Probabilités et statistique

Benjamin JOURDAIN

11 septembre 2013

À Anne

Préface

Ce livre est issu du polycopié du cours de probabilités et statistique de première année de l'École des Ponts ParisTech dont je suis responsable depuis 2002. Du fait de la nature inconnue ou chaotique de leur évolution, de nombreux phénomènes (météorologie, cours de bourse, volume de vente d'une pièce détachée automobile,...) font naturellement l'objet d'une modélisation aléatoire, ce qui explique la place de plus en plus grande accordée aux probabilités dans les formations d'ingénieurs et dans les cursus universitaires. L'objectif du cours est de permettre aux étudiants de comprendre comment construire de tels modèles aléatoires et comment identifier leurs paramètres à partir de données.
À la différence de nombreux cours de probabilités de niveau licence, il ne fait pas appel à la théorie de la mesure. En effet, dans la pédagogie développée à l'École des Ponts, les notions de tribu et de mesurabilité ne sont étudiées qu'en deuxième année pour pouvoir introduire les martingales. En conséquence, le prérequis pour la lecture de ce livre est léger : maîtrise des notions de série et d'intégrale et du calcul matriciel. Et l'accent est mis sur les notions centrales en probabilités et statistique que sont *la loi, l'indépendance, l'espérance, la variance, la fonction caractéristique, les convergences, l'estimateur du maximum de vraisemblance, les intervalles de confiance et les tests d'hypothèses* plutôt que sur les fondements théoriques de ces disciplines.
Des exercices sont insérés au cœur des chapitres pour permettre aux étudiants de mettre en application les différents concepts au fur et à mesure de leur introduction. Mais des exercices et problèmes en nombre plus important sont également réunis à la fin de chaque chapitre. Certains font l'objet d'une correction dans le chapitre 10. Enfin, après chaque chapitre, un résumé d'une page environ reprend les notions importantes qui viennent d'être développées.

Après un chapitre introductif sur les espaces de probabilité finis où les calculs se ramènent à du dénombrement, l'espérance et ses propriétés, dont la linéarité, sont présentées en détail au chapitre 2 dans le cadre des variables aléatoires discrètes. La généralisation de la linéarité de l'espérance au cas des vecteurs aléatoires à densité est énoncée sans preuve dans le chapitre 3. Ces deux chapitres précisent également la notion de loi d'une variable aléatoires et fournissent les outils nécessaires (loi marginale, formule de changement de variable pour les intégrales multidimensionnelles) pour déterminer la loi d'une variable aléatoire d'intérêt dans un modèle probabiliste spécifique.
Le chapitre 4 est consacré aux techniques permettant de simuler sur ordinateur les variables aléatoires discrètes et à densité introduites auparavant. La simulation sur ordinateur permet de mieux appréhender les deux grands théorèmes limites de la théorie des probabilités qui forment le cœur du chapitre 5 : *la loi forte des grands nombres et le théorème de la limite centrale*. Les différentes notions de convergence de variables aléatoires et leurs liens font également l'objet d'un traitement détaillé dans ce chapitre. L'objectif est que les étudiants acquièrent suffisamment de maîtrise sur les théorèmes limites pour bien comprendre ensuite les propriétés asymptotiques des estimateurs, intervalles de confiance

et tests d'hypothèses dans la partie statistique du cours. Le théorème de la limite centrale explique le rôle fondamental en théorie des probabilités de la loi gaussienne et plus généralement des vecteurs gaussiens, auxquels le chapitre 6 est consacré.

En raison de son importance, le modèle gaussien sert d'exemple clé dans toute la partie statistique du livre. Le chapitre 7 introduit l'estimation de paramètres dans le modèle statistique paramétrique. L'accent est mis sur l'estimateur du maximum de vraisemblance et ses propriétés et sur la construction d'intervalles de confiance permettant de mesurer la précision de l'estimation. La notion de test d'hypothèses est présentée sur l'exemple du modèle gaussien dans le chapitre 8 qui explique également comment vérifier si des données sont issues d'une loi de probabilité fixée. Le livre s'achève sur un chapitre consacré à la régression linéaire qui fournit un cadre pour l'étude de l'influence de certains facteurs explicatifs sur des grandeurs mesurées ou des données expérimentales.

Remerciements

Je tiens à remercier les membres de l'équipe enseignante du cours de probabilités et statistique de l'École des Ponts, Aurélien Alfonsi, Mohamed Ben Alaya, Anne Dutfoy, Michel de Lara, Julien Guyon, Tony Lelièvre, Jean-Michel Marin, Mohamed Sbai et Alain Toubol qui ont apporté de nombreuses améliorations à ce livre par leurs remarques et qui ont contribué à la compilation de corrigés d'exercices du chapitre 10. Je suis également très reconnaissant à Jean-François Delmas pour les emprunts qu'il m'a permis de faire au polycopié [6] de son cours de première année à l'ENSTA et au recueil d'exercices de son cours de statistique de seconde année à l'École des Ponts. Je dois beaucoup à Jean-Philippe Chancelier pour son aide précieuse concernant l'utilisation des logiciels Latex et Scilab. Ma gratitude va encore à tous les membres de l'équipe de probabilités appliquées du CERMICS et en particulier à Bernard Lapeyre pour nos discussions sur la pédagogie, qui, je l'espère, ont trouvé leur prolongement dans ce livre. Je tiens à remercier tous mes collègues du CERMICS pour l'ambiance de travail conviviale et stimulante qui règne au sein de ce laboratoire. Mes pensées vont enfin à Anne, Erwan et Alexy pour le bonheur que je partage avec eux au quotidien.

Benjamin Jourdain

Table des matières

Chapitre 1

Introduction : probabilité sur un espace fini

Historiquement, le calcul des probabilités s'est développé à partir du XVII$^{\text{e}}$ siècle autour des problèmes de jeux dans des situations où le nombre de cas possibles est fini. Les développements plus récents concernant des espaces non nécessairement finis nécessitent les outils techniques de la théorie de la mesure. Mais on peut introduire simplement sur les espaces finis toutes les notions importantes de probabilités sans avoir besoin de cet outillage.

1.1 Probabilité sur un espace fini, événements

1.1.1 Définitions

On s'intéresse à une expérience aléatoire qui conduit à la réalisation d'un seul résultat parmi un nombre fini de résultats possibles $\omega_1, \omega_2, \ldots, \omega_n$. On note $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ l'ensemble de ces résultats.

Exemple 1.1.1. – Jet d'une pièce à pile où face : $\Omega = \{P, F\}$.
– Jet d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Si on mesure la fréquence d'apparition du résultat ω_k au cours d'un grand nombre de répétitions de l'expérience i.e. on calcule le rapport $F_k = \frac{N_k}{N}$ du nombre N_k d'expériences dont le résultat est ω_k sur le nombre total d'expériences N, on constate qu'elle fluctue de moins en moins. La limite $p_k \geq 0$ de F_k lorsque $N \to +\infty$ correspond à la notion intuitive de probabilité.

On appelle événement une partie A de Ω. La fréquence de A c'est-à-dire la proportion d'expériences dont le résultat est dans A est égale à $\sum_{k:\omega_k \in A} F_k$. On est donc amené à associer la probabilité $\sum_{k:\omega_k \in A} p_k$ à l'événement A.
Comme la fréquence de Ω vaut 1, en passant à la limite, on obtient $\sum_{k=1}^n p_k = 1$.

Définition 1.1.2. *Une probabilité $\mathbb{P}$ sur un ensemble fini $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ est une pondération $p_1, p_2, \ldots, p_n$ des éléments de cet ensemble t.q.*

$$\forall 1 \leq k \leq n,\ p_k \geq 0 \quad \text{et} \quad \sum_{k=1}^n p_k = 1.$$

On attribue à tout événement $A \subset \Omega$ le nombre

$$\mathbb{P}(A) = \sum_{k:\omega_k \in A} p_k$$

qui est appelé probabilité de l'événement A.

Exemple 1.1.3. Jet de deux dés à six faces : $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$ où i désigne la valeur de la face supérieure du premier dé et j celle du second.
Pour des raisons de symétrie (si les dés ne sont pas pipés), on munit Ω de la pondération suivante :

$$\forall 1 \leq i,j \leq 6,\ p_{(i,j)} = \frac{1}{36}.$$

Soit A l'événement : les valeurs des deux dés sont identiques.

$$A = \{(1,1),(2,2),\dots,(6,6)\} \quad \text{et} \quad \mathbb{P}(A) = \sum_{i=1}^{6} p_{(i,i)} = \frac{6}{36} = \frac{1}{6}.$$

On note S la somme des deux dés et $\{S = k\}$ l'événement $\{(i,j) : S(i,j) = k\}$. On a $S(i,j) = i + j$. Donc

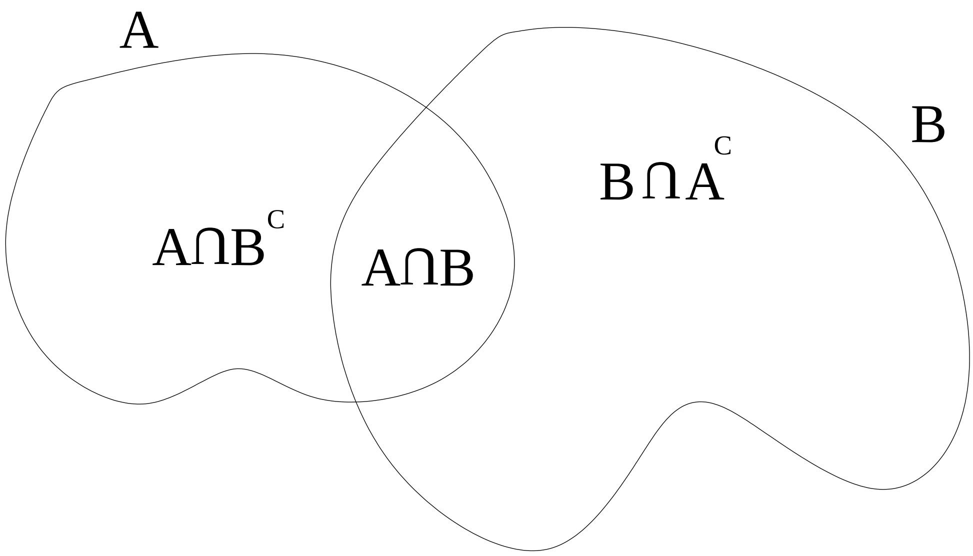
$$\begin{array}{lll}
\{S=2\} & = \{(1,1)\} & \mathbb{P}(S=2) = 1/36 \\
\{S=3\} & = \{(1,2),(2,1)\} & \mathbb{P}(S=3) = 1/18 \\
\{S=4\} & = \{(1,3),(2,2),(3,1)\} & \mathbb{P}(S=4) = 1/12 \\
\{S=5\} & = \{(1,4),(2,3),(3,2),(4,1)\} & \mathbb{P}(S=5) = 1/9 \\
\{S=6\} & = \{(1,5),(2,4),(3,3),(4,2),(5,1)\} & \mathbb{P}(S=6) = 5/36 \\
\{S=7\} & = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} & \mathbb{P}(S=7) = 1/6 \\
\{S=8\} & = \{(2,6),(3,5),(4,4),(5,3),(6,2)\} & \mathbb{P}(S=8) = 5/36 \\
\{S=9\} & = \{(3,6),(4,5),(5,4),(6,3)\} & \mathbb{P}(S=9) = 1/9 \\
\{S=10\} & = \{(4,6),(5,5),(6,4)\} & \mathbb{P}(S=10) = 1/12 \\
\{S=11\} & = \{(5,6),(6,5)\} & \mathbb{P}(S=11) = 1/18 \\
\{S=12\} & = \{(6,6)\} & \mathbb{P}(S=12) = 1/36
\end{array}$$

Terminologie concernant les événements :

- Si $\mathbb{P}(A) = 0$, l'événement A est dit négligeable.
- Si $\mathbb{P}(A) = 1$, il est dit presque sûr.
- On appelle événement contraire de A et on note A^c l'événement $\Omega \setminus A$.
- Si $A, B \subset \Omega$, l'événement A et B (réalisé lorsque A et B le sont) est noté $A \cap B$.
- L'événement A ou B (réalisé lorsque A ou B le sont) est noté $A \cup B$.

Probabilité de l'événement $A \cup B$:

Par définition, $\mathbb{P}(A \cup B) = \sum_{k:\omega_k \in A \cup B} p_k$. Comme $A \cup B$ est égal à l'union disjointe



$(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$,

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \sum_{k:\omega_k \in A\cap B^c} p_k + \sum_{k:\omega_k \in A\cap B} p_k + \sum_{k:\omega_k \in A^c\cap B} p_k \\
&= \left(\sum_{k:\omega_k \in A\cap B^c} p_k + \sum_{k:\omega_k \in A\cap B} p_k \right) + \left(\sum_{k:\omega_k \in A^c\cap B} p_k + \sum_{k:\omega_k \in A\cap B} p_k \right) - \sum_{k:\omega_k \in A\cap B} p_k \\
&= \sum_{k:\omega_k \in A} p_k + \sum_{k:\omega_k \in B} p_k - \sum_{k:\omega_k \in A\cap B} p_k \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).
\end{aligned}$$

Ainsi

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

Fonction indicatrice :

On appelle fonction indicatrice de l'événement A la fonction $1_A : \Omega \to \{0,1\}$ définie par

$$\forall \omega \in \Omega,\ 1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases}$$

Exercice 1.1.4. Quel est l'événement $\{\omega : 1_A(\omega) \times 1_B(\omega) = 1\}$ que l'on note aussi de façon condensée $\{1_A \times 1_B = 1\}$?
Conclure que

$$\boxed{1_{A\cap B} = 1_A \times 1_B.}$$

Montrer également que

$$\boxed{1_{A^c} = 1 - 1_A \quad \text{et} \quad 1_{A\cup B} = 1_A + 1_B - 1_{A\cap B}.}$$

1.1.2 Probabilités uniformes

Dans le cas où les symétries font que tous les résultats possibles $\omega_1, \omega_2, \dots \omega_n$ jouent le même rôle, ces résultats doivent avoir la même pondération $1/\text{Card }(\Omega)$. On dit alors qu'il sont équiprobables.
On a alors pour tout événement $A \subset \Omega$,

$$\boxed{\mathbb{P}(A) = \sum_{k:\omega_k \in A} \frac{1}{\text{Card }(\Omega)} = \frac{\text{Card }(A)}{\text{Card }(\Omega)}.}$$

Cette probabilité $\mathbb{P}$ s'appelle *probabilité uniforme* sur Ω.

Exemple 1.1.5. Dans le cas du jet de deux dés non pipés, $\Omega = \{(i,j) : \ 1 \leq i, j \leq 6\}$ est muni de la probabilité uniforme.

Remarque 1.1.6. Si on s'intéresse à la somme des deux dés, on peut choisir $\Omega = \{2, 3, 4 \dots, 12\}$, ensemble des valeurs prises par cette somme. Mais faute de propriétés de symétrie, on ne sait pas munir cet espace d'une probabilité naturelle.
Dans l'exemple 1.1.3, en travaillant sur l'espace plus gros $\{(i,j) : \ 1 \leq i, j \leq 6\}$ des couples des valeurs des deux dés muni de la probabilité uniforme, nous avons pu construire la pondération naturelle sur les valeurs de la somme des deux dés. Cette pondération n'a rien d'uniforme.
Cet exemple permet de bien comprendre l'importance du choix de l'espace de probabilité sur lequel on travaille.

Dans le cas des probabilités uniformes, les calculs se ramènent à du dénombrement.

Rappels de dénombrement

On se donne $n, k \in \mathbb{N}^*$ avec $k \leq n$.

- Le nombre de permutations d'un ensemble à n éléments est $n!$.
- De façon plus générale, le nombre d'injections d'un ensemble à k éléments dans un ensemble à n éléments est

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

  Le facteur n (resp. $n-1$,..., resp. $n-k+1$) vient du choix de l'image du 1$^{\text{er}}$ (resp 2$^{\text{e}}$,..., k^{e}) élément.
- Le nombre de parties à k éléments d'un ensemble à n éléments est

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}.}$$

Exercice résolu 1.1.7. *Dans une classe de $n \leq 365$ élèves, quelle est la probabilité de l'événement : "2 élèves au moins sont nés le même jour" que l'on note A ?*

On choisit comme espace de probabilité $\Omega = \{f : [1, n] \to [1, 365]\}$ où pour $1 \leq i \leq n$, $f(i)$ représente le jour d'anniversaire du $i$ème élève dans l'ordre alphabétique.

Même si les naissances ne sont pas vraiment équiréparties au long de l'année, on munit Ω de la probabilité uniforme. On a Card $(\Omega) = 365^n$.
Pour calculer la probabilité de A, on peut calculer la probabilité de l'événement contraire A^c : "tous les élèves ont des dates d'anniversaire différentes". En effet comme $A \cup A^c = \Omega$ et $A \cap A^c = \emptyset$,

$$\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) - \mathbb{P}(A \cap A^c) \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c).$$

On a $A^c = \{f : [1, n] \to [1, 365] \text{ injective}\}$. Donc Card $(A^c) = A_{365}^n$ et

$$\mathbb{P}(A^c) = \frac{\text{Card }(A^c)}{\text{Card }(\Omega)} = \frac{365!}{(365-n)!365^n} = \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365-n+1}{365},$$

et

$$\mathbb{P}(A) = 1 - \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365-n+1}{365}.$$

On peut vérifier que dès que $n \geq 23$, cette probabilité est supérieure à 1/2.

1.2 Probabilité conditionnelle et indépendance

1.2.1 Probabilité conditionnelle

La notion de probabilité conditionnelle permet de prendre en compte l'information dont on dispose (à savoir qu'un événement B est réalisé) pour actualiser la probabilité que l'on donne à un événement A :

Définition 1.2.1. *Soit Ω muni d'une probabilité $\mathbb{P}$ et $A, B \subset \Omega$. La probabilité conditionnelle de l'événement A sachant l'événement B est notée $\mathbb{P}(A|B)$ et définie par*

$$\boxed{\mathbb{P}(A|B) = \begin{cases} \mathbb{P}(A \cap B)/\mathbb{P}(B) & \textit{si } \mathbb{P}(B) > 0 \\ \mathbb{P}(A) & \textit{sinon.} \end{cases}}$$

Remarque 1.2.2. Lorsque l'on sait que l'événement B est réalisé, il est naturel d'affecter à l'événement A un poids proportionnel à $\mathbb{P}(A \cap B)$, ce qui justifie le choix du numérateur dans la définition précédente. Le dénominateur $\mathbb{P}(B) = \mathbb{P}(\Omega \cap B)$ est une constante de normalisation qui assure que $\mathbb{P}(\Omega|B) = 1$.

Exercice résolu 1.2.3. 1. *Dans une famille qui comporte deux enfants, l'un est une fille. On cherche la probabilité que l'autre soit un garçon.*
On choisit $\Omega = \{FF, FG, GF, GG\}$ où par exemple FG signifie que l'aîné des enfants est une fille et le second un garçon.
Cet espace est muni de la probabilité uniforme. On note

$$A = \{\text{un des enfants est un garçon}\} = \{FG, GF, GG\}$$
$$B = \{\text{un des enfants est une fille}\} = \{FF, FG, GF\}.$$

On a $\mathbb{P}(B) = \frac{\text{Card }(B)}{\text{Card }(\Omega)} = \frac{3}{4}$. Comme $A \cap B = \{FG, GF\}$, $\mathbb{P}(A \cap B) = \frac{\text{Card }(A \cap B)}{\text{Card }(\Omega)} = \frac{1}{2}$.
Donc la probabilité recherchée est

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

2. *On suppose maintenant que l'aîné des enfants est une fille. On veut alors connaître la probabilité pour que l'autre soit un garçon.*

   En reprenant la démarche ci-dessus, on obtient que cette probabilité vaut 1/2.

Dans certains problèmes, ce sont les probabilités conditionnelles que l'on connaît naturellement et on est amené à utiliser la définition sous la forme

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

qui se généralise en

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_m) =& \mathbb{P}(A_m|A_1 \cap \ldots \cap A_{m-1}) \\ & \times \mathbb{P}(A_{m-1}|A_1 \cap \ldots \cap A_{m-2}) \ldots \mathbb{P}(A_2|A_1)\mathbb{P}(A_1),\end{aligned}$$

pour m événements $A_1, \ldots, A_m$.

Exercice résolu 1.2.4. *Parmi 10 pièces mécaniques, 4 sont défectueuses. On prend successivement deux pièces au hasard dans le lot (sans remise). Quelle est la probabilité pour que les deux pièces soient correctes.*

On note A_1 l'événement la première pièce est bonne et A_2 l'événement la seconde pièce est bonne.
Comme, au départ, il y a 6 pièces bonnes sur 10, $\mathbb{P}(A_1) = 6/10 = 3/5$. Lorsque l'on a retiré une pièce bonne, il reste 5 pièces bonnes sur 9. D'où $\mathbb{P}(A_2|A_1) = 5/9$. On conclut que la probabilité cherchée est

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) = \frac{5}{9} \times \frac{3}{5} = \frac{1}{3}.$$

On peut retrouver ce résultat en munissant l'espace

$$\Omega = \{\text{sous-ensembles comportant 2 pièces de l'ensemble des 10 pièces}\}$$

de la probabilité uniforme. L'événement dont on cherche la probabilité est

$$A = \{\text{sous-ensembles comportant 2 pièces de l'ensemble des 6 pièces correctes}\}.$$

On a alors

$$\mathbb{P}(A) = \frac{\text{Card }(A)}{\text{Card }(\Omega)} = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{6!\ 8!\ 2!}{10!\ 4!\ 2!} = \frac{6 \times 5}{10 \times 9} = \frac{1}{3}.$$

Enfin le résultat suivant qui porte le nom de formule de Bayes est souvent utile.

Proposition 1.2.5. *Soit $B_1, \ldots, B_m$ une partition de Ω (i.e. des sous-ensembles disjoints de Ω dont la réunion est Ω) et $A \subset \Omega$ t.q. $\mathbb{P}(A) > 0$. Alors pour tout $1 \leq i \leq m$,*

$$\boxed{\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^m \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.}$$

Démonstration : Le numérateur du second membre est égal à $\mathbb{P}(A \cap B_i)$. Le dénominateur vaut $\sum_{j=1}^m \mathbb{P}(A \cap B_j)$ et comme les B_j forment une partition de Ω il est égal à $\mathbb{P}(A)$. Donc le second membre est bien égal à $\mathbb{P}(A \cap B_i)/\mathbb{P}(A)$. □

Exercice 1.2.6. Pour dépister une maladie, on applique un test sanguin. Si le patient est atteint, le test donne un résultat positif dans 99% des cas. Mais le test est également positif pour 2% des personnes en bonne santé. La proportion de personnes malades dans la population soumise au test est de 10^{-3}. Calculer la probabilité pour qu'un patient soit en bonne santé sachant que le résultat de son test est positif.

Exercice 1.2.7. Alors qu'ils ne représentent que 13% de la population, les jeunes de 18 à 24 ans représentent 30% des tués sur la route. À l'aide de ces données vérifier qu'un jeune a 2.87 fois plus de risque de mourir sur la route qu'un autre usager.

1.2.2 Indépendance

Définition 1.2.8. *Soit Ω muni d'une probabilité $\mathbb{P}$. Deux événements A et B sont dits indépendants si*

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).}$$

Remarque 1.2.9. L'indépendance de A et B se caractérise aussi par les relations $\mathbb{P}(A|B) = \mathbb{P}(A)$ ou $\mathbb{P}(B|A) = \mathbb{P}(B)$, c'est-à-dire que la probabilité donnée à l'événement A (resp. B) n'est pas modifiée par l'information que l'événement B (resp. A) est réalisé.

Définition 1.2.10. *m événements $A_1, \dots, A_m$ sont dits indépendants si*

$$\forall I \subset \{1, \dots, m\},\ \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Attention

- Il ne suffit pas que $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_m) = \prod_{i=1}^m \mathbb{P}(A_i)$ pour que les événements soient indépendants.
- Pour que 3 événements soient indépendants, il ne suffit pas qu'il soient 2 à 2 indépendants.

Exemple 1.2.11. Jet de deux pièces à Pile ou Face : $\Omega = \{PP, PF, FP, FF\}$ où par exemple PF signifie que la première pièce donne Pile et la seconde Face. Cet espace est muni de la probabilité uniforme.
On note A l'événement "la première pièce donne Pile", B l'événement "la seconde pièce donne Face" et C l'événement "les deux pièces donnent le même résultat".

$$\begin{array}{rlrll}
A & = \{PP, PF\} & \mathbb{P}(A) & = 1/2 & \\
B & = \{PF, FF\} & \mathbb{P}(B) & = 1/2 & \\
C & = \{PP, FF\} & \mathbb{P}(C) & = 1/2 & \\
A \cap B & = \{PF\} & \mathbb{P}(A \cap B) & = 1/4 & = \mathbb{P}(A)\mathbb{P}(B) \\
A \cap C & = \{PP\} & \mathbb{P}(A \cap C) & = 1/4 & = \mathbb{P}(A)\mathbb{P}(C) \\
B \cap C & = \{FF\} & \mathbb{P}(B \cap C) & = 1/4 & = \mathbb{P}(B)\mathbb{P}(C) \\
A \cap B \cap C & = \emptyset & \mathbb{P}(A \cap B \cap C) & = 0 & \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).
\end{array}$$

Ainsi les événements A, B et C sont 2 à 2 indépendants mais pas indépendants.

1.3 Exercices

Exercice 1.3.1. Deux événements A et B disjoints ($A \cap B = \emptyset$) et de probabilités non nulles peuvent-ils être indépendants ?

Exercice 1.3.2. On tire successivement et sans remise 4 lettres du mot “ATTACHANT”. Quelle est la probabilité d'obtenir “CHAT” ?

Exercice 1.3.3. Eugène et Diogène ont l'habitude de se retrouver chaque semaine autour d'un verre et de décider à pile ou face qui règle l'addition. Eugène se lamente d'avoir payé les quatre dernières additions. Diogène lui propose alors de modifier la règle. Il propose à Eugène de lancer 5 fois la pièce et de ne payer que si apparaît une suite d'au moins 3 piles consécutifs ou de 3 faces consécutifs. Eugène se félicite d'avoir un si bon ami. À tort ou à raison ?

Exercice corrigé 1.3.4. Vous jouez à deux à la roulette russe avec un revolver doté d'un barillet tournant qui comporte six emplacements pour les balles. Chaque fois que l'on presse la détente, le barillet tourne d'un cran. Deux balles sont insérées côte à côte dans le barillet qui est ensuite positionné au hasard. Votre adversaire place le premier le canon du revolver contre sa tempe, presse la détente ... et reste en vie. Grand seigneur, il vous propose de faire tourner à nouveau le barillet au hasard avant de tirer à votre tour. Que décidez-vous ?

Exercice 1.3.5. Une personne rentre chez elle après une soirée un peu trop arrosée. Elle ne sait plus laquelle des n clés qui se trouvent dans sa poche ouvre la porte de son domicile. Elle essaie donc les clés successivement. Déterminer la probabilité pour que la k-ième clé soit la bonne ($1 \leq k \leq n$).

Exercice 1.3.6. On note T la valeur obtenue lors du jet d'un dé à douze faces (numérotées de 1 à 12) et S le total obtenu lors du jet de deux dés à six faces (numérotées de 1 à 6). Calculez $\mathbb{P}(S > T)$, $\mathbb{P}(S = T)$ et $\mathbb{P}(S < T)$. Dans un jeu à deux joueurs où celui qui obtient le plus grand total gagne, on vous propose de choisir entre le dé à douze faces et les deux dés à six faces. Quel est votre choix ?

Exercice 1.3.7. Soient $A_1, A_2, ..., A_n$ des événements. Montrer la formule du crible (ou formule de Poincaré)

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} p_k \qquad \text{où} \qquad p_k = \sum_{1 \leq i_1 < i_2 < ... < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k})$$

On pourra commencer par établir que pour $n \geq 2$,

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

Exercice corrigé 1.3.8. L'inspecteur chargé d'une enquête criminelle est à un certain stade convaincu à 60% de la culpabilité d'un suspect. Une nouvelle pièce à conviction permet soudain d'affirmer que le criminel est gaucher. Or 7% des individus dans la population sont gauchers. Comment l'inspecteur doit-il réapprécier la culpabilité du suspect, s'il se trouve que le suspect est gaucher ?

Exercice 1.3.9. Trois chasseurs tirent en même temps sur un éléphant lors d'un safari. La bête meurt atteinte par deux balles. On estime la valeur d'un chasseur par sa probabilité d'atteindre la cible en un coup. Ces probabilités sont respectivement $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{3}{4}$. Trouver pour chacun des chasseurs, la probabilité d'avoir raté l'éléphant.

Exercice 1.3.10. On dispose d'une carte comportant 2 faces rouges, d'une carte comportant une face rouge et une face blanche et d'une carte comportant 2 faces blanches. Une des trois cartes est tirée au sort et une des faces de cette carte (également choisie au hasard) est exposée. Cette face est rouge. On vous demande de parier sur la couleur de la face cachée. Choisissez-vous rouge ou blanc ?

Exercice 1.3.11. La couleur des yeux d'une personne est déterminée par 2 gènes dont l'un est transmis par le père et l'autre par la mère. On suppose qu'il y a deux formes possibles pour les gènes : la forme B (bleue) et la forme M (marron). La forme B est récessive c'est à dire qu'une personne qui a le génotype BM ou MB a les yeux marrons. Vos parents ont tous deux les yeux marrons mais votre sœur a les yeux bleus. Quelle est la probabilité pour que vous ayez les yeux marrons ? On suppose en plus que votre conjoint a les yeux bleus tandis que votre premier enfant a les yeux marrons. Quelle est la probabilité pour que votre second enfant ait les yeux bleus ?

1.4 Résumé

- Soit $A, B \subset \Omega$.
$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$
Pour $B = A^c$, cette égalité se récrit
$$\boxed{\mathbb{P}(A) = 1 - \mathbb{P}(A^c).}$$
- Pour $A, B \subset \Omega$, on appelle probabilité conditionnelle de l'événement A sachant l'événement B le nombre
$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.}$$
- Les événements $A_1, \ldots, A_m$ sont dits indépendants si
$$\forall I \subset [1, m],\ \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$
En particulier les événements A et B sont dits indépendants si
$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).}$$

Chapitre 2

Variables aléatoires discrètes

2.1 Espace de probabilité

Dans le cas d'un espace Ω fini, nous avons défini un événement comme une partie quelconque de Ω. Mais si on souhaite modéliser le temps de première obtention de Pile dans une suite de jets d'une pièce à Pile ou Face, on choisit naturellement $\Omega = \{P, F\}^{\mathbb{N}^*}$, ensemble qui n'est pas dénombrable (un ensemble est dénombrable s'il existe une surjection de $\mathbb{N}$ sur cet ensemble). Lorsque Ω n'est pas dénombrable, pour pouvoir construire une probabilité qui satisfasse des propriétés intuitives, il est nécessaire de restreindre les événements que l'on considère à une sous-classe de $\mathcal{P}(\Omega)$ appelée *tribu*.
Il est naturel de demander que si A et B sont dans cette sous-classe alors A^c, $A \cup B$ et $A \cap B$ le sont aussi. Dans la définition d'une tribu, on demande même la stabilité par union ou intersection dénombrable :

Définition 2.1.1. *Une tribu $\mathcal{A}$ sur Ω est une classe de parties de Ω qui vérifie les trois propriétés suivantes :*

i) $\emptyset, \Omega \in \mathcal{A}$.

ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

iii) *Si $(A_i)_{i \in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ alors $\bigcup_{i \in I} A_i$ et $\bigcap_{i \in I} A_i$ sont dans $\mathcal{A}$.*

On appelle événements les éléments de $\mathcal{A}$.

Exemple 2.1.2. – $\{\emptyset, \Omega\}$ est la plus petite tribu sur Ω. On l'appelle *tribu grossière.*
- $\mathcal{P}(\Omega)$ est la plus grosse tribu sur Ω. On l'appelle *tribu discrète.*
- Si $A \subset \Omega$, $\{\emptyset, A, A^c, \Omega\}$ est une tribu sur Ω.

Définition 2.1.3. *Soit Ω muni d'une tribu $\mathcal{A}$. On appelle probabilité sur $(\Omega, \mathcal{A})$ une application $\mathbb{P} : \mathcal{A} \to [0, 1]$ qui vérifie*

i) $\mathbb{P}(\Omega) = 1$.

ii) *Si $(A_i)_{i \in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints ($\forall i \neq j \in I,\ A_i \cap A_j = \emptyset$) alors*

$$\mathbb{P}\Big(\bigcup_{i \in I} A_i\Big) = \sum_{i \in I} \mathbb{P}(A_i).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ *s'appelle un espace de probabilité, la propriété ii) la* σ*-additivité.*

Dans toute la suite de ce cours, nous travaillerons sans le spécifier nécessairement systématiquement sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Nous pourrons également considérer sans inquiétude que toutes les parties de Ω qui nous intéressent sont des événements (i.e. sont dans $\mathcal{A}$). En effet, s'il est difficile de construire une probabilité sur Ω muni de la tribu discrète $\mathcal{P}(\Omega)$ dès que Ω n'est pas dénombrable, en revanche on peut le faire sur Ω muni d'une tribu suffisamment grosse pour qu'elle contienne tous les sous-ensembles de Ω définis sans l'aide de l'axiome du choix.
Enfin, en utilisant la propriété de σ-additivité ii), on peut facilement vérifier que l'on a toujours pour tout couple d'événements A et B :

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

Les définitions de la probabilité conditionnelle de A sachant B et celle de l'indépendance des événements données au chapitre précédent restent également valables. Comme nous pouvons maintenant être amenés à considérer des familles infinies d'événements, nous précisons simplement qu'une telle famille est indépendante si toute sous-famille finie l'est.

2.2 Variables aléatoires discrètes

2.2.1 Rappel sur les manipulations de séries

Si pour tout $k \in \mathbb{N}$, $a_k \geq 0$ alors la somme partielle $\sum_{k=1}^{n} a_k$ admet lorsque $n \to +\infty$ une limite dans $\mathbb{R} \cup \{+\infty\}$ notée $\sum_{k \in \mathbb{N}} a_k$. En outre, on peut changer l'ordre des termes de la série sans changer la limite : pour toute bijection $\sigma : \mathbb{N} \to \mathbb{N}$, $\sum_{k \in \mathbb{N}} a_{\sigma(k)} = \sum_{k \in \mathbb{N}} a_k$.
Si la série de terme général $(a_k)_{k \in \mathbb{N}}$ est absolument convergente au sens où $\sum_{k \in \mathbb{N}} |a_k| < +\infty$, alors elle convergente au sens où $\sum_{k=1}^{n} a_k$ admet lorsque $n \to +\infty$ une limite finie notée $\sum_{k \in \mathbb{N}} a_k$ et pour toute bijection $\sigma : \mathbb{N} \to \mathbb{N}$, $\sum_{k \in \mathbb{N}} a_{\sigma(k)} = \sum_{k \in \mathbb{N}} a_k$.
Soit maintenant F un ensemble dénombrable et $(a_x)_{x \in F} \in \mathbb{R}^F$. Si F est fini, la définition de $\sum_{x \in F} a_x$ ne pose aucun problème. Si F est infini, alors il existe une bijection $f : \mathbb{N} \to F$ et $\sum_{k \in \mathbb{N}} |a_{f(k)}|$ ne dépend pas de cette bijection puisque si $g : \mathbb{N} \to F$ désigne une autre bijection alors $\sigma = f^{-1} \circ g : \mathbb{N} \to \mathbb{N}$ est une bijection et $\sum_{k \in \mathbb{N}} |a_{g(k)}| = \sum_{k \in \mathbb{N}} |a_{f(\sigma(k))}| = \sum_{k \in \mathbb{N}} |a_{f(k)}|$. On pose donc $\sum_{x \in F} |a_x| = \sum_{k \in \mathbb{N}} |a_{f(k)}|$ pour f bijection quelconque de $\mathbb{N}$ sur F. Dans le cas où $\sum_{x \in F} |a_x| < +\infty$, la famille $(a_x)_{x \in F}$ est dite sommable et de même que précédemment, on peut définir $\sum_{x \in F} a_x$ comme $\sum_{k \in \mathbb{N}} a_{f(k)}$ où f est une bijection quelconque de $\mathbb{N}$ sur F.

Théorème 2.2.1. *Soit F et G deux ensembles dénombrables.*
- *Si $(a_{(x,y)})_{(x,y) \in F \times G}$ est une famille de nombres* **positifs***, alors*

$$\boxed{\sum_{(x,y) \in F \times G} a_{(x,y)} = \sum_{x \in F} \left(\sum_{y \in G} a_{(x,y)} \right) = \sum_{y \in G} \left(\sum_{x \in F} a_{(x,y)} \right).}$$

  Cela signifie que les trois membres sont soit simultanément finis et égaux soit simultanément égaux à $+\infty$.
- *Si $(a_{(x,y)})_{(x,y) \in F \times G}$ est sommable au sens où $\sum_{(x,y) \in F \times G} |a_{(x,y)}| < +\infty$, alors l'égalité précédente reste vraie.*

2.2.2 Définition

Définition 2.2.2. *On appelle variable aléatoire discrète une application $X : \Omega \to F$ où F est un ensemble dénombrable. Pour $x \in F$, on note de façon concise $\{X = x\}$ l'événement $\{\omega \in \Omega : X(\omega) = x\}$. La famille des nombres $(\mathbb{P}(X = x))_{x\in F}$ s'appelle la loi de X.*

Remarque 2.2.3.
- En toute rigueur, pour pouvoir considérer $(\mathbb{P}(X = x))_{x\in F}$, il faut imposer dans la définition que $\forall x \in F$, $\{X = x\} \in \mathcal{A}$.
- En général l'ensemble F est égal à $\mathbb{N}$ ou $\mathbb{Z}$ ou à une partie de $\mathbb{Z}$.
- Notons que $(\{X = x\})_{x\in F}$ est une famille dénombrable d'événements deux à deux disjoints t.q. $\bigcup_{x\in F}\{X = x\} = \Omega$. Donc par la propriété de σ-additivité,

$$\sum_{x\in F} \mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{x\in F}\{X = x\}\right) = \mathbb{P}(\Omega) = 1.$$

Exemple 2.2.4.
- Dans le cas du jet de deux dés, la somme S des deux dés est une variable aléatoire discrète à valeurs dans $F = \{2, 3, \dots, 12\}$ dont nous avons calculé la loi dans l'exemple 1.1.3 :

$$\begin{array}{ll} \mathbb{P}(S = 2) = \mathbb{P}(S = 12) = 1/36 & \mathbb{P}(S = 3) = \mathbb{P}(S = 11) = 1/18 \\ \mathbb{P}(S = 4) = \mathbb{P}(S = 10) = 1/12 & \mathbb{P}(S = 5) = \mathbb{P}(S = 9) = 1/9 \\ \mathbb{P}(S = 6) = \mathbb{P}(S = 8) = 5/36 & \mathbb{P}(S = 7) = 1/6. \end{array}$$

  Il faut noter que la loi de S est la probabilité naturelle dont on doit munir l'ensemble $\{2, 3, \dots, 12\}$ lorsque l'on s'intéresse à la somme de deux dés.
- Soit $A \subset \Omega$ un événement. Sa fonction indicatrice $1_A : \Omega \to \{0, 1\}$ définie par

$$\forall \omega \in \Omega,\ 1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

  est une variable aléatoire discrète de loi :

$$\mathbb{P}(1_A = 1) = \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(1_A = 0) = 1 - \mathbb{P}(A).$$

2.2.3 Indépendance

Définition 2.2.5.
- *Deux variables aléatoires discrètes X et Y à valeurs respectivement dans F et G sont dites indépendantes si*

$$\boxed{\forall x \in F,\ \forall y \in G,\ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y).}$$

- *n variables aléatoires discrètes $X_1, X_2, \dots, X_n$ à valeurs respectivement dans F_1, $F_2, \dots, F_n$ sont dites indépendantes si*

$$\boxed{\forall x_1 \in F_1, \dots,\ \forall x_n \in F_n,\ \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).} \tag{2.1}$$

- *Une famille quelconque de variables aléatoires discrètes est dite indépendante si toute sous-famille finie l'est.*

Exemple 2.2.6. Jet de 2 dés : $\Omega = \{(i,j) : \ 1 \leq i,j \leq 6\}$ muni de la probabilité uniforme. Si on note X_1 la valeur du premier dé et X_2 celle du second, on a $X_1(i,j) = i$ et $X_2(i,j) = j$. On vérifie facilement que

$$\forall 1 \leq i \leq 6,\ \mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = i) = \frac{1}{6},$$

ce qui n'a rien de surprenant. Comme

$$\forall 1 \leq i,j \leq 6,\ \mathbb{P}(X_1 = i, X_2 = j) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = \mathbb{P}(X_1 = i) \times \mathbb{P}(X_2 = j),$$

les variables X_1 et X_2 sont indépendantes.

