ce paramètre par :

$$\eta(\theta) = P \left\{ \sum_{i=1}^n X_i \geq C \mid \theta > \theta_0 \right\}$$

Pour le problème de test suivant :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

il n'existe pas de test UPP. La région critique W de ce test s'obtient par réunion des régions critiques W_1 et W_2 des deux tests précédents, pour le même risque de première espèce $\frac{\alpha_0}{2}$.

Exemple 8.4

Dans l'exemple précédent, la région critique W_1 obtenue était l'ensemble des points (x_1, \dots, x_n) tels que :

$$S_n \geq C_1$$

avec $2\frac{C_1}{\theta_0}$ qui est le fractile d'ordre $1 - \frac{\alpha_0}{2}$ de la loi du khi-deux à $2n$ degrés de liberté. De même, pour le test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$ la région critique W_2 obtenue est définie par :

$$S_n \leq C_2$$

avec $2\frac{C_2}{\theta_0}$ qui est le fractile d'ordre $\frac{\alpha_0}{2}$ de la loi du khi-deux à $2n$ degrés de liberté. La région critique de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ est $W = W_1 \cup W_2$. Il est plus facile de définir ici la région d'acceptation de l'hypothèse nulle par la condition :

$$C_2 \leq S_n \leq C_1$$

Pour le problème de test suivant :

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

on suppose que la loi P_θ est à rapport des vraisemblances monotone. Cela signifie qu'il existe une statistique $T_n = T_n(x_1, \dots, x_n)$ telle que le rapport des vraisemblances :

$$\frac{L(x_1, \dots, x_n; \theta)}{L(x_1, \dots, x_n; \theta')}$$

s'exprime comme une fonction croissante de T_n pour toutes les valeurs de θ et θ' qui vérifient l'inégalité $\theta > \theta'$. Dans ces conditions, on utilise le théorème suivant.

Théorème 8.2

Théorème de Lehmann

Il existe un un test UPP dont la région critique W est l'ensemble des points (x_1, \dots, x_n) tels que :

$$T_n(x_1, \dots, x_n) > k$$

où la valeur de la constante k est déterminée par le risque fixé $\alpha_0 = P_\theta(W | \theta = \theta_0)$.

Exemple 8.5

Dans l'exemple 8.2, nous avons obtenu comme rapport des vraisemblances :

$$\frac{L(x_1, \dots, x_n; \theta)}{L(x_1, \dots, x_n; \theta')} = \left(\frac{\theta'}{\theta}\right)^n \exp\left[\frac{\theta - \theta'}{\theta\theta'} \sum_{i=1}^n x_i\right]$$

Pour toutes les valeurs de θ et θ' qui vérifient l'inégalité $\theta > \theta'$, c'est une fonction croissante de $T_n = \sum_{i=1}^n x_i$. Donc, par application du théorème de Lehmann pour $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$, le test UPP a pour région critique l'ensemble des

points (x_1, \dots, x_n) tels que :

$$\sum_{i=1}^n x_i > k$$

La constante k est déterminée par la condition :

$$\alpha_0 = P \left\{ \sum_{i=1}^n X_i > k \mid \theta = \theta_0 \right\}$$

4 Test d'indépendance du khi-deux

Pour tester l'indépendance de deux caractères X et Y , qualitatifs ou quantitatifs (répartis alors en classes), à respectivement r et s modalités, on relève le nombre n_{ij} d'individus d'une population de taille $n = \sum_{i=1}^r \sum_{j=1}^s n_{ij}$ qui possèdent simultanément la modalité i , $1 \leq i \leq r$, du caractère X et la modalité j , $1 \leq j \leq s$, du caractère Y . Soit p_{ij} la probabilité théorique correspondante, pour un individu tiré au hasard dans la population, de posséder simultanément ces deux modalités i et j . Les probabilités marginales sont $p_{i.} = \sum_{j=1}^s p_{ij}$ et $p_{.j} = \sum_{i=1}^r p_{ij}$. L'indépendance de ces deux caractères se traduit par l'hypothèse nulle $H_0 : p_{ij} = p_{i.} \times p_{.j}$. Pour tester cette hypothèse contre l'alternative $H_1 : p_{ij} \neq p_{i.} \times p_{.j}$, on utilise la statistique :

$$D_n = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_i n_j / n)^2}{n_i n_j / n} = n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i n_j} - 1 \right)$$

Sa loi asymptotique, sous H_0 , est la loi du khi-deux à $(r-1)(s-1)$ degrés de liberté. On a noté $n_{i.} = \sum_{j=1}^s n_{ij}$ et $n_{.j} = \sum_{i=1}^r n_{ij}$ les effectifs marginaux.

La région critique de ce test est de la forme :

$$D_n \geq C$$

Pour un risque de première espèce $\alpha = P(D_n \geq C \mid H_0)$, la valeur de C est approximée par le fractile d'ordre $1-\alpha$ de la loi $\chi_{(r-1)(s-1)}^2$.

Exemple 8.6

Pour comparer l'efficacité de deux médicaments semblables, mais de prix très différents, la Sécurité sociale a effectué une enquête sur les guérisons obtenues avec ces deux traitements. Les résultats sont présentés dans le tableau suivant :

	Médicament cher	Médicament bon marché	
Guérisons	156	44	200
Non-guérisons	44	6	50
	200	50	250

On calcule la valeur de la statistique :

$$D_n = 250 \left(\frac{156^2}{4 \cdot 10^4} + 2 \frac{44^2}{10^4} + \frac{6^2}{25 \cdot 10^2} - 1 \right) = 2,5$$

Pour un risque de première espèce $\alpha = 0,05$ le fractile d'ordre $1 - \alpha$ de la loi χ_1^2 a pour valeur $C = 3,84$. Puisque la valeur observée de la statistique D_n est inférieure, on accepte l'hypothèse nulle d'indépendance du taux de guérison et du coût du médicament.

5 Compléments

Construire un test consiste à établir une **règle de décision**, c'est-à-dire une application ϕ de \mathbb{R}^n dans $\{0,1\}$. Pour toute réalisation (x_1, \dots, x_n) de l'échantillon telle que $\phi(x_1, \dots, x_n) = 1$ on rejette l'hypothèse H_0 . Pour toute réalisation (x_1, \dots, x_n) de l'échantillon telle que $\phi(x_1, \dots, x_n) = 0$ on accepte l'hypothèse H_0 . La fonction de test ϕ est donc la fonction indicatrice de la région critique W : $\phi = \mathbf{1}_W$. Pour un test entre deux hypothèses simples $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$, le risque de première espèce est $\alpha = E_{\theta_0} [\phi(X_1, \dots, X_n)]$ et la puissance $\eta = E_{\theta_1} [\phi(X_1, \dots, X_n)]$.

Quand ϕ est à valeurs dans $\{0,1\}$ le test est dit **pur** ou **déterministe**. Dans le cas d'une loi discrète il n'est pas toujours possible de trouver une région critique W de probabilité exacte α . Dans ce cas, si on souhaite que le risque de première espèce soit exactement égal à α , il faut utiliser un **test mixte**. La fonction ϕ est cette fois à valeurs dans $[0,1]$ et on accepte toujours H_1 pour $\phi = 1$ et H_0 pour $\phi = 0$. Pour $\phi = p \in]0,1[$ on effectue un tirage au sort où on accepte l'hypothèse H_1 avec une probabilité p .

En présence d'hypothèses multiples $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$, le risque de première espèce est une fonction de θ définie par $\alpha = E_{\theta}(\phi)$ pour $\theta \in \Theta_0$. On définit alors la **taille** τ du test par :

$$\tau = \sup_{\theta \in \Theta_0} E_{\theta}(\phi)$$

Le test est dit de **niveau** α si $\tau \leq \alpha$. La puissance du test est la fonction de θ définie par $\eta = E_{\theta}(\phi)$ pour $\theta \in \Theta_1$. Le graphe de la fonction définie pour tout θ de Θ par $\phi(\theta) = 1 - E_{\theta}(\phi)$ s'appelle **courbe d'efficacité**.

Un test ϕ de niveau α est dit **uniformément plus puissant** qu'un autre test ϕ' de même niveau α si, pour tout $\theta \in \Theta_1$ on a $E_{\theta}(\phi) \geq E_{\theta}(\phi')$.

Un test de niveau α est dit **sans biais** si, pour tout $\theta \in \Theta_1$, on a $E_{\theta}(\phi) \geq \alpha$.

La famille exponentielle à un paramètre dont la densité s'écrit sous la forme :

$$f(x; \theta) = a(\theta)b(x) \exp[\alpha(\theta)T(x)]$$

est à rapport des vraisemblances monotone si la fonction α est une fonction monotone du paramètre θ . Si α est croissante, on retient la statistique $T_n = \sum_{i=1}^n T(x_i)$, et si α est décroissante, on retient la statistique $T_n = -\sum_{i=1}^n T(x_i)$.

En dehors du cas particulier où on se donne une loi de probabilité *a priori* sur les valeurs du paramètre et où on utilise la méthode de Bayes, c'est la méthode de Neyman et Pearson qui est utilisée pour effectuer un test d'hypothèses. Le théorème de Neyman et Pearson fournit l'outil essentiel pour bâtir un test, c'est-à-dire définir la région où on va rejeter l'hypothèse nulle retenue. Il est bien entendu essentiel de connaître les définitions précises des différents concepts intervenant dans les tests statistiques. Notons que les deux hypothèses à confronter ne sont pas équivalentes et que l'hypothèse nulle est privilégiée, étant choisie comme celle dont le rejet à tort est le plus préjudiciable. Notons enfin l'importance du choix de la valeur du risque de première espèce, dont peut dépendre la conclusion tirée d'un échantillon donné.



Entraînez-vous

- 8.1** Une machine produit des pièces dont une proportion θ présente des défauts qui conduisent à classer ces pièces en second choix. En cas de bon fonctionnement la valeur est $\theta = 0,1$. Si la machine se dérègle, la valeur passe à $\theta = 0,2$. On estime que la probabilité que la machine soit bien réglée est $p_0 = 0,6$. On peut souscrire un contrat d'entretien qui permet de maintenir la machine bien réglée. Le coût est de 0,4 euros par pièce produite. Chaque pièce de second choix est vendue 6 euros de moins que les autres. Après contrôle des 100 premières pièces produites, on constate que 13 d'entre elles doivent être classées en second choix. Quelle décision est-on conduit à prendre si on utilise la méthode de Bayes ?
- 8.2** On dispose d'un échantillon (X_1, \dots, X_n) d'une v.a. X qui suit une loi normale d'espérance inconnue m et d'écart type connu $\sigma = 1$ pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : m = 1 \\ H_1 : m = 1,5 \end{cases}$$

Déterminer la région critique du test de Neyman et Pearson et calculer sa puissance dans le cas où $n = 25$ et $\alpha = 0,05$. Quelle devrait être la taille d'échantillon minimum pour que cette puissance soit supérieure à 0,90 ?

- 8.3** On dispose d'un échantillon (X_1, \dots, X_n) d'une v.a. X qui suit une loi normale centrée, d'écart type inconnu σ pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : \sigma = 1 \\ H_1 : \sigma = 2 \end{cases}$$

Déterminer la région critique du test de Neyman et Pearson et calculer sa puissance dans le cas où $n = 15$ et $\alpha = 0,05$.

- 8.4** Le revenu annuel des individus d'une population est distribué selon une loi de Pareto de densité :

$$f(x; \theta) = \begin{cases} \frac{\theta}{x^{\theta+1}} & \text{si } x \geq 1 \\ 0 & \text{sinon} \end{cases}$$

On dispose d'un échantillon (X_1, \dots, X_n) de cette loi pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : \theta = \frac{4}{3} \\ H_1 : \theta = \frac{8}{5} \end{cases}$$

Déterminer la région critique du test de Neyman et Pearson et calculer sa puissance dans le cas où $n = 400$ et $\alpha = 0,05$.

- 8.5** Soit X une variable aléatoire dont la densité a pour expression, pour $x > \lambda$:

$$f(x) = \frac{1}{\theta} \exp\left(-\frac{x-\lambda}{\theta}\right)$$

et nulle sinon, où θ et λ sont deux paramètres strictement positifs.

1. Déterminer l'estimateur $\hat{\lambda}_n$ du paramètre λ par la méthode du maximum de vraisemblance construit à partir d'un échantillon (X_1, \dots, X_n) de X et étudier ses propriétés.
2. Déterminer l'estimateur $\hat{\theta}_n$ du paramètre θ par la méthode du maximum de vraisemblance et étudier ses propriétés. Déterminer sa loi asymptotique.
3. Déterminer la région critique du test :

$$\begin{cases} H_0 : \lambda = \lambda_0 \\ H_1 : \lambda = \lambda_1 \end{cases}$$

pour un risque de première espèce donné α , sachant que $\lambda_1 > \lambda_0$. Déterminer le risque de seconde espèce β et la puissance η .

4. Déterminer la région critique du test :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

pour un risque de première espèce donné α , sachant que $\theta_1 > \theta_0$. Pour déterminer approximativement le seuil critique de ce test on utilisera l'estimateur $\hat{\lambda}_n$ et la loi asymptotique de $\hat{\theta}_n$. Calculer de la même façon une valeur approchée de la puissance de ce test.

5. Déterminer la région critique du test :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

Ce test est-il UPP ? Peut-on calculer sa puissance ?

6. Déterminer la région critique du test :

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Appliquer le théorème de Lehmann pour établir que ce test est UPP de niveau α . Montrer que le risque de première espèce α est une fonction croissante de θ qui est maximum pour $\theta = \theta_0$.

- 8.6** On dispose d'un échantillon de taille $n = 15$ d'une v.a. de loi normale centrée et de variance $\frac{1}{\theta}$ pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : \theta = 1 \\ H_1 : \theta > 1 \end{cases}$$

Déterminer la région critique d'un test UPP de risque de première espèce α et préciser sa fonction puissance. Calculer cette puissance dans le cas où $n = 15$, pour $\theta = 3$ et $\alpha = 0,05$.

- 8.7** Le poids indiqué par une balance, lorsqu'on effectue la pesée d'un poids étalonné à 100 g, est une v.a. de loi normale d'espérance 100. Si la balance est bien réglée, l'écart type a pour valeur $\sigma = 5$ et sinon cette valeur est inconnue, avec $\sigma > 5$. Déterminer la région critique d'un test UPP de risque de première espèce $\alpha = 0,05$ basé sur un échantillon de $n = 10$ pesées et préciser sa fonction puissance.

8.8 Le nombre de pannes mensuelles d'un ascenseur est une v.a. qui suit une loi de Poisson de paramètre $\lambda = 2$. Après avoir souscrit un contrat d'entretien, on pense que la valeur du paramètre doit diminuer. Préciser la règle de décision à l'issue de six mois de contrat.

8.9 On dispose d'un échantillon (X_1, \dots, X_n) d'une v.a. X qui suit une loi normale d'espérance inconnue m et d'écart type connu $\sigma = 1$ pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : m \leq 3 \\ H_1 : m > 3 \end{cases}$$

Déterminer la région critique du test le plus puissant de niveau $\alpha = 0,05$ dans le cas où $n = 100$ et préciser sa fonction puissance.

8.10 On dispose d'un échantillon de taille $n = 12$ d'une v.a. X qui suit une loi normale d'espérance inconnue m et d'écart type inconnu σ pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : m \leq 6 \\ H_1 : m > 6 \end{cases}$$

Déterminer la région critique d'un test de niveau $\alpha = 0,05$. Peut-on déterminer sa fonction puissance ?

8.11 La durée de vie d'un certain matériel est une v.a. X qui suit une loi de Weibull de densité :

$$f(x; \theta) = \begin{cases} \frac{1}{2\theta\sqrt{x}} \exp\left(-\frac{\sqrt{x}}{\theta}\right) & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

On dispose d'un échantillon (X_1, \dots, X_n) de cette loi pour choisir entre les deux hypothèses :

$$\begin{cases} H_0 : \theta \leq 1 \\ H_1 : \theta > 1 \end{cases}$$

Déterminer la région critique du test le plus puissant de niveau $\alpha = 0,01$ dans le cas où $n = 15$ et préciser sa fonction puissance.

8.12 Pour sa fabrication, un industriel utilise des machines de deux constructeurs différents. Après six mois d'utilisation, il constate que sur les 80 machines de type A, 50 ne sont jamais tombées en panne, alors que pour le type B la proportion est de 40 sur 60. Peut-on considérer que ces deux types de machines sont équivalents ?

8.13 Deux séries d'observations effectuées à des dates différentes, sur des échantillons de tailles respectives $n_1 = 41$ et $n_2 = 61$, ont conduit à des valeurs respectives de moyennes empiriques et d'écart types empiriques $\bar{x} = 785, s_x = 1,68$, $\bar{y} = 788$ et $s_y = 1,40$. Peut-on considérer que ces deux échantillons proviennent de la même loi normale ?

8.14 Le tableau ci-après donne la répartition par taille, en cm, de 2 700 salariés français, de sexe masculin, par catégories socio-professionnelles (CSP) :

	Ouvriers	Employés	Cadres	Total
Moins de 165 cm	325	66	22	413
De 165 à moins de 170 cm	488	110	51	649
De 170 à moins de 175 cm	636	158	123	917
175 cm et plus	451	146	124	721
Total	1 900	480	320	2 700

Au vu de ce tableau, peut-on considérer que la taille est indépendante de la CSP ?

● Solutions



8.1 Nous allons formaliser le problème en construisant le modèle statistique. Nous allons introduire une variable indicatrice qui prend la valeur 1 si la pièce produite est de second choix. C'est une variable de Bernoulli définie par :

$$X = \begin{cases} 1 & \text{si second choix } (\theta) \\ 0 & \text{sinon } (1 - \theta) \end{cases}$$

Nous sommes en présence de deux hypothèses, $H_0 : \theta = 0,1$ avec une probabilité *a priori* $p_0 = 0,6$ et $H_1 : \theta = 0,2$ avec une probabilité $p_1 = 1 - p_0 = 0,4$. Nous devons choisir entre la décision D_0 de ne pas souscrire le contrat d'entretien et la décision D_1 de le souscrire. Le coût moyen par pièce produite associé à la décision D_0 est de 6θ , qui correspond au manque à gagner, car θ représente la valeur moyenne de X . Si on prend la décision D_1 de souscrire le contrat d'entretien, le coût est de 0,4 par pièce, plus le manque à gagner de $6 \times 0,1$ puisque la proportion de second choix est alors toujours de 10 %. Tout ceci est résumé dans le tableau de coûts suivant :

	D_0	D_1	Probabilité <i>a priori</i>
$H_0 : \theta = 0,1$	0,6	1	0,6
$H_1 : \theta = 0,2$	1,2	1	0,4

La vraisemblance de l'échantillon (X_1, \dots, X_n) de cette loi de Bernoulli s'écrit :

$$L(x_1, \dots, x_n; \theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$$

avec $x_i \in \{0, 1\}$, $1 \leq i \leq n$. On note L_0 la vraisemblance pour $\theta = 0,1$ et L_1 pour $\theta = 0,2$. Les espérances *a posteriori* des coûts sont :

$$E[C(D_0)] = 0,6\pi_0 + 1,2\pi_1 \text{ et } E[C(D_1)] = \pi_0 + \pi_1 = 1$$

La règle de Bayes conduit à souscrire le contrat si $E[C(D_0)] > E[C(D_1)]$ soit :

$$\begin{aligned} 0,6\pi_0 + 1,2\pi_1 > 1 &\iff 0,36L_0 + 0,48L_1 > 0,6L_0 + 0,4L_1 \\ &\iff 8L_1 > 24L_0 \\ &\iff \left(\ln 2 - \ln \frac{8}{9}\right) \sum_{i=1}^n x_i > \ln 3 - n \ln \frac{8}{9} \\ &\iff \sum_{i=1}^n x_i > \frac{\ln 3 - n \ln \frac{8}{9}}{\ln \frac{9}{4}} \end{aligned}$$

Pour un échantillon de taille $n = 100$, on décide de souscrire le contrat d'entretien si $\sum_{i=1}^{100} x_i > 15,9$ soit, puisque les x_i ne prennent que des valeurs entières, $\sum_{i=1}^{100} x_i \geq 16$. Ayant observé $\sum_{i=1}^{100} x_i = 13$, la règle de Bayes conduit à ne pas souscrire le contrat d'entretien.

8.2 La vraisemblance s'écrit :

$$L(x_1, \dots, x_n; m) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - m)^2\right\}$$

La forme de la région critique, donnée par le théorème de Neyman et Pearson, est $L_0/L_1 \leq k$, ce qui en passant aux logarithmes conduit à l'inégalité :

$$-\frac{1}{2} \sum_{i=1}^n \left[(x_i - 1)^2 - (x_i - 1,5)^2 \right] \leq \ln k$$

Elle est équivalente, après développement du crochet, à :

$$-\sum_{i=1}^n x_i + \frac{5n}{4} \leq 2 \ln k$$

Pour l'instant, on cherche seulement à déterminer la forme de la région critique et l'expression des constantes qui apparaissent dans cette inégalité est sans intérêt. On remplace donc cette inégalité par $\sum_{i=1}^n x_i \geq k_1$, qui lui est équivalente. La région critique est donc définie à partir de la valeur de la somme des observations. Cependant, comme c'est la loi de la moyenne empirique \bar{X}_n qui est connue, il est préférable de faire intervenir cette moyenne dans la définition de la région critique W , qui sera donc l'ensemble des points (x_1, \dots, x_n) tels que :

$$\bar{x}_n \geq C$$

Il reste à préciser la valeur de cette constante C , qui sera le seuil de cette région, à partir du risque de première espèce α qui est fixé. Il représente la probabilité de cette région, lorsque l'hypothèse nulle est réalisée, et par conséquent, la condition qui définit C s'écrit :

$$\alpha = P(W | m = 1) = P(\bar{X}_n \geq C | m = 1)$$

Sous H_0 , la moyenne empirique \bar{X}_n suit la loi $N(1, 1/\sqrt{n})$, donc en centrant et réduisant on obtient la condition :

$$\alpha = P\left(U \geq \frac{C-1}{1/\sqrt{n}}\right)$$

où U est une v.a. de loi $N(0, 1)$. Ainsi la constante C est définie par :

$$C = 1 + \frac{u}{\sqrt{n}}$$

où u est le fractile d'ordre $1-\alpha$ de la loi $N(0, 1)$. Pour un risque $\alpha = 0,05$ on lit dans la table 2 le fractile $u = 1,6449$ d'où une région critique définie pour $n = 25$ par :

$$W = \{(x_1, \dots, x_{25}) / \bar{x}_{25} \geq 1,33\}$$

Dans ce cas, la puissance du test est la probabilité de cette région dans l'hypothèse alternative, soit :

$$\eta = P(\bar{X}_{25} \geq 1,33 | m = 1,5) = P\left(U \geq \frac{1,33 - 1,5}{1/\sqrt{25}}\right) = P(U \geq -0,85) = 0,8023$$

Pour un risque $\alpha = 0,05$ la condition $\eta \geq 0,90$ conduit à choisir une taille d'échantillon n telle que :

$$\eta = P\left(\bar{X}_n \geq 1 + \frac{u}{\sqrt{n}} \mid m = 1,5\right) = P\left(U \geq u - \frac{\sqrt{n}}{2}\right) \geq 0,90$$

ce qui est équivalent $1,6449 - \sqrt{n}/2 \leq -1,2816$ soit $\sqrt{n} \geq 5,853$ donc une taille d'échantillon supérieure à $n_0 = 35$.