Remarque 2.2.7.
- Il peut sembler étonnant que dans la définition de l'indépendance de n variables aléatoires discrètes, contrairement à celle de l'indépendance des événements (définition 1.2.10), on ne regarde pas de propriétés portant sur des sous-ensembles des n variables. La raison est la suivante : en sommant par exemple sur $x_n \in F_n$ l'égalité (2.1), et en utilisant la propriété de σ-additivité on obtient

$$\mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n \in F_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_{n-1} = x_{n-1})\mathbb{P}(X_n \in F_n)$$
$$\text{i.e. } \mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \prod_{i=1}^{n-1} \mathbb{P}(X_i = x_i),$$

  c'est-à-dire une relation analogue à (2.1) mais dans laquelle la variable X_n a disparu. De façon plus générale pour $1 \leq d < n$, $\mathbb{P}(X_1 = x_1, \ldots, X_d = x_d) = \prod_{i=1}^{d} \mathbb{P}(X_i = x_i)$.
- La remarque précédente permet de montrer que si les variables aléatoires discrètes $X_1, \ldots, X_n$ sont indépendantes, pour $1 \leq d < n$, les deux variables aléatoires discrètes $(X_1, \ldots, X_d)$ et $(X_{d+1}, \ldots, X_n)$ sont indépendantes : en effet,

$$\begin{aligned}\mathbb{P}((X_1, \ldots, X_d) = (x_1, \ldots, x_d))&\mathbb{P}((X_{d+1}, \ldots, X_n) = (x_{d+1}, \ldots, x_n)) \\ &= \prod_{i=1}^{d} \mathbb{P}(X_i = x_i) \prod_{j=d+1}^{n} \mathbb{P}(X_j = x_j) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).\end{aligned}$$

  Ce résultat se généralise bien sûr de la façon suivante : $\forall m \in \{1, \ldots, n-1\}$, $\forall 1 \leq d_1 < d_2 < \ldots < d_m < n$, les variables aléatoires discrètes $(X_1, \ldots, X_{d_1})$, $(X_{d_1+1}, \ldots, X_{d_2}), \ldots, (X_{d_{m-1}+1}, \ldots, X_{d_m})$ et $(X_{d_m+1}, \ldots, X_n)$ sont indépendantes.

2.2.4 Lois discrètes usuelles

Ce sont des lois qui portent sur $F \subset \mathbb{N}$.

Loi de Bernoulli de paramètre $p \in [0,1]$

On dit que X suit la loi de Bernoulli de paramètre p et on note $X \sim \mathcal{B}(p)$ si :

$$\boxed{\mathbb{P}(X = 1) = p \ \text{ et } \ \mathbb{P}(X = 0) = 1 - p.}$$

On a alors

$$\forall x \in \{0,1\},\ \mathbb{P}(X = x) = p^x(1-p)^{1-x} \quad (\text{convention : } 0^0 = 1). \tag{2.2}$$

Exemple 2.2.8. Si A est un événement, $1_A \sim \mathcal{B}(\mathbb{P}(A))$.

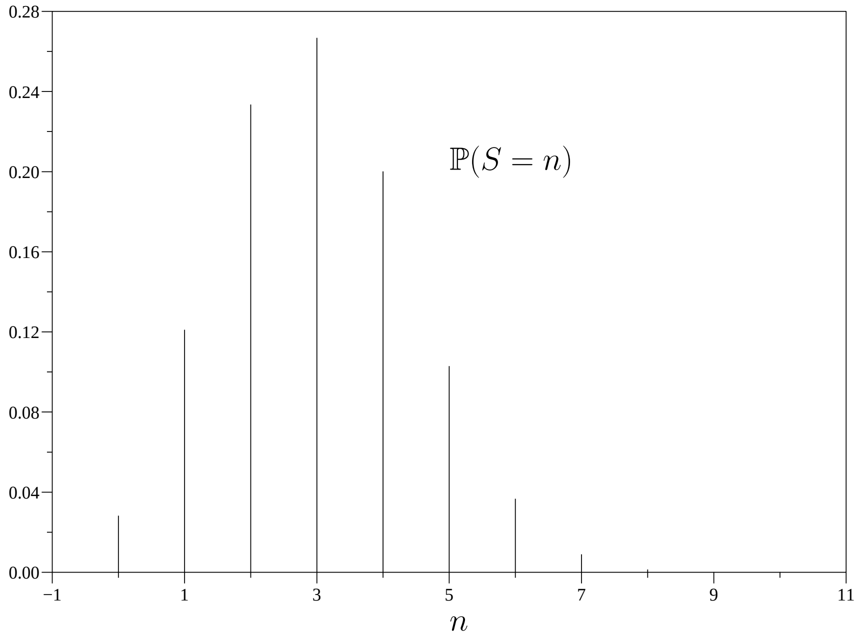


FIGURE 2.1 – Loi binomiale de paramètres $n = 10$ et $p = 0.3$.

Loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$

C'est la loi de la somme $S = X_1 + \ldots + X_n$ de n variables de Bernoulli de paramètre p indépendantes $X_1, \ldots, X_n$. On a alors pour $k \in F = \{0, \ldots, n\}$,

$$\begin{aligned}
\mathbb{P}(S = k) = \mathbb{P}(X_1 + \ldots + X_n = k) &= \mathbb{P}\left(\bigcup_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} \{X_1 = x_1, \ldots, X_n = x_n\}\right) \\
&= \sum_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \text{ par } \sigma\text{-additivité}, \\
&= \sum_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n) \text{ par indépendance des } X_i, \\
&= \sum_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} p^{x_1 + \ldots + x_n} (1 - p)^{n - x_1 - \ldots - x_n} \text{ d'après (2.2)}, \\
&= p^k (1 - p)^{n-k} \text{Card}\left(\{(x_1, \ldots, x_n) \in \{0, 1\}^n : \; x_1 + \ldots + x_n = k\}\right) \\
&= \binom{n}{k} p^k (1 - p)^{n-k}.
\end{aligned}$$

Le cardinal de $\{(x_1, \ldots, x_n) \in \{0, 1\}^n : \; x_1 + \ldots + x_n = k\}$ est $\binom{n}{k}$ car l'application de cet ensemble dans celui des parties de $\{1, \ldots, n\}$ à k éléments qui à $(x_1, \ldots, x_n)$ associe

$\{i : x_i = 1\}$ est une bijection.

$$\boxed{\text{Si } \forall 0 \leq k \leq n,\ \mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k},\ \text{on note } S \sim \mathcal{B}(n,p).}$$

La loi binomiale de paramètres $n = 10$ et $p = 0.3$ est représentée sur la figure 2.1.

Loi de Poisson de paramètre $\lambda > 0$

On dit que N suit la loi de Poisson de paramètre $\lambda > 0$ et on note $N \sim \mathcal{P}(\lambda)$ si

$$\boxed{\forall n \in \mathbb{N},\ \mathbb{P}(N = n) = \exp(-\lambda)\frac{\lambda^n}{n!}.}$$

La loi de Poisson de paramètre $\lambda = 2.5$ est représentée sur la figure 2.2.

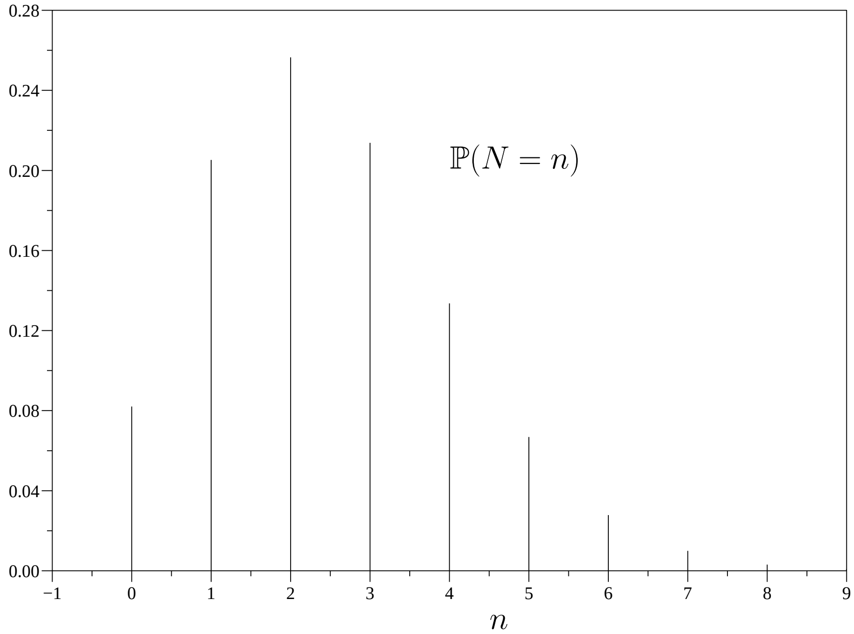


FIGURE 2.2 – Loi de Poisson de paramètre $\lambda = 2.5$.

Loi géométrique de paramètre $p \in]0, 1]$

C'est la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes où la probabilité de succès est p.
Une telle suite se modélise à l'aide d'une suite $(X_i)_{i\geq 1}$ de variables indépendantes et identiquement distribuées (abréviation : I.I.D.) suivant la loi de Bernoulli de paramètre p. L'événement "la ième expérience est un succès" s'écrit alors $\{X_i = 1\}$ et le temps T de premier succès est donné par

$$T = \inf\{i \geq 1 :\ X_i = 1\}.$$

Pour $k \geq 1$, en utilisant l'indépendance des X_i on obtient

$$\begin{aligned}\mathbb{P}(T = k) &= \mathbb{P}(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1)\\ &= \mathbb{P}(X_1 = 0) \ldots \mathbb{P}(X_{k-1} = 0)\mathbb{P}(X_k = 1) = (1-p)^{k-1}p.\end{aligned}$$

$$\boxed{\text{Si } \forall k \in \mathbb{N}^*,\ \mathbb{P}(T = k) = p(1-p)^{k-1},\ \text{on note } T \sim \mathcal{G}eo(p).}$$

On vérifie que

$$\mathbb{P}(T \in \mathbb{N}^*) = \sum_{k \in \mathbb{N}^*} \mathbb{P}(T = k) = p \sum_{k \in \mathbb{N}^*} (1-p)^{k-1} = p \sum_{l \in \mathbb{N}} (1-p)^l = p\, \frac{1}{1-(1-p)} = 1.$$

La loi géométrique de paramètre $p = 0.39$ est représentée sur la figure 2.3.

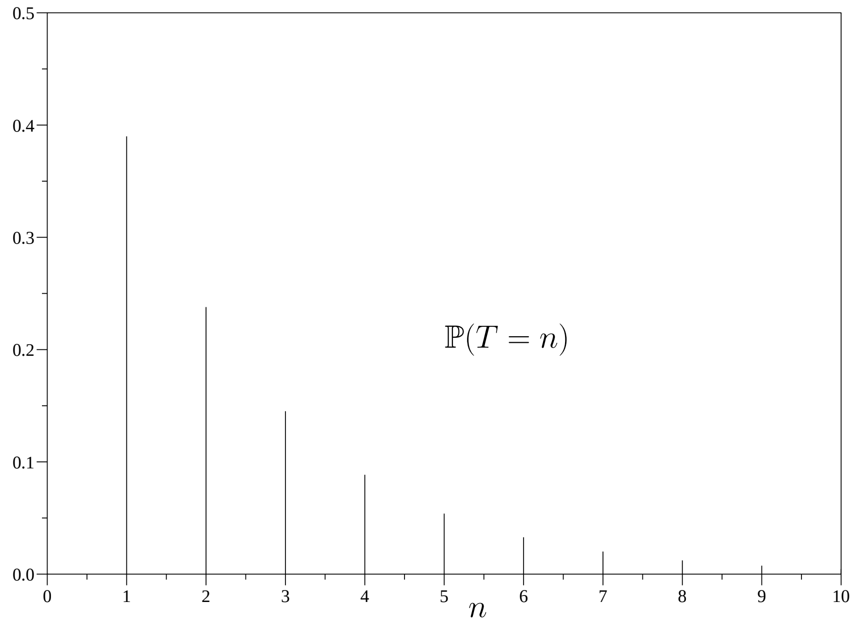


FIGURE 2.3 – Loi géométrique de paramètre $p = 0.39$.

Exercice 2.2.9. Soit $S \sim \mathcal{G}eo(p)$ et $T \sim \mathcal{G}eo(q)$ deux variables indépendantes. On cherche la loi de $Z = \min(S, T)$.

1. Pour $k \in \mathbb{N}^*$, calculer $\mathbb{P}(S \geq k)$.
2. En déduire $\mathbb{P}(Z \geq k)$.
3. Quelle est la loi de Z ?
4. Retrouver ce résultat en raisonnant en termes de temps de premier succès.

2.2.5 Loi marginale

Si X est une variable aléatoire discrète à valeurs dans F et Y une variable discrète à valeur dans G alors, comme le produit de deux ensembles dénombrables est dénombrable,

(X, Y) est une variable aléatoire discrète à valeurs dans $F \times G$. Mais la connaissance de la loi de X et de la loi de Y ne suffit pas à connaître la loi de (X, Y). Il faut rajouter de l'information comme par exemple le caractère indépendant de X et Y qui entraîne $\mathbb{P}((X, Y) = (x, y)) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ pour obtenir la loi du couple.

Exemple 2.2.10. Soit $X \sim \mathcal{B}(1/2)$. Alors il est facile de voir que $Y = 1 - X \sim \mathcal{B}(1/2)$. Ainsi Y a même loi que X, ce que l'on note aussi $Y \overset{\mathcal{L}}{=} X$.
Donc les premières coordonnées des couples (X, X) et (X, Y) ont même loi de même que les secondes coordonnées. Mais les deux couples n'ont bien sûr pas la même loi :

$$\mathbb{P}((X, Y) = (1, 0)) = \mathbb{P}(X = 1) = \frac{1}{2} \neq 0 = \mathbb{P}((X, X) = (1, 0)).$$

En revanche, si l'on connaît la loi du couple discret (X, Y) on en déduit la loi de X (et celle de Y) par la formule dite de la *loi marginale* :

Proposition 2.2.11. *Soit (X, Y) un couple discret à valeurs dans $F \times G$. Alors*

$$\boxed{\forall x \in F,\ \mathbb{P}(X = x) = \sum_{y \in G} \mathbb{P}(X = x, Y = y).}$$

On somme sur les valeurs prises par la variable Y dont on souhaite se débarrasser.

Démonstration : Le résultat se déduit de la propriété de σ-additivité en remarquant que $\{X = x\}$ est l'union disjointe de la famille dénombrable $\{X = x, Y = y\}_{y \in G}$. $\square$

Remarque 2.2.12. Si la loi de (X, Y) se met sous forme produit i.e.

$$\forall (x, y) \in F \times G,\ \mathbb{P}(X = x, Y = y) = c\mu(x)\nu(y),$$

alors les variables X et Y sont indépendantes de lois respectives $(\mu(x)/\sum_{x' \in F} \mu(x'))_{x \in F}$ et $(\nu(y)/\sum_{y' \in G} \nu(y'))_{y \in G}$.
En effet, comme $\sum_{(x', y') \in F \times G} \mathbb{P}(X = x', Y = y') = 1$,

$$c = 1 \Big/ \left(\left(\sum_{x' \in F} \mu(x') \right) \left(\sum_{y' \in G} \nu(y') \right) \right).$$

La formule de la loi marginale permet de vérifier que $\mathbb{P}(X = x) = \mu(x)/\sum_{x' \in F} \mu(x')$ et $\mathbb{P}(Y = y) = \nu(y)/\sum_{y' \in G} \nu(y')$ et on a bien $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$.

2.3 Espérance et variance

2.3.1 Espérance

Définition 2.3.1. *Soit $X : \Omega \to F \subset \mathbb{R}$ une variable aléatoire discrète à valeurs réelles. Elle est dite intégrable si $\sum_{x \in F} |x|\mathbb{P}(X = x) < +\infty$. Dans ce cas, on définit son espérance $\mathbb{E}(X)$ par*

$$\boxed{\mathbb{E}(X) = \sum_{x \in F} x\mathbb{P}(X = x).}$$

Remarque 2.3.2. – Le caractère intégrable et l'espérance d'une variable aléatoire ne dépendent que de sa loi : $X \stackrel{\mathcal{L}}{=} Y \Rightarrow \mathbb{E}(X) = \mathbb{E}(Y)$.

– Si on note $|F| = \{|x| : x \in F\}$, alors

$$\sum_{x\in F} |x|\mathbb{P}(X = x) = \sum_{y\in|F|} (|y|\mathbb{P}(X = y) + |-y|\mathbb{P}(X = -y)) = \sum_{y\in|F|} y\mathbb{P}(|X| = y).$$

Donc X est intégrable si et seulement si $|X|$ l'est et dans ce cas, $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

– $\mathbb{E}(1) = 1$. Plus généralement, l'espérance d'une constante est égale à cette constante.

– Soit A un événement. Sa fonction indicatrice qui est à valeurs dans $\{0, 1\}$ est bien sûr intégrable et

$$\boxed{\mathbb{E}(1_A) = 1 \times \mathbb{P}(1_A = 1) + 0 \times \mathbb{P}(1_A = 0) = \mathbb{P}(A).}$$

Ainsi l'espérance de la fonction indicatrice d'un événement est égale à la probabilité de cet événement. On peut donc dire que la notion d'espérance prolonge celle de probabilité.

Exemple 2.3.3. – Si $X \sim \mathcal{B}(p)$,

$$\mathbb{E}(X) = 1 \times \mathbb{P}(X = 1) + 0 \times \mathbb{P}(X = 0) = p.$$

– Si $Y \sim \mathcal{P}(\lambda)$,

$$\mathbb{E}(Y) = \sum_{n\in\mathbb{N}} n \exp(-\lambda)\frac{\lambda^n}{n!} = \lambda \exp(-\lambda) \sum_{n\in\mathbb{N}^*} \frac{\lambda^{n-1}}{(n-1)!} = \lambda \exp(-\lambda) \sum_{k\in\mathbb{N}} \frac{\lambda^k}{k!} = \lambda.$$

– Si $Z \sim \mathcal{G}eo(p)$,

$$\mathbb{E}(Z) = \sum_{n\in\mathbb{N}^*} np(1-p)^{n-1} = p(-f'(p)) \text{ où } f(x) = \sum_{k\in\mathbb{N}} (1-x)^k = \frac{1}{x}.$$

$$\text{D'où } \mathbb{E}(Z) = p\,\frac{1}{p^2} = \frac{1}{p}.$$

Le fait que le temps moyen d'attente du premier succès décroisse avec la probabilité de succès est parfaitement intuitif.

Propriétés 2.3.4. *1. Linéarité : si X et Y sont deux variables discrètes à valeurs réelles intégrables et $\lambda \in \mathbb{R}$, alors $X + \lambda Y$ est intégrable et*

$$\boxed{\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y).}$$

2. Condition suffisante d'intégrabilité : Si X et Y deux variables aléatoires discrètes réelles t.q. $\mathbb{P}(|X| \leq |Y|) = 1$ et Y est intégrable, alors X l'est aussi.

3. Positivité : si X est une variable aléatoire discrète intégrable et presque sûrement positive au sens où $\mathbb{P}(X \geq 0) = 1$ alors

- $\mathbb{E}(X) \geq 0$.
- $\mathbb{E}(X) = 0 \Longrightarrow \mathbb{P}(X = 0) = 1$ *(On dit alors que X est presque sûrement nulle).*

4. Croissance : si X et Y sont deux variables intégrables t.q. $\mathbb{P}(X \geq Y) = 1$ alors $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

Exercice 2.3.5. Soit $Y \sim \mathcal{B}(n,p)$. Montrer sans calcul que $\mathbb{E}(Y) = np$.

Démonstration :
Linéarité : Si F et G désignent les ensembles dénombrables dans lesquels X et Y prennent leurs valeurs, alors $Z = X + \lambda Y$ est une variable aléatoire discrète à valeurs dans $H = \{x + \lambda y : (x,y) \in F \times G\}$. En utilisant le fait que pour $z \in H$, l'événement $\{Z = z\}$ est l'union disjointe de la famille dénombrable $\{X = x, Y = y\}_{(x,y)\in F\times G: x+\lambda y = z}$, on obtient par σ-additivité que

$$\mathbb{P}(Z = z) = \sum_{\substack{(x,y)\in F\times G\\ x+\lambda y = z}} \mathbb{P}(X = x, Y = y) = \sum_{(x,y)\in F\times G} 1_{\{x+\lambda y = z\}} \mathbb{P}(X = x, Y = y). \tag{2.3}$$

Vérifions d'abord l'intégrabilité de Z : en utilisant (2.3) et le Théorème de Fubini 2.2.1, on a

$$\begin{aligned}
\sum_{z\in H} |z| \mathbb{P}(Z = z) &= \sum_{z\in H} |z| \sum_{(x,y)\in F\times G} 1_{\{x+\lambda y = z\}} \mathbb{P}(X = x, Y = y)\\
&= \sum_{z\in H} \sum_{(x,y)\in F\times G} 1_{\{x+\lambda y = z\}} |x + \lambda y| \mathbb{P}(X = x, Y = y)\\
&= \sum_{(x,y)\in F\times G} \left(\sum_{z\in H} 1_{\{x+\lambda y = z\}} \right) |x + \lambda y| \mathbb{P}(X = x, Y = y)\\
&= \sum_{(x,y)\in F\times G} |x + \lambda y| \mathbb{P}(X = x, Y = y)\\
&\le \sum_{(x,y)\in F\times G} (|x| + |\lambda||y|) \mathbb{P}(X = x, Y = y)\\
&= \sum_{x\in F} |x| \sum_{y\in G} \mathbb{P}(X = x, Y = y) + |\lambda| \sum_{y\in G} |y| \sum_{x\in F} \mathbb{P}(X = x, Y = y)\\
&= \sum_{x\in F} |x| \mathbb{P}(X = x) + |\lambda| \sum_{y\in G} |y| \mathbb{P}(Y = y) \text{ d'après la proposition 2.2.11.}
\end{aligned}$$

Le dernier membre est fini par intégrabilité de X et Y.
Pour montrer que $\mathbb{E}(Z) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$, on reprend le calcul que l'on vient d'effectuer en enlevant les valeurs absolues.
C.S. d'intégrabilité :
Soit X et Y deux variables aléatoires discrètes réelles à valeurs respectives dans F et G t.q. $\mathbb{P}(|X| \le |Y|) = 1$. Alors $\forall (x,y) \in F \times G$ t.q. $|x| > |y|$, $\mathbb{P}(X = x, Y = y) = 0$. En utilisant la formule de la loi marginale, on en déduit que

$$\begin{aligned}
\sum_{x\in F} |x| \mathbb{P}(X = x) &= \sum_{(x,y)\in F\times G} |x| \mathbb{P}(X = x, Y = y)\\
&\le \sum_{(x,y)\in F\times G} |y| \mathbb{P}(X = x, Y = y) = \sum_{y\in G} |y| \mathbb{P}(Y = y).
\end{aligned}$$

Donc si Y est intégrable, alors X l'est aussi.

Positivité et croissance :
Si X est intégrable, par définition $\mathbb{E}(X) = \sum_{x\in F} x\mathbb{P}(X = x)$. Si en outre $\mathbb{P}(X \geq 0) = 1$ alors

$$\mathbb{P}(X < 0) = \sum_{\substack{x\in F\\x<0}} \mathbb{P}(X = x) = 0 \quad \text{d'où } \mathbb{E}(X) = \sum_{\substack{x\in F\\x>0}} x\mathbb{P}(X = x).$$

Ainsi $\mathbb{E}(X) \geq 0$.
On en déduit en outre que si $\mathbb{E}(X) = 0$, alors $\forall x \in F\cap]0, +\infty[$, $\mathbb{P}(X = x) = 0$, ce qui entraîne que $\mathbb{P}(X > 0) = \sum_{\substack{x\in F\\x>0}} \mathbb{P}(X = x) = 0$ puis $\mathbb{P}(X = 0) = \mathbb{P}(X \geq 0) - \mathbb{P}(X > 0) = 1$.
Lorsque X et Y intégrables vérifient $\mathbb{P}(X \geq Y) = 1$, alors $\mathbb{P}(X - Y \geq 0) = 1$ ce qui entraîne $\mathbb{E}(X - Y) \geq 0$. On en déduit par linéarité de l'espérance que

$$\mathbb{E}(X) = \mathbb{E}(X - Y) + \mathbb{E}(Y) \geq \mathbb{E}(Y).$$

□

Le résultat suivant qui exprime l'espérance de $f(X)$ en fonction de la loi de X est très utile en pratique.

Théorème 2.3.6. *Soit $X : \Omega \to F$ une variable aléatoire discrète et $f : F \to \mathbb{R}$. Alors la variable $f(X)$ est intégrable si et seulement si $\sum_{x\in F} |f(x)|\mathbb{P}(X = x) < +\infty$ et alors,*

$$\boxed{\mathbb{E}(f(X)) = \sum_{x\in F} f(x)\mathbb{P}(X = x).}$$

Remarque 2.3.7.
- Si f est bornée sur F alors $\sum_{x\in F} |f(x)|\mathbb{P}(X = x) \leq \sup_{x\in F} |f(x)|$ et $f(X)$ est intégrable.
- $\mathbb{E}(f(X))$ ne dépend que de la loi de X : $X \stackrel{\mathcal{L}}{=} Y \Rightarrow \mathbb{E}(f(X)) = \mathbb{E}(f(Y))$.

Démonstration : La variable aléatoire $Y = f(X)$ est discrète à valeurs dans $f(F)$.

$$\sum_{y\in f(F)} |y|\mathbb{P}(Y = y) = \sum_{y\in f(F)} |y| \sum_{\substack{x\in F\\f(x)=y}} \mathbb{P}(X = x) = \sum_{x\in F} |f(x)|\mathbb{P}(X = x).$$

D'où la caractérisation de l'intégrabilité de $f(X)$. Dans le cas où cette variable aléatoire est effectivement intégrable, on peut reprendre le calcul précédent en enlevant les valeurs absolues pour obtenir l'expression désirée de son espérance. □

La caractérisation suivante de l'indépendance au travers des espérances fait intervenir un produit. D'une façon très générale l'indépendance permet d'effectuer des factorisations et il est bon d'avoir en tête le lien "indépendance = produit".

Proposition 2.3.8. *Soient X et Y deux variables aléatoires discrètes à valeurs respectivement dans F et G.*

1. *Si X et Y sont indépendantes alors pour toutes fonctions $f : F \to \mathbb{R}$ et $g : G \to \mathbb{R}$ t.q. $f(X)$ et $g(Y)$ sont intégrables, alors $f(X)g(Y)$ est intégrable et*

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

2. Inversement, si pour toutes fonctions $f : F \to \mathbb{R}$ et $g : G \to \mathbb{R}$ bornées, $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$, alors X et Y sont indépendantes.

Démonstration :
1. Supposons $f(X)$ et $g(Y)$ intégrables avec X et Y indépendantes. L'indépendance entraîne que

$$\sum_{(x,y)\in F\times G} |f(x)g(y)|\mathbb{P}(X=x, Y=y) = \sum_{(x,y)\in F\times G} |f(x)||g(y)|\mathbb{P}(X=x)\mathbb{P}(Y=y)$$
$$= \left(\sum_{x\in F} |f(x)|\mathbb{P}(X=x)\right)\left(\sum_{y\in G} |g(y)|\mathbb{P}(Y=y)\right).$$

Le dernier membre est fini par intégrabilité de $f(X)$ et $g(Y)$ et d'après le théorème 2.3.6. Toujours d'après ce résultat, on en déduit que $f(X)g(Y)$ est intégrable. En reprenant le calcul qui précède sans valeurs absolues, on conclut que l'espérance du produit est égale au produit des espérances.
2. Pour $x \in F$ et $y \in G$, on introduit les fonctions bornées $f(w) = 1_{\{w=x\}}$ et $g(z) = 1_{\{z=y\}}$. En utilisant la remarque 2.3.2, l'égalité de l'énoncé se récrit

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

□

Exercice 2.3.9. Montrer que si X et Y sont deux variables aléatoires discrètes indépendantes à valeurs respectives dans F et G, et $\varphi : F \to F'$, $\psi : G \to G'$ alors les variables $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

2.3.2 Variance

Définition 2.3.10. – *Soit $X : \Omega \to F \subset \mathbb{R}$ une variable aléatoire discrète à valeurs réelles. Alors X est dite de carré intégrable si X^2 est intégrable i.e. si*

$$\sum_{x\in F} x^2\mathbb{P}(X=x) < +\infty.$$

Dans ce cas, on définit la variance de X par

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

– *La racine carrée de la variance est appelée écart-type.*

Remarque 2.3.11. – Si X est de carré intégrable alors comme $|X| \leq (1+X^2)/2$, d'après la condition suffisante d'intégrabilité donnée dans les propriétés 2.3.4, cette variable est intégrable ce qui donne un sens à $\mathbb{E}(X)$. En outre, par linéarité de l'espérance, $(X-\mathbb{E}(X))^2 = X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2$ est intégrable. Ainsi la variance est bien définie.
– La variance et l'écart-type mesurent l'étalement de la variable aléatoire X autour de son espérance : plus ils sont grands et plus X est étalée.

Exercice 2.3.12. – Montrer que

$$\boxed{\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.} \tag{2.4}$$

– Montrer que $\forall a, b \in \mathbb{R},\ \mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

Remarque 2.3.13. – L'expression (2.4) et la remarque 2.3.7 impliquent que la variance d'une variable aléatoire ne dépend que de sa loi : $X \stackrel{\mathcal{L}}{=} Y \Longrightarrow \mathrm{Var}(X) = \mathrm{Var}(Y)$.
– Comme par positivité de l'espérance $\mathrm{Var}(X) \geq 0$, l'égalité (2.4) entraîne que $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$.

Dans l'énoncé suivant, l'hypothèse d'indépendance est essentielle. Pour une fois, elle ne permet pas une factorisation mais une sommation :

Proposition 2.3.14. *Soit $X_1, \ldots, X_n$ des variables aléatoires de carré intégrable. Alors $X_1 + \ldots + X_n$ est de carré intégrable et*

$$\boxed{\textit{si les } X_i \textit{ sont } \textbf{indépendantes}\textit{, alors } \mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i).}$$

Démonstration : Le fait que la somme $X_1 + \ldots + X_n$ est de carré intégrable découle de l'inégalité $(X_1 + \ldots + X_n)^2 \leq n(X_1^2 + \cdots + X_n^2)$. Par linéarité de l'espérance,

$$\begin{aligned}\mathrm{Var}(X_1 + \ldots + X_n) &= \mathbb{E}\left(\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i))\right)^2\right)\\ &= \mathbb{E}\left(\sum_{i,j=1}^n (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right).\end{aligned}$$

Si Y et Z sont deux variables de carré intégrable, comme $|YZ| \leq (Y^2+Z^2)/2$, leur produit YZ est intégrable. Donc chaque terme $(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))$ est intégrable et par linéarité de l'espérance,

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n \mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right).$$

On conclut en remarquant que par **indépendance** des variables $X_1, \ldots, X_n$, pour $i \neq j$,

$$\mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) = \mathbb{E}(X_i - \mathbb{E}(X_i))\mathbb{E}(X_j - \mathbb{E}(X_j)) = 0.$$

□

Exercice 2.3.15. Calculer la variance d'une variable de Bernoulli de paramètre p, d'une variable binomiale de paramètres n et p, d'une variable de Poisson de paramètre λ et d'une variable géométrique de paramètre p.

2.4 Fonction génératrice des variables aléatoires entières

Définition 2.4.1. *Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire discrète à valeurs entières. On appelle fonction génératrice de X la fonction $g_X : [-1, 1] \to \mathbb{R}$ définie par*

$$g_X(s) = \mathbb{E}(s^X) = \sum_{n\in\mathbb{N}} s^n \mathbb{P}(X = n).$$

Exemple 2.4.2. – Si $X \sim \mathcal{B}(n, p)$,

$$g_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{n} s^k \binom{n}{k} p^k(1-p)^{n-k} = (ps + (1-p))^n.$$

– Si $Y \sim \mathcal{G}eo(p)$,

$$g_Y(s) = \sum_{n\in\mathbb{N}^*} s^n p(1-p)^{n-1} = ps \sum_{n\in\mathbb{N}^*} ((1-p)s)^{n-1} = ps \sum_{k\in\mathbb{N}} ((1-p)s)^k = \frac{ps}{1-(1-p)s}.$$

– Si $N \sim \mathcal{P}(\lambda)$,

$$g_N(s) = \sum_{n\in\mathbb{N}} s^n e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n\in\mathbb{N}} \frac{(\lambda s)^n}{n!} = e^{\lambda(s-1)}.$$

Comme $\sum_{n\in\mathbb{N}} \mathbb{P}(X = n) = 1 < +\infty$, la série entière $\sum_{n\in\mathbb{N}} s^n \mathbb{P}(X = n) = g_X(s)$ a un rayon de convergence supérieur ou égal à 1, est C^∞ sur $]-1, 1[$ et continue sur $[-1, 1]$. En particulier si on note $g_X^{(k)}$ sa dérivée d'ordre k, on a

$$\forall k \in \mathbb{N},\ \mathbb{P}(X = k) = \frac{g_X^{(k)}(0)}{k!},$$

ce qui implique la première des propriétés suivantes.

Propriétés 2.4.3. *1. La fonction génératrice d'une variable aléatoire entière caractérise sa loi : si $X, Y : \Omega \to \mathbb{N}$, alors $X \stackrel{\mathcal{L}}{=} Y \Leftrightarrow \forall s \in [-1, 1],\ g_X(s) = g_Y(s)$.*

2. Pour $k \in \mathbb{N}^*$, $\lim_{s\to1^-} g_X^{(k)}(s) = \sum_{n\geq k} n \times (n-1) \times \ldots \times (n-k+1)\mathbb{P}(X = n)$ où les deux membres sont à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. En particulier, X est intégrable ssi $\lim_{s\to1^-} g'_X(s) < +\infty$ et dans ce cas, $\mathbb{E}(X) = \lim_{s\to1^-} g'_X(s)$.

Démonstration : Pour démontrer la seconde assertion, on pose $a_n = 1_{\{n\geq k\}} n \times \ldots \times (n-k+1)\mathbb{P}(X = n)$. Pour $s \in]-1, 1[$, on a alors $g_X^{(k)}(s) = \sum_{n\in\mathbb{N}} a_n s^{n-k}$.
Si $\sum_{n\in\mathbb{N}} a_n = +\infty$, on remarque que

$$\forall n_1 \geq k,\ \forall s \in [2^{-1/(n_1-k)}, 1[,\ g_X^{(k)}(s) \geq \sum_{n=0}^{n_1} a_n s^{n-k} \geq s^{n_1-k} \sum_{n=0}^{n_1} a_n \geq \frac{1}{2} \sum_{n=0}^{n_1} a_n,$$

ce qui assure que $\lim_{s\to1^-} g_X^{(k)}(s) = +\infty$.
Si $\sum_{n\in\mathbb{N}} a_n < +\infty$, $\forall n_1 \geq k$, $\forall s \in [0, 1[$,

$$0 \leq \sum_{n\in\mathbb{N}} a_n - g_X^{(k)}(s) \leq \sum_{n=1}^{n_1} a_n(1 - s^{n-k}) + \sum_{n>n_1} a_n \leq (1 - s^{n_1-k}) \sum_{n\in\mathbb{N}} a_n + \sum_{n>n_1} a_n.$$

On conclut en remarquant que le second terme du membre de droite est arbitrairement petit pour n_1 assez grand tandis qu'à n_1 fixé, le premier terme tend vers 0 lorsque $s \to 1^-$.
□

Un des principaux intérêts de la fonction génératrice est qu'elle permet très facilement de caractériser la loi de la somme de variables aléatoires entières indépendantes. En effet si X et Y sont de telles variables et $s \in [-1,1]$, alors d'après la proposition 2.3.8, $\mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X)\mathbb{E}(s^Y)$, ce qui démontre le résultat suivant.

Proposition 2.4.4. *Si $X : \Omega \to \mathbb{N}$ et $Y : \Omega \to \mathbb{N}$ sont indépendantes alors*

$$\forall s \in [-1,1],\ g_{X+Y}(s) = g_X(s) \times g_Y(s).$$

Exercice résolu 2.4.5. *Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires entières indépendantes et identiquement distribuées et N une variable aléatoire entière indépendante de la suite. On pose*

$$S = \begin{cases} X_1 + \ldots + X_N & si\ N \in \mathbb{N}^* \\ 0 & si\ N = 0 \end{cases}.$$

Exprimer g_S en fonction de g_{X_1} et g_N. En déduire la loi de S lorsque N suit la loi géométrique de paramètre p et les X_i la loi géométrique de paramètre q.

La somme définissant S est doublement aléatoire puisque les X_i sont aléatoires mais l'indice terminal de sommation N est aussi aléatoire. Pour effectuer le calcul, on va décomposer l'espérance donnant la fonction génératrice de S sur les valeurs prises par cet indice en remarquant que $\sum_{n\in\mathbb{N}} 1_{\{N=n\}} = 1$ puis utiliser les propriétés d'indépendance entre les différentes variables aléatoires. Ainsi,

$$\begin{aligned} g_S(s) &= \mathbb{E}\left(\sum_{n\in\mathbb{N}} 1_{\{N=n\}} s^S\right) = \mathbb{P}(N=0) + \sum_{n\in\mathbb{N}^*} \mathbb{E}\left(1_{\{N=n\}} \prod_{i=1}^{n} s^{X_i}\right) \\ &= \sum_{n\in\mathbb{N}} \mathbb{P}(N=n) g_{X_1}(s)^n = g_N(g_{X_1}(s)). \end{aligned}$$

Dans le cas particulier évoqué dans l'énoncé, $g_N(s) = \frac{ps}{1-(1-p)s}$, $g_{X_1}(s) = \frac{qs}{1-(1-q)s}$ et

$$g_S(s) = \frac{\frac{pqs}{1-(1-q)s}}{1 - \frac{(1-p)qs}{1-(1-q)s}} = \frac{pqs}{1-(1-q)s-(1-p)qs} = \frac{pqs}{1-(1-pq)s}.$$

On reconnaît la fonction génératrice de la loi géométrique de paramètre pq et on conclut que $S \sim \mathcal{G}eo(pq)$.

2.5 Loi et espérance conditionnelles

Définition 2.5.1. *Soient X et Y deux variables aléatoires discrètes à valeurs respectives dans F et G. Pour $y \in G$, on appelle loi conditionnelle de X sachant $Y = y$ la famille des nombres $(\mathbb{P}(X = x|Y = y))_{x\in F}$.*

Remarque 2.5.2. D'après la définition 1.2.1, on a

$$\mathbb{P}(X=x|Y=y)=\begin{cases}\dfrac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)} & \text{si } \mathbb{P}(Y=y)>0\\ \mathbb{P}(X=x) & \text{sinon.}\end{cases}$$

Comme d'après la proposition 2.2.11, $\sum_{x\in F}\mathbb{P}(X=x,Y=y)=\mathbb{P}(Y=y)$, lorsque $\mathbb{P}(Y=y)>0$, on a

$$\sum_{x\in F}\mathbb{P}(X=x|Y=y)=\sum_{x\in F}\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\frac{1}{\mathbb{P}(Y=y)}\sum_{x\in F}\mathbb{P}(X=x,Y=y)=1.$$

Dans le cas où $\mathbb{P}(Y=y)=0$, cette propriété reste vraie puisque $\sum_{x\in F}\mathbb{P}(X=x|Y=y)=\sum_{x\in F}\mathbb{P}(X=x)=1$.