8.3 Pour effectuer ce test, nous disposons d'un échantillon d'une loi normale centrée, d'écart type $\sigma = 1$ sous H_0 et $\sigma = 2$ sous H_1 . Le rapport des vraisemblances a donc pour expression :

$$\frac{L_0}{L_1} = 2^n \exp\left(\frac{1}{8} \sum_{i=1}^n x_i^2 - \frac{1}{2} \sum_{i=1}^n x_i^2\right) = 2^n \exp\left(-\frac{3}{8} \sum_{i=1}^n x_i^2\right)$$

La région critique est donc définie par le théorème de Neyman et Pearson comme l'ensemble des points (x_1, \dots, x_n) tels que :

$$\sum_{i=1}^n x_i^2 \geq C$$

La valeur de la constante C est déterminée par la condition :

$$\alpha = P\left(\sum_{i=1}^n X_i^2 \geq C \mid H_0\right)$$

Sous l'hypothèse nulle H_0 , $\sum_{i=1}^n X_i^2$ suit la loi χ_n^2 , donc C est le fractile d'ordre $1-\alpha$ de cette loi. Pour $\alpha = 0,05$ on lit dans la table 5 la valeur $C = 25$. La puissance du test est :

$$\eta = P\left(\sum_{i=1}^n X_i^2 \geq C \mid H_1\right)$$

Sous l'hypothèse H_1 , $\sum_{i=1}^n \frac{X_i^2}{4}$ suit la loi χ_n^2 , donc :

$$\eta = P\left(\frac{1}{4} \sum_{i=1}^n X_i^2 \geq \frac{C}{4}\right) = 0,975$$

par lecture de la table 5. Notons qu'ici le risque de seconde espèce $\beta = 0,025$ est inférieur au risque de première espèce $\alpha = 0,05$. Pour respecter la logique de Neyman et Pearson, il faudrait donc choisir un risque de première espèce plus faible.

8.4 L'expression de la vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \theta x_i^{-\theta-1} = \theta^n \left(\prod_{i=1}^n x_i\right)^{-\theta-1}$$

pour $x_i \geq 1, 1 \leq i \leq n$.

Le rapport des vraisemblances, L_0 sous H_0 , et L_1 sous H_1 , s'écrit donc :

$$\frac{L_0}{L_1} = \left(\frac{\theta_0}{\theta_1}\right)^n \left(\prod_{i=1}^n x_i\right)^{\theta_1 - \theta_0}$$

Le théorème de Neyman et Pearson, qui permet d'obtenir le test de puissance maximum, définit la région critique par la condition $\frac{L_0}{L_1} \leq k$, qui est donc ici équivalente, en prenant les logarithmes, à :

$$(\theta_1 - \theta_0) \sum_{i=1}^n \ln x_i \leq k_1$$

Puisque $\theta_1 > \theta_0$, cette région critique W est définie par la condition :

$$\sum_{i=1}^n \ln x_i \leq C$$

La valeur du seuil C est déterminée par :

$$\alpha = P(W | H_0) = P\left(\sum_{i=1}^n \ln X_i \leq C \mid \theta = \frac{4}{3}\right)$$

La taille d'échantillon permet ici d'utiliser le théorème central limite que nous allons appliquer à $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$, ayant posé $Y_i = \ln X_i, 1 \leq i \leq n$. Il nous faut pour cela déterminer la loi de la v.a. $Y = \ln X$. Sa fonction de répartition est définie par :

$$G(y) = P(Y < y) = P(\ln X < y) = P(X < e^y) = F(e^y)$$

pour $y > 0$. On obtient la densité par dérivation :

$$g(y) = e^y f(e^y) = \theta e^{-\theta y}$$

On reconnaît la loi exponentielle de paramètre θ , donc $E(Y) = \frac{1}{\theta}$ et $V(Y) = \frac{1}{\theta^2}$.

D'après le théorème central limite :

$$\sqrt{n} \frac{\bar{Y}_n - 1/\theta}{1/\theta} \xrightarrow{\text{loi}} N(0, 1)$$

La constante C est alors déterminée approximativement par :

$$\alpha = P\left(\bar{Y}_n \leq \frac{C}{n}\right) = P\left\{\sqrt{n}(\theta_0 \bar{Y}_n - 1) \leq \sqrt{n}\left(\theta_0 \frac{C}{n} - 1\right)\right\}$$

Le fractile d'ordre 0,05 de la loi $N(0, 1)$ a pour valeur $-1,645$ donc la valeur approchée du seuil C est définie par la condition :

$$\sqrt{n}\left(\frac{4C}{3n} - 1\right) = -1,645$$

Soit, pour $n = 400$, $C = 275$. La puissance est déterminée par :

$$\begin{aligned} \eta &= P(W | H_1) = P\left(\sum_{i=1}^n \ln X_i \leq C \mid \theta = \frac{8}{5}\right) = P\left(\bar{Y}_n \leq \frac{C}{n}\right) \\ &= P\left\{\sqrt{n}(\theta_1 \bar{Y}_n - 1) \leq \sqrt{n}\left(\theta_1 \frac{C}{n} - 1\right)\right\} \end{aligned}$$

Sa valeur approchée est définie par $\eta = \Phi(2, 02) = 0,98$, où Φ est la f.r. de la loi $N(0, 1)$. Notons que le risque de seconde espèce $\beta = 0,02$ est inférieur au risque de première espèce $\alpha = 0,05$.

8.5 1. La vraisemblance est :

$$L(x_1, \dots, x_n; \theta, \lambda) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum_{i=1}^n x_i\right) e^{n\lambda/\theta}$$

à condition que tous les x_i soient supérieurs à λ , ce qui est équivalent à la condition $m_n = \min\{x_1, \dots, x_n\} > \lambda$. La vraisemblance est une fonction croissante de λ avant m_n et nulle au-delà, donc atteint son maximum pour $\lambda = m_n$. L'estimateur du maximum de vraisemblance est donc :

$$\hat{\lambda}_n = m_n = \min\{X_1, \dots, X_n\}$$

On peut établir que la v.a. $Z = n \frac{m_n - \lambda}{\theta}$ suit une loi exponentielle, avec $E(Z) = V(Z) = 1$. On en déduit :

$$E(m_n) = \lambda + \frac{\theta}{n} \text{ et } V(m_n) = \frac{\theta^2}{n^2}$$

C'est donc un estimateur asymptotiquement sans biais et convergent.

2. La log-vraisemblance a pour expression :

$$\ln L(x_1, \dots, x_n; \theta, \lambda) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n (x_i - \lambda)$$

Cette fonction est dérivable pour tout $\theta > 0$ avec :

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n (x_i - \lambda)$$

La dérivée s'annule pour $\theta = \sum_{i=1}^n (x_i - \lambda) / n$; la dérivée seconde est négative pour cette valeur :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n (x_i - \lambda)$$

L'estimateur du maximum de vraisemblance est donc :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i - \hat{\lambda}_n = \bar{X}_n - m_n$$

Cet estimateur est asymptotiquement sans biais :

$$E(\hat{\theta}_n) = E(X) - E(m_n) = \theta - \frac{\theta}{n}$$

résultat déduit de la loi de m_n .

D'après la loi des grands nombres, \bar{X}_n converge vers $E(X)$. Nous pouvons établir que m_n converge vers λ , donc $\hat{\theta}_n$ converge vers $\theta + \lambda - \lambda = \theta$.

D'après ce qui précède :

$$E\left[\sqrt{n}(m_n - \lambda)\right] = \frac{\theta}{\sqrt{n}} \text{ et } V\left[\sqrt{n}(m_n - \lambda)\right] = \frac{\theta^2}{n}$$

Ces deux moments tendent vers 0 avec $1/n$ donc :

$$\sqrt{n} (m_n - \lambda) \xrightarrow{p} 0$$

Nous pouvons écrire :

$$\sqrt{n} (\hat{\theta}_n - \theta) = \sqrt{n} (\bar{X}_n - \theta - \lambda) - \sqrt{n} (m_n - \lambda)$$

Le théorème central-limite et le résultat précédent nous permettent de conclure :

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{loi} N(0, \theta)$$

3. La région critique de ce test est construite à partir de l'estimateur du maximum de vraisemblance de λ . Comme $\lambda_1 > \lambda_0$, elle est définie comme l'ensemble des points (x_1, \dots, x_n) tels que $m_n > C$. La constante C étant définie par :

$$\alpha = P(m_n > C | H_0) = \exp -n \frac{C - \lambda_0}{\theta}$$

soit $C = \lambda_0 - (\theta \ln \alpha)/n$. Le risque de seconde espèce est :

$$\beta = P(m_n < C | H_1) = 1 - \exp -n \frac{C - \lambda_1}{\theta} = 1 - \alpha \exp -n \frac{\lambda_0 - \lambda_1}{\theta}$$

La puissance s'en déduit par $\eta = 1 - \beta$.

4. On utilise le théorème de Neyman et Pearson et on forme le rapport des vraisemblances :

$$\frac{L_0}{L_1} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp \left\{ -\frac{1}{\theta_0} \sum_{i=1}^n (x_i - \lambda) + \frac{1}{\theta_1} \sum_{i=1}^n (x_i - \lambda) \right\}$$

En prenant le logarithme, on définit la région critique par l'inégalité :

$$\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right) \sum_{i=1}^n (x_i - \lambda) < k$$

Comme $\theta_1 > \theta_0$, ceci est équivalent à :

$$\sum_{i=1}^n (x_i - \lambda) > k'$$

La valeur de λ étant inconnue, on doit utiliser une approximation en remplaçant ce paramètre par son estimateur :

$$\hat{\theta}_n = \bar{x}_n - m_n > C$$

La constante C étant définie approximativement par :

$$\alpha = P(\hat{\theta}_n > C | H_0) = P\left(\sqrt{n} \frac{\hat{\theta}_n - \theta_0}{\theta_0} > \sqrt{n} \frac{C - \theta_0}{\theta_0}\right)$$

On retient donc comme valeur approchée du seuil :

$$C = \theta_0 + u \frac{\theta_0}{\sqrt{n}}$$

où u est le fractile d'ordre $1 - \alpha$ de la loi $N(0, 1)$.

La puissance se calcule alors par :

$$\eta = P(\hat{\theta}_n > C | H_1) = P\left(\sqrt{n}\frac{\hat{\theta}_n - \theta_1}{\theta_1} > \sqrt{n}\frac{C - \theta_1}{\theta_1}\right)$$

en utilisant à nouveau la loi asymptotique $N(0, 1)$.

5. La région critique précédente dépendait des valeurs respectives de θ dans les deux hypothèses. Le test de région d'acceptation définie par :

$$C_1 < \hat{\theta}_n < C_2$$

n'est donc pas UPP. Les valeurs approchées des seuils critiques sont :

$$C_1 = \theta_0 - u\frac{\theta_0}{\sqrt{n}} \quad \text{et} \quad C_2 = \theta_0 + u\frac{\theta_0}{\sqrt{n}}$$

où u est le fractile d'ordre $1 - \alpha/2$ de la loi $N(0, 1)$. On ne peut pas calculer la puissance de ce test, l'hypothèse alternative étant multiple.

6. Pour $\theta' > \theta$ le rapport des vraisemblances $L_{\theta'}/L_{\theta}$ est une fonction croissante de $\hat{\theta}_n$. D'après le théorème de Lehmann, la région critique du test de la question 4 est celle d'un test UPP pour les hypothèses confrontées ici. Son risque de première espèce est défini par :

$$\alpha(\theta) = P(\hat{\theta}_n > C | H_0) = P\left(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\theta} > \sqrt{n}\frac{C - \theta}{\theta}\right)$$

Cette fonction atteint son maximum, d'après le théorème de Lehmann, pour $\theta = \theta_0$. Si on utilise la loi asymptotique, le seuil $\sqrt{n}(C/\theta - 1)$ est une fonction décroissante de θ , donc α est bien une fonction croissante de θ .

8.6 L'hypothèse alternative étant une hypothèse multiple, on détermine d'abord la région critique du test entre les deux hypothèses simples :

$$\begin{cases} H_0 : \theta = 1 \\ H_1 : \theta = \theta_1 \end{cases}$$

avec θ_1 quelconque, mais vérifiant $\theta_1 > 1$. L'expression de la vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \sqrt{\frac{\theta}{2\pi}} \exp\left(-\frac{\theta}{2}x_i^2\right) = \left(\frac{\theta}{2\pi}\right)^{n/2} \exp\left(-\frac{\theta}{2} \sum_{i=1}^n x_i^2\right)$$

Le rapport des vraisemblances, L_0 sous H_0 , et L_1 sous H_1 , s'écrit donc :

$$\frac{L_0}{L_1} = \left(\frac{1}{\theta_1}\right)^{n/2} \exp -\frac{1}{2}(1 - \theta_1) \sum_{i=1}^n x_i^2$$

Le théorème de Neyman et Pearson, qui permet d'obtenir le test de puissance maximum, définit la région critique par la condition $\frac{L_0}{L_1} \leq k$, qui est donc ici équivalente, en prenant les logarithmes, à :

$$-\frac{1}{2}(1 - \theta_1) \sum_{i=1}^n x_i^2 \leq k_1$$

Puisque $\theta_1 > 1$, cette région critique W est définie par la condition :

$$\sum_{i=1}^n x_i^2 \leq C$$

La valeur du seuil C est déterminée par :

$$\alpha = P(W | H_0) = P\left(\sum_{i=1}^n X_i^2 \leq C | \theta = 1\right)$$

Sous H_0 , les v.a. X_i sont normales, centrées, réduites et indépendantes, donc $\sum_{i=1}^n X_i^2$ suit une loi du khi-deux à n degrés de liberté. La valeur du seuil C est donc celle du fractile d'ordre $1 - \alpha$ de la loi χ_n^2 . La région critique est indépendante de la valeur choisie pour θ_1 , donc elle est aussi celle du test UPP pour l'alternative $H_1 : \theta > 1$. La fonction puissance est définie pour $\theta > 1$ par :

$$\eta(\theta) = P(W | H_1) = P\left(\sum_{i=1}^n X_i^2 \leq C | \theta > 1\right)$$

Elle se calcule à partir de la fonction de répartition F_n de la loi χ_n^2 par :

$$\eta(\theta) = F_n(\theta C)$$

Pour $n = 15$ et $\alpha = 0,05$ on lit dans la table 5 la valeur $C = 7,26$. Pour $\theta = 3$ la puissance vaut alors $\eta(3) = F_{15}(21, 78) \simeq 0,9$.

8.7 L'hypothèse alternative étant une hypothèse multiple, on détermine d'abord la région critique du test entre les deux hypothèses simples :

$$\begin{cases} H_0 : \sigma = 5 \\ H_1 : \sigma = \sigma_1 \end{cases}$$

avec σ_1 quelconque, mais vérifiant $\sigma_1 > 5$. L'expression de la vraisemblance est :

$$\begin{aligned} L(x_1, \dots, x_n; \sigma) &= \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - 100)^2\right\} \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - 100)^2\right\} \end{aligned}$$

Le rapport des vraisemblances, L_0 sous H_0 , et L_1 sous H_1 , s'écrit donc :

$$\frac{L_0}{L_1} = \left(\frac{\sigma_1}{5}\right)^n \exp\left\{-\frac{1}{2}\left(\frac{1}{25} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n (x_i - 100)^2\right\}$$

Le théorème de Neyman et Pearson, qui permet d'obtenir le test de puissance maximum, définit la région critique par la condition $\frac{L_0}{L_1} \leq k$, qui est donc ici équivalente, en prenant les logarithmes, à :

$$-\frac{1}{2}\left(\frac{1}{25} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n (x_i - 100)^2 \leq k_1$$

Puisque $\sigma_1^2 > 25$, cette région critique W est définie par la condition :

$$\sum_{i=1}^n (x_i - 100)^2 \geq C$$

La valeur du seuil C est déterminée par :

$$\alpha = P(W|H_0) = P\left(\sum_{i=1}^n (X_i - 100)^2 \geq C | \sigma = 5\right)$$

Sous H_0 , les v.a. $X_i - 100$ sont normales, centrées, d'écart type $\sigma = 5$ et indépendantes, donc $\sum_{i=1}^n (X_i - 100)^2 / 25$ suit une loi du khi-deux à n degrés de liberté. La valeur de $C/25$ est donc celle du fractile d'ordre $1-\alpha$ de la loi χ_n^2 . La région critique est indépendante de la valeur choisie pour σ_1 , donc elle est aussi celle du test UPP pour l'alternative $H_1 : \sigma > 5$. La fonction puissance est définie pour $\sigma > 5$ par :

$$\eta(\sigma) = P(W|H_1) = P\left(\sum_{i=1}^n (X_i - 100)^2 \geq C | \sigma > 5\right)$$

Elle se calcule à partir de la fonction de répartition F_n de la loi χ_n^2 par :

$$\eta(\sigma) = 1 - F_n\left(\frac{C}{\sigma^2}\right)$$

Pour $n = 10$ et $\alpha = 0,05$ on lit dans la table 5 la valeur $C = 25 \times 18,3 = 457,5$.

8.8 Il s'agit de tester l'hypothèse $H_0 : \lambda = 2$ contre l'hypothèse $H_1 : \lambda < 2$ à partir d'un échantillon de taille $n = 6$. On effectue d'abord le test d'hypothèses simples $H_0 : \lambda = 2$ contre $H_1 : \lambda = \lambda_1$ avec λ_1 quelconque, mais vérifiant $\lambda_1 < 2$. Le rapport des vraisemblances, L_0 sous H_0 , et L_1 sous H_1 , s'écrit :

$$\frac{L_0}{L_1} = e^{-n(2-\lambda_1)} \left(\frac{2}{\lambda_1}\right)^{\sum_{i=1}^n x_i}$$

Le théorème de Neyman et Pearson, qui permet d'obtenir le test de puissance maximum, définit la région critique par la condition $\frac{L_0}{L_1} \leq k$, qui est donc ici équivalente, en prenant les logarithmes, à :

$$(\ln 2 - \ln \lambda_1) \sum_{i=1}^n x_i \leq k_1$$

Puisque $\lambda_1 < 2$, la région critique W est définie par la condition :

$$\sum_{i=1}^n x_i \leq C$$

La valeur du seuil C est déterminée par :

$$\alpha = P(W|H_0) = P\left(\sum_{i=1}^n X_i \leq C | \lambda = 2\right)$$

C'est donc le fractile d'ordre α de la loi de Poisson de paramètre $2n$. Pour $n = 6$ et $\alpha = 0,05$ on lit $C = 6$. La région critique est indépendante de la valeur choisie pour λ_1 , donc elle est aussi celle du test UPP pour l'alternative $H_1 : \lambda < 2$. Au bout de six mois, on considère que le contrat a été efficace si on observe un nombre de pannes inférieur ou égal à 6.

8.9 Les deux hypothèses sont multiples. On détermine d'abord la région critique du test entre les deux hypothèses simples :

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m = m_1 \end{cases}$$

avec m_0 et m_1 quelconques, mais vérifiant $m_0 \leq 3$ et $m_1 > 3$. La vraisemblance s'écrit :

$$L(x_1, \dots, x_n; m) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - m)^2 \right\}$$

La forme de la région critique, donnée par le théorème de Neyman et Pearson, est $L_0/L_1 \leq k$, ce qui en passant aux logarithmes conduit à l'inégalité :

$$-\frac{1}{2} \sum_{i=1}^n [(x_i - m_0)^2 - (x_i - m_1)^2] \leq \ln k$$

Elle est équivalente, après développement du crochet, à :

$$(m_1 - m_0)\bar{x}_n \geq k_1$$

ayant posé $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$. Cette région critique est indépendante des valeurs de m_0 et m_1 qui vérifient toujours $m_1 - m_0 > 0$. C'est donc celle du test UPP pour $H_0 : m \leq 3$ contre $H_1 : m > 3$ et se définit par :

$$\bar{x}_n \geq C$$

On peut aussi appliquer le théorème de Lehmann. Le rapport des vraisemblances s'écrit :

$$\frac{L(x_1, \dots, x_n; \theta)}{L(x_1, \dots, x_n; \theta')} = \exp \left[-\frac{n}{2}(\theta^2 - \theta'^2) + n(\theta - \theta')\bar{x}_n \right]$$

C'est une fonction croissante de \bar{x}_n pour $\theta > \theta'$ donc $\bar{x}_n > C$ définit la région critique d'un test UPP. La valeur de la constante C est déterminée par :

$$\alpha = P(W | m = 3) = P(\bar{X}_n \geq C | m = 3)$$

Pour $m = 3$, la moyenne empirique \bar{X}_n suit la loi $N(3, 1/\sqrt{n})$, donc en centrant et réduisant on obtient la condition :

$$\alpha = P \left(U \geq \frac{C-3}{1/\sqrt{n}} \right)$$

où U est une v.a. de loi $N(0, 1)$. Ainsi la constante C est définie par :

$$C = 3 + \frac{u}{\sqrt{n}}$$

où u est le fractile d'ordre $1-\alpha$ de la loi $N(0, 1)$. Pour un risque $\alpha = 0,05$ on lit dans la table 2 le fractile $u = 1,6449$ d'où une région critique définie pour $n = 100$ par :

$$W = \{(x_1, \dots, x_{100})/\bar{x}_{100} \geq 3,16\}$$

La fonction puissance du test est :

$$\begin{aligned} \eta(m) &= P\left(\bar{X}_n \geq 3 + \frac{u}{\sqrt{n}} \mid m > 3\right) = P\left\{U \geq \sqrt{n}(3 - m) + 1,645\right\} \\ &= 1 - \Phi\left[1,645 - \sqrt{n}(m - 3)\right] \end{aligned}$$

où Φ est la f.r.de la loi $N(0, 1)$.

- 8.10** Les hypothèses en présence sont de la même forme que dans l'exercice précédent. Mais ici, pour $m = 6$, la loi de la moyenne empirique \bar{X}_n est inconnue et nous ne pouvons pas déterminer le seuil. Nous allons utiliser une autre statistique, de loi connue pour $m = 6$, qui sera $\sqrt{n}(\bar{X}_n - 6)/S_n$, de loi de Student à $n - 1$ degrés de liberté. La région critique est définie par la condition :

$$\sqrt{n} \frac{\bar{x}_n - 6}{s_n} > C$$

On détermine la valeur de C par :

$$\alpha = P\left(\sqrt{n} \frac{\bar{X}_n - 6}{S_n} > C \mid m = 6\right)$$

Pour $n = 12$ et $\alpha = 0,025$ on lit dans la table 6 la valeur $C = 2,201$. On ne peut pas calculer la fonction puissance qui est définie par :

$$\eta(m) = P\left(\sqrt{n} \frac{\bar{X}_n - 6}{S_n} > C \mid m > 6\right) = P\left(\sqrt{n} \frac{\bar{X}_n - m}{S_n} > C - \sqrt{n} \frac{m - 6}{S_n}\right)$$

puisque $C - \sqrt{n} \frac{m - 6}{S_n}$ est une v.a

- 8.11** Nous allons appliquer le théorème de Lehmann. L'expression de la vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \frac{1}{2\theta\sqrt{x_i}} \exp\left(-\frac{\sqrt{x_i}}{\theta}\right) = \frac{1}{(2\theta)^n} \left(\prod_{i=1}^n \frac{1}{\sqrt{x_i}}\right) \exp\left(-\frac{1}{\theta} \sum_{i=1}^n \sqrt{x_i}\right)$$

pour $x_i > 0, 1 \leq i \leq n$.