Proposition 2.5.3. *Les variables X et Y sont indépendantes si et seulement si la loi conditionnelle de X sachant $Y=y$ ne dépend pas de $y\in G$.*

Démonstration : Il est immédiat de vérifier la condition nécessaire. Nous allons donc nous concentrer sur la condition suffisante.
Par hypothèse, pour tout $x\in F$, il existe $\mu(x)$ t.q. $\forall y\in G$,

$$\mathbb{P}(X=x|Y=y)=\mu(x). \tag{2.5}$$

À x fixé, en multipliant cette égalité par $\mathbb{P}(Y=y)$ et en sommant sur $y\in G$, on obtient $\mathbb{P}(X=x)=\mu(x)$. Donc (2.5) se récrit

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

□

Définition 2.5.4. *Soient X et Y deux variables aléatoires discrètes à valeurs dans F et G respectivement et $f:F\times G\to\mathbb{R}$ telle que $f(X,Y)$ est intégrable.*
On appelle espérance conditionnelle de $f(X,Y)$ sachant Y et on note $\mathbb{E}(f(X,Y)|Y)$ la variable aléatoire discrète

$$\boxed{\mathbb{E}(f(X,Y)|Y)=\psi(Y)\quad \text{où}\ \ \forall y\in G,\ \psi(y)=\sum_{x\in F}f(x,y)\mathbb{P}(X=x|Y=y).}$$

Lorsque X est à valeurs réelles intégrable, pour le choix $f(x,y)=x$, on obtient le

Cas particulier : $\boxed{\mathbb{E}(X|Y)=\psi(Y)\quad \text{où}\ \ \forall y\in G,\ \psi(y)=\sum_{x\in F}x\mathbb{P}(X=x|Y=y).}$

Remarque 2.5.5. – Notons que dans le cas général,

$$\sum_{y\in G}\left(\sum_{x\in F}|f(x,y)|\mathbb{P}(X=x|Y=y)\right)\mathbb{P}(Y=y)=\sum_{x\in F,y\in G}|f(x,y)|\mathbb{P}(X=x,Y=y)$$
$$=\mathbb{E}|f(X,Y)|<+\infty. \tag{2.6}$$

On en déduit que l'ensemble $A=\left\{y\in G:\sum_{x\in F}|f(x,y)|\mathbb{P}(X=x|Y=y)=+\infty\right\}$ sur lequel $\psi(y)$ n'est pas défini vérifie $\mathbb{P}(Y\in A)=\sum_{y\in G}1_A(y)\mathbb{P}(Y=y)=0$. Ainsi les variables aléatoires $\psi(Y)$ et donc $\mathbb{E}(f(X,Y)|Y)$ sont bien définies avec probabilité 1.

– Lorsque X et Y sont indépendantes, $\mathbb{E}(f(X,Y)|Y) = \psi(Y)$ où $\psi(y) = \mathbb{E}(f(X,y))$ pour tout $y \in G$.

Proposition 2.5.6. *On suppose que $f(X,Y)$ est intégrable. Pour toute fonction $g : G \to \mathbb{R}$ telle que $f(X,Y)g(Y)$ est intégrable, la variable aléatoire $\mathbb{E}(f(X,Y)|Y)g(Y)$ est intégrable et on a*

$$\mathbb{E}\left(\mathbb{E}(f(X,Y)|Y)g(Y)\right) = \mathbb{E}(f(X,Y)g(Y)).$$

Démonstration : L'intégrabilité de $\mathbb{E}(f(X,Y)|Y)g(Y)$ se démontre en remarquant que pour $y \in G$, $|\psi(y)g(y)| \le |g(y)| \sum_{x\in F} |f(x,y)|\mathbb{P}(X = x|Y = y)$ puis en raisonnant comme dans (2.6). En outre,

$$\begin{aligned}\mathbb{E}\left(\mathbb{E}(f(X,Y)|Y)g(Y)\right) &= \sum_{y\in G} g(y)\psi(y)\mathbb{P}(Y=y)\\ &= \sum_{y\in G} g(y)\left(\sum_{x\in F} f(x,y)\mathbb{P}(X=x|Y=y)\right)\mathbb{P}(Y=y)\\ &= \sum_{x\in F, y\in G} f(x,y)g(y)\mathbb{P}(X=x,Y=y) = \mathbb{E}(f(X,Y)g(Y)).\end{aligned}$$

□

Corollaire 2.5.7. *Si la variable $f(X,Y)$ est intégrable, alors l'espérance conditionnelle $\mathbb{E}(f(X,Y)|Y)$ l'est aussi et $\mathbb{E}\left(\mathbb{E}(f(X,Y)|Y)\right) = \mathbb{E}(f(X,Y))$. En outre, si $f(X,Y)$ est de carré intégrable, $\mathbb{E}(f(X,Y)|Y)$ l'est aussi et $\mathrm{Var}(\mathbb{E}(f(X,Y)|Y)) \le \mathrm{Var}(f(X,Y))$.*

Démonstration : La première assertion s'obtient en prenant $g \equiv 1$ dans l'égalité de la proposition 2.5.6. Supposons maintenant $f(X,Y)$ de carré intégrable. Par l'inégalité de Cauchy-Schwarz et comme $\sum_{x\in F} \mathbb{P}(X = x|Y = y) = 1$,

$$\left(\sum_{x\in F} f(x,y)\sqrt{\mathbb{P}(X=x|Y=y)}\times\sqrt{\mathbb{P}(X=x|Y=y)}\right)^2 \le \sum_{x\in F} f^2(x,y)\mathbb{P}(X=x|Y=y)\times 1.$$

Donc $\mathbb{E}(f(X,Y)|Y)^2 \le \mathbb{E}(f^2(X,Y)|Y)$. Comme $\mathbb{E}(f^2(X,Y)|Y)$ est intégrable d'espérance égale à $\mathbb{E}(f^2(X,Y))$, on déduit des propriétés 2.3.4 que $\mathbb{E}(f(X,Y)|Y)^2$ est intégrable et que

$$\mathbb{E}\left(\mathbb{E}(f(X,Y)|Y)^2\right) \le \mathbb{E}(f^2(X,Y)).$$

On conclut en soustrayant $\left(\mathbb{E}(\mathbb{E}(f(X,Y)|Y))\right)^2 = \left(\mathbb{E}(f(X,Y))\right)^2$ à cette inégalité. □

Exercice résolu 2.5.8. *Soient $X_1, \ldots, X_n$ des variables aléatoires I.I.D. (indépendantes et identiquement distribuées) suivant la loi de Bernoulli de paramètre $p \in]0,1[$ et $S = X_1 + \ldots + X_n$ leur somme. Pour $s \in [0,n]$, donner la loi conditionnelle de X_1 sachant $S = s$ et calculer $\mathbb{E}(X_1|S)$.*

Soit $x \in \{0,1\}$.

$$\begin{aligned}\mathbb{P}(X_1 = x|S = s) &= \frac{\mathbb{P}(X_1 = x, S = s)}{\mathbb{P}(S = s)} = \frac{\mathbb{P}(X_1 = x, X_2 + \ldots + X_n = s - x)}{\mathbb{P}(S = s)}\\ &= \frac{\mathbb{P}(X_1 = x)\mathbb{P}(X_2 + \ldots + X_n = s - x)}{\mathbb{P}(S = s)} \text{ par indépendance des } X_i.\end{aligned}$$

On a $\mathbb{P}(X_1 = x) = p^x(1-p)^{1-x}$.
La variable S suit la loi binomiale $\mathcal{B}(n,p)$ tandis que $X_2 + \ldots + X_n \sim \mathcal{B}(n-1,p)$. Donc $\mathbb{P}(S = s) = \binom{n}{s}p^s(1-p)^{n-s}$ et

$$\mathbb{P}(X_2 + \ldots + X_n = s - x) = \begin{cases} 0 \text{ si } s - x = n \text{ ou } s - x = -1, \\ \binom{n-1}{s-x}p^{s-x}(1-p)^{n-1-s+x} \text{ sinon.}\end{cases}$$

Donc

$$\mathbb{P}(X_1 = x|S = s) = \begin{cases} 0 \text{ si } s - x = n \text{ ou } s - x = -1 \\ \frac{\binom{n-1}{s-x}}{\binom{n}{s}} = \frac{(n-1)!}{n!}\frac{s!}{(s-x)!}\frac{(n-s)!}{(n-1-s+x)!} = \frac{1}{n}s^x(n-s)^{1-x} \text{ sinon}\end{cases}.$$

Ainsi pour $x \in \{0,1\}$, $\mathbb{P}(X_1 = x|S = s) = \left(\frac{s}{n}\right)^x\left(\frac{n-s}{n}\right)^{1-x}$.
On conclut que la loi conditionnelle de X_1 sachant $S = s$ est la loi de Bernoulli de paramètre s/n, ce qui est assez intuitif. Il faut remarquer qu'elle ne dépend pas du paramètre p de départ.
Enfin, comme $0 \times \mathbb{P}(X_1 = 0|S = s) + 1 \times \mathbb{P}(X_1 = 1|S = s) = s/n$, $\mathbb{E}(X_1|S) = S/n$.

2.6 Exercices

Exercice 2.6.1. Soit N une variable aléatoire qui suit la loi de Poisson de paramètre $\lambda > 0$. Calculer $\mathbb{E}\left(\frac{1}{N+1}\right)$.

Exercice 2.6.2. Soient X et Y des variables indépendantes qui suivent des lois binomiales de paramètres respectifs (n,p) et (m,p). Quelle est la loi de la somme $S = X + Y$?

Exercice 2.6.3. Dans la population, une proportion $p << 1$ d'individus est touchée par une maladie. Lors du test sanguin qui permet de détecter cette maladie, on groupe les échantillons de sang par lots de $n \geq 2$: si le test est négatif, les n patients du lot sont sains tandis que si le test est positif, on pratique n tests individuels pour déterminer quels sont les patients atteints. Le coût d'un test sera pris égal à 1.

1. Montrer que le coût moyen par individu est égal à $C(n) = 1 + \frac{1}{n} - (1-p)^n$. En faisant un développement au premier ordre en utilisant $np << 1$, trouver la valeur de n qui minimise ce coût.
2. Dans le cas où $p = 0.01$, calculer cette valeur optimale ainsi que l'économie réalisée par rapport au cas $n = 1$ où on teste le sang de chaque patient séparément.

Exercice corrigé 2.6.4. Un composant électronique a une durée de vie X qu'on mesure en nombre entier d'unités de temps. On fait l'hypothèse que, à chaque unité de temps, ce composant a une probabilité $p \in]0,1[$ de tomber en panne, de sorte que $X \sim \mathcal{G}eo(p)$. On considère un autre composant dont la durée de vie Y est indépendante de X et de même loi. On pose

$$S = \min(X,Y) \text{ et } T = |X - Y|.$$

1. Que représentent S et T ?
2. Calculer $\mathbb{P}(S = s \text{ et } T = t)$ pour $s \geq 1$ et $t \geq 0$ (distinguer $t = 0$ de $t \geq 1$).
3. En déduire les lois de S et T puis $\mathbb{E}(T)$. Quel est le nom de la loi de S ?
4. Les variables aléatoires S et T sont-elles indépendantes ?

Exercice 2.6.5. 1. Pour $n \in \mathbb{N}^*$, $0 \leq p \leq 1$ et $k \in \mathbb{N}$, on note $b_{n,p}(k)$ la probabilité pour qu'une variable qui suit la loi binomiale de paramètre (n, p) vaille k. De manière analogue, pour $\lambda > 0$ et $k \in \mathbb{N}$, on note $p_\lambda(k)$ la probabilité pour qu'une variable qui suit la loi de Poisson de paramètre λ vaille k.
Montrer que si $n \to +\infty$ et $p \to 0$ avec $np \to \lambda$ la loi de binomiale de paramètre (n, p) converge vers la loi de Poisson de paramètre λ au sens suivant :

$$\forall k \in \mathbb{N},\ \lim b_{n,p}(k) = p_\lambda(k).$$

Indication : commencer par montrer que pour $k \in \mathbb{N}$ fixé, $\lim_{n\to+\infty} \frac{n!}{(n-k)!n^k} = 1$.

2. Si X est une variable discrète distribuée suivant la loi de Poisson de paramètre λ, calculer $P(X = 1|X \geq 1)$.
Dans une ville de 200000 personnes, un voleur a laissé une empreinte digitale de type t. La police arrête un suspect dont les empreintes digitales sont de type t. Sachant que la probabilité pour qu'une personne ait des empreintes de ce type est 5.10^{-6}, est-il raisonnable de condamner le suspect sur la foi de ce seul indice ?
3. Aux jeux olympiques de Sydney, 300 médailles d'or ont été mises en jeu. En faisant l'hypothèse que “le nombre de médailles d'or remportées par un pays est proportionnel à sa population” que l'on traduit mathématiquement en supposant que ce nombre est une variable binomiale de paramètre ($n = 300, p =$ population du pays/population mondiale), calculer la probabilité pour que le nombre de médailles d'or remportées par les athlètes français soit supérieur ou égal à 10 (la population française sera prise égale à 50 millions et la population mondiale à 5 milliards). Que concluez-vous du fait que la France a remporté 13 médailles d'or ?

Exercice 2.6.6. Une urne contient n_1 boules blanches et n_2 boules noires.

1. On choisit simultanément au hasard $n \leq n_1 + n_2$ boules dans l'urne et on note N le nombre de boules blanches obtenues. Pour $k \in \{0, \ldots, n\}$, calculer $h_{n,n_1,n_2}(k) = \mathbb{P}(N = k)$.
2. On tire successivement n boules sans remise dans l'urne. Vérifier que la loi du nombre de boules blanches obtenues est la même qu'à la question précédente. Cette loi porte le nom de loi hypergéométrique de paramètre (n, n_1, n_2).
3. Déterminer la limite de $h_{n,n_1,n_2}(k)$ lorsque $\min(n_1, n_2) \to +\infty$ avec $\frac{n_1}{n_1+n_2} \to p$.

Exercice 2.6.7. On suppose que le nombre N de clients pendant une journée dans un grand magasin suit la loi de Poisson de paramètre $\lambda > 0$.
Chaque client a une probabilité p de se faire voler son portefeuille et ce indépendamment des autres clients, ce que l'on modélise à l'aide d'une suite $(X_i)_{i\geq1}$ de variables I.I.D. suivant la loi de Bernoulli de paramètre p indépendante de N : $X_i = 1$ si le ième client se fait dépouiller.

1. Exprimer le nombre V de clients volés en fonction de N et des X_i.

2. Déterminer la loi de $(V, N - V)$. En déduire la loi de V, celle de $N - V$. Ces deux variables aléatoires sont-elles indépendantes ?

Exercice 2.6.8. On considère une suite $(X_i)_{i\geq 1}$ de variables de Bernoulli de paramètre p $(0 < p < 1)$ indépendantes (schéma de Bernoulli) et on s'intéresse aux temps d'apparition

$$T_{10} = \inf\{i \geq 2,\ (X_{i-1}, X_i) = (1, 0)\} \quad \text{et} \quad T_{11} = \inf\{i \geq 2,\ (X_{i-1}, X_i) = (1, 1)\}$$

des séquences "10" et "11".

1. Soit $T_1 = \inf\{i \geq 1,\ X_i = 1\}$. Calculer $\mathbb{P}(T_1 = k, T_{10} - T_1 = l)$ et en déduire l'espérance de T_{10}.
2. Calculer $\mathbb{P}(T_{11} = 2)$, puis montrer en distinguant les cas $X_1 = 0$, $(X_1, X_2) = (1, 0)$ et $(X_1, X_2) = (1, 1)$ que $\forall k \geq 3$,

$$\mathbb{P}(T_{11} = k) = (1 - p)\mathbb{P}(T_{11} = k - 1) + p(1 - p)\mathbb{P}(T_{11} = k - 2).$$

   En déduire l'espérance de T_{11}.
3. On suppose $p = 1/2$. Comparer les espérances de T_{10} et T_{11}. En remarquant que $\{T_{11} < T_{10}\} = \{X_{T_1+1} = 1\}$ et en découpant cet événement suivant les valeurs prises par T_1, calculer $\mathbb{P}(T_{11} < T_{10})$. Commentaires ?

Exercice 2.6.9.
1. Soit X une variable aléatoire discrète qui suit la loi géométrique de paramètre p $(0 < p < 1)$.
   Montrer que X "n'a pas de mémoire", c'est-à-dire que la loi conditionnelle de $X - n_0$ sachant $X > n_0$ ne dépend pas de $n_0 \geq 0$.
2. Inversement, caractériser la loi d'une variable aléatoire à valeurs dans $\mathbb{N}^*$ qui "n'a pas de mémoire". On pourra pour cela montrer par récurrence que

$$\forall k \geq 1,\ \mathbb{P}(X \geq k) = \mathbb{P}(X \geq 2)^{k-1}.$$

Exercice 2.6.10. Est-il possible de piper deux dés à 6 faces indépendants de façon à ce que leur somme soit uniformément répartie sur $\{2, \ldots, 12\}$?
Indication : on pourra commencer par montrer que si cela est possible, alors la fonction génératrice associée au résultat de chacun des dés admet au moins deux racines réelles.

Exercice 2.6.11. Soit X et Y deux variables aléatoires indépendantes distribuées respectivement suivant les lois de Poisson de paramètre $\lambda > 0$ et de paramètre $\mu > 0$. On note $S = X + Y$.

1. Déterminer la loi de S.
2. Pour tout $s \in \mathbb{N}$ déterminer la loi conditionnelle de X sachant $S = s$.
3. Donner $\mathbb{E}(X|S)$.

Problème corrigé 2.6.12. Soit $A_1, \cdots, A_n$ une suite d'événements.

1. Montrer que $1_{A_1 \cup \ldots \cup A_n} = 1 - \prod_{i=1}^n (1 - 1_{A_i})$.
2. En déduire la formule de Poincaré :

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Application : Une personne écrit à n correspondants des lettres personnelles, met chaque lettre dans une enveloppe, ferme les enveloppes, puis écrit les adresses au hasard. On s'intéresse au nombre X_n de lettres qui parviennent à leur destinataire. Pour $1 \leq i \leq n$, on note A_i l'événement : la lettre i arrive à son destinataire.

3. Préciser l'espace fini Ω choisi pour modéliser le problème ainsi que la probabilité $\mathbb{P}$ dont il est muni.
   Pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, calculer $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$.
4. En déduire $\mathbb{P}(X_n > 0)$ puis $\mathbb{P}(X_n = 0)$. Quel est le nombre de permutations σ de $\{1, \ldots, n\}$ sans point fixe i.e. telles que $\forall i \in \{1, \ldots, n\}$, $\sigma(i) \neq i$?
5. En déduire que pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, $\mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \dfrac{(n-k)!}{n!} \displaystyle\sum_{l=0}^{n-k} \dfrac{(-1)^l}{l!}$ puis donner la loi de X_n.
6. Vérifier que $k\mathbb{P}(X_n = k) = \mathbb{P}(X_n = k-1) - \dfrac{(-1)^{n-(k-1)}}{(k-1)!(n-(k-1))!}$ pour k dans $\{1, \ldots, n\}$. En déduire que $\mathbb{E}(X_n) = 1$. Retrouver ce résultat en remarquant que $X_n = \sum_{i=1}^n 1_{A_i}$.
7. Soit X une variable de loi de Poisson de paramètre 1. Vérifier que pour tout entier k, $\lim_{n\to+\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$.

Problème corrigé 2.6.13. Pour fidéliser ses clients, une marque de chocolat place dans chaque tablette une pièce d'un puzzle. Le puzzle est composé de n morceaux distincts. Le morceau qui se trouve dans une tablette est supposé suivre une loi uniforme sur les n morceaux possibles. Le client est supposé acheter les différentes tablettes au hasard. On s'intéresse au nombre N d'achats à réaliser pour obtenir l'ensemble des morceaux du puzzle.

Événements

1. Pour $0 \leq m \leq n-1$, que vaut $\mathbb{P}(N > m)$?
2. On suppose maintenant $m \geq n$.
   (a) Pour $1 \leq k \leq n$, on désigne par A_k^m l'événement “la pièce k n'a toujours pas été obtenue au bout de m achats”. Calculer pour $k_1, \ldots, k_r \in \{1, \ldots, n\}$ des entiers distincts, la probabilité $\mathbb{P}(A_{k_1}^m \cap \ldots \cap A_{k_r}^m)$ (préciser l'espace fini Ω choisi pour modéliser les pièces obtenues dans les m tablettes et la probabilité dont il est muni).
   (b) Montrer que $\{N > m\} = \displaystyle\bigcup_{k=1}^n A_k^m$.
   (c) En déduire que
   $$\mathbb{P}(N > m) = \sum_{r=1}^n (-1)^{r+1} \binom{n}{r} \left(1 - \frac{r}{n}\right)^m.$$
3. En remarquant que pour tout $l \in \mathbb{N}$, $l = \displaystyle\sum_{m=0}^{+\infty} 1_{\{l>m\}}$, montrer que
   $$\mathbb{E}(N) = \sum_{m=0}^{+\infty} \mathbb{P}(N > m).$$
   En déduire une expression pour cette espérance.

4. Donner la loi de N.

Variables aléatoires

La première pièce ayant été découverte dans la première tablette ($N_1 = 1$), on note N_2 le nombre d'achats supplémentaires nécessaires à l'obtention d'une seconde pièce, puis N_3 le nombre d'achats supplémentaires nécessaires à l'obtention d'une troisième, et ainsi de suite.

1. Exprimer N en fonction de $N_1, N_2, \ldots, N_n$.
2. On note X_i le numéro de la i-ème pièce de puzzle obtenue. Pour $m \geq 1$, montrer que
$$\{N_2 = m\} = \bigcup_{i=1}^{n} \{X_1 = i, X_2 = i, \ldots, X_m = i, X_{m+1} \neq i\}.$$
3. Par hypothèse, les X_i sont des variables aléatoires indépendantes, de loi uniforme sur $\{1, \ldots, n\}$. En déduire la loi de N_2.
4. Pour $m_2, m_3 \geq 1$, exprimer l'événement $\{N_2 = m_2, N_3 = m_3\}$ en fonction des X_i. Calculer sa probabilité et en déduire l'indépendance de N_2 et N_3.
5. Justifier intuitivement que les N_i sont des variables de loi géométrique et donner leurs paramètres.
6. En déduire une autre expression de l'espérance de N.
7. Conclure que
$$\sum_{r=1}^{n-1} (-1)^{r+1} \binom{n}{r} \frac{1}{r} \left(1 - \frac{r}{n}\right)^n = \sum_{i=2}^{n} \frac{1}{i}.$$

2.7 Résumé

- **Variable aléatoire discrète** $X : \Omega \to F$ où F est dénombrable.
  La famille $(\mathbb{P}(X = x))_{x\in F}$ s'appelle la **loi** de X.

- Soit X et Y deux variables discrètes à valeurs respectives dans F et G.
  - **Loi marginale :** $\forall x \in F,\ \mathbb{P}(X = x) = \sum_{y\in G} \mathbb{P}(X = x, Y = y)$.
  - **Loi conditionnelle** de X sachant $Y = y$:
  $$\left(\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}\right)_{x\in F}$$
  - **Indépendance** si $\forall x \in F,\ \forall y \in G,\ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$
  $\iff$ la loi conditionnelle de X sachant $Y = y$ ne dépend pas de $y \in G$.

- **Espérance et variance** $X : \Omega \to F \subset \mathbb{R}$ est dite
  - intégrable si $\sum_{x\in F} |x|\mathbb{P}(X = x) < +\infty$ et alors $\boxed{\mathbb{E}(X) = \sum_{x\in F} x\mathbb{P}(X = x).}$
  - de carré intégrable si $\sum_{x\in F} x^2\mathbb{P}(X = x) < +\infty$ et alors
  $$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

  Propriétés :
  - Pour tout événement A, $\boxed{\mathbb{P}(A) = \mathbb{E}(1_A).}$
  - Linéarité : $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$.
  - Croissance : $\mathbb{P}(X \geq Y) = 1 \Longrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$.
  - X et Y indépendantes $\Longrightarrow \mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$.
  - $X_1, \ldots, X_n$ indépendantes $\Longrightarrow \mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i)$.

- **Espérance de** $f(X)$ **:** $f(X)$ est intégrable ssi $\sum_{x\in F} |f(x)|\mathbb{P}(X = x) < +\infty$ et alors
$$\boxed{\mathbb{E}(f(X)) = \sum_{x\in F} f(x)\mathbb{P}(X = x).}$$

- **Fonction génératrice de** $X : \Omega \to \mathbb{N}$ **:** $s \in [-1, 1] \to g_X(s) = \mathbb{E}(s^X)$ caractérise la loi de X. Si $Y : \Omega \to \mathbb{N}$ est indépendante de X, $g_{X+Y} = g_X \times g_Y$.

- **Espérance conditionnelle de** $f(X, Y)$ **sachant** Y **(où** $f(X, Y)$ **intégrable) :**
$$\boxed{\mathbb{E}(f(X, Y)|Y) = \psi(Y) \quad \text{où} \quad \forall y \in G,\ \psi(y) = \sum_{x\in F} f(x, y)\mathbb{P}(X = x|Y = y).}$$

  $\forall g : G \to \mathbb{R}$, t.q. $f(X, Y)g(Y)$ intég, $\mathbb{E}\left(\mathbb{E}(f(X, Y)|Y)g(Y)\right) = \mathbb{E}(f(X, Y)g(Y))$.

- **Lois discrètes usuelles :**

Nom	**Loi**	$\mathbb{E}(X)$	$\mathrm{Var}(X)$	$g_X(s)$
Bernoulli $\mathcal{B}(p)$	$\mathbb{P}(X=1) = p = 1 - \mathbb{P}(X=0)$	p	$p(1-p)$	$1-p+ps$
binomiale $\mathcal{B}(n,p)$	$\forall 0 \leq k \leq n,\ \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(1-p+ps)^n$
Poisson $\mathcal{P}(\lambda)$	$\forall n \in \mathbb{N},\ \mathbb{P}(X=n) = \exp(-\lambda)\frac{\lambda^n}{n!}$	λ	λ	$e^{\lambda(s-1)}$
géométrique $\mathcal{G}eo(p)$	$\forall n \in \mathbb{N}^*,\ \mathbb{P}(X=n) = p(1-p)^{n-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{ps}{1-(1-p)s}$

Chapitre 3

Variables aléatoires à densité

Si on s'intéresse à un pneu de vélo crevé, en prenant la valve comme origine des angles, intuitivement, la probabilité pour que l'abscisse angulaire du point de crevaison se trouve dans l'intervalle $[\theta_1, \theta_2]$ où $0 \leq \theta_1 \leq \theta_2 \leq 2\pi$ est égale à $(\theta_2 - \theta_1)/2\pi$. On ne peut pas modéliser cette abscisse angulaire à l'aide d'une variable aléatoire discrète Θ à valeurs dans $F \subset [0, 2\pi]$ [1]. En revanche, le cadre des variables aléatoires à densité est tout à fait adapté à la situation que nous venons de décrire.

Dans les calculs relatifs à ces variables aléatoires, les sommes manipulées dans le cas des variables aléatoires discrètes sont remplacées par des intégrales. Et lorsque l'on s'intéresse à un vecteur aléatoire à densité en dimension n, il est nécessaire de manipuler des intégrales multiples posées sur $\mathbb{R}^n$.

3.1 Manipulation d'intégrales multiples

3.1.1 Théorème de Fubini

Théorème 3.1.1. *Soit* $f : \mathbb{R}^2 \to \mathbb{R}$.

- *Si* f *est* **positive** ,

$$\int_{\mathbb{R}^2} f(x,y)dxdy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dx \right) dy.$$

  Cela signifie que les trois termes sont soit simultanément finis et égaux soit simultanément égaux à $+\infty$.
- *Si* f *est* **intégrable** *au sens où* $\int_{\mathbb{R}^2} |f(x,y)|dxdy < +\infty$ *alors l'égalité ci-dessus reste vraie.*

Remarque 3.1.2.
- Pour vérifier l'intégrabilité de f, on peut bien sûr appliquer le théorème à la fonction positive $|f|$.
- Ce résultat se généralise en dimension supérieure i.e. pour le calcul de l'intégrale d'une fonction de n variables positive ou intégrable, l'ordre dans lequel on effectue les intégrations sur chacune de ces variables est sans importance.

1. Soit en effet θ t.q. $\mathbb{P}(\Theta = \theta) > 0$. Pour $\varepsilon \leq \pi\mathbb{P}(\Theta = \theta)/2$, on a $\mathbb{P}(\Theta \in [\theta - \varepsilon, \theta + \varepsilon]) \geq \mathbb{P}(\Theta = \theta) > 2\varepsilon/2\pi$.

Exemple 3.1.3. Soit $f : [0,+\infty[\to [0,+\infty[$.

$$\begin{aligned}\int_{[0,+\infty[\times[0,+\infty[} \frac{f(x+y)}{x+y}\,dxdy &= \int_0^{+\infty}\left(\int_0^{+\infty}\frac{f(x+y)}{x+y}dy\right)dx\\ &= \int_0^{+\infty}\left(\int_x^{+\infty}\frac{f(z)}{z}dz\right)dx\\ &= \int_0^{+\infty}\left(\int_0^{+\infty}1_{\{z\geq x\}}\frac{f(z)}{z}dz\right)dx\\ &= \int_0^{+\infty}\frac{f(z)}{z}\left(\int_0^{+\infty}1_{\{z\geq x\}}dx\right)dz\\ &= \int_0^{+\infty}f(z)dz.\end{aligned}$$

Notons au passage l'intérêt d'utiliser la fonction indicatrice $1_{\{z\geq x\}}$ pour éviter de faire des erreurs sur le domaine d'intégration lors de l'échange des intégrales en x et en z.

3.1.2 Changement de variables

Soit φ une bijection continuement différentiable ainsi que son inverse φ^{-1} d'un ouvert O de $\mathbb{R}^d$ sur un ouvert O' de $\mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ bornée et $g : \mathbb{R}^d \to \mathbb{R}$ intégrable.

$$\boxed{\int_O f(\varphi(x))g(x)\,dx = \int_{O'} f(y)g(\varphi^{-1}(y))\,|\mathrm{Jac}\,\varphi^{-1}(y)|\,dy,}$$

où

$$\mathrm{Jac}\,\varphi^{-1}(y) = \mathrm{Det}\left(\frac{\partial(\varphi^{-1})_i}{\partial y_j}(y),\ 1\leq i,j\leq d\right) = 1\Big/\mathrm{Det}\left(\frac{\partial\varphi_i}{\partial x_j}(\varphi^{-1}(y)),\ 1\leq i,j\leq d\right).$$

Remarque 3.1.4. En dimension $d=1$, lorsque $O =]a,b[$ avec $a<b$ et φ est strictement décroissante alors $O' =]\varphi(b),\varphi(a)[$ et φ^{-1} est décroissante si bien que $|\mathrm{Jac}\,\varphi^{-1}(y)| = -(\varphi^{-1})'(y)$. Ainsi le second membre s'écrit

$$\int_{\varphi(b)}^{\varphi(a)} f(y)\,g(\varphi^{-1}(y))\,(-(\varphi^{-1})'(y))dy = \int_{\varphi(a)}^{\varphi(b)} f(y)\,g(\varphi^{-1}(y))\,(\varphi^{-1})'(y)dy,$$

et on retrouve bien le changement de variables usuel dans $\mathbb{R}$.

Exercice résolu 3.1.5. *Calculer* $I = \int_{\mathbb{R}} e^{-\frac{x^2}{2}}dx$.

On va utiliser la formule de changement de variables pour calculer

$$I^2 = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}}\,dx\,dy.$$

On utilise pour cela le changement de variables

$$\varphi : (x,y)\in O = \mathbb{R}^2\setminus\{(x,0) : x\leq 0\} \to (\rho,\theta) = \left(x^2+y^2, 2\arctan\left(\frac{y}{x+\sqrt{x^2+y^2}}\right)\right).$$

Notons que si θ désigne l'angle polaire de (x, y),

$$\frac{y/\sqrt{x^2+y^2}}{1+x/\sqrt{x^2+y^2}} = \frac{\sin\theta}{1+\cos\theta} = \frac{2\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{1+\cos^2\frac{\theta}{2}-\sin^2\frac{\theta}{2}} = \tan\frac{\theta}{2},$$

d'où $2\arctan\left(\frac{y}{x+\sqrt{x^2+y^2}}\right) = \theta$. Ainsi φ est une bijection C^1 ainsi que son inverse de O sur $O' =]0,+\infty[\times]-\pi,\pi[$ et $\varphi^{-1}(\rho,\theta) = (\sqrt{\rho}\ \cos\theta, \sqrt{\rho}\ \sin\theta)$. Pour ne pas se tromper entre le jacobien de φ et de φ^{-1} dans le changement d'élément différentiel, il est éclairant d'utiliser la notation suivante pour la matrice jacobienne

$$\frac{D(x,y)}{D(\rho,\theta)} = \begin{pmatrix} \cos(\theta)/(2\sqrt{\rho}) & -\sqrt{\rho}\ \sin\theta \\ \sin(\theta)/(2\sqrt{\rho}) & \sqrt{\rho}\ \cos\theta \end{pmatrix}.$$

En prenant la valeur absolue du déterminant du second membre on obtient formellement que $\frac{dx\,dy}{d\rho\,d\theta} = \frac{\cos^2\theta+\sin^2\theta}{2} = \frac{1}{2}$ i.e. qu'il faut remplacer $dx\ dy$ par $\frac{1}{2}\ d\rho\ d\theta$.
Comme le volume (ou plutôt la surface) dans $\mathbb{R}^2$ de la demi-droite $\{(x,0) : x \leq 0\}$ est nul, on ne change rien à la valeur de I^2 en restreignant l'intégration à O : $I^2 = \int_O e^{-\frac{x^2+y^2}{2}}\ dx\ dy$.
Par la formule de changement de variable on en déduit que

$$I^2 = \int_{]0,+\infty[\times]-\pi,\pi[} \exp\left(-\frac{\rho}{2}\right)\ \frac{1}{2}\ d\rho\ d\theta = 2\pi.$$

Comme I est l'intégrale d'une fonction positive, on conclut que $I = \sqrt{I^2} = \sqrt{2\pi}$.

En général, dans les problèmes de probabilités, on connaît O et φ et on souhaite transformer une intégrale comme celle du premier membre en une intégrale comme celle du second. Il faut faire attention aux difficultés suivantes :

- la fonction φ n'est pas nécessairement injective sur le domaine O de départ (ex : $O = \mathbb{R}$, $\varphi(x) = x^2$). Pour surmonter cette difficulté, on peut essayer de découper O en sous-domaines sur lesquels φ est injective.
- lorsque φ est injective sur O, il faut bien raisonner par conditions nécessaires et suffisantes pour obtenir le domaine image O' et ne pas se contenter de conditions nécessaires.

Exemple 3.1.6. Si $O =]0,+\infty[\times]0,+\infty[$ et $\varphi(x,y) = (x+y, x-y)$. Un raisonnement hâtif pourrait laisser penser que $O' = \varphi(O)$ est égal à $]0,+\infty[\times\mathbb{R}$ ce qui est faux. Pour déterminer O', on commence par déterminer φ^{-1} en résolvant le système

$$\varphi^{-1}(z,w) = (x,y) \Leftrightarrow (z,w) = \varphi(x,y) \Leftrightarrow \begin{cases} z = x+y \\ w = x-y \end{cases} \Leftrightarrow \begin{cases} x = \frac{z+w}{2} \\ y = \frac{z-w}{2} \end{cases}$$

Ainsi $\varphi^{-1}(z,w) = (\frac{z+w}{2}, \frac{z-w}{2})$. Ensuite on raisonne par conditions nécessaires et suffisantes :

$$(z,w) \in O' \Leftrightarrow \varphi^{-1}(z,w) \in O \Leftrightarrow \begin{cases} \frac{z+w}{2} > 0 \\ \frac{z-w}{2} > 0 \end{cases} \Leftrightarrow z > |w|.$$

Ainsi $O' = \{(z,w) \in \mathbb{R}^2 :\ z > |w|\}$.

Remarque 3.1.7. Pour le calcul de certaines intégrales en dimension d petite, le recours au théorème de Fubini et à des changements de variable en dimension 1 peut constituer une alternative à un changement de variables dans $\mathbb{R}^d$. Ainsi pour $f : \mathbb{R}^2 \to \mathbb{R}$ positive,

$$\begin{aligned}
\int_0^{+\infty}\int_0^{+\infty} f(x+y,x-y)dxdy &= \int_0^{+\infty}\left(\int_0^{+\infty} f(z,z-2y)1_{\{z\geq y\}}dz\right)dy \text{ où } z=x+y,\\
&= \int_0^{+\infty}\left(\int_0^{+\infty} f(z,z-2y)1_{\{z\geq y\}}dy\right)dz \text{ par Fubini,}\\
&= \int_0^{+\infty}\left(\int_0^{z} f(z,z-2y)dy\right)dz\\
&= \int_0^{+\infty}\left(\frac{1}{2}\int_{-z}^{z} f(z,w)dw\right)dz \text{ où } w=z-2y,\\
&= \frac{1}{2}\int_{\mathbb{R}^2} 1_{\{z>|w|\}} f(z,w)dzdw.
\end{aligned}$$

Ce résultat s'obtient également par le changement de variables $\varphi(x,y) = (x+y, x-y)$: le domaine image O' a été déterminé dans l'exemple 3.1.6 ci-dessus et un rapide calcul permet de vérifier que $|\text{Jac}\,\varphi^{-1}| = 1/2$.
Mais dès que la dimension d devient grande ou la transformation φ compliquée, la formule de changement de variables ci-dessus devient un outil incontournable.

3.2 Variables aléatoires réelles à densité

3.2.1 Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Définition 3.2.1. *On dit que la variable aléatoire $X : \Omega \to \mathbb{R}$ possède la densité $p : \mathbb{R} \to \mathbb{R}$ si*

$$\boxed{\forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a < X \leq b) = \int_a^b p(x)dx.} \tag{3.1}$$

où $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

Remarque 3.2.2.
- En toute rigueur pour pouvoir considérer $\mathbb{P}(a < X \leq b)$ pour $a < b \in \bar{\mathbb{R}}$, il faut demander que $\forall a < b \in \bar{\mathbb{R}}$, $\{a < X \leq b\} \in \mathcal{A}$.
- La positivité de $\mathbb{P}$ entraîne qu'une densité de probabilité p est une fonction positive. En outre $\mathbb{P}(X \in \mathbb{R}) = \int_{\mathbb{R}} p(x)dx$ implique $\int_{\mathbb{R}} p(x)dx = 1$. Ainsi une densité de probabilité est une fonction positive d'intégrale 1.
- Pour $x \in \mathbb{R}$ et $n \in \mathbb{N}^*$

$$\mathbb{P}(X = x) \leq \mathbb{P}\left(x - \frac{1}{n} < X \leq x\right) = \int_{x-\frac{1}{n}}^{x} p(x)dx.$$

En prenant la limite $n \to +\infty$ dans le membre de droite, on obtient $\mathbb{P}(X = x) = 0$.

On en déduit que (3.1) est équivalente à

$$\begin{aligned} & \forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a < X < b) = \int_a^b p(x)dx \\ \iff & \forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a \leq X < b) = \int_a^b p(x)dx \\ \iff & \forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a \leq X \leq b) = \int_a^b p(x)dx. \end{aligned}$$

Le fait que $\forall x \in \mathbb{R},\ \mathbb{P}(X = x) = 0$ entraîne aussi par σ-additivité que pour tout sous-ensemble F de $\mathbb{R}$ dénombrable, $\mathbb{P}(X \in F) = \sum_{x \in F} \mathbb{P}(X = x) = 0$, ce qui montre la différence de nature entre variables aléatoires discrètes et variables aléatoires à densité.

– D'un point de vue infinitésimal, on a $\mathbb{P}(X \in [x, x + dx]) \simeq p(x)dx$.

3.2.2 Densités réelles usuelles

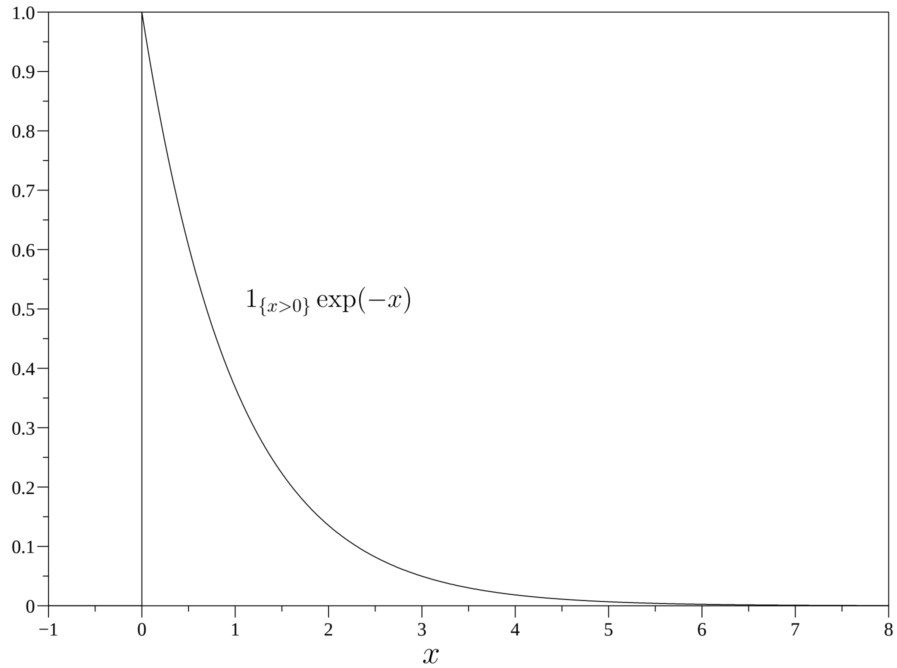


FIGURE 3.1 – Densité de la loi exponentielle de paramètre 1 $\mathcal{E}(1)$.

On dit que X suit la

- **loi uniforme sur** $[a, b]$ **où** $a < b$ et on note $X \sim \mathcal{U}[a, b]$, si X possède la densité

$$\frac{1}{b-a} 1_{[a,b]}(x).$$

- **loi exponentielle de paramètre** $\lambda > 0$ et on note $X \sim \mathcal{E}(\lambda)$ si X possède la densité
$$\lambda e^{-\lambda x} 1_{\{x>0\}}.$$
La densité exponentielle de paramètre 1 est représentée sur la figure 3.1.

- **loi gaussienne (ou normale) de paramètres** $\mu \in \mathbb{R}$ **et** $\sigma^2 > 0$ et on note $X \sim \mathcal{N}_1(\mu, \sigma^2)$ si X possède la densité
$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$
Dans le cas où $\mu = 0$ et $\sigma^2 = 1$, on dit aussi que X suit la loi normale centrée réduite. La densité normale centrée réduite est représentée sur la figure 3.2.

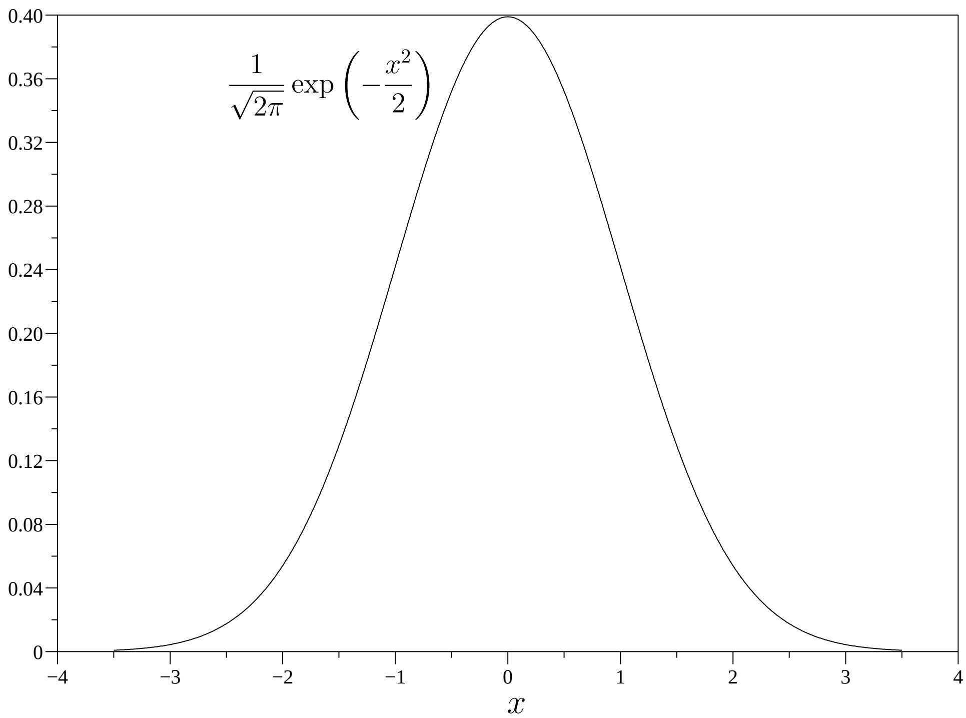


FIGURE 3.2 – Densité de la loi normale centrée réduite $\mathcal{N}(0, 1)$.

- **loi de Cauchy de paramètre** $a > 0$ et on note $X \sim \mathcal{C}(a)$ si X possède la densité
$$\frac{a}{\pi} \frac{1}{x^2 + a^2}.$$
La densité de la loi de Cauchy de paramètre 1 est représentée sur la figure 3.3.

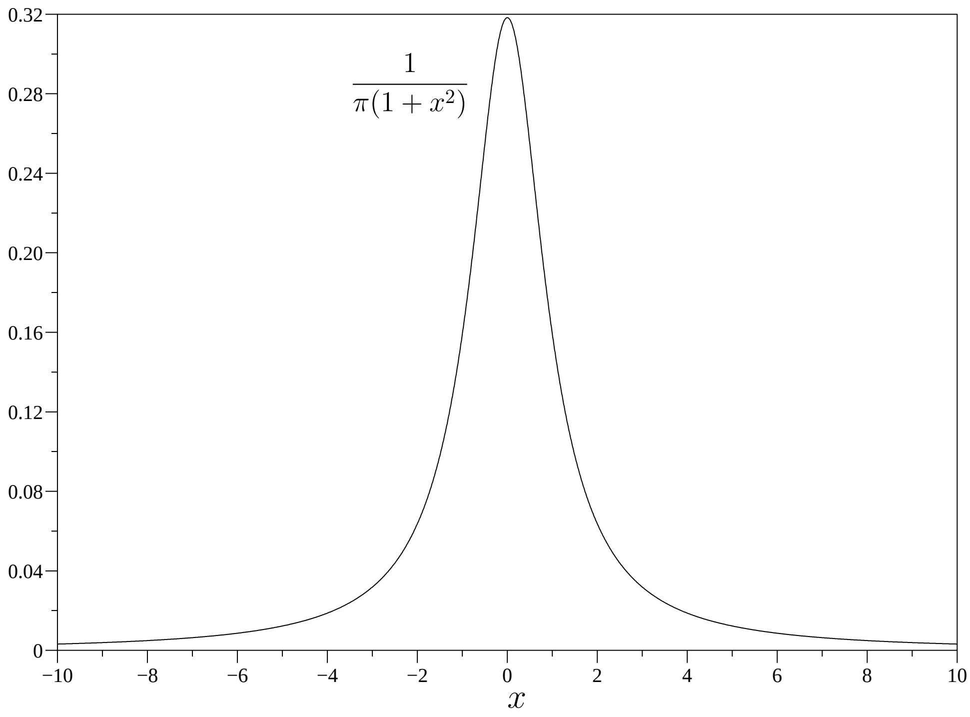


FIGURE 3.3 – Densité de la loi de Cauchy de paramètre 1 $\mathcal{C}(1)$.

3.2.3 Espérance, variance

Définition 3.2.3. *La variable aléatoire $X : \Omega \to \mathbb{R}$ qui possède la densité p est dite*

– *intégrable si $\int_{\mathbb{R}} |x|p(x)dx < +\infty$ et alors on définit son espérance par*

$$\boxed{\mathbb{E}(X) = \int_{\mathbb{R}} x\ p(x)\ dx.}$$

– *de carré intégrable si $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2\ p(x)\ dx < +\infty$ et alors on définit sa variance par*

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

Les propriétés de l'espérance et de la variance sont les mêmes que dans le cas des variables aléatoires discrètes :

Propriétés 3.2.4. *1. L'espérance d'une variable aléatoire X qui possède une densité ne dépend que de cette densité.*

2. Linéarité : $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$.

3. C.S. d'intégrabilité : si $\mathbb{P}(|X| \leq Y) = 1$ et Y est intégrable alors X l'est aussi.

4. Croissance : si X et Y sont intégrables, $\mathbb{P}(X \geq Y) = 1 \Longrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$.

Exercice 3.2.5. Calculer l'espérance et la variance d'une variable uniforme sur $[a, b]$, d'une variable exponentielle de paramètre λ, d'une variable de Cauchy de paramètre a et d'une variable normale centrée réduite.

3.2.4 Fonction de répartition

Définition 3.2.6. *Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle (qui ne possède pas nécessairement une densité). On appelle fonction de répartition de X la fonction $F_X : x \in \mathbb{R} \to \mathbb{P}(X \leq x)$.*

Exemple 3.2.7.
– Si $X \sim \mathcal{B}(p)$, $F_X(x) = (1-p)1_{\{x\geq 0\}} + p1_{\{x \geq 1\}}$.
– Si $T \sim \mathcal{E}(\lambda)$, $F_T(t) = 1_{\{t\geq 0\}}(1 - \exp(-\lambda t))$.

Proposition 3.2.8. *Si la fonction de répartition F_X de la variable aléatoire réelle X est globalement continue et C^1 par morceaux (au sens où il existe un nombre fini de points $x_1 < x_2 < \ldots < x_n$ t.q. F_X est C^1 sur $]-\infty, x_1[$, $]x_1, x_2[$,..., $]x_{n-1}, x_n[$, $]x_n, +\infty[$) alors X possède la densité F'_X.*

Démonstration : Soit $a < b \in \bar{R}$. Comme $\{X \leq b\}$ est l'union disjointe de $\{X \leq a\}$ et de $\{a < X \leq b\}$,

$$\begin{aligned}\mathbb{P}(a < X \leq b) &= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)\\ &= F_X(b) - F_X(a)\\ &= \int_a^b F'_X(x)\,dx.\end{aligned}$$

D'après la définition 3.2.1, on conclut que X possède la densité $F'_X(x)$. □

Exercice 3.2.9. On considère un système constitué de n composants. On suppose que les durées de vie des composants sont des variables exponentielles $T_1, \ldots, T_n$ de paramètres respectifs $\lambda_1, \ldots, \lambda_n > 0$ et qu'elles sont indépendantes ce qui implique en particulier que $\forall (t_1, \ldots, t_n) \in \mathbb{R}^n$, les événements $\{T_1 \leq t_1\}, \ldots, \{T_n \leq t_n\}$ sont indépendants ainsi que les événements $\{T_1 > t_1\}, \ldots, \{T_n > t_n\}$.

1. On suppose que le système est en parallèle i.e. qu'il fonctionne lorsqu'un au moins des composants fonctionne. Exprimer sa durée de vie T en fonction des T_i. Déterminer la fonction de répartition de T et en déduire sa loi.
2. Même question dans le cas où le système est en série i.e. où il fonctionne seulement lorsque tous les composants fonctionnent.

3.3 Vecteurs aléatoires à densité

3.3.1 Définition

Définition 3.3.1. *On dit que le vecteur aléatoire $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ possède la densité $p : \mathbb{R}^d \to \mathbb{R}$ si*

$$\boxed{\forall O \text{ ouvert de } \mathbb{R}^d,\ \mathbb{P}(X \in O) = \int_O p(x)\,dx = \int_{\mathbb{R}^d} 1_O(x_1, \ldots, x_d)\, p(x_1, \ldots, x_d)\, dx_1 \ldots dx_d.}$$

Remarque 3.3.2. – En toute rigueur pour pouvoir considérer $\mathbb{P}(X \in O)$ pour O ouvert, il faut demander que $\{X \in O\}$ soit dans la tribu $\mathcal{A}$.
– Une densité de probabilité p sur $\mathbb{R}^d$ est une fonction positive d'intégrale 1.

Exemple 3.3.3. Si $p_1, \ldots, p_d$ sont des densités de probabilité sur $\mathbb{R}$ alors $p(x) = p_1(x_1) \times p_2(x_2) \times \ldots \times p_d(x_d)$ est une densité de probabilité sur $\mathbb{R}^d$.

La caractérisation suivante par la *méthode de la fonction muette* est très utile en pratique :

Théorème 3.3.4. *Le vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ possède la densité p si et seulement si*

$$\boxed{\forall f : \mathbb{R}^d \to \mathbb{R} \text{ bornée}, \ \mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\, p(x)\, dx.}$$

Il est possible et souvent utile de généraliser la condition nécessaire à des fonctions non bornées.

Remarque 3.3.5. Soit $X : \Omega \to \mathbb{R}^d$ de densité p et $f : \mathbb{R}^d \to \mathbb{R}$. Alors $f(X)$ est intégrable si et seulement si $\int_{\mathbb{R}^d} |f(x)|\, p(x)\, dx < +\infty$ et alors $\mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\, p(x)\, dx$.

3.3.2 Densité marginale

Soit $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ un vecteur aléatoire de densité p et $k < d$. Si O_k est un ouvert de $\mathbb{R}^k$,

$$\begin{aligned}\mathbb{P}((X_1, \ldots, X_k) \in O_k) &= \mathbb{P}\left(X \in O_k \times \mathbb{R}^{d-k}\right) \\ &= \int_{O_k \times \mathbb{R}^{d-k}} p(x)\, dx \\ &= \int_{O_k} \left(\int_{\mathbb{R}^{d-k}} p(x_1, \ldots, x_d)\, dx_{k+1} \ldots dx_d \right) dx_1 \ldots dx_k.\end{aligned}$$

On en déduit que le sous-vecteur $(X_1, \ldots, X_k)$ possède la densité

$$q(x_1, \ldots, x_k) = \int_{\mathbb{R}^{d-k}} p(x_1, \ldots, x_d)\, dx_{k+1} \ldots dx_d.$$

Proposition 3.3.6. *Soit X un vecteur aléatoire qui possède une densité. Alors tout sous-vecteur Y possède la densité* **marginale** *obtenue en intégrant celle de X sur les composantes qui ne figurent pas dans Y.*

3.3.3 Changement de variables

Avant d'énoncer un résultat abstrait, nous allons illustrer cette technique sur un exemple particulier. En effet, dans la pratique, pour éviter les erreurs, il vaut mieux reprendre la démarche explicitée sur cet exemple que d'appliquer le résultat abstrait.

Exercice résolu 3.3.7. *Soit (X, Y) un couple aléatoire de densité $\lambda^2 \exp(-\lambda(x + y))1_{\{x\geq 0\}}1_{\{y\geq 0\}}$. Quelle est la loi de $(Z, W) = (X + Y, X - Y)$?*

Pour cela, on utilise la *méthode de la fonction muette* i.e. on se donne une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ bornée et on cherche à mettre $\mathbb{E}(f(Z, W))$ sous la forme $\int_{\mathbb{R}^2} f(z, w)p(z, w)\,dz\,dw$, ce qui permettra de conclure par la condition suffisante du théorème 3.3.4 que (Z, W) possède la densité p. Mais bien sûr, il faut utiliser ce que l'on connaît c'est à dire la loi de (X, Y) :

$$\mathbb{E}(f(Z, W)) = \mathbb{E}(f(X + Y, X - Y)).$$

Soit $\varphi : (x, y) \in \mathbb{R}^2 \to (x+y, x-y) \in \mathbb{R}^2$. La fonction $g(x, y) = f\circ\varphi(x, y) = f(x+y, x-y)$ est une fonction bornée sur $\mathbb{R}^2$. Donc d'après la condition nécessaire du théorème 3.3.4,

$$\mathbb{E}(g(X, Y)) = \int_{\mathbb{R}^2} g(x, y)\lambda^2 \exp(-\lambda(x + y))1_{\{x\geq 0\}}1_{\{y\geq 0\}}\,dx\,dy,$$

égalité qui se récrit en explicitant g

$$\mathbb{E}(f(X + Y, X - Y)) = \int_{\mathbb{R}^2} f(x + y, x - y)\lambda^2 \exp(-\lambda(x + y))1_{\{x\geq 0\}}1_{\{y\geq 0\}}\,dx\,dy.$$

Ainsi

$$\mathbb{E}(f(Z, W)) = \int_0^{+\infty}\int_0^{+\infty} f(x + y, x - y)\lambda^2 \exp(-\lambda(x + y))dxdy.$$

Pour écrire le second membre comme une intégrale de $f(z, w)$ contre une fonction des variables (z, w) le bon outil est la formule de changement de variables vue au début de ce chapitre. La fonction φ est une bijection C^1 ainsi que son inverse $(x, y) = \varphi^{-1}(z, w) = \left(\frac{z+w}{2}, \frac{z-w}{2}\right)$ de $O =]0, +\infty[\times]0, +\infty[$ sur $O' = \{(z, w) \in \mathbb{R}^2 : \ z > |w|\}$ (voir exemple 3.1.6). Par ailleurs,

$$\frac{D(x, y)}{D(z, w)} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

En prenant la valeur absolue du déterminant du second membre on obtient formellement que $\frac{dx\,dy}{dz\,dw} = \frac{1}{2}$ i.e. qu'il faut remplacer $dx\,dy$ par $\frac{1}{2}dz\,dw$. Ainsi

$$\begin{aligned}\mathbb{E}(f(Z, W)) &= \int_{\{(z,w):z>|w|\}} f(z, w)\lambda^2 \exp\left(-\lambda\left(\frac{z + w}{2} + \frac{z - w}{2}\right)\right)\frac{1}{2}dzdw \\ &= \int_{\mathbb{R}^2} f(z, w)\frac{\lambda^2}{2}\exp(-\lambda z)1_{\{z>|w|\}}dzdw.\end{aligned}$$

On conclut par la condition suffisante du théorème 3.3.4 que (Z, W) possède la densité $\frac{\lambda^2}{2}\exp(-\lambda z)1_{\{z>|w|\}}$.
Par la formule des densité marginales, on en déduit que Z possède la densité

$$\int_{\mathbb{R}} \frac{\lambda^2}{2}\exp(-\lambda z)1_{\{z>|w|\}}dw = 1_{\{z>0\}}\lambda^2 z\exp(-\lambda z),$$

et W la densité

$$\int_{\mathbb{R}} \frac{\lambda^2}{2}\exp(-\lambda z)1_{\{z>|w|\}}dz = \frac{\lambda}{2}\exp(-\lambda|w|).$$

Le résultat abstrait qui se démontre en reprenant la démarche ci-dessus est le suivant :

Proposition 3.3.8. *Soit $X : \Omega \to \mathbb{R}^d$ qui possède la densité $p(x)$ portée par un ouvert O de $\mathbb{R}^d$ au sens où $\int_O p(x)dx = 1$ et φ une bijection de O sur O' C^1 ainsi que son inverse φ^{-1}. Alors le vecteur $Y = \varphi(X)$ possède la densité*

$$q(y) = 1_{O'}(y)\ p(\varphi^{-1}(y))\ |\mathrm{Jac}\ \varphi^{-1}(y)|.$$

3.3.4 Indépendance

Définition 3.3.9. *Les vecteurs aléatoires $X_1 : \Omega \to \mathbb{R}^{d_1}, \ldots, X_n : \Omega \to \mathbb{R}^{d_n}$ qui possèdent respectivement les densités p_1,...,p_n sont dits indépendants si $X = (X_1, \ldots, X_n)$ possède la densité produit $p_1(x_1) \times \ldots \times p_n(x_n)$.*

Remarque 3.3.10. Si $X = (X_1, \ldots, X_n)$ possède la densité $p(x) = cg_1(x_1) \times \ldots \times g_n(x_n)$ alors les vecteurs X_i sont indépendants et possèdent les densités

$$p_i(x_i) = \frac{1}{\int_{\mathbb{R}^{d_i}} g_i(y_i)dy_i} \ g_i(x_i).$$

En effet, comme l'intégrale de la densité p sur $\mathbb{R}^{d_1+\ldots+d_n}$ vaut 1,

$$1/c = \left(\int_{\mathbb{R}^{d_1}} g_1(y_1)dy_1\right) \times \ldots \times \left(\int_{\mathbb{R}^{d_n}} g_n(y_n)dy_n\right).$$

La formule des densités marginales permet de vérifier que la densité de X_i est p_i. Et comme $p(x) = p_1(x_1) \times \ldots \times p_n(x_n)$, les X_i sont indépendants.

La caractérisation suivante de l'indépendance de vecteurs aléatoires qui ne possèdent pas nécessairement des densités est parfois utile :

Proposition 3.3.11. *Soient $X_1 : \Omega \to \mathbb{R}^{d_1}, \ldots, X_n : \Omega \to \mathbb{R}^{d_n}$ des vecteurs aléatoires.*

1. *Si ces vecteurs aléatoires sont indépendants alors pour toutes fonctions $f_1 : \mathbb{R}^{d_1} \to \mathbb{R}, \ldots, f_n : \mathbb{R}^{d_n} \to \mathbb{R}$ telles que $f_1(X_1), \ldots, f_n(X_n)$ sont intégrables alors $f_1(X_1) \times \ldots \times f_n(X_n)$ est intégrable et*

$$\mathbb{E}(f_1(X_1) \times \ldots \times f_n(X_n)) = \prod_{i=1}^{n} \mathbb{E}(f_i(X_i)).$$

2. *Inversement, si pour toutes fonctions $f_1 : \mathbb{R}^{d_1} \to \mathbb{R}, \ldots, f_n : \mathbb{R}^{d_n} \to \mathbb{R}$ bornées, $\mathbb{E}(f_1(X_1) \times \ldots \times f_n(X_n)) = \prod_{i=1}^{n} \mathbb{E}(f_i(X_i))$, alors les vecteurs aléatoires $X_1, \ldots, X_n$ sont indépendants.*

Remarque 3.3.12. Lorsque les vecteurs aléatoires $X_1, \ldots, X_n$ sont indépendants, alors $\forall m \in [1, n]$, $\forall 1 \leq d_1 < d_2 < \ldots < d_m \leq n$, les vecteurs $(X_1, X_2, \ldots, X_{d_1})$, $(X_{d_1+1}, \ldots, X_{d_2})$,...,$(X_{d_{m-1}+1}, \ldots, X_{d_m})$ et $(X_{d_m+1}, \ldots, X_n)$ sont indépendants, propriété démontrée dans la remarque 2.2.7 dans le cas discret.

3.3.5 Covariance

On rappelle que si Y et Z sont deux variables aléatoires de carré intégrable, comme $|YZ| \leq (Y^2 + Z^2)/2$, leur produit YZ est intégrable.

Définition 3.3.13. – *Soit Y et Z deux variables aléatoires réelles de carré intégrable (non nécessairement à densité). On appelle covariance de Y et Z le nombre $\mathrm{Cov}(Y, Z)$ défini par*

$$\mathrm{Cov}(Y, Z) = \mathbb{E}((Y - \mathbb{E}(Y))(Z - \mathbb{E}(Z))).$$

– *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire à valeurs $\mathbb{R}^d$ (non nécessairement à densité) dont les composantes sont de carré intégrable. On appelle matrice de covariance du vecteur X la matrice K^X de dimension $d \times d$ et d'élément courant*

$$K^X_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j)),\ 1 \leq i, j \leq d.$$

Propriétés 3.3.14. *1.* $\mathrm{Cov}(Y, Y) = \mathrm{Var}(Y)$ *et les éléments diagonaux de la matrice de covariance K^X sont les variances des composantes de X.*

2. $\mathrm{Cov}(Y, Z) = \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)$.

3. Si Y et Z sont indépendantes $\mathrm{Cov}(Y, Z) = 0$.

4. La matrice de covariance K^X est symétrique et positive. En outre,

$$\forall u \in \mathbb{R}^d,\ \mathrm{Var}(u.X) = u.K^X u \quad \text{où } u.v \text{ désigne le produit scalaire de } u \text{ et } v \text{ dans } \mathbb{R}^d.$$

En utilisant la propriété 4. pour $u = (1, 1, \ldots, 1)$, on obtient que $\mathrm{Var}(X_1 + \ldots + X_d) = \sum_{i,j=1}^d \mathrm{Cov}(X_i, X_j)$. Avec les propriétés 1. et 3., on en déduit que

Corollaire 3.3.15. *Si $X_1, \ldots, X_d$ sont des variables aléatoires réelles de carré intégrable* **indépendantes**, $\mathrm{Var}(X_1 + \ldots + X_d) = \sum_{i=1}^d \mathrm{Var}(X_i)$.

Démonstration : La preuve des trois premiers est laissée à titre d'exercice.
Soit $u = (u_1, \ldots, u_d) \in \mathbb{R}^d$.

$$\begin{aligned}
\mathrm{Var}(u.X) &= \mathbb{E}\left(\left(\sum_{i=1}^d u_i X_i - \mathbb{E}\left(\sum_{i=1}^d u_i X_i\right)\right)^2\right) \\
&= \mathbb{E}\left(\sum_{i,j=1}^d u_i u_j (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) \\
&= \sum_{i,j=1}^d u_i u_j \mathrm{Cov}(X_i, X_j) \\
&= u.K^X u.
\end{aligned}$$

Comme la variance d'une variable aléatoire est toujours positive, on conclut que $\forall u \in \mathbb{R}^d$, $u.K^X u \geq 0$. □

Remarque 3.3.16. – Notons que lorsque K^X est dégénérée, il existe u_0 non nul tel que $K^X u_0 = 0$. Alors $\mathrm{Var}(u_0.X) = 0$, ce qui entraîne que $u_0.X$ est une constante i.e. X prend ses valeurs dans dans le sous espace affine strict de $\mathbb{R}^d$, $\mathcal{E} = \{x \in \mathbb{R}^d : u_0.x = u_0.\mathbb{E}(X)\}$. Comme $\mathcal{E}$ a une mesure nulle pour la mesure de Lebesgue (mesure de volume) sur $\mathbb{R}^d$, pour toute densité de probabilité p sur $\mathbb{R}^d$, $\int_{\mathbb{R}^d} 1_{\mathcal{E}}(x)p(x)dx = 0$. On en déduit que X ne possède pas de densité. Par contraposée, la matice de covariance d'un vecteur aléatoire X de carré intégrable et qui possède une densité est définie positive.

– La covariance de deux variables aléatoires indépendantes est nulle mais il ne suffit pas que $\mathrm{Cov}(Y, Z) = 0$ pour que Y et Z soient indépendantes.
En effet si $Y \sim \mathcal{U}[-1, 1]$ et $Z = \varepsilon Y$ où ε est une variable aléatoire indépendante de Y t.q. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$ alors $\mathbb{E}(Y) = 0$ et $\mathbb{E}(YZ) = \mathbb{E}(\varepsilon Y^2) = \mathbb{E}(\varepsilon)\mathbb{E}(Y^2) = 0 \times \mathbb{E}(Y^2)$ si bien que $\mathrm{Cov}(Y, Z) = 0$. Mais comme $\varepsilon^2 = 1$,

$$\begin{aligned}
&\mathbb{E}(Y^2 Z^2) = \mathbb{E}(\varepsilon^2 Y^4) = \mathbb{E}(Y^4) = \frac{1}{2}\int_{-1}^{1} y^4 dy = \frac{1}{5}\\
&\mathbb{E}(Y^2) = \frac{1}{2}\int_{-1}^{1} y^2 dy = \frac{1}{3} \text{ et } \mathbb{E}(Z^2) = \mathbb{E}(\varepsilon^2 Y^2) = \mathbb{E}(Y^2) = \frac{1}{3}\\
&\text{si bien que } \mathbb{E}(Y^2 Z^2) = \frac{1}{5} \neq \frac{1}{9} = \mathbb{E}(Y^2)\mathbb{E}(Z^2),
\end{aligned}$$

et les variables Y et Z ne sont donc pas indépendantes.

3.3.6 Loi et espérance conditionnelles

Dans tout ce paragraphe, on considère un couple $(X, Y) : \Omega \to \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ tel que (X, Y) possède la densité $p_{X,Y}(x, y)$ et on note $p_X(x)$ et $p_Y(y)$ les densités marginales respectives de X et de Y.
On souhaite définir la notion de loi conditionnelle de X sachant $Y = y$. Bien sûr, comme pour tout $y \in \mathbb{R}^{d_2}$, $\mathbb{P}(Y = y) = 0$, conditionner par $Y = y$ n'a pas de sens mathématique. C'est pourquoi il faut revenir au point de vue infinitésimal et conditionner par $Y \in [y, y + dy]$. Ainsi formellement,

$$\begin{aligned}
\mathbb{P}(X \in [x, x + dx] | Y \in [y, y + dy]) &= \frac{\mathbb{P}(X \in [x, x + dx], Y \in [y, y + dy])}{\mathbb{P}(Y \in [y, y + dy])}\\
&= \frac{p_{X,Y}(x, y)dxdy}{p_Y(y)dy} = \frac{p_{X,Y}(x, y)}{p_Y(y)}dx.
\end{aligned}$$

Notons que lorsque $p_Y(y) > 0$, comme $p_Y(y) = \int_{\mathbb{R}^{d_1}} p_{X,Y}(x, y)dx$, alors la fonction $x \to p_{X,Y}(x, y)/p_Y(y)$ est une densité de probabilité sur $\mathbb{R}^{d_1}$.

Définition 3.3.17. *Pour $y \in \mathbb{R}^{d_2}$, on appelle densité conditionnelle de X sachant $Y = y$ la densité $p_{X,y}(x)$ donnée par la formule*

$$p_{X,y}(x) = \begin{cases} \dfrac{p_{X,Y}(x, y)}{p_Y(y)} & \textit{si } \; p_Y(y) > 0,\\ p_X(x) & \textit{sinon.}\end{cases}$$

Comme dans le cas discret, on peut traduire l'indépendance de X et de Y sur la loi conditionnelle de X sachant Y.

Proposition 3.3.18. *Les variables X et Y sont indépendantes si et seulement si la densité conditionnelle de X sachant $Y = y$ ne dépend pas de y.*

La preuve s'effectue en remplaçant les sommes par des intégrales dans celle de la proposition 2.5.3.
La notion de loi conditionnelle permet de définir celle d'espérance conditionnelle :

Définition 3.3.19. *Soit $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ telle que $f(X,Y)$ est intégrable.*
On appelle espérance conditionnelle de $f(X,Y)$ sachant Y et on note $\mathbb{E}(f(X,Y)|Y)$ la variable aléatoire

$$\boxed{\mathbb{E}(f(X,Y)|Y) = \psi(Y) \quad \text{où } \forall y \in \mathbb{R}^{d_2},\ \psi(y) = \int_{\mathbb{R}^{d_1}} f(x,y) p_{X,y}(x)dx.}$$

Notons que

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} |f(x,y)| p_{X,y}(x)dx \right) p_Y(y)dxdy = \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} |f(x,y)| p_{X,Y}(x,y)dxdy$$
$$= \mathbb{E}|f(X,Y)| < +\infty.$$

On en déduit que l'ensemble $A = \left\{ y \in \mathbb{R} : \int_{\mathbb{R}^{d_2}} |f(x,y)| p_{X,y}(x)dx = +\infty \right\}$ sur lequel $\psi(y)$ n'est pas défini vérifie $\mathbb{P}(Y \in A) = \int_{\mathbb{R}^{d_2}} 1_A(y) p_Y(y)dy = 0$. Ainsi les variables aléatoires $\psi(Y)$ et donc $\mathbb{E}(f(X,Y)|Y)$ sont bien définies avec probabilité 1.
Dans le cas particulier où X et Y sont indépendantes, on a $\psi(y) = \mathbb{E}(f(X,y))$ pour tout $y \in \mathbb{R}^{d_2}$.
Comme dans la proposition 2.5.6 et le corollaire 2.5.7 relatifs au cas des variables aléatoires discrètes, on peut démontrer en remplaçant les sommes par des intégrales que

Proposition 3.3.20. *On suppose que $f(X,Y)$ est intégrable. Pour toute fonction $g : \mathbb{R}^{d_2} \to \mathbb{R}$ telle que $f(X,Y)g(Y)$ est intégrable, la variable aléatoire $\mathbb{E}(f(X,Y)|Y)g(Y)$ est intégrable et on a*

$$\mathbb{E}\left(\mathbb{E}(f(X,Y)|Y)g(Y)\right) = \mathbb{E}(f(X,Y)g(Y)).$$

En outre, $\mathbb{E}\left(\mathbb{E}(f(X,Y)|Y)\right) = \mathbb{E}(f(X,Y))$. Enfin, si $f(X,Y)$ est de carré intégrable, $\mathbb{E}(f(X,Y)|Y)$ l'est aussi et $\mathrm{Var}(\mathbb{E}(f(X,Y)|Y)) \leq \mathrm{Var}(f(X,Y))$.

Exercice résolu 3.3.21. *Soient U et V deux variables aléatoires uniformes sur $[0,1]$ indépendantes et $Y = U - V$. Donner la loi conditionnelle de U sachant $Y = y$ et calculer $\mathbb{E}(U|Y)$.*

On commence par déterminer la loi de (U,Y). Pour $f : \mathbb{R}^2 \to \mathbb{R}$ bornée, en posant $y = u - v$, on obtient

$$\mathbb{E}(f(U,Y)) = \int_{\mathbb{R}} 1_{\{0<u<1\}} \int_{\mathbb{R}} f(u, u-v) 1_{\{0<v<1\}} dvdu$$
$$= \int_{\mathbb{R}^2} f(u,y) 1_{\{0<u<1\}} 1_{\{y<u<1+y\}} dudy.$$

Donc (U,Y) possède la densité $p_{U,Y}(u,y) = 1_{\{0<u<1\}} 1_{\{y<u<1+y\}}$. On en déduit la densité marginale de Y :

$$p_Y(y) = \begin{cases} \int_0^{1+y} du = 1+y & \text{si } -1 \leq y \leq 0, \\ \int_y^1 du = 1-y & \text{si } 0 \leq y \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

On en déduit que pour y dans $[-1,0]$, $p_{U,y}(u) = 1_{\{0<u<1+y\}}/(1+y)$ c'est-à-dire que la loi conditionnelle de U sachant $Y = y$ est la loi uniforme sur $[0, 1+y]$ d'espérance $(1+y)/2$. Pour y dans $[0,1]$, $p_{U,y}(u) = 1_{\{y<u<1\}}/(1-y)$ c'est-à-dire que la loi conditionnelle de U sachant $Y = y$ est la loi uniforme sur $[y,1]$ d'espérance $(1+y)/2$. Comme $p_Y(y) = 0$ pour y en dehors de $[-1,1]$, il n'est pas utile d'étudier le cas $y \notin [-1,1]$.
On conclut donc que $\mathbb{E}(U|Y) = \dfrac{1+Y}{2}$ formule qui montre bien que si Y est proche de 1, U l'est aussi et que si Y est proche de -1, U est proche de 0.

3.4 Lois béta, gamma, du chi 2, de Student et de Fisher

On note Γ la fonction gamma d'Euler :

$$a > 0 \rightarrow \Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

On vérifie facilement que cette fonction est t.q.

$$\forall a > 0,\ \Gamma(a+1) = a\Gamma(a) \quad \text{et} \quad \forall n \in \mathbb{N}^*,\ \Gamma(n) = (n-1)!$$

On dit que X suit la

- **loi gamma de paramètres** a **et** θ **où** $a, \theta > 0$ et on note $X \sim \Gamma(a, \theta)$ si X possède la densité
$$\frac{\theta^a}{\Gamma(a)}\, x^{a-1}\, e^{-\theta x}\, 1_{\{x>0\}}.$$

- **loi béta de paramètres** a **et** b **où** $a, b > 0$ et on note $X \sim \beta(a, b)$, si X possède la densité
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1} 1_{\{0<x<1\}}.$$

Exemple 3.4.1. – La loi uniforme sur $[0, 1]$ est la loi $\beta(1, 1)$.
– La loi exponentielle de paramètre θ est la loi $\Gamma(1, \theta)$.

Remarque 3.4.2. Le changement de variables $y = \theta x$ entraîne que

$$\int_0^{+\infty} \theta^a\, x^{a-1}\, e^{-\theta x}\, dx = \int_0^{+\infty} y^{a-1}\, e^{-y}\, dy = \Gamma(a).$$

Le fait que $\int_0^1 z^{a-1}(1-z)^{b-1} dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ sera vérifié au cours de la preuve de la proposition 3.4.4 ci-dessous.

Exercice résolu 3.4.3. *Soit $X_1, \ldots, X_n$ des variables aléatoires I.I.D. suivant la loi exponentielle de paramètre $\theta > 0$. Quelle est la loi de $S_n = X_1 + \ldots + X_n$?*

On se donne une fonction muette $f : \mathbb{R} \rightarrow \mathbb{R}$ bornée.

$$\begin{aligned}\mathbb{E}(f(S_n)) &= \mathbb{E}(f(X_1 + \ldots + X_n)) \\ &= \int_0^{+\infty} \ldots \int_0^{+\infty} f(x_1 + \ldots + x_n)\theta^n \exp(-\theta(x_1 + \ldots + x_n)) dx_1 \ldots dx_n,\end{aligned}$$

car par indépendance des X_i la densité de $(X_1, \ldots, X_n)$ est le produit des densités de ces variables. On effectue le changement de variables

$$(y_1, \ldots, y_n) = \varphi(x_1, \ldots, x_n) = (x_1, x_1 + x_2, \ldots, x_1 + \ldots + x_n).$$

La matrice jacobienne de φ est triangulaire avec des 1 sur la diagonale. Donc $dx_1 \dots dx_n = dy_1 \dots dy_n$ et

$$\begin{aligned}\mathbb{E}(f(S_n)) &= \int_{0<y_1<y_2<\dots<y_n} f(y_n)\theta^n \exp(-\theta y_n) dy_1 \dots dy_n \\ &= \int_0^{+\infty} f(y_n)\theta^n \exp(-\theta y_n) \left(\int_{0<y_1<\dots<y_n} dy_1 \dots dy_{n-1} \right) dy_n \\ &= \int_0^{+\infty} f(y_n)\theta^n \exp(-\theta y_n) \frac{y_n^{(n-1)}}{(n-1)!} dy_n.\end{aligned}$$

Donc $S_n \sim \Gamma(n, \theta)$.

Ainsi la somme de n-variables de loi $\Gamma(1, \theta)$ indépendantes suit la loi $\Gamma(n, \theta)$.