Le rapport des vraisemblances s'écrit :

$$\frac{L(x_1, \dots, x_n; \theta)}{L(x_1, \dots, x_n; \theta')} = \left(\frac{\theta'}{\theta}\right)^n \exp\left[\frac{\theta - \theta'}{\theta\theta'} \sum_{i=1}^n \sqrt{x_i}\right]$$

C'est une fonction croissante de $s_n = \sum_{i=1}^n \sqrt{x_i}$ pour $\theta > \theta'$ donc $s_n > C$ définit la région critique d'un test UPP. La valeur de la constante C est déterminée par :

$$\alpha = P(W|\theta_0 = 1) = P\left(\sum_{i=1}^n \sqrt{X_i} \geq C \mid \theta_0 = 1\right)$$

Nous allons d'abord déterminer la loi de la v.a. $Y = \sqrt{X}$ à partir de sa fonction de répartition définie par :

$$G(y) = P(Y < y) = P(\sqrt{X} < y) = P(X < y^2) = F(y^2)$$

pour $y > 0$, F étant la f.r. de X . On obtient la densité par dérivation, soit pour $y > 0$:

$$g(y) = 2yf(y^2) = \frac{1}{\theta} e^{-y/\theta}$$

On reconnaît la loi exponentielle de paramètre $1/\theta$, ou loi $\gamma(1, 1/\theta)$. La v.a. $S_n = \sum_{i=1}^n \sqrt{X_i}$ est la somme de n v.a. indépendantes et de même loi $\gamma(1, 1/\theta)$, donc suit une loi $\gamma(n, 1/\theta)$. On utilise le résultat du chapitre 3 § 2.5 où on avait établi que $2S_n / \theta \rightsquigarrow \chi_{2n}^2$. On a donc ici :

$$\alpha = P\left(\sum_{i=1}^n \sqrt{X_i} \geq C \mid \theta_0 = 1\right) = P(2S_n \geq 2C)$$

Ainsi, $2C$ est le fractile d'ordre $1-\alpha$ de la loi χ_{2n}^2 . Pour $\alpha = 0,01$ et $n = 15$, on lit dans la table 5 la valeur $2C = 50,892$ soit un seuil $C = 25,45$. La fonction puissance du test est définie pour $\theta > 1$ par :

$$\eta(\theta) = P\left(\sum_{i=1}^n \sqrt{X_i} \geq C \mid \theta > 1\right) = P(2S_n/\theta \geq 2C/\theta) = 1 - K(2C/\theta)$$

avec K qui est la f.r. de la loi χ_{2n}^2 .

8.12 Nous allons associer à une machine de type A (resp. B) une variable indicatrice de Bernoulli X (resp. Y) de paramètre p_1 (resp. p_2). On dispose d'échantillons de ces deux lois, d'effectifs respectifs $n_1 = 80$ et $n_2 = 60$, pour effectuer le test :

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases} \Leftrightarrow \begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases}$$

Les moyennes empiriques \bar{X} et \bar{Y} de ces deux échantillons permettent de définir l'estimateur sans biais $\bar{X} - \bar{Y}$ du paramètre à tester $\theta = p_1 - p_2$. Sa loi approchée est une loi normale d'espérance θ et d'écart type inconnu $\sigma = \sqrt{p_1 q_1 / n_1 + p_2 q_2 / n_2}$ où $q_1 = 1 - p_1$ et $q_2 = 1 - p_2$. Sous l'hypothèse nulle, on estime la valeur commune $p = p_1 = p_2$ par la réunion des deux échantillons :

$$\hat{p} = \frac{n_1 \bar{X} + n_2 \bar{Y}}{n_1 + n_2} = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i \right)$$

Cet estimateur permet aussi d'estimer dans ce cas l'écart type $\sigma = \sqrt{p(1-p)(1/n_1 + 1/n_2)}$. On effectue alors le test à l'aide de la v.a. normalisée :

$$\hat{\theta} = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$$

dont on peut admettre, compte tenu des tailles d'échantillon, qu'elle suit approximativement une loi normale standard dans l'hypothèse nulle. La région critique est définie par $|\hat{\theta}| > C$, où on retient comme valeur approchée du seuil C celle qui vérifie $\alpha = P(|U| > C)$, avec U de loi $N(0, 1)$. Pour la valeur standard $\alpha = 0,05$ la région critique est définie par $|\hat{\theta}| > 1,96$. On obtient pour ces échantillons

$$\bar{x} = 0,625, \bar{y} = 0,667, \hat{p} = 0,643, \sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)} = 0,08$$

et $|\hat{\theta}| = 0,51$ donc on accepte l'hypothèse nulle d'équivalence des deux types de machines.

8.13 Chaque série d'observations est un échantillon des v.a. X et Y , de lois respectives $N(m_1, \sigma_1)$ et $N(m_2, \sigma_2)$, ces quatre paramètres étant inconnus. Les deux échantillons proviennent de la même loi si les paramètres sont égaux deux à deux. Le premier test à effectuer sur les espérances est :

$$\begin{cases} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{cases} \Leftrightarrow \begin{cases} H_0 : m_1 - m_2 = 0 \\ H_1 : m_1 - m_2 \neq 0 \end{cases}$$

et utilise pour cela l'estimateur sans biais $\bar{X} - \bar{Y}$ de $m_1 - m_2$. Cet estimateur suit une loi normale centrée sous H_0 , mais de variance inconnue $\sigma_1^2/n_1 + \sigma_2^2/n_2$ où n_1 et n_2 sont les effectifs des échantillons respectifs des lois de X et Y . On utilise donc les estimateurs sans biais de σ_1^2 et σ_2^2 qui sont respectivement :

$$S_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \quad \text{et} \quad S_y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$

Si on remplace l'écart type inconnu par son estimateur $\sqrt{S_x^2/n_1 + S_y^2/n_2}$ pour réduire $\bar{X} - \bar{Y}$, on n'obtiendra pas une loi de Student. Il faut pour cela que l'écart type des deux échantillons soit le même. On doit donc faire le test préalable d'égalité des variances :

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \Leftrightarrow \begin{cases} H_0 : \sigma_1^2/\sigma_2^2 = 1 \\ H_1 : \sigma_1^2/\sigma_2^2 \neq 1 \end{cases}$$

On accepte l'hypothèse nulle si le rapport S_x^2/S_y^2 est voisin de 1, soit une région d'acceptation de la forme :

$$a < \frac{S_x^2}{S_y^2} < b$$

Si on répartit le risque α de façon symétrique, les valeurs des constantes a et b sont définies par :

$$\frac{\alpha}{2} = P\left(\frac{S_x^2}{S_y^2} < a | H\right) = P\left(\frac{S_x^2}{S_y^2} > b | H_0\right)$$

avec S_x^2/S_y^2 qui suit une loi de Fisher-Snedecor $F(n_1 - 1, n_2 - 1)$ sous H_0 . Pour la valeur standard $\alpha = 0,05$ on lit dans la table 7 le fractile $a = \frac{1}{1,80} = 0,56$ et $b = 1,74$.

La valeur calculée de S_x^2/S_y^2 pour cet échantillon est 1,44 donc on accepte l'égalité des variances. On retient alors comme estimateur sans biais de la variance commune $\sigma^2 = \sigma_1^2 = \sigma_2^2$:

$$S^2 = \frac{(n_1 - 1) S_x^2 + (n_2 - 1) S_y^2}{n_1 + n_2 - 2}$$

La région critique du test initial est alors définie par :

$$\frac{|\bar{X} - \bar{Y}|}{S\sqrt{1/n_1 + 1/n_2}} > t$$

Le seuil t est déterminé par :

$$\alpha = P\left(\frac{|\bar{X} - \bar{Y}|}{S\sqrt{1/n_1 + 1/n_2}} > t | H_0\right)$$

C'est donc le fractile d'ordre $1-\alpha/2$ de la loi de Student à $n_1 + n_2 - 2$ degrés de liberté. Pour $\alpha = 0,05$ on lit dans la table 6 le fractile $t = 1,984$. Pour cet échantillon, on observe $s = 1,52$ et :

$$\frac{\bar{x} - \bar{y}}{s\sqrt{1/n_1 + 1/n_2}} = -9,77$$

On refuse donc l'égalité des moyennes de ces deux lois.

8.14 Dans l'hypothèse nulle d'indépendance des caractéristiques de taille et de CSP, la répartition de la population est obtenue à partir du produit des effectifs marginaux. Par exemple, le nombre d'ouvriers de moins de 165 cm serait dans ce cas $\frac{1900 \times 413}{2700} = 290,6$, arrondi à 291 puisqu'il s'agit d'effectifs entiers. On aboutit ainsi au tableau suivant :

	Ouvriers	Employés	Cadres	Total
Moins de 165 cm	291	73	49	413
De 165 à moins de 170 cm	457	115	77	649
De 170 à moins de 175 cm	645	163	109	917
175 cm et plus	507	128	86	721
Total	1900	479	321	2700

La région critique est de la forme $D_n \geq C$ où C a comme valeur approchée le fractile d'ordre $1-\alpha$ de la loi du khi-deux à six degrés de liberté, pour un test de risque de première espèce α . Pour $\alpha = 0,05$ on lit dans la table 5 la valeur $C = 12,6$. La valeur de la statistique utilisée pour ce test est ici $D_n = 58,2$ donc on rejette l'hypothèse d'indépendance. La conclusion serait la même pour un risque $\alpha = 0,01$, le seuil étant $C = 16,8$.

Tables statistiques

Table 1 : Fonction de répartition de la loi normale centrée réduite

Table 2 : Fractiles de la loi normale centrée réduite

Table 3 : Loi binômiale

Table 4 : Loi de Poisson

Table 5 : Fractiles de la loi χ^2_v

Table 6 : Fractiles de la loi de Student T_v

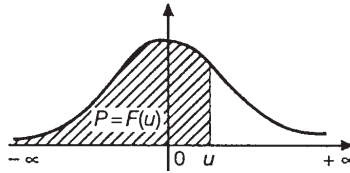
Table 7 : Fractiles d'ordre 0,95 de la loi de Fisher-Snedecor

Table 7 (suite) : Fractiles d'ordre 0,975 de la loi de Fisher-Snedecor

Abaque 1 : Intervalles de confiance pour une proportion (bilatéral de niveau 0,90 ou unilatéral de niveau 0,95)

Abaque 2 : Intervalles de confiance pour une proportion (bilatéral de niveau 0,95 ou unilatéral de niveau 0,975)

Table 1 – Fonction de répartition de la loi normale centrée réduite
 Probabilité $F(u)$ d'une valeur inférieure à u .



u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	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	0,9779	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

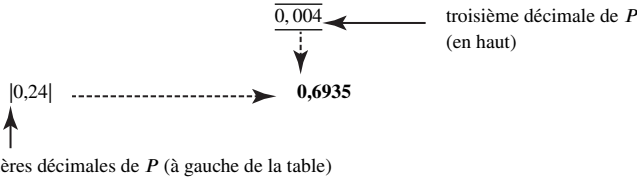
Tables pour les grandes valeurs de u

u	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
$F(u)$	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

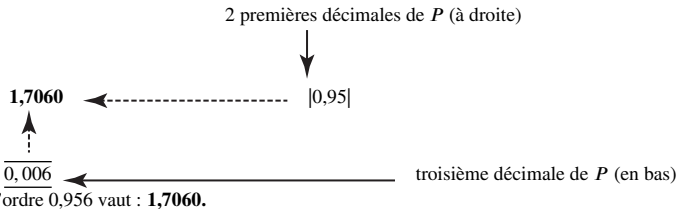
Table 2 – Fractiles d'ordre P de la loi normale centrée réduite.