Ce résultat peut-être déduit de la proposition suivante par un raisonnement de récurrence :

Proposition 3.4.4. *Soit $X \sim \Gamma(a, \theta)$ et $Y \sim \Gamma(b, \theta)$ indépendantes. Alors $S = X + Y$ et $Z = \frac{X}{X+Y}$ sont deux variables aléatoires indépendantes distribuées respectivement suivant les lois $\Gamma(a + b, \theta)$ et $\beta(a, b)$.*

Démonstration : On utilise la méthode de la fonction muette : soit $f : \mathbb{R}^2 \to \mathbb{R}$ bornée.

$$\begin{aligned}\mathbb{E}(f(S, Z)) &= \mathbb{E}\left(f\left(X + Y, \frac{X}{X + Y} \right) \right) \\ &= \int_0^{+\infty} \int_0^{+\infty} f\left(x + y, \frac{x}{x + y} \right) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} y^{b-1} e^{-\theta(x+y)} dx\, dy.\end{aligned}$$

On effectue le changement de variables $(s, z) = \varphi(x, y) = \left(x + y, \frac{x}{x+y} \right)$. On a

$$\left\{ \begin{array}{l} s = x + y \\ z = \frac{x}{x+y} \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x = sz \\ y = s(1 - z) \end{array} \right. \text{ et } \left\{ \begin{array}{l} x > 0 \\ y > 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} sz > 0 \\ s(1 - z) > 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} s > 0 \\ 0 < z < 1 \end{array} \right.$$

$$\frac{D(x, y)}{D(s, z)} = \begin{pmatrix} z & s \\ 1 - z & -s \end{pmatrix}, \text{ d'où } dx\, dy = s\, ds\, dz.$$

On en déduit par la formule de changement de variables que

$$\mathbb{E}(f(S, Z)) = \int_0^{+\infty} \int_0^1 f(s, z) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} (sz)^{a-1} (s(1 - z))^{b-1} e^{-\theta s} s\, dz\, ds.$$

Ainsi (S, Z) possède la densité produit

$$\frac{\theta^{a+b}}{\Gamma(a + b)} s^{a+b-1} e^{-\theta s} 1_{\{s>0\}} \times \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} z^{a-1} (1 - z)^{b-1} 1_{\{0<z<1\}}.$$

Notons que cela démontre au passage que $\int_0^1 z^{a-1}(1 - z)^{b-1} dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. □

Définition 3.4.5. – *On appelle loi du chi 2 à n degrés de liberté et on note $\chi^2(n)$ la loi de $X_1^2 + \dots + X_n^2$ où $X_1, \dots, X_n$ sont n variables normales centrées réduites indépendantes.*

– *On appelle loi de Student de paramètre n et on note $t(n)$ la loi de $\frac{G}{\sqrt{Y/n}}$ où $G \sim \mathcal{N}_1(0,1)$ et $Y \sim \chi^2(n)$ sont indépendantes.*

Proposition 3.4.6. *La loi $\chi^2(n)$ est la loi $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ de densité*

$$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\, y^{\frac{n}{2}-1}\, e^{-\frac{y}{2}} 1_{\{y>0\}}.$$

La loi de Student $t(n)$ est la loi de densité

$$\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \times \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}}.$$

Démonstration : Pour $f : \mathbb{R} \to \mathbb{R}$ bornée et $G \sim \mathcal{N}_1(0,1)$,

$$\begin{aligned}\mathbb{E}(f(G^2)) &= \int_{\mathbb{R}} f(x^2)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx = 2\int_0^{+\infty} f(x^2)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx \\ &= \int_0^{+\infty} f(y)\frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}dy, \quad \text{en posant } y = x^2.\end{aligned}$$

Ainsi $\chi^2(1) \equiv \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$.
Avec la proposition 3.4.4, on conclut par récurrence que $\chi^2(n) \equiv \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Soit maintenant $Y \sim \chi^2(n)$ indépendante de G. En effectuant successivement les changements de variable $x \to t = x/\sqrt{y/n}$ puis $y \to z = \frac{y(1+t^2/n)}{2}$, on obtient

$$\begin{aligned}\mathbb{E}\left(f\left(\frac{G}{\sqrt{Y/n}}\right)\right) &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^{+\infty}\int_{\mathbb{R}} f\left(\frac{x}{\sqrt{y/n}}\right)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx\, y^{\frac{n}{2}-1}e^{-\frac{y}{2}}dy \\ &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})\sqrt{2n\pi}}\int_{\mathbb{R}} f(t)\int_0^{+\infty} y^{\frac{n-1}{2}}e^{-\frac{y(1+t^2/n)}{2}}dydt \\ &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})\sqrt{2n\pi}}\int_0^{+\infty} z^{\frac{n-1}{2}}e^{-z}dz\int_{\mathbb{R}} f(t)\left(\frac{2}{1+t^2/n}\right)^{\frac{n+1}{2}} dt.\end{aligned}$$

On conclut en remarquant que $\int_0^{+\infty} z^{\frac{n-1}{2}}e^{-z}dz = \Gamma(\frac{n+1}{2})$. □

Exercice 3.4.7. Pour $n, m \in \mathbb{N}^*$, on appelle loi de Fisher (ou Fisher-Snedecor) de paramètre (n,m) et on note $\mathcal{F}_{n,m}$ la loi de $F = \frac{Y/n}{Z/m}$ où $Y \sim \chi^2(n)$ et $Z \sim \chi^2(m)$ sont indépendantes. Montrer que F a pour densité

$$\frac{\Gamma(\frac{n+m}{2})n^{\frac{n}{2}}m^{\frac{m}{2}}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \times \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}1_{\{x\geq 0\}}.$$

3.5 Exercices

Exercice 3.5.1. Une cerise est placée sur la circonférence d'un gateau rond que l'on partage en 2 au hasard en pratiquant deux découpes suivant des rayons.
Si on prend la position de la cerise comme origine des angles, les positions U et V des deux coups de couteau sont des variables uniformément réparties sur $[0, 2\pi]$ indépendantes. Exprimer la taille T de la part contenant la cerise, calculer son espérance puis déterminer la probabilité pour cette part soit plus grosse que l'autre. Quelle doit être la décision d'un gourmand qui doit choisir entre la part avec la cerise et la part sans la cerise avant le découpage ?

Exercice 3.5.2. On suppose que la durée de vie d'un équipement A est une variable aléatoire réelle S possédant la densité $\alpha e^{-\alpha s} 1_{\{s \geq 0\}}$, que celle de l'équipement B est une variable T possédant la densité $\beta^2 t e^{-\beta t} 1_{\{t \geq 0\}}$ avec $\alpha, \beta > 0$ et que S et T sont indépendantes.

1. Que vaut $P(S \geq t+s \mid S \geq t)$ pour $s, t \geq 0$? Est-il judicieux de changer l'équipement A au temps t s'il fonctionne toujours ?
2. Quelle est la probabilité pour que l'équipement A tombe en panne avant l'équipement B ?

Exercice 3.5.3. Soit X une variable qui suit la loi de Cauchy de paramètre $a > 0$. Quelle est la loi de $Y = 1/X$?

Exercice 3.5.4. Soit (X, Y) un couple de variables aléatoires normales centrées réduites indépendantes. Déterminer la loi de $(Z, W) = (X/Y, X + Y)$ et en déduire la loi de Z.

Exercice 3.5.5. Soit $v : \mathbb{R} \to \mathbb{R}$ une fonction convexe C^1 et X une variable aléatoire réelle intégrable telle que $v(X)$ est intégrable.

1. Montrer l'inégalité de Jensen : $\boxed{\mathbb{E}(v(X)) \geq v(\mathbb{E}(X))}$. Commenter le cas particulier $v(x) = x^2$.
2. Dans la théorie économique de von Neumann Morgenstern, les investisseurs cherchent à maximiser l'espérance de l'utilité de leur richesse. Un investisseur averse au risque, ce qui se traduit par la concavité de sa fonction d'utilité, préférera-t-il la richesse aléatoire X ou bien la richesse déterministe $\mathbb{E}(X)$?

Exercice 3.5.6. Soit X et Y deux variables aléatoires réelles exponentielles de paramètre λ ($\lambda > 0$) indépendantes.

1. Quelle est la loi de X^2 ?
2. Quelle est celle de
$$V = \begin{cases} X & si \ \ 0 \leq X \leq 1 \\ 2X & si \ \ X > 1 \end{cases}$$
3. Déterminer la loi de
$$(Z, S) = \left(\frac{X}{X+Y}, X + Y \right).$$
Les variables Z et S sont-elles indépendantes ?

Exercice 3.5.7. Les durées de vie S et T de deux machines suivent des lois exponentielles de paramètres respectifs α et β et sont indépendantes. Quelle est la loi du couple $(\min(S, T), |T - S|)$? (On rappelle que $|T - S| = \max(S, T) - \min(S, T)$.) Les variables $\min(S, T)$ et $|T - S|$ sont-elles indépendantes ?

Exercice corrigé 3.5.8. On coupe un bâton de longueur 1 au hasard en trois morceaux : les abscisses U et V des découpes sont supposées indépendantes et uniformément réparties sur $[0,1]$. On veut calculer la probabilité p pour que l'on puisse faire un triangle avec les trois morceaux (on peut faire un triangle avec trois segments de longueur l_1, l_2 et l_3 si et seulement si $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$).

1. Exprimer en fonction de U et V les longueurs respectives L_1, L_2 et L_3 du morceau le plus à gauche, du morceau du milieu et du morceau le plus à droite.
2. Montrer que
$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$
3. Calculer p.

Exercice corrigé 3.5.9. Soient X et Y deux variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi de Pareto de paramètre 2 i.e. qui possèdent la densité $p(x) = \frac{1_{\{x>1\}}}{x^2}$. On pose
$$(Z, W) = \left(\ln(X), 1 + \frac{\ln(Y)}{\ln(X)}\right).$$

1. Quelle est la loi de (Z, W) ? Les variables Z et W sont-elles indépendantes ?
2. Quelle est la loi de W ?
3. Quelle est la loi de Z ?

Exercice corrigé 3.5.10. Soient X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\Gamma(p, 1)$ et $\Gamma(q, 1)$, $p > 0$ et $q > 0$.

1. Déterminer l'espérance et la variance de X.
2. On veut calculer la loi de la variable aléatoire $Z = \frac{X}{Y}$. On dit que Z suit la loi bêta de seconde espèce de paramètres p et q.
   (a) On pose $W = X - Y$. Exprimer en fonction de X et Y l'événement $\{Z < 1, W > 0\}$. En déduire sans calcul sa probabilité. Les variables Z et W sont-elles indépendantes ?
   (b) Déterminer la loi du vecteur (Z, W).
   (c) En déduire la loi de la variable aléatoire Z.
3. Déterminer la loi de $T = \frac{Z}{1+Z}$.

Exercice 3.5.11. Le cycle d'un feu de circulation est le suivant : le feu est vert sur l'intervalle $[0, v]$ et orange ou rouge sur $[v, v + r]$ avec $v, r > 0$. L'instant d'arrivée U d'un automobiliste est supposé uniformément réparti sur le cycle $[0, r + v]$.

1. Exprimer en fonction de U le temps T d'attente de cet automobiliste au feu dans le cas où aucun véhicule ne se trouve devant le feu à l'instant où il arrive.
2. Montrer que pour $f : \mathbb{R} \to \mathbb{R}$ bornée,
$$\mathbb{E}(f(T)) = \frac{v}{r+v} f(0) + \frac{1}{r+v} \int_0^r f(s)ds.$$

3. Déterminer la probabilité $\mathbb{P}(T = 0)$ pour que l'automobiliste passe immédiatement.
4. Montrer que si $t \neq 0$, $\mathbb{P}(T = t) = 0$.
5. Conclure que le temps d'attente T n'est ni une variable aléatoire discrète ni une variable aléatoire à densité.

Exercice 3.5.12. Montrer que la probabilité de tomber sur une fève qui a la forme d'un disque de rayon r lors de la découpe d'un rayon d'une galette des rois de rayon $R > 2r$ est égale à

$$\frac{r^2}{2(R-r)^2} + \frac{r\sqrt{(R-r)^2 - r^2}}{\pi(R-r)^2} + \frac{1}{\pi} \arcsin\left(\frac{r}{R-r}\right).$$

Indication : cette probabilité est égale à celle pour que la fève dont le centre est uniformément réparti sur le disque de rayon $R - r$ touche un rayon fixé de la galette.

Exercice 3.5.13. Pour $\alpha > 0$, soient U de loi uniforme sur $[0, 1]$, ε de loi de Bernoulli de paramètre $\frac{1}{1+\alpha}$ et X de densité $\frac{\alpha}{x^{1+\alpha}} 1_{\{x>1\}}$ indépendantes.

1. Calculer la fonction de répartition de X. En déduire que X a même loi que $U^{-1/\alpha}$.
2. Déterminer la loi de $Y = XU$.
3. Vérifier que $Z = \varepsilon X + (1 - \varepsilon)U$ a même loi que Y.

Exercice 3.5.14. Au cours d'un jeu télévisé, un candidat tire au sort entre 2 enveloppes contenant respectivement les sommes x_1 et x_2 ($x_1 > x_2 > 0$), qu'il ne connaît pas. Après avoir examiné le contenu X de son enveloppe, il a le choix entre conserver ce contenu et effectuer un échange avec l'autre enveloppe. On veut étudier la stratégie suivante : on se donne T une variable exponentielle de paramètre 1 indépendante du tirage au sort et on change d'enveloppe seulement lorsque $T > X$. Calculer la probabilité d'obtenir la somme la plus élevée x_1 en suivant cette stratégie. Est-elle supérieure ou inférieure à $1/2$? Trouvez-vous ce résultat intuitif ?

Exercice 3.5.15. Soit R et Θ, deux variables aléatoires indépendantes uniformément réparties respectivement sur $[0, 1]$ et $[0, 2\pi]$. Le vecteur

$$(X, Y) = (R \cos \Theta, R \sin \Theta)$$

est-il uniformément réparti sur le disque unité $D = \{(x, y) : x^2 + y^2 \leq 1\}$ i.e. possède-t-il la densité $\frac{1}{\pi} 1_D(x, y)$?

Problème corrigé 3.5.16. Soient $\Theta \sim \mathcal{U}[-\pi, \pi]$ et R une variable aléatoire réelle indépendante de Θ possédant une densité p nulle sur $\mathbb{R}_-$. On pose

$$(X, Y) = (R \cos \Theta, R \sin \Theta).$$

1. Déterminer la loi du vecteur (X, Y). Que peut-on dire des lois des vecteurs (X, Y) et (Y, X) ? En déduire que X et Y ont la même densité.
2. Pour $a > 0$, quelle est la loi de $Z = aX/Y$ (on pourra exprimer Z en fonction de Θ) ? En déduire sans calcul la loi de l'inverse d'une variable de Cauchy de paramètre a.
3. Quelle doit être la loi de R pour que le couple (X, Y) soit uniformément distribué sur le disque $D(0, \rho)$ où $\rho > 0$ (i.e. possède une densité constante sur ce disque et nulle en dehors) ?

4. On suppose la densité p de R continue sur $\mathbb{R}_+$ et strictement positive sur $\mathbb{R}_+^*$. On souhaite trouver à quelle condition sur p les variables X et Y sont indépendantes.
   (a) Dans le cas où X et Y sont indépendantes, si q désigne leur densité commune, exprimer $q(x)q(y)$ en fonction de p.
   En commençant par choisir $y = 0$ dans cette relation, établir que pour $x, y \in \mathbb{R}^*$,
   $$\frac{p(|x|)}{2\pi q(0)|x|} \times \frac{p(|y|)}{2\pi q(0)|y|} = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}.$$
   (b) Pour $s > 0$, on pose $f(s) = \ln\left(\frac{p(\sqrt{s})}{2\pi q^2(0)\sqrt{s}}\right)$. Vérifier que f est une fonction continue sur $\mathbb{R}_+^*$ telle que pour $s, t > 0$, $f(t) + f(s) = f(t+s)$.
   En déduire qu'il existe une constante $\sigma > 0$ telle que $p(r) = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}1_{\{r>0\}}$ i.e. que R suit la loi de Rayleigh de paramètre σ^2. Quelle est alors la loi de (X, Y) ?
   La loi de Rayleigh intervient notamment dans l'analyse d'erreur de positionnement du système GPS.

Exercice 3.5.17. Soit X et Y deux variables aléatoires exponentielles de paramètre λ indépendantes et $S = X + Y$. Déterminer la loi conditionnelle de X sachant $S = s$ et en déduire $\mathbb{E}(X|S)$.

3.6 Résumé

- $X : \Omega \to \mathbb{R}^d$ **possède la densité** $p : \mathbb{R}^d \to \mathbb{R}_+$ si

$$\forall D \text{ ouvert de } \mathbb{R}^d,\ \mathbb{P}(X \in D) = \int_D p(x)\, dx = \int_{\mathbb{R}^d} 1_D(x)\, p(x)\, dx.$$

- Soit $(X, Y) : \Omega \to \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ de densité $p_{X,Y}(x, y)$.
  - **Densité marginale :** Y a pour densité $\boxed{p_Y(y) = \int_{\mathbb{R}^{d_1}} p_{X,Y}(x, y)dx.}$
  - **Densité conditionnelle** de X sachant $Y = y$:

$$\boxed{p_{X,y}(x) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.}$$

  - **Indépendance** si $p_{X,Y}(x, y) = p_X(x)p_Y(y)$
    $\iff$ la densité conditionnelle de X sachant $Y = y$ ne dépend pas de $y \in \mathbb{R}^{d_2}$.

- **Espérance, variance :** $X : \Omega \to \mathbb{R}$ de densité p est dite
  - intégrable si $\int_{\mathbb{R}} |x|p(x)dx < +\infty$ et alors $\boxed{\mathbb{E}(X) = \int_{\mathbb{R}} x\, p(x)\, dx.}$
  - de carré intégrable si $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2\, p(x)\, dx < +\infty$ et alors

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

  Propriétés :

  1. Linéarité : $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$.
  2. Croissance : $\mathbb{P}(X \geq Y) = 1 \Longrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$.
  3. $X_1, \ldots, X_n$ **indépendantes** $\Longrightarrow \mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i)$.

- **Fonction muette :** $X : \Omega \to \mathbb{R}^d$ possède la densité p

$$\iff \boxed{\forall f : \mathbb{R}^d \to \mathbb{R} \text{ bornée },\ \mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\, p(x)\, dx.}$$

- **Espérance conditionnelle de $f(X, Y)$ sachant Y (où $f(X, Y)$ intégrable) :**

$$\boxed{\mathbb{E}(f(X, Y)|Y) = \psi(Y) \text{ où } \forall y \in \mathbb{R}^{d_2},\ \psi(y) = \int_{R^{d_1}} f(x, y)p_{X,y}(x)dx.}$$

  $\forall g : \mathbb{R}^{d^2} \to \mathbb{R}$, t.q. $f(X, Y)g(Y)$ intég, $\mathbb{E}\left(\mathbb{E}(f(X, Y)|Y)g(Y)\right) = \mathbb{E}(f(X, Y)g(Y))$.

- **Densités usuelles sur $\mathbb{R}$:**

Nom	**Densité**	$\mathbb{E}(X)$	$\mathrm{Var}(X)$
uniforme $\mathcal{U}[a,b]$	$\frac{1}{b-a}\ 1_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
exponentielle $\mathcal{E}(\lambda)$	$\lambda \exp(-\lambda x) 1_{\{x\geq 0\}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
normale $\mathcal{N}_1(m,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\ \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$	m	σ^2
Cauchy $\mathcal{C}(a)$	$\frac{a}{\pi}\ \frac{1}{x^2+a^2}$	non intég	-

- **Covariance :** si X et Y sont deux variables réelles de carré intégrable,

$$\boxed{\mathrm{Cov}(X,Y) \stackrel{\text{def}}{=} \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)\right) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).}$$

Chapitre 4

Simulation

Nous avons vu à l'occasion de l'exercice 3.5.14 qu'un participant à un jeu télévisé peut augmenter l'espérance de son gain s'il sait simuler une variable aléatoire exponentielle de paramètre 1. Mais l'intérêt de la simulation probabiliste dépasse largement le cadre de cet exemple ludique. Tout d'abord, si l'on s'intéresse à un système dont les variables d'entrée sont aléatoires de loi connue, on peut essayer de déterminer analytiquement la loi des variables de sortie du système lorsque celui-ci n'est pas trop compliqué. Dans les deux chapitres précédents, notamment avec la technique de changement de variables, nous nous sommes intéressé à ce problème qui revient à déterminer la loi de l'image par une fonction connue d'un vecteur aléatoire de loi donnée. Mais lorsque le système est complexe comme par exemple un modèle météorologique, la simulation devient la seule alternative pour obtenir des informations sur les variables de sortie. Enfin, nous verrons au chapitre suivant que la loi forte des grands nombres assure que la moyenne de n variables aléatoires intégrables indépendantes et identiquement distribuées converge lorsque n tend vers l'infini vers l'espérance commune de ces variables. Pour calculer numériquement cette espérance, on peut simuler sur ordinateur une réalisation de ces n variables avec n grand et calculer la moyenne correspondante. C'est le principe de la méthode de Monte Carlo qui est très largement utilisée en physique, en fiabilité mais aussi en mathématiques financières.

Pour effectuer des simulations probabilistes sur ordinateur, on utilise un *générateur de nombres pseudo-aléatoires*. Un tel générateur retourne une suite $(x_n)_n$ de nombres réels compris entre 0 et 1. Ces réels sont calculés par un algorithme déterministe mais imitent une réalisation d'une suite de variables aléatoires indépendantes et indentiquement distribuées suivant la loi uniforme sur $[0, 1]$. Le bon comportement de la suite est vérifié à l'aide de tests statistiques.
Une méthode couramment utilisée pour construire la suite $(x_n)_n$ est la *congruence* : $x_n = y_n/N$ où les y_n sont des entiers compris entre 0 et $N-1$ calculés grâce à la relation de récurrence

$$y_{n+1} = (ay_n + b) \mod (N).$$

Le choix des entiers a, b, N est fait de façon à ce que la période du générateur (toujours plus petite que N) soit aussi grande que possible et que les propriétés de la suite $(x_n)_n$ soient proches de celles d'une réalisation d'une suite de variables I.I.D. suivant la loi uniforme sur $[0, 1]$. Pour vérifier ce dernier point, on utilise des tests statistiques dans l'esprit de ceux qui seront introduits au chapitre 8.

Dans tout ce chapitre, le point de vue est le suivant : on suppose que l'on dispose d'un bon générateur de nombres pseudo-aléatoires et on se demande comment à partir d'une suite $(U_i)_{i\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi uniforme sur $[0,1]$ construire une variable aléatoire de loi donnée, avec une attention particulière pour les lois usuelles introduites dans les chapitres précédents.

Remarque 4.0.1. Il existe des dispositifs qui à partir de phénomènes physiques aléatoires comme les temps de désintégration atomique de matériaux radioactifs permettent de transmettre à l'ordinateur des nombres aléatoires. Mais le coût de ces dispositifs reste prohibitif.

4.1 Simulation de variables aléatoires discrètes

4.1.1 Loi de Bernoulli de paramètre $p \in [0,1]$

Si $U \sim \mathcal{U}([0,1])$ alors

$$\boxed{X = 1_{\{U\leq p\}} \sim \mathcal{B}(p).}$$

En effet X prend les valeurs 0 ou 1 et

$$\mathbb{P}(X=1) = \mathbb{P}(U \leq p) = \int_0^1 1_{\{u\leq p\}} du = p.$$

4.1.2 Loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$

Si $U_1, \ldots, U_n$ sont n variables uniformes sur $[0,1]$ indépendantes alors

$$\boxed{X = 1_{\{U_1\leq p\}} + \ldots + 1_{\{U_n\leq p\}} = \sum_{i=1}^n 1_{\{U_i\leq p\}} \sim \mathcal{B}(n,p).}$$

D'après ce qui précède, les variables $1_{\{U_i\leq p\}},\ 1 \leq i \leq n$ sont des variables de Bernoulli de paramètre p indépendantes. La variable aléatoire X, somme de ces n variables suit donc la loi binomiale de paramètres n et p.

4.1.3 Loi géométrique de paramètre $p \in]0,1]$

C'est la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes avec propabilité de succès p. Ainsi, si les $(U_i)_{i\geq 1}$ sont des variables uniformes sur $[0,1]$ indépendantes

$$\boxed{N = \inf\{i \geq 1 :\ U_i \leq p\} \sim \mathcal{G}eo(p).}$$

Remarque 4.1.1. Nous verrons dans l'exercice 4.3.1 une méthode qui permet de simuler une variable géométrique de paramètre p à l'aide d'une seule variable uniforme sur $[0,1]$. Cette méthode qui repose sur l'interprétation de la loi géométrique comme version discrète de la loi exponentielle est beaucoup plus rapide en terme de temps de calcul.

4.1.4 Simulation suivant une loi discrète quelconque

Nous venons de voir des méthodes spécifiques pour simuler des variables aléatoires suivant des lois discrètes usuelles.
Il est cependant toujours possible d'obtenir une variable qui prend les valeurs $(x_i)_{i\in\mathbb{N}^*}$ avec probabilités respectives $(p_i)_{i\in\mathbb{N}^*}$ (avec les $p_i \geq 0$ t.q. $\sum_{i\in\mathbb{N}^*} p_i = 1$) à l'aide d'une seule variable U uniforme sur $[0,1]$ en posant

$$\boxed{X = x_1 1_{\{U\leq p_1\}} + x_2 1_{\{p_1<U\leq p_1+p_2\}} + \ldots + x_i 1_{\{p_1+\ldots+p_{i-1}<U\leq p_1+\ldots+p_i\}} + \ldots .}$$

Remarque 4.1.2. Pour implémenter cette méthode très générale, il faut programmer une boucle sur i avec comme test d'arrêt $p_1 + \ldots + p_i \geq U$. Cela peut s'avérer coûteux en temps de calcul lorsque la série de terme général p_i converge lentement vers 1.

Remarque 4.1.3. Outre cette méthode générale, pour simuler une variable qui suit la loi de Poisson de paramètre $\lambda > 0$, on peut recourir à l'approche présentée dans l'exercice 4.3.2 qui repose sur les liens entre variables de Poisson et variables exponentielles de même paramètre λ.

4.2 Simulation de variables aléatoires à densité

4.2.1 Loi uniforme sur $[a,b]$ avec $a < b \in \mathbb{R}$

Si U est une variable uniforme sur $[0,1]$ alors

$$\boxed{X = a + (b-a)U \sim \mathcal{U}[a,b].}$$

4.2.2 Méthode d'inversion de la fonction de répartition

Soit p une densité de probabilité sur $\mathbb{R}$ strictement positive et $F(x) = \int_{-\infty}^{x} p(y)dy$ la fonction de répartition associée. Comme F est continue et strictement croissante sur $\mathbb{R}$ et vérifie $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$, elle admet une fonction inverse $F^{-1} :]0,1[\to \mathbb{R}$.

Proposition 4.2.1. *Si $U \sim \mathcal{U}[0,1]$, alors $X = F^{-1}(U)$ possède la densité p.*

Démonstration : Soit $a < b \in \mathbb{R}$. Comme F est strictement croissante, $\{a < X \leq b\} = \{a < F^{-1}(U) \leq b\} = \{F(a) < U \leq F(b)\}$. Donc

$$\mathbb{P}(a < X \leq b) = \int_0^1 1_{\{F(a)<u\leq F(b)\}} du = F(b) - F(a) = \int_a^b p(y)dy,$$

ce qui d'après la définition 3.2.1 signifie que X possède la densité p. □

Exemple 4.2.2. La loi de Cauchy de paramètre $a > 0$ est la loi de densité $p(y) = \frac{a}{\pi(y^2+a^2)}$ sur $\mathbb{R}$. La fonction de répartition associée est

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} p(y)dy = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{(y/a)^2+1}\,\frac{dy}{a} \\ &= \frac{1}{\pi}\int_{-\infty}^{x/a} \frac{1}{1+z^2}dz = \frac{1}{\pi}[\arctan(z)]_{-\infty}^{x/a} = \frac{1}{\pi}\arctan(x/a) + 1/2. \end{aligned}$$

Pour inverser la fonction de répartition, on se donne $u \in]0,1[$ et on cherche x t.q. $F(x) = u$:

$$F(x) = u \Leftrightarrow x = a\tan(\pi(u-1/2)).$$

En conséquence si $U \sim \mathcal{U}[0,1]$, $X = a\tan(\pi(U-1/2)) \sim \mathcal{C}(a)$.
Comme $\pi(U-1/2) \sim \mathcal{U}[-\frac{\pi}{2},\frac{\pi}{2}]$ et comme la fonction tan est périodique de période π, pour $f : \mathbb{R} \to \mathbb{R}$ bornée,

$$\begin{aligned} \mathbb{E}(f(\tan(\pi(U-1/2))) &= \frac{1}{\pi}\int_{-\pi/2}^{0} f(\tan(\pi+x))dx + \frac{1}{\pi}\int_{0}^{\pi/2} f(\tan x)dx \\ &= \frac{1}{\pi}\int_{0}^{\pi} f(\tan x)dx = \mathbb{E}(f(\tan(\pi U))). \end{aligned}$$

Ainsi, $\tan(\pi(U-1/2)) \stackrel{\mathcal{L}}{=} \tan(\pi U)$ et

$$\boxed{X = a\tan(\pi U) \sim \mathcal{C}(a).}$$

Exemple 4.2.3. La loi exponentielle de paramètre $\lambda > 0$ est la loi de densité $\lambda e^{-\lambda z}1_{\{z>0\}}$. La fonction de répartition associée est

$$F(x) = \int_{-\infty}^{x} \lambda e^{-\lambda z}1_{\{z>0\}}dz = 1_{\{x>0\}}(1-e^{-\lambda x}).$$

Pour $u \in]0,1[$, on a

$$F(x) = u \Leftrightarrow x = -\frac{1}{\lambda}\ln(1-u).$$

On en déduit que si $U \sim \mathcal{U}[0,1]$, $-\frac{1}{\lambda}\ln(1-U) \sim \mathcal{E}(\lambda)$.
On peut apporter une simplification en remarquant que $1-U \stackrel{\mathcal{L}}{=} U$ ($1-U$ a même loi que U), ce qui entraîne que $-\frac{1}{\lambda}\ln(1-U) \stackrel{\mathcal{L}}{=} -\frac{1}{\lambda}\ln(U)$ et permet de conclure que

$$\boxed{S = -\frac{1}{\lambda}\ln(U) \sim \mathcal{E}(\lambda).}$$

4.2.3 Méthode polaire pour la loi normale centrée réduite

Proposition 4.2.4. *Soit R de loi exponentielle de paramètre $1/2$ et Θ de loi uniforme sur $[0,2\pi]$ indépendantes. Alors*

$$X = \sqrt{R}\ \cos(\Theta) \quad \text{et} \quad Y = \sqrt{R}\ \sin(\Theta)$$

sont des variables normales centrées réduites (densité $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$) indépendantes.

Démonstration : On applique la méthode de la fonction muette. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ bornée. On a

$$\begin{aligned}\mathbb{E}(f(X,Y)) &= \mathbb{E}(f(\sqrt{R}\,\cos(\Theta), \sqrt{R}\,\sin(\Theta)))\\ &= \frac{1}{4\pi}\int_0^{+\infty}\int_0^{2\pi} f(\sqrt{r}\,\cos(\theta), \sqrt{r}\,\sin(\theta))e^{-r/2}drd\theta.\end{aligned}$$

Le changement de variable $(x,y) = \varphi(r,\theta) = (\sqrt{r}\,\cos(\theta), \sqrt{r}\,\sin(\theta))$ est une bijection C^1 ainsi que son inverse de $]0,+\infty[\times]0,2\pi[$ sur $\mathbb{R}^2 \setminus \{(x,0) : x \geq 0\}$. Sa matrice jacobienne est

$$\frac{D(x,y)}{D(r,\theta)} = \begin{pmatrix} \cos(\theta)/(2\sqrt{r}) & -\sqrt{r}\,\sin(\theta) \\ \sin(\theta)/(2\sqrt{r}) & \sqrt{r}\,\cos(\theta)\end{pmatrix},$$

ce qui entraîne que $dxdy = \frac{1}{2}drd\theta$. On conclut par la formule de changement de variables que

$$\mathbb{E}(f(X,Y)) = \frac{1}{2\pi}\int_{\mathbb{R}^2} f(x,y)e^{-\frac{x^2+y^2}{2}}dxdy.$$

□

D'après ce qui précède si (U_1, U_2) est un couple de variables uniformes sur $[0,1]$ indépendantes : $(-2\ln(U_1), 2\pi U_2) \stackrel{\mathcal{L}}{=} (R,\Theta)$, ce qui entraîne que

$$\left(\sqrt{-2\ln(U_1)}\,\cos(2\pi U_2), \sqrt{-2\ln(U_1)}\,\sin(2\pi U_2))\right) \stackrel{\mathcal{L}}{=} (X,Y).$$

On conclut donc que

$$\boxed{\left\{\begin{array}{l}\sqrt{-2\ln(U_1)}\,\cos(2\pi U_2)\\ \sqrt{-2\ln(U_1)}\,\sin(2\pi U_2)\end{array}\right. \quad \text{sont 2 variables normales centrées réduites indépendantes.}}$$

On en déduit facilement par changement de variable que

Corollaire 4.2.5. *Soit $\mu, \sigma \in \mathbb{R}$ et U_1, U_2 deux variables aléatoires uniformes sur $[0,1]$ indépendantes. Alors $\mu + \sigma\sqrt{-2\ln(U_1)}\cos(2\pi U_2) \sim \mathcal{N}_1(\mu, \sigma^2)$.*

4.2.4 Méthode du rejet

On souhaite simuler une variable aléatoire qui possède la densité p sur $\mathbb{R}^d$ dans le cas où il existe une densité q sur $\mathbb{R}^d$ suivant laquelle on sait simuler et une constante $k > 0$ t.q.

$$\forall x \in \mathbb{R}^d,\ p(x) \leq kq(x).$$

En intégrant cette inégalité sur $\mathbb{R}^d$ on obtient que nécessairement $k \geq 1$.
On se donne sur Ω une suite $(Y_i, U_i)_{i\geq 1}$ I.I.D. avec Y_1 de densité q et U_1 uniforme sur $[0,1]$ indépendantes.

Le principe de la méthode du rejet est le suivant :

1. Tant que $kU_i(\omega) > \frac{p(Y_i)}{q(Y_i)}(\omega)$ cela signifie que le rapport au point $Y_i(\omega)$ entre la densité p suivant laquelle on souhaite simuler et la densité q de Y_i est en certain sens trop petit et on rejette le couple d'indice i (d'où le nom de la méthode) et on passe au couple d'indice $i+1$.

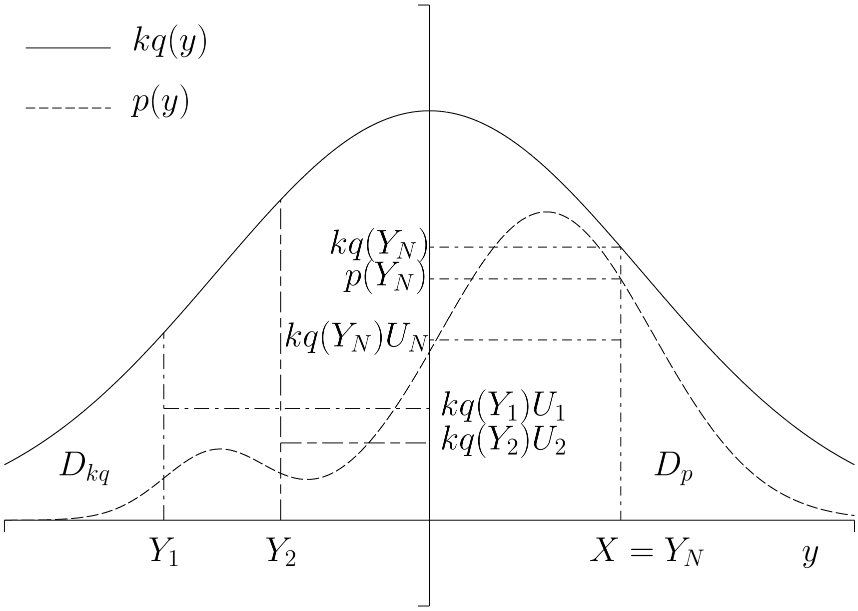


FIGURE 4.1 – Méthode du rejet.

2. Si $kq(Y_i)U_i(\omega) \leq p(Y_i)(\omega)$, on pose $X(\omega) = Y_i(\omega)$.

Cette méthode est illustrée sur la figure 4.1. D'un point de vue plus formalisé on pose

$$N(\omega) = \inf\{i \geq 1 : \; kq(Y_i)U_i(\omega) \leq p(Y_i)(\omega)\} \;\text{ et }\; X(\omega) = Y_{N(\omega)}(\omega).$$

Théorème 4.2.6. *La variable aléatoire N suit la loi géométrique de paramètre $1/k$ (elle est donc finie avec probabilité 1). Elle est indépendante du couple $(Y_N, kq(Y_N)U_N)$ qui est uniformément réparti sur $D_p = \{(x,z) \in \mathbb{R}^d \times \mathbb{R} : \; 0 \leq z \leq p(x)\}$.*
En particulier $X = Y_N$ possède la densité p.

Démonstration : Commençons par déterminer la loi de $(Y_i, kq(Y_i)U_i)$. Pour cela on se donne une fonction muette $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ bornée :

$$\begin{aligned}\mathbb{E}(f(Y_i, kq(Y_i)U_i)) &= \int_{\mathbb{R}^d} \left(\int_0^1 f(y, kq(y)u)q(y)du \right) dy = \int_{\mathbb{R}^d} \left(\frac{1}{k} \int_0^{kq(y)} f(y,z)dz \right) dy \\ &= \frac{1}{k} \int_{\mathbb{R}^d} \int_{\mathbb{R}} f(y,z) 1_{\{0 \leq z \leq kq(y)\}} dydz. \qquad (4.1)\end{aligned}$$

Ainsi $(Y_i, kq(Y_i)U_i))$ est une variable aléatoire uniformément répartie sur

$$D_{kq} = \{(y,z) \in \mathbb{R}^d \times \mathbb{R} : \; 0 \leq z \leq kq(y)\}$$

qui a pour volume k ($\int_{\mathbb{R}^d} q(y)dy = 1$).

L'événement $kq(Y_i)U_i \leq p(Y_i)$ s'écrit également $(Y_i, kq(Y_i)U_i) \in D_p$ où $D_p = \{(x,z) \in$

$\mathbb{R}^d \times \mathbb{R} : 0 \leq z \leq p(x)\} \subset D_{kq}$ a pour volume 1.
On en déduit que la probabilité pour que $kq(Y_i)U_i \leq p(Y_i)$ est égale au rapport des volumes soit $1/k$. Ainsi la variable aléatoire N est le temps de premier succès dans une suite d'expériences indépendantes avec probabilité de succès $1/k$ et suit la loi géométrique de paramètre $1/k$. En particulier, elle est finie avec probabilité 1 c'est-à-dire que

$$\mathbb{P}\left(\sum_{n\in\mathbb{N}^*} 1_{\{N=n\}} = 1\right) = 1. \tag{4.2}$$

On se donne maintenant des fonctions muettes bornées $g : \mathbb{N}^* \to \mathbb{R}$ et $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$. Pour traiter le caractère aléatoire de l'indice N dans le calcul de $\mathbb{E}(g(N)f(Y_N, kq(Y_N)U_N))$, on utilise (4.2) qui permet de décomposer l'espérance suivant les valeurs prises par N :

$$\begin{aligned}
\mathbb{E}\left(g(N)f(Y_N, kq(Y_N)U_N)\right) &= \mathbb{E}\left(\sum_{n\in\mathbb{N}^*} 1_{\{N=n\}} g(N)f(Y_N, kq(Y_N)U_N)\right) \\
&= \sum_{n\in\mathbb{N}^*} \mathbb{E}\left(1_{\{N=n\}} g(n)f(Y_n, kq(Y_n)U_n)\right) \quad \text{par linéarité de l'espérance} \\
&= \sum_{n\in\mathbb{N}^*} g(n)\mathbb{E}\left(1_{\{kq(Y_1)U_1 > p(Y_1)\}} \dots 1_{\{kq(Y_{n-1})U_{n-1} > p(Y_{n-1})\}} 1_{\{kq(Y_n)U_n \leq p(Y_n)\}} f(Y_n, kq(Y_n)U_n)\right) \\
&= \sum_{n\in\mathbb{N}^*} g(n)\mathbb{P}(kq(Y_1)U_1 > p(Y_1))^{n-1} \mathbb{E}\left(1_{\{kq(Y_n)U_n \leq p(Y_n)\}} f(Y_n, kq(Y_n)U_n)\right)
\end{aligned}$$

car les couples (Y_i, U_i) sont I.I.D.. En utilisant (4.1), on en déduit que

$$\begin{aligned}
&\mathbb{E}(g(N)f(Y_N, kq(Y_N)U_N)) \\
&\qquad = \sum_{n\in\mathbb{N}^*} g(n)\left(1-\frac{1}{k}\right)^{n-1} \frac{1}{k}\int_{\mathbb{R}^d}\int_{\mathbb{R}} 1_{\{0\leq z\leq p(y)\}} f(y,z) 1_{\{0\leq z\leq kq(y)\}} dzdy \\
&\qquad = \sum_{n\in\mathbb{N}^*} g(n)\frac{1}{k}\left(1-\frac{1}{k}\right)^{n-1} \int_{\mathbb{R}^d}\int_{\mathbb{R}} f(y,z) 1_{\{0\leq z\leq p(y)\}} dzdy.
\end{aligned}$$

Pour $g \equiv 1$, on obtient que $\mathbb{E}\left(f(Y_N, kq(Y_N)U_N)\right) = \int_{\mathbb{R}^d}\int_{\mathbb{R}} f(y,z)1_{\{0\leq z\leq p(y)\}}dzdy$, ce qui assure que le couple $(Y_N, kq(Y_N)U_N)$ est uniformément réparti sur D_p. On en déduit la densité de $X = Y_N$ par la formule des densités marginales :

$$\int_{\mathbb{R}} 1_{\{0\leq z\leq p(x)\}} dz = p(x).$$

En outre,

$$\mathbb{E}\left(g(N)f(Y_N, kq(Y_N)U_N)\right) = \mathbb{E}(g(N))\mathbb{E}\left(f(Y_N, kq(Y_N)U_N)\right)$$

et le couple $(Y_N, kq(Y_N)U_N)$ est indépendant de N d'après la Proposition 3.3.11. □

Remarque 4.2.7. Le nombre moyen de tirages (Y_i, U_i) nécessaires pour générer X est $\mathbb{E}(N) = k$. Pour diminuer le temps de calcul, on a donc bien sûr intérêt à privilégier le choix d'une densité q telle $\sup_{x\in\mathbb{R}^d} \frac{p(x)}{q(x)}$ soit aussi petit que possible et à poser $k = \sup_{x\in\mathbb{R}^d} \frac{p(x)}{q(x)}$.

Exercice résolu 4.2.8. *On souhaite simuler une variable aléatoire qui possède la loi gamma de paramètres* $3/2$ *et* 1 *i.e. qui possède la densité*

$$p(x) = \frac{2}{\sqrt{\pi}}\sqrt{x}\, e^{-x} 1_{\{x\geq 0\}}$$

par la méthode du rejet avec $q(x) = \lambda e^{-\lambda x} 1_{\{x \geq 0\}}$. Quel est le choix de λ optimal ? Que vaut alors $\mathbb{E}(N)$?

On veut trouver $\lambda^* \in]0, +\infty[$ qui minimise $f(\lambda) = \sup_{x \geq 0} \frac{\sqrt{x} e^{(\lambda-1)x}}{\lambda}$.
Pour $\lambda \geq 1$, $f(\lambda) = +\infty$. Déterminons maintenant $f(\lambda)$ pour $\lambda \in]0, 1[$. Le supremum en x est réalisé en $x(\lambda)$ qui annule la dérivée i.e. qui vérifie

$$\left(\frac{1}{2\sqrt{x}} + (\lambda - 1)\sqrt{x} \right) e^{(\lambda-1)x} = 0,$$

ce qui entraîne $x(\lambda) = \frac{1}{2(1-\lambda)}$.
Ainsi $\forall \lambda \in]0, 1[$, $f(\lambda) = 1/\sqrt{2e\lambda^2(1-\lambda)}$ et pour minimiser $f(\lambda)$ il suffit de maximiser $\lambda^2(1-\lambda)$ ce qui conduit à $\lambda^* = 2/3$.
En prenant $k = \frac{2}{\sqrt{\pi}} f(2/3) = 3^{3/2}/\sqrt{2\pi e}$, on a $\mathbb{E}(N) = k \sim 1,257$.

4.3 Exercices

Exercice 4.3.1. Pour $x \in \mathbb{R}$, on note $\lfloor x \rfloor$ la partie entière de x.
Soit X une variable exponentielle de paramètre $\lambda > 0$. Quelle est la loi de $N = 1 + \lfloor X \rfloor$?
Pour $p \in]0, 1[$, en déduire λ t.q.

$$1 + \lfloor -\ln(U)/\lambda \rfloor \sim \mathcal{G}eo(p).$$

Exercice 4.3.2.
1. Soit $(X_i)_{i \geq 1}$ une suite de variables exponentielles de paramètre $\lambda > 0$ indépendantes. Pour $n \in \mathbb{N}^*$, calculer $\mathbb{P}(X_1 + \ldots + X_n \leq 1 < X_1 + \ldots + X_{n+1})$. Quelle est la loi de

$$N = \sum_{n \in \mathbb{N}^*} n 1_{\{X_1 + \ldots + X_n \leq 1 < X_1 + \ldots + X_{n+1}\}} ?$$

2. En déduire que si $(U_i)_{i \geq 1}$ est une suite de variables I.I.D. suivant la loi uniforme sur $[0, 1]$,

$$\inf\{n \in \mathbb{N} : \; U_1 \times U_2 \times \ldots \times U_{n+1} < e^{-\lambda}\}$$

suit la loi de Poisson de paramètre λ.

Exercice 4.3.3. Soit (X, Y) un vecteur uniformément réparti sur le disque D centré en 0 et de rayon 1 (densité $1_{\{x^2+y^2 \leq 1\}}/\pi$).

1. Quelle est la loi de

$$(R, \Theta) = \left(X^2 + Y^2 \;\; , \;\; 2 \arctan \left(\frac{Y}{X + \sqrt{X^2 + Y^2}} \right) \right) ?$$

2. Soit

$$(Z, W) = \left(X \sqrt{\frac{-2\ln(X^2 + Y^2)}{X^2 + Y^2}}, Y \sqrt{\frac{-2\ln(X^2 + Y^2)}{X^2 + Y^2}} \right).$$

Exprimer (Z, W) en fonction de (R, Θ) et en déduire sans calcul la loi de (Z, W).

3. Comment simuler un couple uniformément réparti sur le disque D à partir d'une suite I.I.D. de couples uniformément répartis sur le carré $[-1,1]^2$. En déduire une méthode pour simuler des variables normales centrées réduites qui ne fait pas appel aux fonctions trigonométriques.

Exercice 4.3.4. Soit $a, b > 0$.

1. Montrer que si (U, V) suit la loi uniforme sur le domaine

$$D = \{(u, v) \in \mathbb{R}^2,\ u > 0,\ v > 0,\ u^{\frac{1}{a}} + v^{\frac{1}{b}} < 1\},$$

(i.e. (U, V) possède la densité $1_{\{(u,v)\in D\}}/|D|$ où $|D|$ est la surface de D) alors

$$X = \frac{U^{\frac{1}{a}}}{U^{\frac{1}{a}} + V^{\frac{1}{b}}}$$

suit la loi béta de paramètres a et b.
Indication : il suffit de montrer que X possède une densité proportionnelle à $1_{\{0<x<1\}} x^{a-1}(1-x)^{b-1}$.

2. En déduire une méthode pour simuler une variable aléatoire de loi béta.

Exercice corrigé 4.3.5. Soit $((X_i, Y_i))_{i\geq 1}$ une suite I.I.D. avec X_1 et Y_1 variables exponentielles de paramètre 1 indépendantes. On se donne également indépendamment de cette suite une variable ε t.q. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$ et on pose

$$N = \inf\{i \geq 1 :\ 2Y_i \geq (1 - X_i)^2\} \ \text{ et } \ Z = \varepsilon X_N.$$

1. Quelle est la loi de N ? Donner $\mathbb{E}(N)$.
2. Quelle est la loi de Z ?
3. En déduire une méthode pour simuler une variable distribuée suivant la loi normale centrée réduite.

Exercice 4.3.6. Soient $a, \theta > 0$ et $(U_i)_{i\geq 1}$ une suite I.I.D. suivant la loi uniforme sur $[0, 1]$. On note respectivement $\lfloor a \rfloor$ et $\alpha = a - \lfloor a \rfloor$ la partie entière et la partie fractionnaire de a.

1. Soit Y une variable aléatoire distribuée suivant la loi $\Gamma(\alpha, 1)$ (par convention $Y = 0$ si $\alpha = 0$) indépendante de la suite $(U_i)_{i\geq 1}$. Vérifier que $X = \frac{1}{\theta}\left(Y - \ln\left(\prod_{i=1}^{\lfloor a \rfloor} U_i\right)\right)$ suit la loi $\Gamma(a, \theta)$.
2. On suppose $\alpha > 0$. Vérifier qu'il est possible de simuler Y par rejet sous la densité $q(y) = \frac{\alpha e}{\alpha + e}\left(y^{\alpha-1} 1_{\{0<y\leq 1\}} + e^{-y} 1_{\{y>1\}}\right)$ avec un nombre d'itérations d'espérance égale à $\frac{\alpha+e}{\alpha e \Gamma(\alpha)}$. Expliciter comment la méthode d'inversion de la fonction de répartition permet de simuler suivant la densité q.

4.4 Résumé

Dans tout ce résumé, on note U une variable de loi uniforme sur $[0,1]$ et $(U_i)_{i\geq 1}$ une suite de variables I.I.D. suivant cette loi.

Loi	**Méthode de simulation**
Bernoulli $\mathcal{B}(p)$	$1_{\{U\leq p\}}$
binomiale $\mathcal{B}(n,p)$	$\sum_{i=1}^{n} 1_{\{U_i\leq p\}}$
géométrique $\mathcal{G}eo(p)$	$1+\lfloor \ln(U)/\ln(1-p)\rfloor$ où $\lfloor x\rfloor$ désigne la partie entière de x
Poisson $\mathcal{P}(\lambda)$	$\min\{n\in\mathbb{N} :\ U_1\times U_2\times\ldots\times U_{n+1}\leq e^{-\lambda}\}$
uniforme $\mathcal{U}[a,b]$	$a+(b-a)U$
exponentielle $\mathcal{E}(\lambda)$	$-\frac{1}{\lambda}\ln(U)$
Cauchy $\mathcal{C}(a)$	$a\tan(\pi U)$
normale $\mathcal{N}_1(\mu,\sigma^2)$	$\mu+\sigma\sqrt{-2\ln(U_1)}\ \cos(2\pi U_2)$

Chapitre 5

Convergence et théorèmes limites

Dans ce chapitre, nous allons introduire différentes notions de convergence pour une suite de variables aléatoires. Après avoir étudié les liens entre ces différentes notions, nous énoncerons deux théorèmes limites très importants des probabilités :

- la loi forte des grands nombres qui énonce la convergence de la moyenne empirique $\frac{1}{n}\sum_{j=1}^{n} X_j$ d'une suite $(X_j)_{j\geq 1}$ de variables aléatoires réelles indépendantes, identiquement distribuées et intégrables vers $\mathbb{E}(X_1)$ lorsque $n \to +\infty$.

- le théorème de la limite centrale qui indique à quelle vitesse cette convergence a lieu sous l'hypothèse supplémentaire que les X_j sont de carré intégrable.

5.1 Convergence

Définition 5.1.1. *Pour $n \to +\infty$, on dit qu'une suite $(X_n)_{n\geq 1}$ de variables aléatoires à valeurs $\mathbb{R}^d$ converge vers la variable aléatoire X à valeurs $\mathbb{R}^d$*

- *presque sûrement si $\mathbb{P}(X_n \longrightarrow X) = \mathbb{P}(\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\}) = 1$, c'est-à-dire si les fonctions X_n définies sur Ω convergent ponctuellement sur un sous-ensemble de Ω de probabilité 1 vers la fonction X.*
- *en probabilité si $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| \geq \varepsilon) \longrightarrow_{n\to+\infty} 0$.*
- *dans L^1 si les variables X_n, X sont intégrables et $\mathbb{E}|X_n - X| \longrightarrow_{n\to+\infty} 0$.*
- *dans L^2 (ou en moyenne quadratique) si les variables X_n, X sont de carré intégrable et $\mathbb{E}(|X_n - X|^2) \longrightarrow_{n\to+\infty} 0$.*

Remarque 5.1.2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge dans L^1 vers X. Alors $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$. En effet, comme $X_n - X \leq |X_n - X|$, par linéarité et croissance de l'espérance, $\mathbb{E}(X_n) - \mathbb{E}(X) = \mathbb{E}(X_n - X) \leq \mathbb{E}|X_n - X|$. Par symétrie, $\mathbb{E}(X) - \mathbb{E}(X_n) \leq \mathbb{E}|X_n - X|$. Ainsi $|\mathbb{E}(X_n) - \mathbb{E}(X)| \leq \mathbb{E}|X_n - X|$, ce qui permet de conclure.

Pour établir les liens entre ces différents types de convergence, nous utiliserons les deux résultats suivants. Le premier qui porte le nom de théorème de convergence dominée et que nous ne démontrerons pas est l'analogue du théorème de convergence dominée pour les intégrales.

Théorème 5.1.3 (de convergence dominée)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge presque sûrement vers X. On suppose de plus que la suite est dominée au sens où il existe une variable aléatoire Y* **intégrable** *telle que*

$$\forall n \geq 1,\ \mathbb{P}(|X_n| \leq Y) = 1.$$

Alors X est intégrable et $(X_n)_n$ converge dans L^1 vers X ce qui entraîne en particulier que

$$\boxed{\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X).}$$

Proposition 5.1.4 (Quelques inégalités)**.** *Soit X et Y deux variables aléatoires réelles.*
Inégalité de Markov : *Si $\mathbb{E}|X| < +\infty$, alors*

$$\boxed{\forall a > 0,\ \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}|X|}{a}.}$$

Inégalité de Bienaymé-Tchebychev : *Si $\mathbb{E}(X^2) < +\infty$, alors*

$$\boxed{\forall a > 0,\ \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(X^2)}{a^2}.}$$

Inégalité de Cauchy-Schwarz : *Si les variables X et Y sont de carré intégrable, alors*

$$\boxed{|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)}\ \sqrt{\mathbb{E}(Y^2)}.}$$

Démonstration : Comme $\forall x \in \mathbb{R},\ 1_{\{|x|\geq a\}} \leq \frac{|x|}{a}$ (voir la figure 5.1),

$$\mathbb{P}\left(1_{\{|X|\geq a\}} \leq \frac{|X|}{a}\right) = 1$$

et on obtient l'inégalité de Markov en utilisant la propriété de croissance de l'espérance et l'égalité $\mathbb{E}(1_{\{|X|\geq a\}}) = \mathbb{P}(|X| \geq a)$.
L'inégalité de Bienaymé-Tchebychev s'obtient par la même méthode en remarquant que $\forall x \in \mathbb{R},\ 1_{\{|x|\geq a\}} \leq \frac{x^2}{a^2}$ (voir la figure 5.1).
Pour montrer l'inégalité de Cauchy-Schwarz, on constate que pour tout $\lambda \in \mathbb{R}$,

$$\mathbb{E}(X^2) + 2\lambda\mathbb{E}(XY) + \lambda^2\mathbb{E}(Y^2) = \mathbb{E}((X + \lambda Y)^2) \geq 0.$$

On en déduit que le discriminant $4(\mathbb{E}(XY))^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2)$ de ce polynôme du second degré en λ est négatif ou nul, ce qui fournit le résultat souhaité. □

Proposition 5.1.5.
- *La convergence L^2 implique la convergence L^1 qui elle-même implique la convergence en probabilité.*
- *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge dans L^2 vers X. Alors $\mathbb{E}(X_n)$, $\mathbb{E}(X_n^2)$ et $\mathrm{Var}(X_n)$ convergent resp. vers $\mathbb{E}(X)$, $\mathbb{E}(X^2)$ et $\mathrm{Var}(X)$.*
- *La convergence presque sûre entraîne la convergence en probabilité.*

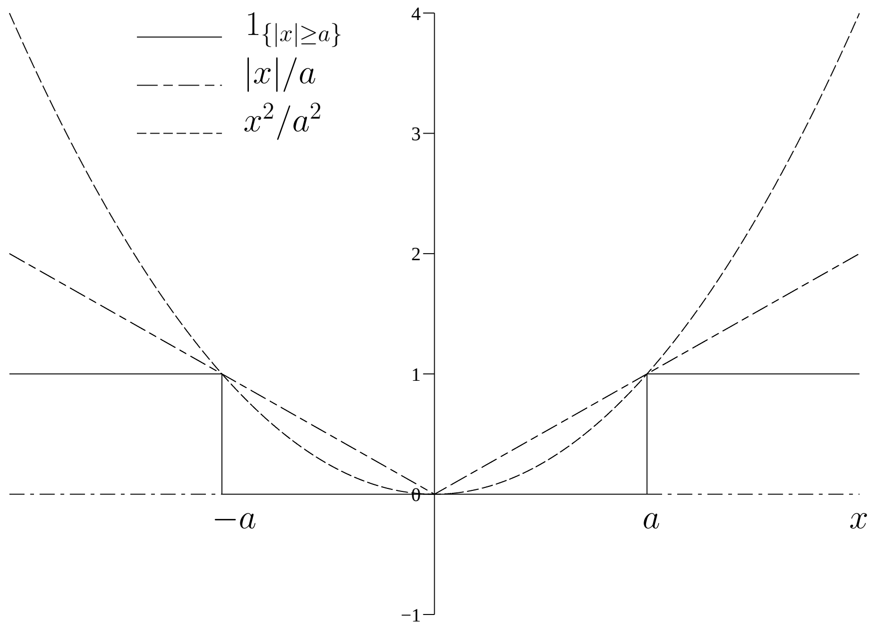


FIGURE 5.1 – Majoration de $1_{\{|x|\geq a\}}$.

Démonstration : *Convergence $L^2 \Rightarrow$ convergence L^1 :*
Cela découle du fait que toute variable aléatoire de carré intégrable est intégrable et de l'inégalité $\mathbb{E}|X_n - X| \leq \sqrt{\mathbb{E}(|X_n - X|^2)}$ (qui se déduit elle-même par exemple de la positivité de la variance de $|X_n - X|$).

Convergence $L^1 \Rightarrow$ convergence en probabilité :
Cela découle de l'inégalité de Markov $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\varepsilon}$ pour $\varepsilon > 0$.

Convergence $L^2 \Rightarrow$ convergence des espérances et variances :
D'après le premier point et la remarque 5.1.2, il suffit de vérifier que $\mathbb{E}(X_n^2)$ converge vers $\mathbb{E}(X^2)$.
Par l'inégalité de Cauchy-Schwarz, $\mathbb{E}(X_n X) \leq \sqrt{\mathbb{E}(X_n^2)}\sqrt{\mathbb{E}(X^2)}$.
Donc

$$\mathbb{E}((X_n - X)^2) = \mathbb{E}(X_n^2) - 2\mathbb{E}(X_n X) + \mathbb{E}(X^2) \geq (\sqrt{\mathbb{E}(X_n^2)} - \sqrt{\mathbb{E}(X^2)})^2.$$

Ainsi $\sqrt{\mathbb{E}(X_n^2)}$ converge vers $\sqrt{\mathbb{E}(X^2)}$ et on conclut par continuité de $x \to x^2$.

Convergence presque sûre $\Rightarrow$ convergence en probabilité :
Soit $(X_n)_n$ une suite qui converge presque sûrement vers X et $\varepsilon > 0$. La suite $|X_n - X|$ converge presque sûrement vers 0.
Comme la fonction $1_{\{|x|\geq\varepsilon\}}$ est continue en 0, on en déduit que $Y_n = 1_{\{|X_n - X|\geq\varepsilon\}}$ converge presque sûrement vers 0. Les variables Y_n sont dominées par la constante 1 qui est bien sûr intégrable. Donc le théorème de convergence dominée 5.1.3 implique que

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

□

5.2 Lois des grands nombres

Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles indépendantes indentiquement distribuées (ce que l'on note aussi I.I.D.). Les lois des grands nombres portent sur le comportement de la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ lorsque $n \to +\infty$.

5.2.1 Loi faible des grands nombres

On suppose que les X_j sont de carré intégrable (i.e. que $\mathbb{E}(X_1^2) < +\infty$).

$$\begin{aligned}
\mathbb{E}\left((\bar{X}_n - \mathbb{E}(X_1))^2\right) &= \frac{1}{n^2}\mathbb{E}\left(\left(\sum_{j=1}^n X_j - \mathbb{E}\left(\sum_{j=1}^n X_j\right)\right)^2\right) = \frac{1}{n^2}\mathrm{Var}(\sum_{j=1}^n X_j) \\
&= \frac{1}{n^2}\sum_{j=1}^n \mathrm{Var}(X_j) \quad \text{par indépendance des } X_j, \\
&= \frac{\mathrm{Var}(X_1)}{n} \longrightarrow_{n\to+\infty} 0
\end{aligned}$$

On en déduit

Proposition 5.2.1. *La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ d'une suite $(X_j)_{j\geq 1}$ de variables aléatoires réelles I.I.D. et de carré intégrable converge dans L^2 (et donc dans L^1 et en probabilité) vers l'espérance commune des X_j lorsque $n \to +\infty$.*

5.2.2 Loi forte des grands nombres

Le résultat précédent est supplanté par la loi forte des grands nombres qui indique que la convergence a lieu presque sûrement. La convergence presque sûre est illustrée par la figure 5.2.

Théorème 5.2.2 (Loi forte des grands nombres)**.** *La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ d'une suite $(X_j)_{j\geq 1}$ de variables aléatoires réelles I.I.D.* **intégrables** *converge presque sûrement et dans L^1 vers l'espérance commune des X_j lorsque $n \to +\infty$.*

$$\boxed{\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n X_j \to \mathbb{E}(X_1)\right) = 1.}$$

Démonstration : La démonstration de ce résultat est délicate (voir par exemple [3] p. 52) et nous nous contenterons de montrer la convergence presque sûre sous l'hypothèse

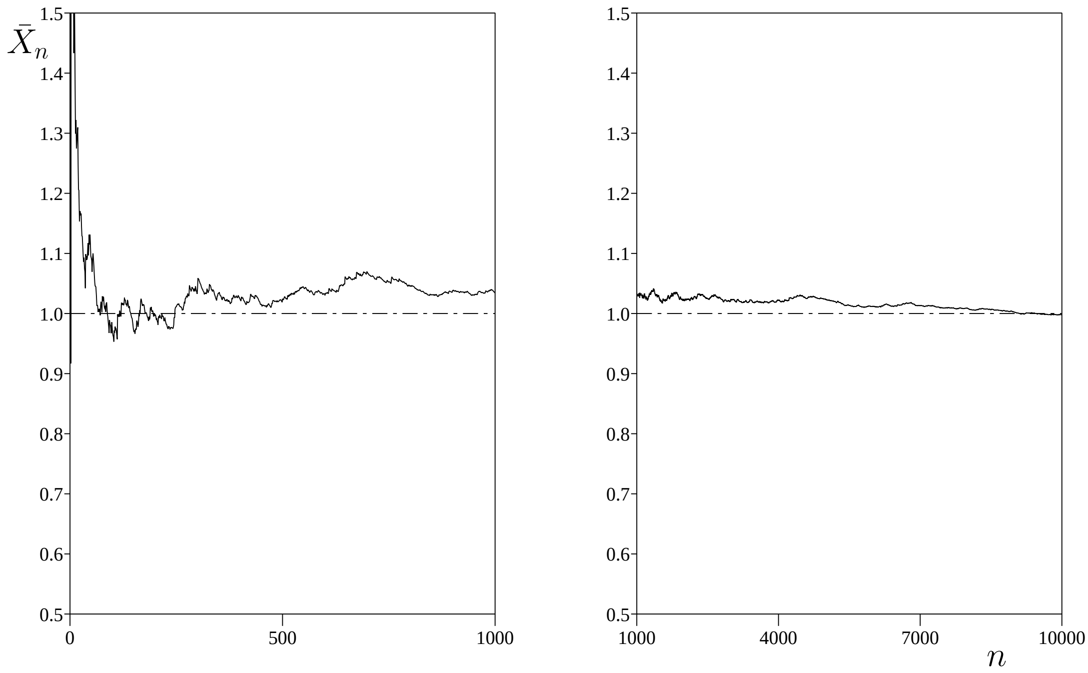


FIGURE 5.2 – Évolution d'une réalisation de $\bar{X}_n$ avec n lorsque les X_j sont I.I.D. suivant la loi exponentielle de paramètre 1 ($\mathbb{E}(X_1) = 1$).

forte $\mathbb{E}(X_1^4) < +\infty$.
Par linéarité de l'espérance,

$$\mathbb{E}\left((\bar{X}_n - \mathbb{E}(X_1))^4\right)$$
$$= \frac{1}{n^4} \sum_{i_1,i_2,i_3,i_4=1}^{n} \mathbb{E}\left((X_{i_1} - \mathbb{E}(X_{i_1}))(X_{i_2} - \mathbb{E}(X_{i_2}))(X_{i_3} - \mathbb{E}(X_{i_3}))(X_{i_4} - \mathbb{E}(X_{i_4}))\right).$$

Parmi les n^4 espérances qui figurent dans la somme au second membre beaucoup sont nulles : en effet dès que l'un des quatre indices i_k est différent des trois autres, par indépendance des X_j, l'espérance est égale à l'espérance du produit des termes correspondant aux trois autres indices par $\mathbb{E}(X_{i_k} - \mathbb{E}(X_{i_k})) = 0$.
Il reste donc seulement les contributions des cas où
- les 4 indices sont égaux soit n termes en $\mathbb{E}((X_1 - \mathbb{E}(X_1))^4)$
- 2 des indices prennent une valeur et les 2 autres une autre valeur soit $3n(n-1)$ termes en $\mathbb{E}\left((X_1 - \mathbb{E}(X_1))^2(X_2 - \mathbb{E}(X_2))^2\right) = (\mathrm{Var}(X_1))^2$ (3 pour le choix de l'indice i_k, $k = 2, 3$ ou 4, égal au premier indice i_1, $n(n-1)$ pour le choix de deux valeurs différentes parmi n).

Ainsi

$$\mathbb{E}((\bar{X}_n - \mathbb{E}(X_1))^4 = \frac{1}{n^3}\mathbb{E}((X_1 - \mathbb{E}(X_1))^4) + \frac{3(n-1)}{n^3}(\mathrm{Var}(X_1))^2.$$

On en déduit qu'il existe $C > 0$ t.q. $\forall n \geq 1$, $\mathbb{E}\Big((\bar{X}_n - \mathbb{E}(X_1))^4\Big) \leq \frac{C}{n^2}$ puis que

$$\mathbb{E}\left(\sum_{n\geq 1}((\bar{X}_n - \mathbb{E}(X_1))^4\right) = \sum_{n\geq 1}\mathbb{E}\left(((\bar{X}_n - \mathbb{E}(X_1))^4\right) < +\infty.$$

$$\text{Donc } \mathbb{P}\left(\sum_{n\geq 1}(\bar{X}_n - \mathbb{E}(X_1))^4 < +\infty\right) = 1.$$

Ainsi avec probabilité 1, $((\bar{X}_n - \mathbb{E}(X_1))^4)_n$ converge vers 0 comme terme général d'une série convergente. □

Lorsque la variance commune des X_j est non nulle, $\mathbb{E}\left((\bar{X}_n - \mathbb{E}(X_1))^2\right) = \mathrm{Var}(X_1)/n$ n'est pas le terme général d'une série sommable et on ne peut pas reprendre la démonstration précédente de la convergence presque sûre de $\bar{X}_n$ en supposant seulement les X_j de carré intégrable. Néanmoins, il est possible de l'adapter et cela fait l'objet de l'exercice suivant. Démontrer cette convergence presque sûre en supposant seulement les X_j intégrables est beaucoup plus délicat.

Exercice 5.2.3. On suppose $\mathbb{E}(X_1^2) < +\infty$.

1. Montrer que $\bar{X}_{n^2}$ converge presque sûrement vers $\mathbb{E}(X_1)$ lorsque $n \to +\infty$.
2. Vérifier que
$$\max_{n^2+1\leq k\leq (n+1)^2-1} |\bar{X}_{n^2} - \bar{X}_k| \leq \frac{2}{n}|\bar{X}_{n^2}| + \frac{1}{n^2}\sum_{j=n^2+1}^{(n+1)^2-1}|X_j|.$$
3. En remarquant que $\left(\frac{1}{n^2}\sum_{j=n^2+1}^{(n+1)^2-1}|X_j|\right)^2 \leq \frac{2}{n^3}\sum_{j=n^2+1}^{(n+1)^2-1}X_j^2$, vérifier que $\frac{1}{n^2}\sum_{j=n^2+1}^{(n+1)^2-1}|X_j|$ converge presque sûrement vers 0 lorsque $n \to +\infty$.
4. Conclure que $\bar{X}_k$ converge presque sûrement vers $\mathbb{E}(X_1)$ lorsque $k \to +\infty$.

Exercice 5.2.4. Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles I.I.D. et de carré intégrable, $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ et $V_n = \frac{1}{n}\sum_{j=1}^n X_j^2 - (\bar{X}_n)^2$.

1. Montrer que V_n converge presque sûrement vers $\mathrm{Var}(X_1)$.
2. Montrer que $\mathbb{E}(V_n) = \frac{n-1}{n}\mathrm{Var}(X_1)$. Comme l'espérance de V_n est différente de $\mathrm{Var}(X_1)$, on dit que V_n est un estimateur biaisé de $\mathrm{Var}(X_1)$. En revanche, $S_n^2 = \frac{n}{n-1}V_n$ est un estimateur sans biais.

5.3 Fonction caractéristique et convergence en loi

5.3.1 Fonction caractéristique

Définition 5.3.1. *Soit X un vecteur aléatoire à valeurs $\mathbb{R}^d$. On appelle fonction caractéristique de X la fonction*

$$\boxed{\Phi_X : u \in \mathbb{R}^d \to \Phi_X(u) = \mathbb{E}\left(e^{iu.X}\right) \in \mathbb{C}.}$$

Remarque 5.3.2. – $\Phi_X(0,\ldots,0) = 1$.
- $\forall u \in \mathbb{R}^d,\ \Phi_X(-u) = \overline{\Phi_X(u)}$.
- La fonction caractéristique de X ne dépend que de la loi de X :

$$X \overset{\mathcal{L}}{=} Y \Longrightarrow \Phi_X \equiv \Phi_Y.$$

- Si Φ_X est intégrable au sens où $\int_{\mathbb{R}^d} |\Phi_X(u)|du < +\infty$, alors X possède la densité obtenue par inversion de Fourier :

$$x \in \mathbb{R}^d \to p(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iu.x}\Phi_X(u)du.$$

Le tableau suivant donne les fonctions caractéristiques associées aux lois réelles usuelles :

Loi	**fonction caractéristique**
Bernoulli $\mathcal{B}(p)$	$(1-p) + pe^{iu}$
binomiale $\mathcal{B}(n,p)$	$((1-p) + pe^{iu})^n$
géométrique $\mathcal{G}eo(p)$	$\frac{pe^{iu}}{1-(1-p)e^{iu}}$
Poisson $\mathcal{P}(\lambda)$	$\exp(\lambda(e^{iu}-1))$
uniforme $\mathcal{U}[a,b]$	$\frac{\sin((b-a)u/2)}{(b-a)u/2}\ e^{iu\frac{a+b}{2}}$
exponentielle $\mathcal{E}(\lambda)$	$\frac{\lambda}{\lambda-iu}$
Cauchy $\mathcal{C}(a)$	$e^{-a\|u\|}$
gaussienne $\mathcal{N}_1(\mu,\sigma^2)$	$e^{iu\mu-\frac{\sigma^2u^2}{2}}$

Pour les six premières lois, le calcul est un exercice sans difficulté (le faire). Le cas de la loi de Cauchy est traité un peu plus loin dans l'exercice 5.3.4.

Calcul de Φ_X pour $X \sim \mathcal{N}_1(\mu,\sigma^2)$:

On commence par vérifier par changement de variable que $G = \frac{1}{\sigma}(X-\mu) \sim \mathcal{N}_1(0,1)$. Comme

$$\Phi_X(u) = \mathbb{E}\left(e^{iuX}\right) = \mathbb{E}\left(e^{iu(\sigma G+\mu)}\right) = e^{iu\mu}\mathbb{E}\left(e^{iu\sigma G}\right) = e^{iu\mu}\Phi_G(u\sigma),$$

il suffit de calculer Φ_G pour en déduire Φ_X.
On a

$$\Phi_G(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux}e^{-\frac{x^2}{2}}dx.$$

Par dérivation sous le signe intégral, on en déduit que

$$\begin{aligned}
\Phi'_G(u) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ixe^{iux} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(-i \int_{\mathbb{R}} (iu - x) e^{iux} e^{-\frac{x^2}{2}} dx - u \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2}} dx \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(-i \left[e^{iux - \frac{x^2}{2}} \right]_{-\infty}^{+\infty} - u \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2}} dx \right) \\
&= -u\Phi_G(u).
\end{aligned}$$

On résout donc l'équation différentielle $\Phi'_G(u) = -u\Phi_G(u)$ avec la condition initiale $\Phi_G(0) = 1$ (cf remarque 5.3.2). On obtient $\Phi_G(u) = e^{-\frac{u^2}{2}}$ puis on en déduit Φ_X.

Nous avons remarqué que la fonction caractéristique d'une variable aléatoire ne dépend que de sa loi. Il s'avère en fait que comme son nom le laisse penser :

Proposition 5.3.3. *La fonction caractéristique caractérise la loi c'est à dire que*

$$\forall u \in \mathbb{R}^d,\ \Phi_X(u) = \Phi_Y(u) \Longrightarrow X \stackrel{\mathcal{L}}{=} Y \text{ i.e. } \forall f : \mathbb{R}^d \to \mathbb{R} \text{ bornée } \mathbb{E}(f(X)) = \mathbb{E}(f(Y)).$$

Nous ne démontrerons pas ce résultat qui intuitivement vient du fait que les exponentielles complexes $x \to e^{iu.x}$ forment un ensemble *suffisamment gros* de fonctions bornées pour que lorsque l'égalité $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$ est vérifiée pour toute fonction f dans cet ensemble alors elle l'est pour toute fonction f bornée. Dans le cas où Φ_X est intégrable, c'est une conséquence de la formule d'inversion donnée dans la remarque 5.3.2. Nous renvoyons à [2] p.139 ou [9] p.107 pour une démonstration dans le cas général.

Exercice 5.3.4. Soit T une variable aléatoire exponentielle de paramètre $a > 0$ et ε une variable indépendante t.q. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. On pose

$$X = \varepsilon T.$$

1. Déterminer la loi de X (on pourra décomposer suivant les valeurs prises par ε).
2. Calculer sa fonction caractéristique Φ_X.
3. En appliquant la formule d'inversion donnée dans la remarque 5.3.2, en déduire la fonction caractéristique d'une variable aléatoire qui suit la loi de Cauchy $\mathcal{C}(a)$.
4. En déduire la loi de la moyenne empirique $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^n Y_j$ d'une suite $(Y_j)_{j\geq 1}$ de variables aléatoires I.I.D. suivant la loi de Cauchy $\mathcal{C}(a)$.

Le résultat suivant qui illustre le fait que la régularité de la fonction caractéristique d'une variable aléatoire est reliée à l'intégrabilité de cette variable, nous servira pour établir le théorème de la limite centrale.

Lemme 5.3.5. *Soit X une variable aléatoire réelle de carré intégrale (i.e. telle que $\mathbb{E}(X^2) < +\infty$). Alors sa fonction caractéristique Φ_X est C^2 et admet le développement limité suivant en 0 :*

$$\Phi_X(u) = 1 + iu\mathbb{E}(X) - \frac{u^2}{2}\mathbb{E}(X^2) + o(u^2).$$

Démonstration : Soit $u \in \mathbb{R}$.

$$\frac{1}{h}\left(e^{i(u+h)X} - e^{iuX}\right) \xrightarrow{h\to 0} iXe^{iuX} \text{ presque sûrement.}$$

Comme $\forall a, b \in \mathbb{R},\ |e^{ia} - e^{ib}| \leq |b-a|$,

$$\left|\frac{1}{h}\left(e^{i(u+h)X} - e^{iuX}\right)\right| \leq |X|.$$

La variable aléatoire X étant de carré intégrable, $|X|$ est intégrable et par le théorème de convergence dominée 5.1.3, on obtient que

$$\lim_{h\to 0} \mathbb{E}\left(\frac{1}{h}\left(e^{i(u+h)X} - e^{iuX}\right)\right) = \mathbb{E}(iXe^{iuX}).$$

Ainsi $\lim_{h\to 0} \frac{1}{h}(\Phi_X(u+h) - \Phi_X(u)) = \mathbb{E}(iXe^{iuX})$ et la fonction Φ_X est dérivable de dérivée $\Phi'_X(u) = \mathbb{E}(iXe^{iuX})$. En appliquant à nouveau le théorème de convergence dominée, on peut montrer que Φ'_X est continue et même C^1 avec $\Phi''_X(u) = -\mathbb{E}(X^2 e^{iuX})$.
Ainsi Φ_X est C^2 avec $\Phi'_X(0) = i\mathbb{E}(X)$ et $\Phi''_X(0) = -\mathbb{E}(X^2)$. Le développement limité en 0 s'en déduit par exemple en utilisant la formule de Taylor avec reste intégral. □

5.3.2 Convergence en loi

Définition 5.3.6. *On dit que la suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X à valeurs $\mathbb{R}^d$ et on note $X_n \xrightarrow{\mathcal{L}} X$ si*

$$\boxed{\forall f : \mathbb{R}^d \to \mathbb{R} \ \textbf{continue bornée},\ \mathbb{E}(f(X_n)) \longrightarrow_{n\to+\infty} \mathbb{E}(f(X)).}$$

Exemple 5.3.7.
- Pour $n \in \mathbb{N}^*$, on suppose que $\forall 1 \leq k \leq n$, $\mathbb{P}(U_n = k/n) = 1/n$. Soit $f : \mathbb{R} \to \mathbb{R}$ continue bornée. La convergence des sommes de Riemann vers l'intégrale entraîne que

  $$\mathbb{E}(f(U_n)) = \frac{1}{n}\sum_{k=1}^{n} f(k/n) \to_{n\to+\infty} \int_0^1 f(u)du = \mathbb{E}(f(U)),$$

  où U est une variable uniforme sur $[0,1]$. Ainsi la suite $(U_n)_n$ converge en loi vers $U \sim \mathcal{U}[0,1]$.
- Soit maintenant pour $n \in \mathbb{N}^*$, X_n une variable aléatoire uniformément répartie sur $[0, 1/n]$. Alors pour f continue bornée,

  $$\mathbb{E}(f(X_n)) = n\int_0^{1/n} f(x)dx \to_{n\to+\infty} f(0).$$

  Donc la suite $(X_n)_n$ converge en loi vers X telle que $\mathbb{P}(X = 0) = 1$.

- Pour $n \in \mathbb{N}$, soit T_n une variable exponentielle de paramètre $\lambda_n > 0$. On suppose que la suite $(\lambda_n)_n$ converge vers $\lambda > 0$. Alors comme pour $f : \mathbb{R} \to \mathbb{R}$ continue bornée,

$$\forall n \in \mathbb{N},\ \forall x \geq 0,\ |f(x)\lambda_n e^{-\lambda_n x}| \leq g(x) = |f(x)|(\sup_n \lambda_n)e^{-(\inf_n \lambda_n)x},$$

  où la fonction g est intégrable sur $[0, +\infty[$, par le théorème de convergence dominée,

$$\mathbb{E}(f(T_n)) = \int_0^{+\infty} f(x)\lambda_n e^{-\lambda_n x} dx$$

  converge lorsque $n \to +\infty$ vers $\int_0^{+\infty} f(x)\lambda e^{-\lambda x} dx = \mathbb{E}(f(T))$ où T suit la loi exponentielle de paramètre λ. Ainsi $(T_n)_n$ converge en loi vers $T \sim \mathcal{E}(\lambda)$.

Propriétés 5.3.8. *Soit $(X_n)_n$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ qui converge en loi vers X et $\varphi : \mathbb{R}^d \to \mathbb{R}^q$ une fonction continue. Alors la suite $(\varphi(X_n))_n$ converge en loi vers $\varphi(X)$.*

Démonstration : Soit $g : \mathbb{R}^q \to \mathbb{R}$ continue bornée. La fonction $g \circ \varphi : \mathbb{R}^d \to \mathbb{R}$ est continue bornée. Donc la convergence en loi de $(X_n)_n$ vers X entraîne que

$$\lim_{n\to+\infty} \mathbb{E}(g(\varphi(X_n)) = \mathbb{E}(g(\varphi(X))).$$

□

Du fait que les exponentielles complexes constituent à nouveau un sous-ensemble assez gros de l'ensemble des fonctions continues bornées (pour se ramener à des fonctions à valeurs réelles, il suffit de séparer partie réelle et partie imaginaire), on a la caractérisation suivante de la convergence en loi

Théorème 5.3.9. *La suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X à valeurs $\mathbb{R}^d$ si et seulement si la fonction caractéristique de X_n converge ponctuellement vers la fonction caractéristique de X i.e.*

$$\boxed{X_n \xrightarrow{\mathcal{L}} X \iff \forall u \in \mathbb{R}^d,\ \Phi_{X_n}(u) \to \Phi_X(u).}$$

Pour la démonstration de ce théorème, nous renvoyons à [2] p.179 ou à [9] p.163.