Lecture de la table 2 des fractiles.

Si $P < 0,50$: le fractile est négatif, donc on ajoute le signe – à la valeur lue dans la table.



Si $P > 0,50$: on lit normalement.



P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	0,99
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0749	2,0537	0,98
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808	0,97
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	0,96
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	0,95
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	0,94
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	0,93
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	0,92
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	0,91
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816	0,90
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	0,89
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	0,88
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	0,87
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	0,86
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0450	1,0407	1,0364	0,85
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	0,84
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	0,83
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	0,82
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	0,81
0,19	0,8779	0,8742	0,8705	0,8669	0,8633	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	0,80
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	0,79
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	0,78
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	0,77
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	0,76
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	0,75
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464	0,6433	0,74
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158	0,6128	0,73
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858	0,5828	0,72
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563	0,5534	0,71
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273	0,5244	0,70
	0,010	0,009	0,008	0,007	0,006	0,005	0,004	0,003	0,002	0,001	0,000	P

Table 2 (suite) – Fractiles d'ordre P de la loi normale centrée réduite.

P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987	0,4959	0,69
0,31	0,4959	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705	0,4677	0,68
0,32	0,4677	0,4649	0,4621	0,4593	0,4565	0,4538	0,4510	0,4482	0,4454	0,4427	0,4399	0,67
0,33	0,4399	0,4372	0,4344	0,4316	0,4289	0,4261	0,4234	0,4207	0,4179	0,4152	0,4125	0,66
0,34	0,4125	0,4097	0,4070	0,4043	0,4016	0,3989	0,3961	0,3934	0,3907	0,3880	0,3853	0,65
0,35	0,3853	0,3826	0,3799	0,3772	0,3745	0,3719	0,3692	0,3665	0,3638	0,3611	0,3585	0,64
0,36	0,3585	0,3558	0,3531	0,3505	0,3478	0,3451	0,3425	0,3398	0,3372	0,3345	0,3319	0,63
0,37	0,3319	0,3292	0,3266	0,3239	0,3213	0,3186	0,3160	0,3134	0,3107	0,3081	0,3055	0,62
0,38	0,3055	0,3029	0,3002	0,2976	0,2950	0,2924	0,2898	0,2871	0,2845	0,2819	0,2793	0,61
0,39	0,2793	0,2767	0,2741	0,2715	0,2689	0,2663	0,2637	0,2611	0,2585	0,2559	0,2533	0,60
0,40	0,2533	0,2508	0,2482	0,2456	0,2430	0,2404	0,2378	0,2353	0,2327	0,2301	0,2275	0,59
0,41	0,2275	0,2250	0,2224	0,2198	0,2173	0,2147	0,2121	0,2096	0,2070	0,2045	0,2019	0,58
0,42	0,2019	0,1993	0,1968	0,1942	0,1917	0,1891	0,1866	0,1840	0,1815	0,1789	0,1764	0,57
0,43	0,1764	0,1738	0,1713	0,1687	0,1662	0,1637	0,1611	0,1586	0,1560	0,1535	0,1510	0,56
0,44	0,1510	0,1484	0,1459	0,1434	0,1408	0,1383	0,1358	0,1332	0,1307	0,1282	0,1257	0,55
0,45	0,1257	0,1231	0,1206	0,1181	0,1156	0,1130	0,1105	0,1080	0,1055	0,1030	0,1004	0,54
0,46	0,1004	0,0979	0,0954	0,0929	0,0904	0,0878	0,0853	0,0828	0,0803	0,0778	0,0753	0,53
0,47	0,0753	0,0728	0,0702	0,0677	0,0652	0,0627	0,0602	0,0577	0,0552	0,0527	0,0502	0,52
0,48	0,0502	0,0476	0,0451	0,0426	0,0401	0,0376	0,0351	0,0326	0,0301	0,0276	0,0251	0,51
0,49	0,0251	0,0226	0,0201	0,0175	0,0150	0,0125	0,0100	0,0075	0,0050	0,0025	0,0000	0,50
	0,010	0,009	0,008	0,007	0,006	0,005	0,004	0,003	0,002	0,001	0,000	P

Grandes valeurs de u

P	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
u_P	3,7190	4,2649	4,7534	5,1993	5,6120	5,9978

Table 3 – Loi binômiale.

Probabilités cumulées

n	c	Probabilités cumulées $\Pr(X \leq c) = \sum_{k=0}^{k=c} C_n^k p^k (1-p)^{n-k}$									
		p = 1%	p = 2%	p = 3%	p = 4%	p = 5%	p = 6%	p = 7%	p = 8%	p = 9%	p = 10%
10	0	0,9044	0,8171	0,7374	0,6648	0,5987	0,5386	0,4840	0,4344	0,3894	0,3486
	1	0,9957	0,9838	0,9655	0,9418	0,9139	0,8824	0,8483	0,8121	0,7746	0,7361
	2	0,9999	0,9991	0,9972	0,9938	0,9885	0,9812	0,9717	0,9599	0,9460	0,9298
	3	1	1	0,9999	0,9996	0,9990	0,9980	0,9964	0,9942	0,9912	0,9872
	4			1	1	0,9999	0,9998	0,9997	0,9994	0,9990	0,9984
	5					1	1	1	1	0,9999	0,9999
	6									1	1
20	0	0,8179	0,6676	0,5438	0,4420	0,3585	0,2901	0,2342	0,1887	0,1516	0,1216
	1	0,9831	0,9401	0,8802	0,8103	0,7358	0,6605	0,5869	0,5169	0,4516	0,3917
	2	0,9990	0,9929	0,9790	0,9561	0,9245	0,8850	0,8390	0,7879	0,7334	0,6769
	3		0,9994	0,9973	0,9926	0,9841	0,9710	0,9529	0,9294	0,9007	0,8670
	4		1	0,9997	0,9990	0,9974	0,9944	0,9893	0,9817	0,9710	0,9568
	5			1	0,9999	0,9997	0,9991	0,9981	0,9962	0,9932	0,9887
	6				1	1	0,9999	0,9997	0,9994	0,9987	0,9976
	7						1	1	0,9999	0,9998	0,9996
	8								1	1	0,9999
	9										1
30	0	0,7397	0,5455	0,4010	0,2939	0,2146	0,1563	0,1134	0,0820	0,0591	0,0424
	1	0,9639	0,8794	0,7731	0,6612	0,5535	0,4555	0,3694	0,2958	0,2343	0,1837
	2	0,9967	0,9783	0,9399	0,8831	0,8122	0,7324	0,6488	0,5654	0,4855	0,4114
	3	0,9998	0,9971	0,9881	0,9694	0,9392	0,8974	0,8450	0,7842	0,7175	0,6474
	4	0,9999	0,9996	0,9982	0,9937	0,9844	0,9685	0,9447	0,9126	0,8723	0,8245
	5	1	1	0,9997	0,9989	0,9967	0,9921	0,9838	0,9707	0,9519	0,9268
	6			1	0,9999	0,9994	0,9983	0,9960	0,9918	0,9848	0,9742
	7				1	0,9999	0,9997	0,9992	0,9980	0,9959	0,9922
	8					1	0,9999	0,9999	0,9996	0,9910	0,9980
	9						1	1	0,9999	0,9998	0,9995
	10								1	1	0,9999
	11										1
50	0	0,6050	0,3642	0,2181	0,1299	0,0769	0,0453	0,0266	0,0155	0,0090	0,0052
	1	0,9106	0,7358	0,5553	0,4005	0,2794	0,1900	0,1265	0,0827	0,0532	0,0338
	2	0,9862	0,9216	0,8106	0,6767	0,5405	0,4162	0,3108	0,2260	0,1605	0,1117
	3	0,9984	0,9822	0,9372	0,8609	0,7604	0,6473	0,5327	0,4253	0,3303	0,2503
	4	0,9999	0,9968	0,9832	0,9510	0,8964	0,8206	0,7290	0,6290	0,5277	0,4312
	5	1	0,9995	0,9963	0,9856	0,9622	0,9224	0,8650	0,7919	0,7072	0,6161
	6		0,9999	0,9993	0,9964	0,9882	0,9711	0,9417	0,8981	0,8404	0,7702
	7			0,9999	0,9992	0,9968	0,9906	0,9780	0,9562	0,9232	0,8779
	8			1	0,9999	0,9992	0,9973	0,9927	0,9834	0,9672	0,9421
	9				1	0,9998	0,9993	0,9978	0,9944	0,9875	0,9755
	10					1	0,9998	0,9994	0,9983	0,9957	0,9906
	11						1	0,9999	0,9995	0,9987	0,9968
	12							1	0,9999	0,9996	0,9990
	13								1	0,9999	0,9997
	14									1	0,9999
15										1	

Table 4 – Loi de Poisson.

k	Probabilités individuelles $Pr(k) = e^{-m} \frac{m^k}{k!}$								
	m = 0,1	m = 0,2	m = 0,3	m = 0,4	m = 0,5	m = 0,6	m = 0,7	m = 0,8	m = 0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4		0,0003	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003

k	Probabilités individuelles $Pr(k) = e^{-m} \frac{m^k}{k!}$								
	m = 1,0	m = 1,5	m = 2,0	m = 2,5	m = 3,0	m = 3,5	m = 4,0	m = 4,5	m = 5,0
0	0,3679	0,2231	0,1353	0,0821	0,0498	0,0302	0,0183	0,0111	0,0067
1	0,3679	0,3347	0,2707	0,2052	0,1494	0,1057	0,0733	0,0500	0,0337
2	0,1839	0,2510	0,2707	0,2565	0,2240	0,1850	0,1465	0,1125	0,0842
3	0,0613	0,1255	0,1804	0,2138	0,2240	0,2158	0,1954	0,1687	0,1404
4	0,0153	0,0471	0,0902	0,1336	0,1680	0,1888	0,1954	0,1898	0,1755
5	0,0031	0,0141	0,0361	0,0668	0,1008	0,1322	0,1563	0,1708	0,1755
6	0,0005	0,0035	0,0120	0,0278	0,0504	0,0771	0,1042	0,1281	0,1462
7	0,0001	0,0008	0,0034	0,0099	0,0216	0,0385	0,0595	0,0824	0,1044
8		0,0001	0,0009	0,0031	0,0081	0,0169	0,0298	0,0463	0,0653
9			0,0002	0,0009	0,0027	0,0066	0,0132	0,0232	0,0363
10				0,0002	0,0008	0,0023	0,0053	0,0104	0,0181
11					0,0002	0,0007	0,0019	0,0043	0,0082
12					0,0001	0,0002	0,0006	0,0016	0,0034
13						0,0001	0,0002	0,0006	0,0013
14							0,0001	0,0002	0,0005
15								0,0001	0,0002
16									0,0001

Table 5 – Fractiles d'ordre P de la loi χ^2_v .

$v \backslash P$	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,7	0,75	0,9	0,95	0,975	0,99	0,995	0,999
1	0,000	0,0002	0,001	0,004	0,016	0,102	0,455	1,074	1,32	2,706	3,841	5,02	6,635	7,88	10,827
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,408	2,77	4,605	5,991	7,38	9,210	10,60	13,815
3	0,072	0,115	0,216	0,352	0,584	1,21	2,366	3,665	4,11	6,251	7,815	9,35	11,345	13,00	16,266
4	0,207	0,297	0,484	0,711	1,064	1,92	3,357	4,878	5,39	7,779	9,488	11,14	13,277	15,00	18,467
5	0,412	0,554	0,831	1,15	1,610	2,67	4,351	6,064	6,63	9,236	11,070	12,83	15,086	16,86	20,515
6	0,676	0,872	1,24	1,64	2,204	3,45	5,348	7,231	7,84	10,645	12,592	14,45	16,812	18,65	22,457
7	0,989	1,24	1,69	2,17	2,833	4,25	6,346	8,383	9,04	12,017	14,067	16,01	18,475	20,37	24,322
8	1,34	1,64	2,18	2,73	3,490	5,07	7,344	9,524	10,22	13,362	15,507	17,53	20,090	22,03	26,125
9	1,73	2,09	2,70	3,33	4,168	5,90	8,343	10,656	11,39	14,684	16,919	19,02	21,666	23,66	27,877
10	2,16	2,56	3,25	3,94	4,865	6,73	9,342	11,781	12,55	15,987	18,307	20,48	23,209	25,25	29,588
11	2,60	3,05	3,82	4,57	5,578	7,58	10,341	12,899	13,70	17,275	19,675	21,92	24,725	26,82	31,264
12	3,07	3,57	4,40	5,23	6,304	8,43	11,340	14,011	14,85	18,549	21,026	23,34	26,217	28,35	32,909
13	3,57	4,11	5,01	5,89	7,042	9,30	12,340	15,119	15,98	19,812	22,362	24,74	27,688	29,87	34,528
14	4,07	4,66	5,63	6,57	7,790	10,16	13,339	16,222	17,12	21,064	23,685	26,12	29,141	31,37	36,123
15	4,60	5,23	6,26	7,26	8,547	11,03	14,339	17,322	18,25	22,307	24,996	27,49	30,578	32,85	37,697
16	5,14	5,81	6,91	7,96	9,312	11,91	15,338	18,418	19,37	23,542	26,296	28,85	32,000	34,31	39,252
17	5,70	6,41	7,56	8,67	10,085	12,79	16,338	19,511	20,49	24,769	27,587	30,19	33,409	35,76	40,790
18	6,26	7,01	8,23	9,39	10,865	13,67	17,338	20,601	21,60	25,989	28,869	31,53	34,805	37,19	42,312
19	6,84	7,63	8,91	10,12	11,651	14,56	18,338	21,689	22,72	27,204	30,144	32,85	36,191	38,62	43,820
20	7,43	8,26	9,59	10,85	12,443	15,45	19,337	22,775	23,83	28,412	31,410	34,17	37,566	40,03	45,315
21	8,03	8,90	10,28	11,56	13,240	16,34	20,337	23,858	24,93	29,615	32,671	35,48	38,932	41,43	46,797
22	8,64	9,54	10,98	12,34	14,041	17,24	21,337	24,939	26,04	30,813	33,924	36,78	40,289	42,83	48,268
23	9,26	10,20	11,69	13,09	14,848	18,13	22,337	26,018	27,14	32,007	35,172	38,08	41,638	44,21	49,728
24	9,89	10,86	12,40	13,85	15,659	19,03	23,337	27,096	28,24	33,196	36,415	39,36	42,980	45,59	51,179
25	10,56	11,52	13,12	14,61	16,47	19,940	24,34	28,172	29,340	34,38	37,652	40,65	44,314	46,96	52,620
26	11,20	12,20	13,84	15,38	17,29	20,840	25,34	29,246	30,430	35,56	38,885	41,92	45,642	48,32	54,052
27	11,84	12,88	14,57	16,15	18,114	21,75	26,336	30,319	31,53	36,741	40,113	43,19	46,963	49,67	55,476
28	12,49	13,56	15,31	16,93	18,939	22,66	27,336	31,391	32,62	37,916	41,337	44,46	48,278	51,02	56,893
29	13,15	14,26	16,05	17,71	19,768	23,56	28,236	32,461	33,71	39,087	42,557	45,72	49,588	52,36	58,302
30	13,82	14,95	16,79	18,49	20,599	24,48	29,336	33,530	34,80	40,256	43,773	46,98	50,892	53,70	59,703
40	20,73	22,16	24,43	26,51	29,05	33,66	39,34	44,16	45,62	51,80	55,76	59,34	63,69	66,78	73,44
50	28,01	29,71	32,36	34,76	37,69	42,94	49,33	54,72	56,33	63,17	67,50	71,42	76,15	79,50	86,69
60	35,55	37,48	40,48	43,19	46,46	52,29	59,33	65,23	66,98	74,40	79,08	83,30	88,38	91,96	99,63
70	43,29	45,44	48,76	51,74	55,33	61,70	69,33	75,69	77,58	85,53	90,53	95,02	100,42	104,22	112,34
80	51,18	53,54	57,15	60,39	64,28	71,14	79,33	86,12	88,13	96,58	101,88	106,63	112,33	116,33	124,86
90	59,21	61,75	65,65	69,13	73,29	80,62	89,33	96,52	98,64	107,56	113,14	118,14	124,12	128,31	137,22
100	67,34	70,06	74,22	77,93	82,36	90,13	99,33	106,91	109,14	118,50	124,34	129,56	135,81	140,18	149,46

Table 6 – Fractiles d'ordre P de la loi de Student T_v .

$v \backslash P$	0,60	0,70	0,80	0,90	0,95	0,975	0,990	0,995	0,999	0,9995
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	0,256	0,530	0,853	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	0,255	0,529	0,852	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	0,255	0,529	0,852	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	0,255	0,529	0,851	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,261	3,496
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
70	0,254	0,527	0,847	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
90	0,254	0,526	0,846	1,291	1,662	1,987	2,368	2,632	3,183	3,402
100	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

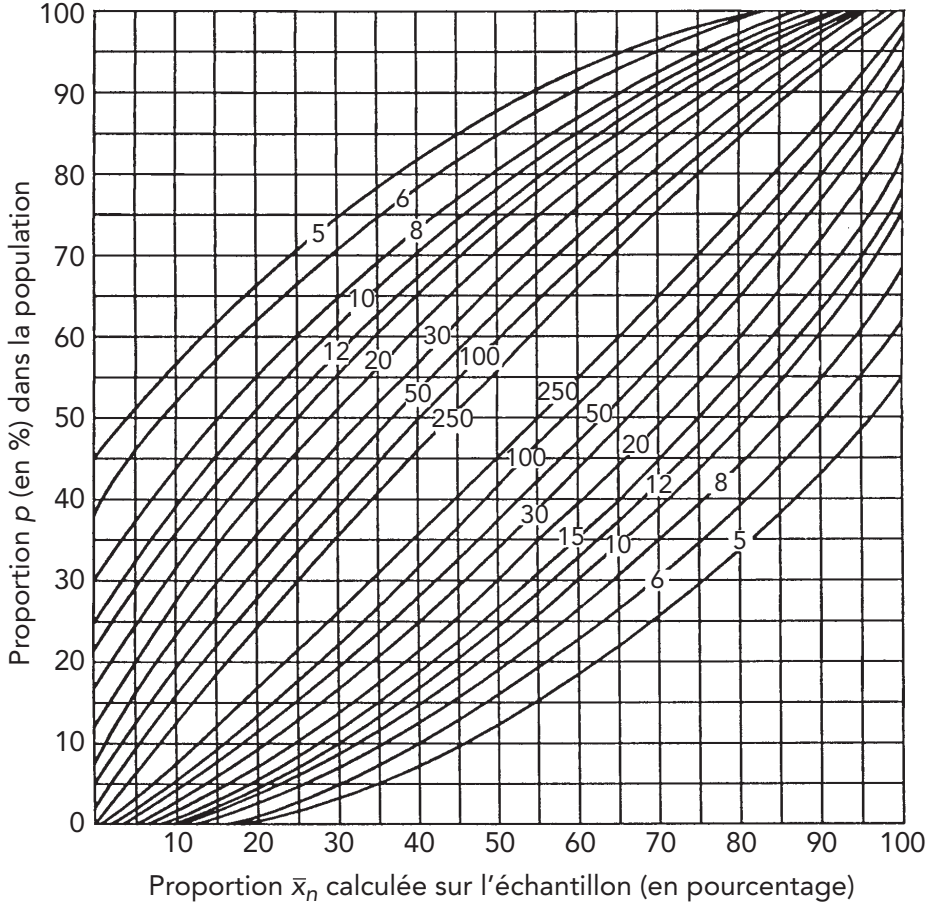
Table 7 – Fractiles d'ordre 0,95 de la loi de Fisher-Snedecor $F(v_1, v_2)$.

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
1	161	200	216	225	230	234	237	239	242	244	248	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,79	8,74	8,66	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	5,96	5,91	5,80	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,74	4,68	4,56	4,46	4,43	4,41	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,06	4,00	3,87	3,77	3,74	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,64	3,57	3,44	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,35	3,28	3,15	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,14	3,07	2,94	2,83	2,79	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	2,98	2,91	2,77	2,66	2,62	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,85	2,79	2,65	2,53	2,49	2,46	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,75	2,69	2,54	2,43	2,38	2,35	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,67	2,60	2,46	2,34	2,30	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,60	2,53	2,39	2,27	2,22	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,54	2,48	2,33	2,20	2,16	2,12	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,49	2,42	2,28	2,15	2,11	2,07	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,45	2,38	2,23	2,10	2,06	2,02	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,41	2,34	2,19	2,06	2,02	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,38	2,31	2,16	2,03	1,98	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,35	2,28	2,12	1,99	1,95	1,91	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,32	2,25	2,10	1,96	1,92	1,88	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,30	2,23	2,07	1,94	1,89	1,85	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,27	2,20	2,05	1,91	1,86	1,82	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,25	2,18	2,03	1,89	1,84	1,80	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,24	2,16	2,01	1,87	1,82	1,78	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,22	2,15	1,99	1,85	1,80	1,76	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,20	2,13	1,97	1,84	1,79	1,74	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,19	2,12	1,96	1,82	1,77	1,73	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,18	2,10	1,94	1,81	1,75	1,71	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,16	2,09	1,93	1,79	1,74	1,70	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,08	2,00	1,84	1,69	1,64	1,59	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	1,99	1,92	1,75	1,59	1,53	1,48	1,39
120	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,91	1,83	1,65	1,49	1,42	1,36	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,83	1,75	1,57	1,39	1,32	1,24	1,00