Corollaire 5.3.10. *Si la suite $(X_n)_n$ converge en probabilité vers X alors elle converge en loi vers X.*

Démonstration : Soit $u \in \mathbb{R}^d$ et $\varepsilon > 0$. On a

$$\begin{aligned}\left|e^{iu.X_n} - e^{iu.X}\right| &= \left|e^{iu.X_n} - e^{iu.X}\right| 1_{\{|X_n - X| \geq \varepsilon\}} + \left|e^{iu.X_n} - e^{iu.X}\right| 1_{\{|X_n - X| < \varepsilon\}} \\ &\leq 2 \times 1_{\{|X_n - X| \geq \varepsilon\}} + \left|e^{iu.X_n} - e^{iu.X}\right| 1_{\{|X_n - X| < \varepsilon\}}.\end{aligned}$$

Comme $\forall a, b \in \mathbb{R},\ |e^{ia} - e^{ib}| \leq |b - a|$, on en déduit que

$$\left|e^{iu.X_n} - e^{iu.X}\right| \leq 2 \times 1_{\{|X_n - X| \geq \varepsilon\}} + |u|\varepsilon 1_{\{|X_n - X| < \varepsilon\}} \leq 2 \times 1_{\{|X_n - X| \geq \varepsilon\}} + |u|\varepsilon$$

puis que

$$|\Phi_{X_n}(u) - \Phi_X(u)| = |\mathbb{E}(e^{iu.X_n} - e^{iu.X})| \leq \mathbb{E}|e^{iu.X_n} - e^{iu.X}| \leq |u|\varepsilon + 2\mathbb{P}(|X_n - X| \geq \varepsilon).$$

Le premier terme du second membre est arbitrairement petit (uniformément en n) pour ε petit tandis qu'à ε fixé le second terme converge vers 0 lors $n \to +\infty$ par définition de la convergence en probabilité. Ainsi $\forall u \in \mathbb{R}^d$, $\lim_{n\to+\infty} \Phi_{X_n}(u) = \Phi_X(u)$ et on conclut en utilisant le théorème 5.3.9. □

Remarque 5.3.11. Si la suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X, alors

$$\mathbb{E}(f(X_n)) \longrightarrow_{n\to+\infty} \mathbb{E}(f(X))$$

pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ bornée dont l'ensemble des points de discontinuité $\mathcal{D}_f$ vérifie $\mathbb{P}(X \in \mathcal{D}_f) = 0$.

Théorème 5.3.12 (de Slutsky)**.** *Soit $(X_n, Y_n)_n$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ telle que $(X_n)_n$ converge en loi (ou en probabilité ou presque sûrement) vers une constante $a \in \mathbb{R}^{d_1}$ et $(Y_n)_n$ converge en loi vers Y. Alors $(X_n, Y_n)_n$ converge en loi vers (a, Y). En particulier lorsque $d_1 = d_2 = 1$, $(X_nY_n)_n$ converge en loi vers aY et lorsque $d_1 = d_2$, $(X_n + Y_n)_n$ converge en loi vers $a + Y$.*

Démonstration : Nous allons utiliser la caractérisation de la convergence en loi donnée par le théorème 5.3.9. Soit $(u, v) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

$$\begin{aligned}|\Phi_{(X_n,Y_n)}(u, v) - \Phi_{(a,Y)}(u, v)| &= \left|\mathbb{E}\left((e^{iu.X_n} - e^{iu.a})e^{iv.Y_n}\right) + e^{iu.a}\mathbb{E}(e^{iv.Y_n} - e^{iv.Y})\right| \\ &\leq \mathbb{E}\left|e^{iu.X_n} - e^{iu.a}\right| + |\Phi_{Y_n}(v) - \Phi_Y(v)|.\end{aligned}$$

La convergence en loi de Y_n vers Y entraîne que le second terme converge vers 0 lorsque $n \to +\infty$. La continuité et la bornitude de $x \in \mathbb{R}^{d_1} \to f(x) = |e^{iu.x} - e^{iu.a}|$ et la convergence en loi de X_n vers a entraîne que le premier terme converge vers $f(a) = 0$, ce qui permet de conclure que (X_n, Y_n) converge en loi vers (a, Y).
On en déduit les cas particuliers par la propriété 5.3.8 en remarquant que $(x, y) \in \mathbb{R}\times\mathbb{R} \to xy$ et $(x, y) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \to x + y$ sont des fonctions continues. □

Exercice résolu 5.3.13. *Soit $\lambda > 0$ et pour tout $n \geq \lambda$, X_n une variable géométrique de paramètre λ/n. Montrer que les variables $Y_n = X_n/n$ convergent en loi vers une limite que l'on précisera.*
On se donne $U \sim \mathcal{U}[0, 1]$. D'après l'exercice 4.3.1, $1 + [\frac{\ln(U)}{\ln(1-\lambda/n)}] \stackrel{\mathcal{L}}{=} X_n$.
Donc $\frac{1}{n}(1 + [\frac{\ln(U)}{\ln(1-\lambda/n)}]) \stackrel{\mathcal{L}}{=} Y_n$. Pour $n \to +\infty$, $\frac{\ln(U)}{\ln(1-\lambda/n)}$ est équivalent à $-\frac{n}{\lambda}\ln(U)$. On en déduit que pour $n \to +\infty$, $\frac{1}{n}(1 + [\frac{\ln(U)}{\ln(1-\lambda/n)}])$ converge presque sûrement et donc en probabilité et en loi (voir proposition 5.1.5 et corollaire 5.3.10) vers $-\frac{1}{\lambda}\ln(U)$. Pour $f : \mathbb{R} \to \mathbb{R}$ continue bornée, on a donc

$$\mathbb{E}(f(Y_n)) = \mathbb{E}\left(f\left(\frac{1}{n}\left(1 + \left[\frac{\ln(U)}{\ln(1 - \lambda/n)}\right]\right)\right)\right) \longrightarrow_{n\to+\infty} \mathbb{E}\left(f\left(-\frac{1}{\lambda}\ln(U)\right)\right).$$

D'après l'exemple 4.2.3, $-\frac{1}{\lambda}\ln(U) \sim \mathcal{E}(\lambda)$. On conclut donc que la suite $(Y_n)_n$ converge en loi vers Y qui suit la loi exponentielle de paramètre λ.
Il faut du recul pour pouvoir trouver la solution qui précède mais on peut parvenir à la même conclusion sans astuce en calculant la fonction caractéristique de Y_n et en utilisant le théorème 5.3.9 (le faire).

5.4 Le théorème de la limite centrale

5.4.1 Enoncé et preuve du résultat

Le théorème de la limite centrale donne la vitesse à laquelle la convergence a lieu dans loi forte des grands nombres, sous l'hypothèse supplémentaire d'intégrabilité des X_j^2 :

Théorème 5.4.1 (de la limite centrale)**.** *Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées telles que $\mathbb{E}(X_1^2) < +\infty$ et $\sigma = \sqrt{\mathrm{Var}(X_1)} > 0$. On note $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ la moyenne empirique.*
Alors pour $n \to +\infty$,

$$\boxed{\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Y \sim \mathcal{N}_1(0,1).}$$

Cette convergence en loi est illustrée par la figure 5.3.

Remarque 5.4.2.
– Formellement, on récrit $\bar{X}_n \overset{\mathcal{L}}{\simeq} \mathbb{E}(X_1) + \frac{\sigma}{\sqrt{n}} Y$, ce qui indique que la convergence dans la loi forte des grands nombres a lieu à la vitesse $\frac{\sigma}{\sqrt{n}}$. On peut aussi en déduire que la loi de $\sum_{i=1}^n X_i = n\bar{X}_n$ est proche de $\mathcal{N}_1(n\mathbb{E}(X_1), n\sigma^2)$.
– Ce résultat explique la place fondamentale des variables aléatoires gaussiennes (ou normales) en probabilités et en statistiques.
– Notons que la renormalisation effectuée est telle que

$$\forall n \in \mathbb{N}^*,\ \mathrm{Var}\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1))\right) = \frac{n}{\sigma^2}\mathrm{Var}((\bar{X}_n - \mathbb{E}(X_1))) = \frac{1}{n\sigma^2}\sum_{j=1}^n \mathrm{Var}(X_j) = 1.$$

Démonstration : On utilise la fonction caractéristique. Soit $u \in \mathbb{R}$.

$$\begin{aligned}
\Phi_{\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1))}(u) &= \mathbb{E}\left(e^{iu\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n (X_j - \mathbb{E}(X_j))}\right) \\
&= \prod_{j=1}^n \mathbb{E}\left(e^{iu\frac{1}{\sigma\sqrt{n}}(X_j - \mathbb{E}(X_j))}\right) \quad \text{par indépendance des } X_j, \\
&= \left(\mathbb{E}\left(e^{iu\frac{1}{\sigma\sqrt{n}}(X_1 - \mathbb{E}(X_1))}\right)\right)^n \quad \text{car les } X_j \text{ ont même loi,} \\
&= \left(\Phi_{X_1 - \mathbb{E}(X_1)}\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n.
\end{aligned}$$

Comme $\mathbb{E}(X_1 - \mathbb{E}(X_1)) = 0$ et $\mathbb{E}((X_1 - \mathbb{E}(X_1))^2) = \sigma^2$, d'après le lemme 5.3.5, pour v au voisinage de 0, $\Phi_{X_1 - \mathbb{E}(X_1)}(v) = 1 - \frac{\sigma^2}{2}v^2 + o(v^2)$. Donc pour n grand,

$$\Phi_{X_1 - \mathbb{E}(X_1)}\left(\frac{u}{\sigma\sqrt{n}}\right) = 1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right).$$

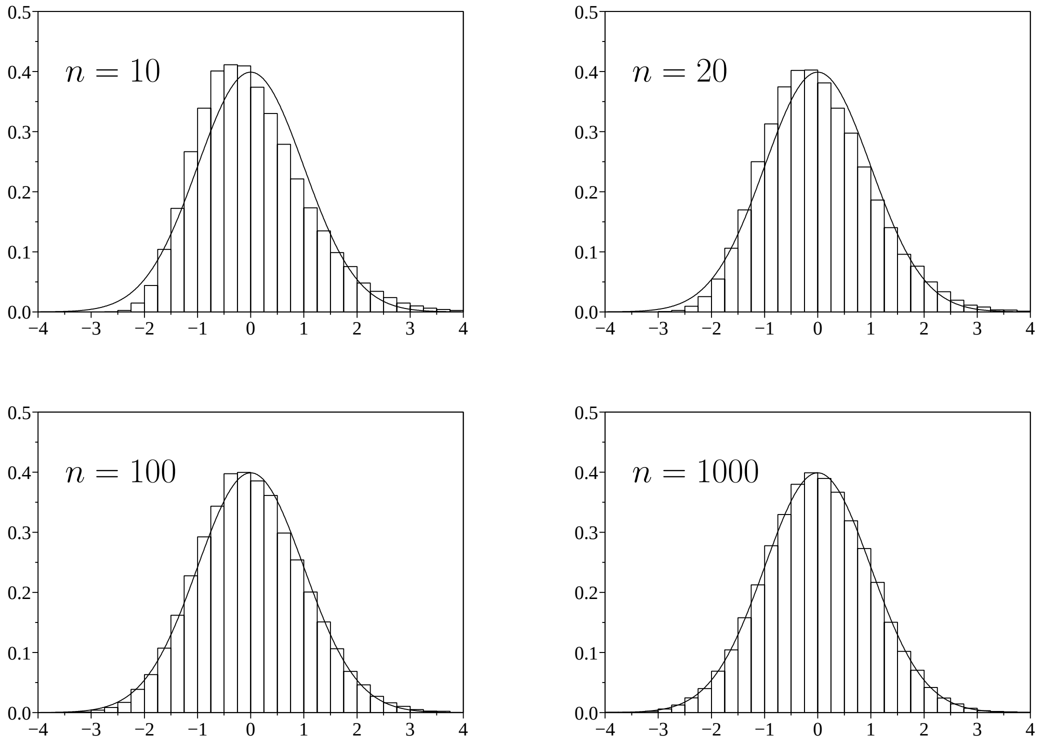


FIGURE 5.3 – Comparaison de la densité de la loi normale centrée réduite avec des histogrammes de 50 000 réalisations de $\sqrt{n}(\bar{X}_n - 1)$ lorsque les X_j sont I.I.D. suivant la loi exponentielle de paramètre 1 ($\mathbb{E}(X_1) = \mathrm{Var}(X_1) = 1$).

Ainsi

$$\Phi_{\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1))}(u) = \left(1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \longrightarrow_{n\to+\infty} e^{-\frac{u^2}{2}} = \Phi_Y(u).$$

On conclut par le théorème 5.3.9. $\square$

Remarque 5.4.3. À la fin de la preuve précédente, le terme $o\left(\frac{1}{n}\right)$ est de la forme $\frac{\varepsilon_n}{n}$ où la suite $(\varepsilon_n)_n$ telle que $\lim_{n\to+\infty} |\varepsilon_n| = 0$ est à valeurs complexes. Pour justifier proprement le passage à la limite, on peut remarquer que pour $n \geq \frac{u^2}{2}$,

$$\left|\left(1 - \frac{u^2}{2n} + \frac{\varepsilon_n}{n}\right)^n - \left(1 - \frac{u^2}{2n}\right)^n\right| = \left|\sum_{k=1}^{n} \binom{n}{k} \frac{\varepsilon_n^k}{n^k}\left(1 - \frac{u^2}{2n}\right)^{n-k}\right| \leq \sum_{k=1}^{n} \frac{n!}{n^k(n-k)!} \times \frac{|\varepsilon_n|^k}{k!}$$
$$\leq \sum_{k\geq 1} \frac{|\varepsilon_n|^k}{k!} = 1 - e^{|\varepsilon_n|} \longrightarrow_{n\to+\infty} 0.$$

5.4.2 Intervalle de confiance dans la méthode de Monte-Carlo

Le principe de la méthode de Monte-Carlo est le suivant : pour évaluer numériquement l'espérance d'une variable aléatoire réelle X intégrable, on génère une réalisation d'une suite $(X_j)_{j\geq 1}$ de variables aléatoires indépendantes et de même loi que X (par exemple en utilisant les techniques de simulation présentées dans le chapitre 4) et on approche $\mathbb{E}(X)$ par la moyenne emprique $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$.

La loi forte des grands nombres justifie la convergence de cette méthode. Comme nous allons le voir, si X est de carré intégrable, le théorème de la limite centrale permet de se faire une idée de la précision avec laquelle $\mathbb{E}(X)$ est évaluée.

Si on se donne $a > 0$, la fonction $f(x) = 1_{\{|x|\leq a\}}$ est bornée et continue en dehors des points $-a$ et a. Malgré ces discontinuités, comme pour $Y \sim \mathcal{N}_1(0,1)$, $\mathbb{P}(Y = -a) = \mathbb{P}(Y = a) = 0$, d'après la remarque 5.3.11, la convergence en loi de $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X))$ vers Y entraîne la convergence de

$$\mathbb{P}\left(\bar{X}_n - \frac{\sigma a}{\sqrt{n}} \leq \mathbb{E}(X) \leq \bar{X}_n + \frac{\sigma a}{\sqrt{n}}\right) = \mathbb{E}\left(f\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X))\right)\right)$$

vers $\mathbb{E}(f(Y))$ pour $n \to +\infty$.
C'est pourquoi dès que n est supérieur à 20, on fait l'approximation

$$\mathbb{P}\left(\bar{X}_n - \frac{\sigma a}{\sqrt{n}} \leq \mathbb{E}(X) \leq \bar{X}_n + \frac{\sigma a}{\sqrt{n}}\right) \simeq \mathbb{E}(f(Y)) = \mathbb{E}(1_{\{|Y|\leq a\}}) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-\frac{x^2}{2}}dx.$$

On utilise le plus couramment les valeurs de a pour lesquelles l'intégrale vaut $0,95$ ou $0,99$ soit respectivement $a = 1,96$ ou $a = 2,58$.

Ainsi la probabilité pour que la valeur $\mathbb{E}(X)$ que l'on cherche à calculer se trouve dans l'intervalle $\left[\bar{X}_n - \frac{1{,}96\sigma}{\sqrt{n}}, \bar{X}_n + \frac{1{,}96\sigma}{\sqrt{n}}\right]$ (resp. $\left[\bar{X}_n - \frac{2{,}58\sigma}{\sqrt{n}}, \bar{X}_n + \frac{2{,}58\sigma}{\sqrt{n}}\right]$) est de l'ordre de $0,95$ (resp. $0,99$). On parle alors d'intervalle de confiance à 95% (resp. 99%).

Dans la pratique, le plus souvent, la valeur de l'écart type σ est inconnue. En effet calculer $\mathrm{Var}(X) = \sigma^2$ est un problème au moins aussi compliqué que de calculer $\mathbb{E}(X)$. On remplace donc σ par $\sqrt{V_n}$ où $V_n = \frac{1}{n}\sum_{j=1}^n X_j^2 - (\bar{X}_n)^2$ converge presque sûrement vers $\mathrm{Var}(X)$ d'après l'exercice 5.2.4. Pour évaluer V_n, il est simplement nécessaire de stocker la somme des carrés des X_i en plus de la somme des X_i au cours de la simulation : cela ne demande que très peu de calculs supplémentaires.

On obtient alors l'intervalle de confiance à 95% :

$$\boxed{\left[\bar{X}_n - \frac{1,96\sqrt{V_n}}{\sqrt{n}}, \bar{X}_n + \frac{1,96\sqrt{V_n}}{\sqrt{n}}\right]}.$$

Lorsque l'on emploie la méthode de Monte-Carlo, il faut systématiquement déterminer un intervalle de confiance puisque cela permet de se faire une idée de la précision de la valeur obtenue tout en nécessitant très peu de calculs supplémentaires. Bien sûr, dans le cas où la largeur de l'intervalle $\frac{2a\sqrt{V_n}}{\sqrt{n}}$ est grande par rapport à $|\bar{X}_n|$ cela signifie qu'il ne faut pas accorder beaucoup de crédit à la valeur $\bar{X}_n$ calculée. On peut alors augmenter le nombre n de tirages pour augmenter la précision.

Remarque 5.4.4. On a $\sqrt{\frac{n}{V_n}}(\bar{X}_n - \mathbb{E}(X_1)) = \frac{\sigma}{\sqrt{V_n}}\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1))$. Sous les hypothèses du théorème 5.4.1, $\frac{\sigma}{\sqrt{V_n}}$ converge presque sûrement vers 1 tandis que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1))$ converge en loi vers $Y \sim \mathcal{N}_1(0,1)$. On déduit alors du théorème de Slutsky 5.3.12 que $\sqrt{\frac{n}{V_n}}(\bar{X}_n - \mathbb{E}(X_1))$ converge en loi vers Y, ce qui justifie l'approximation de σ par $\sqrt{V_n}$ faite plus haut dans la donnée de l'intervalle de confiance.

5.5 Exercices

Exercice corrigé 5.5.1. On considère une machine qui tourne à une cadence de n cycles par unité de temps et qui a une probabilité $p \in]0, 1[$ de tomber en panne lors de chaque cycle et ce indépendamment des autres cycles. Soit N le numéro du cycle où se produit la première panne.

1. Quelle est la loi de N ? Que représente $T = N/n$?
   Soit τ l'espérance de T. Exprimer τ en fonction de p et n.
2. Dans le cas où n est grand, on souhaite approcher la variable discrète T par une variable aléatoire à densité. Pour cela on va effectuer le passage à la limite $n \to +\infty$ en choisissant $p(n)$ de façon à ce que l'espérance de $T(n)$ reste égale à τ.
   (a) Que vaut $p(n)$?
   (b) Calculer la fonction caractéristique de $T(n)$.
   (c) En déduire que cette variable converge en loi vers une limite que l'on précisera.

Exercice 5.5.2. Soit Y une variable uniformément répartie sur $[-1, 1]$, Z une variable normale centrée réduite et $(X_n)_{n\geq 1}$ une suite de variables indépendantes t.q.

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

$$\text{On pose} \quad Y_n = \sum_{k=1}^{n} \frac{X_k}{2^k} \qquad Z_n = \frac{1}{n} \sum_{k=1}^{n} \sqrt{2k}\, X_k.$$

1. Calculer la fonction caractéristique de Y_n. En remarquant que
   $$\sin(\lambda/2^n) \prod_{k=1}^{n} \cos(\lambda/2^k) = \sin(\lambda)/2^n,$$
   en déduire que la suite Y_n converge en loi vers Y.
2. Calculer la fonction caractéristique Φ_{Z_n} de Z_n. Donner le développement limité à l'ordre 2 à l'origine de la fonction $x \in]-\frac{\pi}{2}, \frac{\pi}{2}[\to \ln(\cos(x))$. En déduire pour $u \in \mathbb{R}$ la limite de $\ln(\Phi_{Z_n}(u))$ lorsque n tend vers l'infini et conclure que la suite Z_n converge en loi vers Z.

Exercice 5.5.3. Soit U_i une suite de variables I.I.D. suivant la loi uniforme sur $[0, 1]$ et $\alpha > 0$.

1. Montrer que la suite $X_n = (U_1 \times \ldots \times U_n)^{\frac{\alpha}{n}}$ converge presque sûrement et donner sa limite.
   Indication : on pourra s'intéresser à $\ln(X_n)$.
2. On pose $Y_n = e^{\alpha\sqrt{n}}(U_1 \times \ldots \times U_n)^{\alpha/\sqrt{n}}$.
   Quel est le comportement de la suite $\frac{1}{\alpha}\ln(Y_n)$ pour $n \to +\infty$?
   En déduire que la suite Y_n converge en loi et déterminer la loi de sa limite.

Exercice corrigé 5.5.4. Chaque jour, dans une certaine ville, 100 personnes ont besoin d'un examen radioscopique. Pour préserver le libre choix, n centres d'imagerie sont installés dans cette ville. On admet que les patients choisissent indifféremment l'un ou l'autre centre d'imagerie.
Soit N le nombre de clients journaliers dans un centre d'imagerie.

1. Quelle est la probabilité qu'un client choisisse le centre d'imagerie considéré ?
2. Montrer que N peut s'écrire $N = \sum_{i=1}^{i=100} X_i$ où les $(X_i)_{1\leq i\leq n}$ sont n variables aléatoires indépendantes et distribuées suivant la loi de Bernoulli de même paramètre p que l'on déterminera.
3. Quelle est la loi de N ?
4. On donne que si Y suit la loi normale centrée réduite, alors $\mathbb{P}(Y \leq 2) = 98\%$. En utilisant le théorème de la limite centrale, déterminer quelle capacité $c(n)$ chaque centre d'imagerie doit avoir pour être capable de répondre à la demande avec une probabilité de 98% ? Cas où $n = 2$, $n = 3$, $n = 4$?
5. Quel est le coût de la concurrence : quelle surcapacité $s(n)$ la concurrence entraîne-t-elle par rapport à une situation où chaque centre se verrait affecter un même nombre de clients ? Cas où $n = 2$, $n = 3$, $n = 4$?

Exercice 5.5.5. 1. Soit X une variable de Bernoulli de paramètre $p \in [0, 1]$. Pour quelle valeur de p la variance de X est-elle maximale ? Que vaut-elle alors ?

2. Aux Etats-Unis sur 4 000 000 de naissance annuelles, on observe un ratio de 1 048 garçons pour 1 000 filles. Donner un intervalle de confiance à 99% pour la probabilité p qu'un bébé soit un garçon. Que pensez-vous de l'hypothèse d'équilibre des naissances : $p = 0.5$?
3. Lors du second tour de l'élection présidentielle, un sondage est effectué à la sortie des urnes sur un échantillon de 1 000 personnes. Le candidat A recueille a% (a proche de 50) des suffrages des personnes interrogées. A l'aide du théorème de la limite centrale, donner un intervalle de confiance à 95% pour le score S_A réalisé par ce candidat. A partir de quelle valeur de $|a - 50|$, peut-on se baser sur le sondage pour connaître le résultat de l'élection ?

Exercice 5.5.6. On souhaite estimer une probabilité de la forme $p = \mathbb{P}(X \in A)$ avec p très petite (X désigne par exemple le vecteur des paramètres de fonctionnement d'une centrale nucléaire et A l'ensemble des valeurs de ces paramètres conduisant à la fusion du cœur de la centrale). On se donne pour cela une suite $(X_i)_{i\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi de X et on note $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \in A\}}$.

1. Quel est le comportement asymptotique de $\hat{p}_n$ pour $n \to +\infty$? Donner la variance de $\hat{p}_n$.

On suppose maintenant que l'on sait simuler indépendamment des X_i une suite $(Y_i)_{i\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi conditionnelle de X sachant $X \in B$ où B contient A. En particulier $\mathbb{P}(Y_i \in A) = \mathbb{P}(X \in A|X \in B)$. On note $\hat{q}_n = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \in B\}}$ et $\hat{r}_n = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i \in A\}}$.

2. Quel est le comportement asymptotique de $\hat{q}_n \times \hat{r}_n$ pour $n \to +\infty$? Calculer l'espérance et la variance de $\hat{q}_n \times \hat{r}_n$. Expliquer pourquoi, lorsque $\mathbb{P}(X \in B)$ et $\mathbb{P}(X \in A|X \in B)$ sont du même ordre, $\hat{q}_n \times \hat{r}_n$ approche p beaucoup plus précisément que $\hat{p}_n$.

Exercice 5.5.7. Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables indépendantes et identiquement distribuées suivant la loi de Cauchy de paramètre $a > 0$. On pose $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$.

1. Quelle est la loi de $\bar{X}_n$? La suite $(\bar{X}_n)_{n\in\mathbb{N}^*}$ converge-t-elle en loi ?
2. Quelle est la loi de $\bar{X}_{n+p} - \bar{X}_n$? Quel est le comportement de $\mathbb{P}(|\bar{X}_{2n} - \bar{X}_n| \geq 1)$ lorsque $n \to +\infty$?

3. Montrer que si la suite $(Y_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers Y alors $\forall\epsilon > 0$, $\lim_{\min(n,m)\to+\infty} \mathbb{P}(|Y_n - Y_m| \geq \epsilon) = 0$. La suite $(\bar{X}_n)_{n\in\mathbb{N}^*}$ converge-t-elle en probabilité ?

Exercice 5.5.8. Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes d'espérance 1 qui prennent deux valeurs a et b $(0 < a < 1 < b)$ avec probabilité p et $1-p$ respectivement $(0 < p < 1)$. On pose $Y_n = \prod_{i=1}^n X_i$.

1. Vérifier que pour $x \in]0,1[\cup]1,+\infty[$, $\ln(x) < x - 1$ et en déduire que $\mathbb{E}(\ln(X_1)) < 0$. En étudiant le comportement de $\frac{1}{n}\ln(Y_n)$ lorsque $n \to +\infty$, montrer que la suite Y_n converge presque sûrement vers une limite que l'on précisera. Un joueur joue tous les jours le dixième de sa fortune à pile ou face avec un ami. Quelle est l'évolution de cette fortune au bout d'un temps très long ?
2. Calculer $\mathbb{E}(Y_n)$. La suite Y_n converge-t-elle dans L^1 ?

Exercice 5.5.9. Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires de même loi que X. On suppose d'abord que X est intégrable et à valeurs dans $\mathbb{N}$.

1. Montrer que $\mathbb{E}(X) = \sum_{j=1}^{+\infty} \mathbb{P}(X \geq j)$.
2. En déduire que pour $m \in \mathbb{N}^*$, $\sum_{k=1}^{+\infty} \mathbb{P}\left(\frac{X_k}{k} \geq \frac{1}{m}\right) < +\infty$ puis que $\frac{X_k}{k}$ converge presque sûrement vers 0 lorsque k tend vers l'infini.
3. Conclure que $\frac{1}{n}\max(X_1,\ldots,X_n)$ converge presque sûrement vers 0 lorsque n tend vers l'infini.

On suppose maintenant que X est une variable aléatoire réelle t.q. $\mathbb{E}(\max(X,0)) < +\infty$.

4. Montrer que $\frac{1}{n}\max(X_1,\ldots,X_n)$ converge presque sûrement vers 0 lorsque n tend vers l'infini.

Exercice 5.5.10. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue et S_n^x une variable distribuée suivant la loi binomiale de paramètres n et x où $n \in \mathbb{N}^*$ et $x \in [0,1]$. Montrer que

$$\lim_{n\to+\infty} \sup_{x\in[0,1]} |\mathbb{E}(f(x) - f(S_n^x/n))| = 0.$$

(On pourra séparer l'espérance en deux termes suivant que $|S_n^x/n - x| > \alpha$ ou non, utiliser l'uniforme continuité de f et majorer $\mathbb{P}(|S_n^x/n - x| > \alpha)$ uniformément en x en utilisant l'inégalité de Bienaymé-Tchebychev.)
Comme $\mathbb{E}(f(S_n^x/n)) = \sum_{k=0}^n f(k/n)\binom{n}{k}x^k(1-x)^{n-k}$, toute fonction continue sur $[0,1]$ est limite uniforme d'une suite de polynômes (théorème de Weierstrass).

Exercice corrigé 5.5.11. 1. Calculer la fonction caractéristique d'une variable X suivant la loi de Poisson $\mathcal{P}(\lambda)$. Calculer $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.

2. Soit $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ deux variables de Poisson indépendantes. Quelle est la loi de leur somme ?
3. Soit $X_n \sim \mathcal{P}(n)$. Montrer que pour $n \to +\infty$, $Y_n = (X_n - n)/\sqrt{n}$ converge en loi vers une limite que l'on précisera. Commenter ce résultat à la lumière du théorème de la limite centrale.
4. Quelle est la limite de la suite $u_n = e^{-n}\sum_{k=0}^n \frac{n^k}{k!}$ lorsque $n \to +\infty$?

Exercice 5.5.12. Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles I.I.D. et de carré intégrable t.q. $\frac{X_1+X_2}{\sqrt{2}} \stackrel{\mathcal{L}}{=} X_1$.

1. Quelle est l'espérance commune des X_i ?
2. Pour $k \in \mathbb{N}^*$, que peut-on dire de la loi $\frac{X_1+\ldots+X_{2^k}}{2^{k/2}}$?
3. En déduire que les X_j sont des variables gaussiennes.

Exercice 5.5.13. Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles I.I.D. et de carré intégrable.

1. Montrer que $\left(\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1)), \frac{1}{\sqrt{n}}\sum_{j=n+1}^{2n}(X_j - \mathbb{E}(X_1))\right)$ converge en loi vers (Y_1, Y_2) avec Y_1 et Y_2 I.I.D. de loi à préciser.
2. En déduire que $\sqrt{2n}(\bar{X}_{2n} - \mathbb{E}(X_1)) - \sqrt{n}(\bar{X}_n - \mathbb{E}(X_1))$ converge en loi vers Y_1. Conclure que, lorsque $\mathrm{Var}(X_1) > 0$, $\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1))$ ne converge pas en probabilité.

Exercice 5.5.14. Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles indépendantes et indentiquement distribuées suivant une densité $p(x)$ paire, bornée, continue en 0 et telle que $p(0) > 0$

1. Quelle est la loi de $Y_k = \frac{1}{X_k}$?
2. Exprimer à l'aide de p la fonction caractéristique commune des Y_k.
3. On rappelle que $\int_0^{+\infty} \frac{1-\cos(z)}{z^2}\,dz = \frac{\pi}{2}$. Montrer que la suite $\frac{1}{n}\sum_{k=1}^n Y_k$ converge en loi vers Y qui suit la loi de Cauchy de paramètre $\pi p(0)$.
4. En déduire à l'aide de la remarque 5.3.11 que la moyenne harmonique $n/\sum_{k=1}^n \frac{1}{X_k}$ des X_k converge en loi vers une limite que l'on précisera.

Exercice 5.5.15. Soit $Y_n \sim \chi^2(n)$ pour $n \in \mathbb{N}^*$. Le but de cet exercice est de vérifier que, lorsque $n \to +\infty$, $\sqrt{2Y_n} - \sqrt{2n-1}$ converge en loi vers $G \sim \mathcal{N}_1(0,1)$, résultat qui sert en statistique pour évaluer la fonction de répartition de Y_n lorsque $n > 30$ (voir paragraphe 11.3).
Pour cela on se donne $(G_j)_{j\geq 1}$ une suite de variables aléatoires I.I.D. suivant la loi normale centrée réduite et pour $n \in \mathbb{N}^*$, on pose $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ où $X_j = G_j^2$.

1. Soit $n \in \mathbb{N}^*$. Vérifier que
$$\sqrt{2Y_n} - \sqrt{2n} \overset{\mathcal{L}}{=} 2g(\bar{X}_n) \times \sqrt{\frac{n}{2}}(\bar{X}_n - 1),$$
où
$$g(x) = \begin{cases} \frac{\sqrt{x}-1}{x-1} & \text{si } x \neq 1, \\ \frac{1}{2} & \text{sinon.} \end{cases}$$
2. Donner les comportements asymptotiques de $g(\bar{X}_n)$ et de $\sqrt{\frac{n}{2}}(\bar{X}_n - 1)$ lorsque $n \to +\infty$ et en déduire que $\sqrt{2Y_n} - \sqrt{2n} \xrightarrow{\mathcal{L}} G$.
3. Vérifier que $\sqrt{2n} - \sqrt{2n-1}$ converge vers 0 et conclure que $\sqrt{2Y_n} - \sqrt{2n-1} \xrightarrow{\mathcal{L}} G$.

Problème corrigé 5.5.16. Un automobiliste emprunte tous les jours le même trajet qui comporte un feu tricolore pour se rendre à son travail. Comme le trajet est peu encombré, lorsque le feu est rouge, l'automobiliste peut redémarrer dès que le feu passe au vert. Mais pour faire passer le temps, il se demande quelle est la durée θ pendant laquelle le feu reste rouge. On note $(X_i)_{i\geq 1}$ ses durées d'attente successives au feu lorsque celui-ci est rouge et on suppose ces variables I.I.D. suivant la loi uniforme $\mathcal{U}[0,\theta]$. Pour $n \geq 1$, on note $Z_n = \max_{1\leq i\leq n} X_i$.

1. Montrer que $\mathbb{P}(X_1 = X_2) = 0$.
2. Montrer que plus généralement $\mathbb{P}(\exists 1 \leq i \neq j \leq n, X_i = X_j) = 0$. En déduire que pour toute fonction f bornée sur $\mathbb{R}$, on a :

$$\mathbb{E}(f(Z_n)) = \sum_{i=1}^{n} \mathbb{E}\left(f(X_i) 1_{\{\forall j \neq i,\, X_j < X_i\}}\right) = n\mathbb{E}\left(f(X_1) 1_{\{X_2 < X_1, \ldots, X_n < X_1\}}\right).$$

3. En déduire que Z_n suit la densité $\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1} 1_{\{0 \leq z \leq \theta\}}$.
4. Calculer $\mathbb{E}(Z_n)$ et $\mathrm{Var}(Z_n)$. En conclure que Z_n converge presque sûrement vers θ (*indication* : on rappelle (voir preuve du théorème 5.2.2) que si $(\tilde{Z}_n)_{n \geq 1}$ est une suite de variables aléatoires telle que $\sum_{n \geq 1} \mathbb{E}(\tilde{Z}_n^2) < +\infty$, alors la suite $(\tilde{Z}_n)_{n \geq 1}$ converge presque sûrement vers 0).
5. Soit f une fonction bornée sur $\mathbb{R}$. Montrer que :

$$\mathbb{E}\left(f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right) = \int_0^n f(u)\left(1 - \frac{u}{n}\right)^{n-1} du$$

On rappelle que pour $0 < v < 1$, $\ln(1 - v) \leq -v$. En déduire que pour $n \geq 2$, $\left(1 - \frac{u}{n}\right)^{n-1} 1_{\{0 < u \leq n\}} \leq e^{-u/2} 1_{\{u > 0\}}$ et conclure que :

$$\lim_{n \to \infty} \mathbb{E}\left(f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right) = \int_0^\infty f(u) e^{-u}\, du$$

6. En déduire que $\frac{n}{\theta}(\theta - Z_n)$ converge en loi et identifier la limite.
7. Montrer que $M_n = \frac{2}{n}(X_1 + \cdots + X_n)$ converge presque sûrement vers θ. À quelle vitesse a lieu cette convergence ?
8. L'automobiliste doit-il préférer Z_n ou M_n pour estimer rapidement θ ?

5.6 Résumé

• **Diagramme des convergences :**

$$\begin{tikzcd}
\begin{array}{c} L^2 : \\ \mathbb{E}(|X_n - X|^2) \to 0 \end{array} \arrow[d, Rightarrow] & & \\
\begin{array}{c} L^1 : \\ \mathbb{E}|X_n - X| \to 0 \end{array} \arrow[dr, Rightarrow] & & \begin{array}{c} \text{presque sûre :} \\ \mathbb{P}(X_n \to X) = 1 \end{array} \arrow[dl, Rightarrow] \\
& \begin{array}{c} \text{Proba : } \forall \varepsilon > 0, \\ \mathbb{P}(|X_n - X| \geq \varepsilon) \to 0 \end{array} \arrow[d, Rightarrow] & \\
& \begin{array}{c} \text{Loi : } \forall f \textbf{continue} \text{ bornée,} \\ \mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)) \end{array} &
\end{tikzcd}$$

• **Fonction caractéristique :** $\boxed{\Phi_X(u) = \mathbb{E}(e^{iu.X}).}$

– caractérise la loi :

$$\boxed{\Phi_X \equiv \Phi_Y \Longleftrightarrow X \stackrel{\mathcal{L}}{=} Y \Longleftrightarrow \forall f \text{ bornée, } \mathbb{E}(f(X)) = \mathbb{E}(f(Y)).}$$

– caractérise la convergence en loi :

$$\boxed{\forall u,\ \Phi_{X_n}(u) \longrightarrow \Phi_X(u) \Leftrightarrow X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall f \textbf{continue} \text{ bornée, } \mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)).}$$

• **Comportement de la moyenne empirique** $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ d'une suite de variables aléatoires réelles I.I.D. $(X_j)_{j\geq 1}$:

Loi forte des grands nombres : $\boxed{\mathbb{E}|X_1| < +\infty \Longrightarrow \mathbb{P}(\bar{X}_n \longrightarrow \mathbb{E}(X_1)) = 1.}$

Théorème de la limite centrale :

$$\boxed{\mathbb{E}(X_1^2) < +\infty \text{ et } \mathrm{Var}(X_1) \neq 0 \Longrightarrow \sqrt{\frac{n}{\mathrm{Var}(X_1)}}\,(\bar{X}_n - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Y \sim \mathcal{N}_1(0,1).}$$

Chapitre 6

Vecteurs gaussiens

D'après le théorème de la limite centrale, la loi de la somme d'un grand nombre de variables aléatoires indépendantes et identiquement distribuées est proche d'une loi gaussienne (voir remarque 5.4.2). Ce résultat reste vrai sous certaines conditions lorsque les variables ne sont pas nécessairement identiquement distribuées (voir exercice 5.5.2 question 2). Ainsi lorsqu'une grandeur est obtenue en sommant des contributions indépendantes, il est naturel de la modéliser par une variable aléatoire gaussienne. Cela explique la place très importante des gaussiennes en probabilités. De manière analogue, pour la modélisation de grandeurs multidimensionnelles, l'utilisation des vecteurs gaussiens est justifiée par le théorème de la limite centrale multidimensionnel 6.2.8.

6.1 Définition, construction

6.1.1 Définition

Par extension de la définition donnée au paragraphe 3.2.2, on dit que la variable aléatoire réelle Y suit la loi gaussienne d'espérance $m \in \mathbb{R}$ et de variance nulle si $\mathbb{P}(Y = m) = 1$. On note alors $Y \sim \mathcal{N}_1(m, 0)$.

Définition 6.1.1. *On dit qu'un vecteur aléatoire $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ est un vecteur gaussien si toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne (ou normale) réelle i.e. si pour tout $u \in \mathbb{R}^d$, la variable aléatoire réelle $u.X$ est une variable aléatoire gaussienne.*

Proposition 6.1.2. *Le vecteur $X = (X_1, \ldots, X_d)$ est gaussien si et seulement si pour tout $u \in \mathbb{R}^d$, $\Phi_X(u) = e^{iu.\mu - \frac{1}{2}u.\Gamma u}$ où $\mu \in \mathbb{R}^d$ et $\Gamma \in \mathbb{R}^{d\times d}$ est une matrice symétrique positive.*
Dans ce cas, μ et Γ sont respectivement le vecteur espérance de X ($\mu_i = \mathbb{E}(X_i)$) et sa matrice de covariance ($\Gamma_{ij} = \mathrm{Cov}(X_i, X_j)$). On dit alors que X suit la loi gaussienne en dimension d d'espérance μ et de matrice de covariance Γ, ce que l'on note $X \sim \mathcal{N}_d(\mu, \Gamma)$.