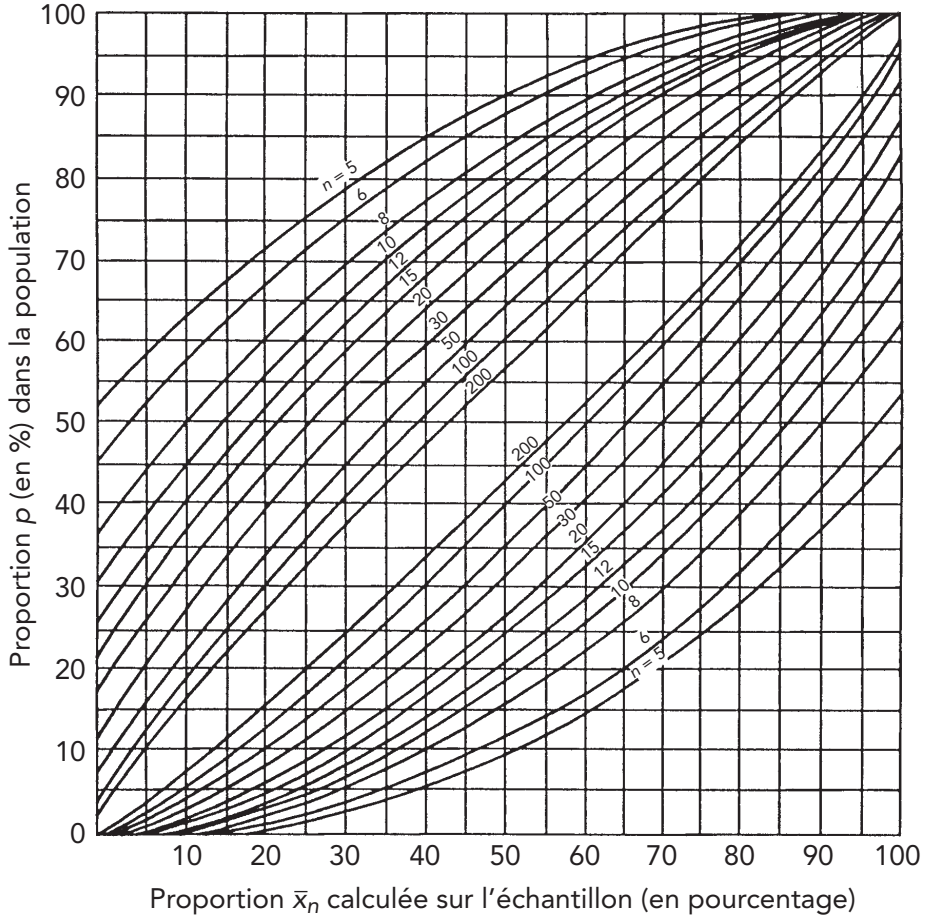
Table 7 (suite) – Fractiles d'ordre 0,975 de la loi de Fisher-Snedecor $F(v_1, v_2)$.

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
1	648	800	864	900	922	937	948	957	969	977	993	1006	1010	1013	1018
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,4	14,3	14,2	14,0	14,0	14,0	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,84	8,75	8,56	8,41	8,36	8,32	8,26
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,62	6,52	6,33	6,18	6,12	6,08	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,46	5,37	5,17	5,01	4,96	4,92	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,76	4,67	4,47	4,31	4,25	4,21	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,30	4,20	4,00	3,84	3,78	3,74	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	3,96	3,87	3,67	3,51	3,45	3,40	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,72	3,62	3,42	3,26	3,20	3,15	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,53	3,43	3,23	3,06	3,00	2,96	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,37	3,28	3,07	2,91	2,85	2,80	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,25	3,15	2,95	2,78	2,72	2,67	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,15	3,05	2,84	2,67	2,61	2,56	2,49
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,06	2,96	2,76	2,58	2,52	2,47	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	2,99	2,89	2,68	2,51	2,45	2,40	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,92	2,82	2,62	2,44	2,38	2,33	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,87	2,77	2,56	2,38	2,32	2,27	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,82	2,72	2,51	2,33	2,27	2,22	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,77	2,68	2,46	2,29	2,22	2,17	2,09
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,73	2,64	2,42	2,25	2,18	2,13	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,70	2,60	2,39	2,21	2,14	2,09	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,67	2,57	2,36	2,18	2,11	2,06	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,64	2,54	2,33	2,15	2,08	2,02	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,61	2,51	2,30	2,12	2,05	2,00	1,91
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,59	2,49	2,28	2,09	2,03	1,97	1,88
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,57	2,47	2,25	2,07	2,00	1,94	1,85
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,55	2,45	2,23	2,05	1,98	1,92	1,83
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,53	2,43	2,21	2,03	1,96	1,90	1,81
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,51	2,41	2,20	2,01	1,94	1,88	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,39	2,29	2,07	1,88	1,80	1,74	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,27	2,17	1,94	1,74	1,67	1,60	1,48
120	5,15	3,80	3,22	2,89	2,67	2,51	2,39	2,30	2,15	2,05	1,82	1,61	1,52	1,45	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,05	1,94	1,71	1,48	1,39	1,30	1,00

Abaque 1 – Intervalles de confiance pour une proportion p .
 Intervalle bilatéral de niveau de confiance 0,90
 Intervalles unilatéraux de niveau de confiance 0,95



Abaque 2 – Intervalles de confiance pour une proportion p .
Intervalle bilatéral de niveau de confiance 0,95
Intervalles unilatéraux de niveau de confiance 0,975



Index

A

Abaque, 208, 278
adéquation (tests d'), 145
algèbre, 7, 8, 29
aplatissement, 52
arrangement, 20
asymétrie, 52

B

Bayes (formule de), 16, 30
Bayes (méthode de), 251
Bayes (règle de décision de), 251
Bernoulli (loi de), 68, 89
Bernoulli (théorème de), 172
Biais, 194
Bienaymé-Tchebychev (inégalité de), 166, 171, 195
Boole (inégalité de), 10, 181

C

Cauchy (loi de), 143, 150, 221
central limite (théorème), 174, 189, 208
coïncidence, 61
combinaison, 21
combinaison avec répétition, 23
converge en loi, 172, 174
converge en probabilité, 167
convergence presque complète, 181
convergence presque sûre, 180
convolution, 85–87, 110, 111, 119, 129
correction de continuité, 178
corrélation, 116
courbe d'efficacité, 258
covariance, 45, 108, 116, 120, 127, 130, 148

D

densité, 48
diffuse (loi), 47
Dirac (loi de), 67, 89, 124, 173, 174
durée de vie, 80, 226

E

écart absolu moyen, 214
échantillon, 148
échantillon gaussien, 150
échantillon ordonné, 154
efficace (estimateur), 199, 222

ensemble fondamental, 5, 25
équiprobabilité, 11, 19, 71
équitable, 41
erreur de première espèce, 250
erreur de seconde espèce, 250
espérance conditionnelle, 107
estimateur du maximum de vraisemblance, 201
étendue, 156
événement, 5, 7
événement certain, 7
événement élémentaire, 6
événement impossible, 7
événements incompatibles, 7
exhaustive (statistique), 217
exhaustive minimale (statistique), 219
exponentielle (famille), 219

F

Fisher (information de), 198, 207, 214, 232, 233, 242
Fisher (théorème de), 149, 157
Fisher-Snedecor (loi de), 143, 150, 160, 276
fonction gamma, 85
fonction pivotale, 213, 214
fonction puissance, 255, 261
formule de la probabilité totale, 16
fractile, 78, 84, 150, 151, 160, 179, 185
Fréchet-Darmonis-Cramer-Rao (inégalité de), 197, 216

G

Glivenko-Cantelli (théorème de), 169, 182

H

hypothèse alternative, 252
hypothèse multiple, 255
hypothèse nulle, 250, 252
hypothèse simple, 253

I

indépendance, 17, 33, 127
indépendance mutuelle, 18
intervalle interquartile, 84

J

jacobien, 118, 129
Jensen (inégalité de), 167

K

Kolmogorov (distance de), 154
 Kolmogorov-Smirnov (test de), 154, 155
 Koopman (théorème de), 222

L

Laplace (loi de), 119
 Lebesgue (mesure de), 55, 129
 Lehmann (théorème de), 256, 261, 273, 274
 Lehmann-Scheffé (théorème de), 223
 loi, 47
 loi bêta, 87
 loi binômiale, 69, 89, 97, 99, 110, 176
 loi binômiale négative, 76, 90
 loi conditionnelle, 115, 135
 loi de Poisson, 111
 loi des événements rares, 176
 loi des grands nombres, 171, 175, 183, 242
 loi du khi-deux, 86, 128, 133, 179, 210
 loi exponentielle, 48, 79, 232, 254
 loi forte des grands nombres, 181
 loi gamma, 85, 179
 loi géométrique, 76, 184, 231
 loi hypergéométrique, 71, 98, 111, 178
 loi log-normale, 88, 96, 226
 loi marginale, 42, 106, 131
 loi multinomiale, 122
 loi normale, 209
 loi sans mémoire, 95, 101
 loi uniforme, 77, 95, 132
 loi uniforme discrète, 11

M

Markov (inégalité de), 166
 matrice de variances-covariances, 120, 123, 124, 130
 médiane, 45, 58, 63
 mesurable (application), 55, 128
 méthode des moments, 202, 225, 229
 méthode du maximum de vraisemblance, 200, 225, 227
 mixte (loi), 47, 62
 moment empirique, 146, 176, 202
 moment factoriel, 70, 75, 90
 moyenne empirique, 146, 148, 175, 193
 moyenne quadratique, 184
 moyenne quadratique (convergence en), 170
 multiple, 253

N

Neyman et Pearson (méthode de), 252
 Neyman et Pearson (théorème de), 253
 Neyman-Fisher (théorème de), 218
 niveau, 258
 non-causalité, 25

P

parente (loi), 145, 151, 154, 160, 163, 191
 Pareto (loi de), 89, 260
 partition, 24
 Pascal (loi de), 76, 95, 97
 permutation, 20, 22
 Poincaré (formule de), 10, 33
 Poisson (loi de), 95, 136, 160, 178, 262
 probabilité, 5, 9
 probabilité conditionnelle, 13
 probabilité image, 36
 puissance, 253, 254, 258

R

Rao-Blackwell (théorème de), 223
 rapport des vraisemblances monotones, 256, 258
 région critique, 250
 région d'acceptation, 250
 règle de Bayes, 264
 règle de décision, 250, 258
 régression, 107, 117, 126, 130
 risque de première espèce, 252, 259, 261
 risque de seconde espèce, 252, 261

S

seuil critique, 151, 249
 Slutsky (théorème de), 170, 173
 statistique d'ordre, 156
 statistique exhaustive et complète, 219
 strict (estimateur), 193
 Student (loi de), 149, 179, 210
 support, 55, 156

T

taille, 258
 taux de panne, 95
 test d'indépendance du khi-deux, 257
 test d'adéquation, 151
 test du khi-deux, 152, 162
 test le plus puissant, 262

test (sans biais), 258
test mixte, 258
test pur, 258
test uniformément le plus puissant (UPP),
255
tribu, 9
tribu borélienne, 55

V

v.a. certaine, 44
v.a. indicatrice, 40, 43

valeurs extrêmes, 156
variable certaine, 37
variable indicatrice, 37, 38, 192
variance conditionnelle, 108, 118
variance empirique, 43, 147, 148, 175, 210
vecteur aléatoire, 120
vecteur normal, 124, 126, 127
vraisemblance, 197

W

Weibull (loi de), 262

Collection ÉCO SUP

Régis Bourdonnais, *Économétrie*, 10^e éd. 2018.

Régis Bourdonnais, Michel Terraza, *Analyse des séries temporelles*, 4^e éd. 2016.

Christophe Chauvet, *Politique économique*, 2016.

Marie Delaplace, *Monnaie et financement de l'économie*, 5^e éd. 2017.

Ghislain Deleplace, *Histoire de la pensée économique*, 3^e éd. 2018.

Stéphane Dees, *Macroéconomie financière*, 2019.

Bernard Goldfarb, Catherine Pardoux, *Introduction à la méthode statistique*, 7^e éd. 2013.

Bernard Guillochon, Annie Kawecky, Frédéric Peltrault, Baptiste Venet, *Économie internationale*, 8^e éd. 2016.

Naïla Hayek, Jean-Pierre Leca, *Mathématiques pour l'économie*, 5^e éd. 2015.

Florence Huart, *Économie des finances publiques*, 2^e éd. 2016.

Pascal de Lima, *Économie bancaire et croissance économique*, 2012.

Frédéric Poulon, *Économie générale*, 8^e éd. 2015.

Marc Raffinot, *Économie du développement*, 2015.

Murat Yildizoglu, *Introduction à la théorie des jeux*, 2^e éd. 2011.