Ainsi la fonction caractéristique et donc la loi d'un vecteur gaussien ne dépendent que de son espérance et de sa matrice de covariance.
Démonstration :
Condition nécessaire : Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien à valeurs $\mathbb{R}^d$. Les X_i

étant des variables gaussiennes réelles, elles sont de carré intégrable. Le vecteur espérance μ et la matrice de covariance Γ de X sont donc bien définis.
Soit $u \in \mathbb{R}^d$. Par linéarité de l'espérance on a :

$$\mathbb{E}(u.X) = u_1\mathbb{E}(X_1) + \cdots + u_d\mathbb{E}(X_n) = u.\mathbb{E}(X),$$

et d'après les propriétés 3.3.14,

$$\mathrm{Var}\,(u.X) = \sum_{i=1}^{d}\sum_{j=1}^{d} u_i u_j \mathrm{Cov}(X_i, X_j) = u.\Gamma u.$$

Donc $u.X \sim \mathcal{N}_1(u.\mathbb{E}(X), u.\Gamma u)$ et

$$\Phi_X(u) = \mathbb{E}\left(e^{iu.X}\right) = \Phi_{u.X}(1) = e^{i\mathbb{E}(u.X)-\frac{1}{2}\mathrm{Var}(u.X)} = e^{iu.\mu-\frac{1}{2}u.\Gamma u}.$$

<u>Condition suffisante :</u> Inversement, si le vecteur $X = (X_1, \ldots, X_d)$ a pour fonction caractéristique $v \in \mathbb{R}^d \to \Phi_X(v) = e^{iv.\mu-\frac{1}{2}v.\Gamma v}$ où $\mu \in \mathbb{R}^d$ et Γ est une matrice $d \times d$ symétrique et positive, alors pour $u \in \mathbb{R}^d$,

$$\forall t \in \mathbb{R},\ \Phi_{u.X}(t) = \mathbb{E}(e^{it(u.X)}) = \mathbb{E}(e^{i(tu).X}) = \Phi_X(tu) = e^{it(u.\mu)-\frac{t^2}{2}u.\Gamma u}.$$

Donc $\forall u \in \mathbb{R}^d$, $u.X \sim \mathcal{N}_1(u.\mu, u.\Gamma u)$. Ainsi X est un vecteur gaussien.
En laissant u décrire la base canonique de $\mathbb{R}^d$, on vérifie que $\forall 1 \leq j \leq d$, $\mathbb{E}(X_j) = \mu_j$.
Puis en utilisant la bilinéarité de la covariance, on obtient que pour $1 \leq i, j \leq d$

$$2\mathrm{Cov}(X_i, X_j) = \mathrm{Var}(X_i + X_j) - \mathrm{Var}(X_i) - \mathrm{Var}(X_j) = (\Gamma_{ii} + 2\Gamma_{ij} + \Gamma_{jj}) - \Gamma_{ii} - \Gamma_{jj} = 2\Gamma_{ij}.$$

Ainsi μ et Γ sont respectivement le vecteur espérance et la matrice de covariance de X.□

Exemple 6.1.3. Si pour $1 \leq j \leq d$, $G_j \sim \mathcal{N}_1(\mu_j, \sigma_j^2)$ et ces variables aléatoires sont indépendantes, alors pour $G = (G_1, \ldots, G_d)$ et $u \in \mathbb{R}^d$,

$$\Phi_G(u) = \mathbb{E}(e^{iu.G}) = \mathbb{E}\left(\prod_{j=1}^{d} e^{iu_jG_j}\right) = \prod_{j=1}^{d}\mathbb{E}\left(e^{iu_jG_j}\right) = \prod_{j=1}^{d} e^{iu_j\mu_j-\frac{1}{2}u_j^2\sigma_j^2} = e^{iu.\mu-\frac{1}{2}u.\mathrm{Diag}(\sigma^2)u}$$

où $\mathrm{Diag}(\sigma^2)$ est la matrice diagonale d'éléments diagonaux $\mathrm{Diag}(\sigma^2)_{jj} = \sigma_j^2$ et μ le vecteur de coordonnées μ_j. Ainsi $G \sim \mathcal{N}_d(\mu, \mathrm{Diag}(\sigma^2))$.
Dans le cas particulier ou les G_j sont centrées réduites, $G \sim \mathcal{N}_d(0, I_d)$ où $I_d \in \mathbb{R}^{d\times d}$ est la matrice identité.

Remarque 6.1.4. Il ne suffit pas que X_1 et X_2 soient des variables gaussiennes réelles pour que (X_1, X_2) soit un vecteur gaussien. Ce résultat est vrai dans le cas particulier où X_1 et X_2 sont indépendantes mais pas en toute généralité comme l'illustre l'exercice 6.3.2.

6.1.2 Stabilité du caractère gaussien par transformation linéaire

Si X est un vecteur aléatoire à valeurs $\mathbb{R}^d$ et $Y = a + MX$ pour un vecteur (constant) $a \in \mathbb{R}^n$ et une matrice M de taille $n \times d$, alors toute combinaison linéaire des coordonnées de Y est combinaison linéaire des coordonnées de X à une constante près : pour $v \in \mathbb{R}^n$, $v.Y = v.a + (M^*v).X$ où M^* désigne la transposée de la matrice M. Donc si X est gaussien, Y l'est aussi.

Exercice 6.1.5. Vérifier que $\mathbb{E}(Y) = a + M\mathbb{E}(X)$ et que la matrice de covariance Λ de Y s'exprime en fonction de celle Γ de X par la relation $\Lambda = M\Gamma M^*$.

Exemple 6.1.6. Si $X = (X_1, \ldots, X_d)$ est un vecteur gaussien à valeurs $\mathbb{R}^d$ alors pour $k \leq d$, le vecteur $(X_1, \ldots, X_k)$ est gaussien (de même que tout vecteur obtenu à partir de X en enlevant certaines coordonnées). La moyenne empirique $\frac{1}{d}\sum_{i=1}^d X_i$ est une gaussienne réelle.

6.1.3 Construction d'un vecteur gaussien de loi $\mathcal{N}_n(\mu, \Lambda)$

Soit $\mu \in \mathbb{R}^n$ et Λ une matrice de covariance $n \times n$ i.e. une matrice symétrique positive. Il existe une matrice $A \in \mathbb{R}^{n\times d}$ telle que :

$$AA^* = \Lambda.$$

On peut par exemple choisir $d = n$ et construire A comme la racine carrée de la matrice symétrique et positive Λ en la diagonalisant.

D'après ce qui précède, si $G = (G_1, \ldots, G_d)$ est un vecteur de variables aléatoires gaussiennes centrées réduites indépendantes alors $\mu + AG$ suit la loi $\mathcal{N}_n(\mu, \Lambda)$.

On débouche ainsi sur l'algorithme suivant pour simuler un vecteur gaussien de loi $\mathcal{N}_n(\mu, \Lambda)$:

Simulation d'un vecteur gaussien :
- Calculer $A \in \mathbb{R}^{n\times d}$ t.q. $AA^* = \Lambda$,
- tirer un vecteur de gaussiennes centrées réduites indépendantes $G = (G_1, \ldots, G_d)$,
- retourner le vecteur $\mu + AG$.

6.2 Propriétés des vecteurs gaussiens

6.2.1 Vecteurs gaussiens et indépendance

Si les coordonnées du vecteur aléatoire $X = (X_1, \ldots, X_d)$ sont indépendantes et de carré intégrable, alors sa matrice de covariance est diagonale. En effet pour $i \neq j$, $\mathbb{E}(X_iX_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ i.e. $\mathrm{Cov}(X_i, X_j) = 0$.
Dans le cas où X est un vecteur gaussien, le caractère diagonal de la matrice de covariance s'avère une condition suffisante d'indépendance. Cela vient du fait que la loi d'un vecteur gaussien ne dépend que de son espérance et de sa matrice de covariance.

Proposition 6.2.1. *Les coordonnées d'un vecteur gaussien $X = (X_1, \ldots, X_d)$ sont indépendantes si et seulement si sa matrice de covariance Γ est diagonale.*

Démonstration : Nous avons déjà démontré la condition nécessaire. Supposons que la matrice de covariance Γ est diagonale et notons μ le vecteur espérance de X. D'après

l'exemple 6.1.3, si $Y_1, \ldots, Y_d$ sont des variables gaussiennes indépendantes d'espérance et variance respectives $\mu_1, \Gamma_{11}, \ldots, \mu_d, \Gamma_{dd}$ alors le vecteur $Y = (Y_1, \ldots, Y_d)$ suit la loi $\mathcal{N}_d(\mu, \Gamma)$. Donc X et Y ont même loi, et en particulier pour $1 \leq j \leq d$, X_j a même loi que Y_j.
Si pour $1 \leq j \leq d$, $f_j : \mathbb{R} \to \mathbb{R}$ est une fonction bornée, on a donc d'une part $\mathbb{E}\left(\prod_{j=1}^d f_j(X_j)\right) = \mathbb{E}\left(\prod_{j=1}^d f_j(Y_j)\right)$ et d'autre part pour $1 \leq j \leq d$, $\mathbb{E}(f_j(X_j)) = \mathbb{E}(f_j(Y_j))$. En combinant ces égalités avec la condition nécessaire de la proposition 3.3.11 vérifiée par les variables Y_j qui sont indépendantes, on obtient que

$$\mathbb{E}\left(\prod_{j=1}^d f_j(X_j)\right) = \mathbb{E}\left(\prod_{j=1}^d f_j(Y_j)\right) = \prod_{j=1}^d \mathbb{E}(f_j(Y_j)) = \prod_{j=1}^d \mathbb{E}(f_j(X_j)).$$

D'après la condition suffisante de la proposition 3.3.11, les coordonnées X_j sont des variables indépendantes. $\square$

Exercice 6.2.2. Soit $Y_1, \ldots, Y_n$ des vecteurs aléatoires à valeurs respectivement dans $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_n}$ t.q. le vecteur $Y = (Y_1, \ldots, Y_n)$ est gaussien (cela implique en particulier que les Y_i sont des vecteurs gaussiens). Montrer que les vecteurs $Y_1, \ldots, Y_n$ sont indépendants si et seulement si la matrice de covariance Γ de Y est diagonale par blocs au sens où :

$$\forall (j,k) \notin \bigcup_{i=1}^n [d_1 + \ldots + d_{i-1} + 1, d_1 + \ldots + d_i]^2,\ \Gamma_{jk} = 0.$$

Le résultat suivant est crucial pour l'étude statistique du modèle gaussien.

Théorème 6.2.3 (de Cochran). *Soit $Y = (Y_1, \ldots, Y_n) \sim \mathcal{N}_n(0, I_n)$ et $E_1, \ldots, E_p$ une décomposition de $\mathbb{R}^n$ en p sous-espaces orthogonaux de dimensions respectives $d_1, \ldots, d_p$ ($d_1 + \ldots + d_p = n$). Alors les projections orthogonales Y_{E_j}, $1 \leq j \leq p$ de Y sur ces sous-espaces sont indépendantes et $|Y_{E_j}|^2 \sim \chi^2(d_j)$.*

Démonstration : Soit $(e_1, \ldots, e_{d_1})$ une base orthornormée de E_1, $(e_{d_1+1}, \ldots, e_{d_1+d_2})$ une base orthonormée de E_2 et plus généralement $(e_{d_1+\ldots+d_{j-1}+1}, \ldots, e_{d_1+\ldots+d_j})$ une base orthonormée de E_j. Alors $(e_1, \ldots, e_n)$ constitue une base orthornormée de $\mathbb{R}^n$. La matrice de passage [1] O qui permet d'exprimer les coordonnées dans cette nouvelle base à partir des coordonnées dans la base canonique est une matrice orthogonale : $OO^* = I_n$. On en déduit que le vecteur $(e_1.Y, \ldots, e_n.Y)$ qui est égal à OY est distribué suivant la loi $\mathcal{N}_n(0, I_n)$. Ainsi les variables aléatoires $e_i.Y$, $1 \leq i \leq n$ sont indépendantes et identiquement distribuées suivant la loi normale centrée réduite. On conclut en remarquant que

$$\forall j \in \{1, \ldots, p\},\ Y_{E_j} = \sum_{i=d_1+\ldots+d_{j-1}+1}^{d_1+\ldots+d_j} (e_i.Y)e_i \quad \text{et} \quad |Y_{E_j}|^2 = \sum_{i=d_1+\ldots+d_{j-1}+1}^{d_1+\ldots+d_j} (e_i.Y)^2.$$

$\square$

On en déduit le corollaire suivant :

1. $O = \begin{pmatrix} e_1^* \\ \vdots \\ e_n^* \end{pmatrix}$

Corollaire 6.2.4. *Soit $X_1, \ldots, X_n$ des variables aléatoires réelles I.I.D. suivant la loi $\mathcal{N}_1(\mu, \sigma^2)$ avec $\sigma^2 > 0$.*
La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et l'estimateur sans biais de la variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont indépendants avec $\bar{X}_n \sim \mathcal{N}_1\left(\mu, \frac{\sigma^2}{n}\right)$ et $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$. En particulier le rapport $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}$ suit la loi de Student de paramètre $n-1$.

Remarque 6.2.5. Comme $\bar{X}_n \sim \mathcal{N}_1\left(\mu, \frac{\sigma^2}{n}\right)$, $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}_1(0,1)$. En remplaçant l'écart-type σ au dénominateur par son estimateur S_n, on obtient la loi de Student de paramètre $n-1$. Lorsque $n \to +\infty$, cette loi converge vers la loi normale centrée réduite.

Démonstration : Par indépendance, $X = (X_1, \ldots, X_n)$ est un vecteur gaussien. La moyenne empirique $\bar{X}_n$ qui s'obtient comme combinaison linéaire de ce vecteur est donc une variable gaussienne réelle. Par linéarité, $\mathbb{E}(\bar{X}_n) = \mu$ et par indépendance, $\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$, ce qui assure $\bar{X}_n \sim \mathcal{N}_1(\mu, \frac{\sigma^2}{n})$ et $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}_1(0,1)$.

Le vecteur $Y = \frac{1}{\sigma}(X_1 - \mu, \ldots, X_n - \mu)$ est distribué suivant $\mathcal{N}_n(0, I_n)$. Soit E_1 l'espace vectoriel engendré par $e_1 = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ et $E_2 = E_1^\perp$. On a $e_1.Y = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ et $Y_{E_1} = (e_1.Y)e_1 = \frac{1}{\sigma}(\bar{X}_n - \mu, \ldots, \bar{X}_n - \mu)$. Par ailleurs $Y_{E_2} = Y - Y_{E_1} = \frac{1}{\sigma}(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$, ce qui entraîne que $|Y_{E_2}|^2 = \frac{(n-1)S_n^2}{\sigma^2}$. On déduit alors du théorème de Cochran 6.2.3 que $\bar{X}_n$ est indépendante de S_n^2 et que $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$.

En remarquant que

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)/\sigma}{\sqrt{\frac{(n-1)S_n^2/\sigma^2}{n-1}}}$$

où le numérateur est une variable normale centrée réduite indépendante de la variable $(n-1)S_n^2/\sigma^2 \sim \chi^2(n-1)$ qui figure au dénominateur, on conclut que le rapport suit la loi de Student de paramètre $n-1$. □

6.2.2 Vecteurs gaussiens et convergence en loi

Le résultat suivant qui assure que toute limite en loi (*a fortiori* presque sûre, en probabilité, dans L^1 ou dans L^2 puisque ces convergences impliquent la convergence en loi) de variables gaussiennes réelles est gaussienne, est très utile pour obtenir le caractère gaussien de variables obtenues par un procédé de passage à la limite.

Proposition 6.2.6. *Soit $(X_n)_n$ une suite de variables aléatoires gaussiennes réelles qui converge en loi vers X. Alors X est gaussienne.*

Démonstration : Nous nous contenterons de faire la preuve dans le cas beaucoup plus facile où X_n converge vers X dans L^2 i.e. où $\mathbb{E}((X_n - X)^2) \to 0$.

D'après la Proposition 5.1.5, on a alors convergence de $\mathbb{E}(X_n)$ et $\mathrm{Var}(X_n)$ respectivement vers $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.
On en déduit que pour $u \in \mathbb{R}$, $\mathbb{E}(e^{iuX_n}) = e^{iu\mathbb{E}(X_n)-\frac{u^2}{2}\mathrm{Var}(X_n)} \to e^{iu\mathbb{E}(X)-\frac{u^2}{2}\mathrm{Var}(X)}$. La convergence dans L^2 de X_n vers X entraîne la convergence en loi de X_n vers X et donc d'après le théorème 5.3.9, la convergence de $\mathbb{E}(e^{iuX_n})$ vers $\mathbb{E}(e^{iuX})$. On conclut donc que

$$\forall u \in \mathbb{R},\ \mathbb{E}(e^{iuX}) = e^{iu\mathbb{E}(X)-\frac{u^2}{2}\mathrm{Var}(X)},$$

ce qui implique que X est gaussienne. □

Exercice 6.2.7. Soit $(X_n)_n$ une suite de vecteurs gaussiens à valeurs $\mathbb{R}^d$ qui converge en loi vers un vecteur X.

1. Soit $u \in \mathbb{R}^d$. Montrer que $u.X_n$ converge en loi vers $u.X$. Que peut-on en déduire sur la loi de $u.X$?
2. Conclure que X est gaussien.

Nous allons maintenant énoncer le théorème de la limite centrale pour les vecteurs aléatoires qui se déduit facilement du cas des variables aléatoires réelles.

Théorème 6.2.8 (de la limite centrale multidimensionnel)**.** *Soit $(X_i, i \geq 1)$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ indépendants suivant tous la même loi que X. On suppose que $\mathbb{E}\left(|X|^2\right) < +\infty$ (où $|x|$ désigne la norme euclidienne d'un vecteur de $\mathbb{R}^d$). On note μ et Γ le vecteur espérance et la matrice de covariance de X. Alors*

$$Y_n = \sqrt{n}\left(\frac{1}{n}(X_1 + \cdots + X_n) - \mu\right)$$

converge en loi vers un vecteur gaussien centré de matrice de covariance Γ lorsque n tend vers $+\infty$.

Démonstration : Soit $u \in \mathbb{R}^d$. Les variables aléatoires réelles $(u.X_i, i \geq 1)$ sont I.I.D. d'espérance $u.\mu$ et de variance $u.\Gamma u$ (voir propriétés 3.3.14). Si cette variance est nulle, alors avec probabilité 1 les variables $u.X_i$ sont égales à $u.\mu$, ce qui entraîne que $u.Y_n = 0$ puis que

$$\Phi_{Y_n}(u) = \mathbb{E}(e^{iu.Y_n}) = 1 = e^{-\frac{1}{2}u.\Gamma u}.$$

Lorsque $u.\Gamma u > 0$, on déduit du théorème de la limite centrale 5.4.1 que $\dfrac{u.Y_n}{\sqrt{u.\Gamma u}}$ converge en loi vers G où G suit la loi normale centrée réduite. Donc, par la condition nécessaire du théorème 5.3.9

$$\Phi_{Y_n}(u) = \Phi_{\frac{u.Y_n}{\sqrt{u.\Gamma u}}}(\sqrt{u.\Gamma u}) \to \Phi_G(\sqrt{u.\Gamma u}) = e^{-\frac{1}{2}u.\Gamma u}.$$

On conclut par la condition suffisante de ce même théorème. □

6.3 Exercices

Exercice corrigé 6.3.1. Soit X et Y des variables gaussiennes réelles indépendantes. Donner une condition nécessaire et suffisante pour que les variables $X+Y$ et $X-Y$ soient indépendantes.

Exercice 6.3.2. Soit X une variable gaussienne centrée réduite et ε une variable vérifiant $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$ indépendantes.

1. Quelle est la loi de $Y = \varepsilon X$?
2. Donner la matrice de covariance du vecteur (X, Y).
3. Comparer $\mathbb{E}(X^2Y^2)$ à $\mathbb{E}(X^2)\mathbb{E}(Y^2)$. Les variables X et Y sont-elles indépendantes ?
4. Que peut-on conclure sur le couple (X, Y) ?

Exercice 6.3.3. Soit (X, Y) un couple gaussien de matrice de covariance

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

où $\rho \in [-1, 1]$. Que peut-on dire des variables aléatoires X et $Z = Y - \rho X$?

Exercice 6.3.4. Soit $X_1, \ldots, X_n$ des variables aléatoires gaussiennes centrées réduites indépendantes et a un vecteur non nul de $\mathbb{R}^n$. Donner une condition nécessaire et suffisante sur a pour que le vecteur aléatoire $X - (a.X)a$ et la variable aléatoire $a.X$ soient indépendants.

Exercice corrigé 6.3.5. Soit $X_1, \ldots, X_n$ des variables aléatoires réelles I.I.D. et de carré intégrable telles que la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ et la variable $S_n^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X}_n)^2$ sont indépendantes.

1. Vérifier que $S_n^2 = \displaystyle\frac{1}{n}\sum_{j=1}^n X_j^2 - \frac{1}{n(n-1)}\sum_{\substack{j,k=1 \\ k\neq j}}^n X_j X_k$ et en déduire $\mathbb{E}(S_n^2)$.
2. En utilisant l'hypothèse d'indépendance, exprimer $\mathbb{E}(S_n^2 e^{iun\bar{X}_n})$ en fonction de $\mathrm{Var}(X_1)$ et de la fonction caractéristique commune des X_j que l'on notera Φ_X.
3. Montrer par ailleurs que
$$\mathbb{E}(S_n^2 e^{iun\bar{X}_n}) = \mathbb{E}(X_1^2 e^{iuX_1})(\Phi_X(u))^{n-1} - (\mathbb{E}(X_1 e^{iuX_1}))^2(\Phi_X(u))^{n-2}.$$
4. Remarquer que $\mathbb{E}(X_1 e^{iuX_1})$ et $\mathbb{E}(X_1^2 e^{iuX_1})$ s'expriment simplement à l'aide des deux premières dérivées de Φ_X et en déduire une équation différentielle satisfaite par cette fonction.
5. Poser $f(u) = \Phi_X'(u)/\Phi_X(u)$ et calculer $f'(u)$. En déduire Φ_X puis la loi commune des X_j.

Exercice 6.3.6. Soit $(X_i)_{i\geq 1}$ une suite I.I.D. de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ de carré intégrable et $g : \mathbb{R}^d \to \mathbb{R}$. On note μ et Γ le vecteur espérance et la matrice de covariance communs des X_i et $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

1. On suppose que la fonction g est C^1. En remarquant que

$$\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) = \int_0^1 \nabla g(\alpha \bar{X}_n + (1-\alpha)\mu) d\alpha . \sqrt{n}(\bar{X}_n - \mu)$$

et en étudiant le comportement asymptotique de chacun des deux termes du produit scalaire, vérifier que lorsque $n \to +\infty$, $\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right)$ converge en loi vers une gaussienne centrée de variance $\nabla g(\mu).\Gamma \nabla g(\mu)$.

2. On suppose maintenant simplement que la fonction g est continue et différentiable au point μ i.e. $g(x) - g(\mu) - \nabla g(\mu).(x - \mu) = |x - \mu| \varepsilon(x - \mu)$ où $\lim_{|y| \to 0^+} \varepsilon(y) = 0$. Vérifier que $\sqrt{n}\left(g(\bar{X}_n) - g(\mu) - \nabla g(\mu).(\bar{X}_n - \mu)\right)$ converge en loi vers 0 lorsque $n \to +\infty$. En déduire que la conclusion de la question précédente reste valable.

6.4 Résumé

- $X = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$ est **un vecteur gaussien** si toute combinaison linéaire de ses coordonnées est une variable gaussienne réelle

  $$\iff \Phi_X(u) = e^{iu.m - \frac{1}{2}u.\Gamma u} \quad \text{où } m \in \mathbb{R}^d \text{ et } \Gamma \text{ matrice } d \times d \text{ symétrique positive.}$$

  Alors $m_j = \mathbb{E}(X_j)$, $\Gamma_{ij} = \mathrm{Cov}(X_i, X_j)$ et on note $X \sim \mathcal{N}_d(m, \Gamma)$.

- **Stabilité par transformation affine :** si $X \sim \mathcal{N}_d(m, \Gamma)$, $a \in \mathbb{R}^n$ et $M \in \mathbb{R}^{n \times d}$ alors $a + MX \sim \mathcal{N}_n(a + Mm, M\Gamma M^*)$.

- Les coordonnées d'un vecteur gaussien X sont **indépendantes** si et seulement si sa matrice de covariance Γ est **diagonale**.

- Toute limite en loi d'une suite de vecteurs gaussiens est un vecteur gaussien.

- Soit $X_1, \dots, X_n$ I.I.D. $\sim \mathcal{N}_1(\mu, \sigma^2)$:
  1. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}_1(\mu, \sigma^2/n)$ indep de $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$
  2. $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ et $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t(n-1)$.

- **Théorème de Cochran :** si $Y = (Y_1, \dots, Y_n) \sim \mathcal{N}_n(0, I_n)$ et $E_1, \dots, E_p$ est une décomposition de $\mathbb{R}^n$ en p sous-espaces orthogonaux de dimensions respectives $d_1, \dots, d_p$ $(d_1 + \ldots + d_p = n)$, alors les projections orthogonales Y_{E_j}, $1 \le j \le p$ de Y sur ces sous-espaces sont indépendantes et $|Y_{E_j}|^2 \sim \chi^2(d_j)$.

Chapitre 7

Estimation de paramètres

Le statisticien travaille sur des données (notes de qualité de pièces produites dans une usine, données météorologiques, résultats d'expériences médicales ou physiques,...). Il le fait à la demande d'un interlocuteur qui a des attentes plus ou moins précises. Ces attentes peuvent être de plusieurs types :

- extraire des résumés pertinents des données,
- répondre à une question comme "le réchauffement climatique est-il réel ?",
- prendre une décision comme la mise sur le marché d'un nouveau médicament,
- effectuer une prévision, par exemple sur le résultat d'une élection qui aura lieu prochainement,...

Il élabore un modèle et construit des outils pour répondre aux questions de son interlocuteur dans ce modèle. Il doit bien sûr garder un sens critique vis à vis du modèle qu'il a construit.

Dans le cadre de ce cours, nous ne considérerons que des modèles paramétriques. Il est bien sûr crucial pour le statisticien d'estimer les paramètres au vu des données dont il dispose et d'avoir une idée de la précision de cette estimation. Après avoir défini la notion de modèle paramétrique, nous introduirons les estimateurs. Nous verrons enfin comment évaluer la précision des estimateurs au travers d'intervalles de confiance.

7.1 Modèle paramétrique

On suppose que l'on observe un échantillon $X = (X_1, \ldots, X_n)$ où les variables aléatoires X_i sont indépendantes et identiquement distribuées et que leur loi commune est dans une famille de probabilités $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ où $\Theta \subset \mathbb{R}^d$.

Exemple 7.1.1.
- modèle de Bernoulli : $\mathcal{P} = \{\mathcal{B}(p), p \in [0,1]\}$ pour l'observation des notes de qualité de n pièces mécaniques (0 si pièce correcte, 1 si pièce défectueuse).
- modèle gaussien : $\mathcal{P} = \{\mathcal{N}_1(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$. Le paramètre $\theta = (\mu, \sigma^2)$ est alors bidimensionnel et prend ses valeurs dans $\Theta = \mathbb{R} \times \mathbb{R}_+^*$. Mais on peut également supposer la variance σ^2 connue auquel cas $\theta = \mu$ ou supposer la moyenne μ connue ($\theta = \sigma^2$).

Notation : On notera respectivement $\mathbb{P}_\theta(X \in A)$ et $\mathbb{E}_\theta(f(X))$ la probabilité de l'événement $\{X \in A\}$ et l'espérance de $f(X)$ et lorsque les variables X_i sont I.I.D. suivant P_θ.

Exemple 7.1.2. Dans le modèle gaussien, $\mathbb{E}_{(\mu,\sigma^2)}(X_1) = \mu$.

Dans toute la suite on supposera

- soit que pour tout $\theta \in \Theta$, les X_i sont des variables aléatoires discrètes à valeurs dans F dénombrable. Alors on pose $p(x_1, \theta) = \mathbb{P}_\theta(X_1 = x_1)$ pour $x_1 \in F$ et on a

$$\forall x = (x_1, \ldots, x_n) \in F^n,\ \mathbb{P}_\theta(X = x) = p_n(x, \theta) \ \text{ où } \boxed{p_n(x, \theta) = p(x_1, \theta) \ldots p(x_n, \theta)}.$$

Exemple 7.1.3. Dans le modèle de Bernoulli, $F = \{0, 1\}$ et

$$\forall p \in [0, 1],\ \forall x = (x_1, \ldots, x_n) \in \{0, 1\}^n,\ p_n(x, p) = p^y(1-p)^{n-y} \ \text{ où } \ y = x_1+\ldots+x_n.$$

- soit que les X_i sont des variables aléatoires à valeurs $\mathbb{R}^k$ qui possèdent une densité. On note alors $p(x_1, \theta)$ la densité de X_1 et $\boxed{p_n(x, \theta) = p(x_1, \theta) \ldots p(x_n, \theta)}$ la densité produit de $X = (X_1, \ldots, X_n)$ sous $\mathbb{P}_\theta$.

Exemple 7.1.4. Dans le modèle gaussien

$$p_n(x, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

Définition 7.1.5.

- *Pour $x = (x_1, \ldots, x_n)$ fixé, la fonction $\theta \in \Theta \to p_n(x, \theta)$ s'appelle la vraisemblance de la réalisation x.*
- *On appelle statistique toute variable aléatoire S qui s'écrit comme une fonction de l'échantillon $X = (X_1, \ldots, X_n)$ ne faisant pas intervenir θ.*

Exemple 7.1.6. $\sum_{i=1}^{n} X_i$ et X_1X_n sont des statistiques mais pas θX_1.

7.2 Estimateurs

7.2.1 Définitions

Dans ce paragraphe on s'intéresse à l'estimation de $g(\theta)$ où $g : \Theta \to \mathbb{R}^q$. En pratique, l'estimation de θ correspond au choix $q = d$ et $g(\theta) = \theta$. Lorsque $d \geq 2$, l'estimation des q premières coordonnées de θ avec $q < d$ correspond au choix $g(\theta_1, \ldots, \theta_d) = (\theta_1, \ldots, \theta_q)$.

Définition 7.2.1. *On appelle estimateur de $g(\theta)$ toute statistique Z (construite à partir de l'échantillon X) à valeurs dans l'image $g(\Theta)$ de Θ par g.*

Dans toute la suite, on supposera que Z est de carré intégrable au sens où $\forall \theta \in \Theta$, $\mathbb{E}_\theta(|Z|^2) < +\infty$.

Exemple 7.2.2. Dans le modèle de Bernoulli, $\mathcal{P} = \{\mathcal{B}(p), p \in [0, 1]\}$, la variable aléatoire X_n et la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ sont des estimateurs de p.

La définition suivante introduit quelques notions relatives aux estimateurs

Définition 7.2.3.

- *Un estimateur Z de $g(\theta)$ est dit sans biais si $\forall \theta \in \Theta$, $\mathbb{E}_\theta(Z) = g(\theta)$.*

- *Lorsque g est à valeurs réelles, on appelle risque quadratique de l'estimateur Z la fonction*
$$\theta \in \Theta \to R(Z,\theta) = \mathbb{E}_\theta((Z - g(\theta))^2).$$
- *L'estimateur Z_1 est dit préférable à Z_2 si pour tout $\theta \in \Theta$, $R(Z_1,\theta) \leq R(Z_2,\theta)$.*

Exemple 7.2.4. Dans le modèle gaussien, X_1 et $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ sont des estimateurs sans biais de μ de risques quadratiques respectifs $R(X_1,(\mu,\sigma^2)) = \sigma^2$ et $R(\bar{X}_n,(\mu,\sigma^2)) = \sigma^2/n$ si bien que $\bar{X}_n$ est préférable à X_1.
Notons que comme

$$\mathbb{E}_{(\mu,\sigma^2)}((\bar{X}_n)^2) = \frac{1}{n}\mathbb{E}_{(\mu,\sigma^2)}(X_1^2) + \frac{n-1}{n}\mathbb{E}_{(\mu,\sigma^2)}(X_1X_2) = \frac{\mu^2+\sigma^2}{n} + \frac{n-1}{n}\mu^2 = \mu^2 + \frac{\sigma^2}{n},$$

$(\bar{X}_n)^2$ est un estimateur biaisé de μ^2. Ainsi le caractère sans biais n'est en général pas préservé par des transformations non linéaires.

Lorsque g est à valeurs réelles, comme $(g(\theta) - Z)^2 = (g(\theta) - \mathbb{E}_\theta(Z))^2 + 2(g(\theta) - \mathbb{E}_\theta(Z))(\mathbb{E}_\theta(Z) - Z) + (\mathbb{E}_\theta(Z) - Z)^2$ avec le second terme d'espérance nulle sous $\mathbb{P}_\theta$,

Propriétés 7.2.5. *On a la décomposition suivante du risque quadratique*

$$\boxed{R(Z,\theta) = (g(\theta) - \mathbb{E}_\theta(Z))^2 + \mathrm{Var}_\theta(Z)}$$

en un terme lié au biais plus un terme de variance.

On souhaite bien sûr que lorsque la taille n de l'échantillon augmente, l'estimation soit de plus en plus précise. C'est pourquoi on introduit la définition suivante :

Définition 7.2.6. *Une suite $(Z_n)_{n\geq 1}$ d'estimateurs de $g(\theta)$ avec Z_n construit à partir de $(X_1,\ldots,X_n)$ est dite :*
- *fortement convergente si, pour tout $\theta \in \Theta$, $\mathbb{P}_\theta$ presque sûrement, $Z_n \to g(\theta)$ i.e. $\mathbb{P}_\theta(Z_n \to g(\theta)) = 1$, lorsque $n \to +\infty$.*
- *asymptotiquement normale si pour tout $\theta \in \Theta \subset \mathbb{R}^d$, il existe une matrice de covariance $\Sigma(\theta) \in \mathbb{R}^{q\times q}$ telle que sous $\mathbb{P}_\theta$, $\sqrt{n}(Z_n - g(\theta))$ converge en loi vers $Y \sim \mathcal{N}_q(0,\Sigma(\theta))$ lorsque $n \to +\infty$.*

Remarque 7.2.7.
- Lorsque les conditions de la définition précédente sont vérifiées, par abus de langage, on dit que Z_n est fortement convergent et asymptotiquement normal.
- Lorsque pour tout $\theta \in \Theta$, Z_n converge en probabilité vers $g(\theta)$ sous $\mathbb{P}_\theta$, on dit que l'estimateur Z_n est convergent.

Exemple 7.2.8. Dans le modèle de Bernoulli, la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur de p fortement convergent d'après la loi forte des grands nombres et asymptotiquement normal de variance asymptotique $\mathrm{Var}_p(X_1) = p(1-p)$ d'après le théorème de la limite centrale.

7.2.2 L'Estimateur du Maximum de Vraisemblance

Lors d'une enquête policière, si un suspect de sexe inconnu mesure environ 1.50 m, on aura plutôt tendance à rechercher une femme tandis que s'il mesure environ 1.80 m, on

recherchera plutôt un homme.
La notion de maximum de vraisemblance permet de formaliser cette intuition. On peut modéliser la distribution des tailles (en mètres) féminines par une loi gaussienne d'espérance $\mu_F = 1,62$ et d'écart type $\sigma_F = 0.069$ et celle des tailles masculines par une loi gaussienne d'espérance $\mu_H = 1,76$ et d'écart type $\sigma_H = 0.073$. Les densités de ces deux lois sont représentées sur la figure 7.1. Lorsque l'on connait la taille x d'un suspect, on

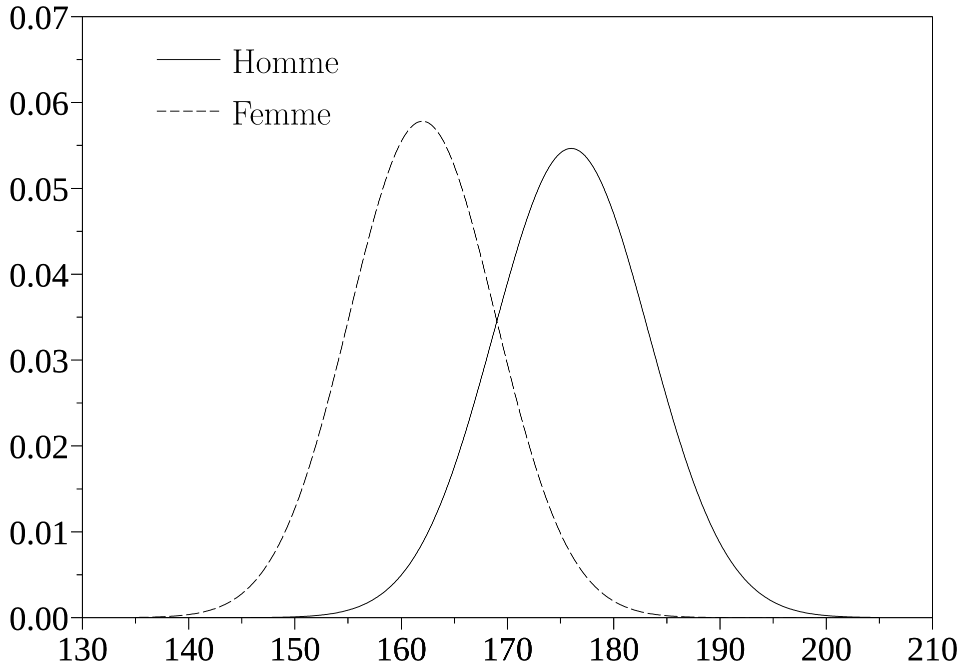


FIGURE 7.1 – Densités des lois gaussiennes modélisant la taille des hommes et celle des femmes.

pourra supposer que ce suspect est une femme si la densité des tailles féminines prise en x est supérieure à celle des tailles masculines et vice et versa.
D'un point de vue mathématique, dans le modèle $\{\mathcal{N}(\mu_\theta, \sigma_\theta^2), \theta \in \{F, H\}\}$, lorsque l'on observe la taille x d'un individu, on peut estimer le sexe de cet individu en choisissant $\theta \in \Theta = \{F, H\}$ qui maximise la vraisemblance $\theta \to \frac{1}{\sigma_\theta \sqrt{2\pi}} e^{-\frac{(x-\mu_\theta)^2}{2\sigma_\theta^2}}$.

Définition 7.2.9. *On suppose que pour toute réalisation $x = (x_1, \ldots, x_n)$ de l'échantillon $X = (X_1, \ldots, X_n)$, il existe une unique valeur $\theta_n(x) \in \Theta$ qui maximise la vraisemblance de la réalisation x : $p_n(x, \theta_n(x)) = \max_{\theta \in \Theta} p_n(x, \theta)$. Alors la statistique $\hat{\theta}_n = \theta_n(X)$ est appelée Estimateur du Maximum de Vraisemblance de θ.*

Remarque 7.2.10. Dans le cas d'un modèle discret, $p_n(x, \theta) = \mathbb{P}_\theta(X = x)$ et $\theta_n(x)$ est la valeur du paramètre θ dans Θ qui maximise la probabilité d'observer x sous $\mathbb{P}_\theta$.

Comme la fonction logarithme est strictement croissante, il revient au même de maximiser la vraisemblance $\theta \to p_n(x, \theta)$ ou de maximiser la log-vraisemblance $\theta \to l_n(x, \theta)$

définie par

$$\boxed{l_n(x,\theta) = \ln(p_n(x,\theta)).}$$

On pose également $\boxed{l(x_1,\theta) = \ln(p(x_1,\theta))} = l_1(x_1,\theta)$. Les calculs sont parfois plus simples avec la log-vraisemblance notamment parce que $\boxed{l_n(x,\theta) = \sum_{i=1}^n l(x_i,\theta)}$.

Exercice 7.2.11. On se place dans le modèle $\mathcal{P} = \{\mathcal{E}(\theta), \theta > 0\}$.

1. Déterminer l'EMV $\hat{\theta}_n$ de θ et montrer qu'il est fortement convergent.
2. Remarquer que sous $\mathbb{P}_\theta$, $X_1 + \ldots + X_n \sim \Gamma(n,\theta)$. En déduire $\mathbb{E}_\theta\left(\frac{1}{X_1+\ldots+X_n}\right)$ pour $n \geq 2$. L'EMV est-il sans biais ?
3. Remarquer que $\sqrt{n}(\hat{\theta}_n - \theta) = \frac{\theta}{\bar{X}_n}\sqrt{n}\left(\frac{1}{\theta} - \bar{X}_n\right)$. Vérifier que $\left(\frac{\theta}{\bar{X}_n}, \sqrt{n}\left(\frac{1}{\theta} - \bar{X}_n\right)\right)$ converge en loi vers une limite à préciser et conclure que l'EMV est asymptotiquement normal de variance asymptotique à préciser.

Exemple 7.2.12. Le cas du modèle gaussien $\mathcal{P} = \{\mathcal{N}_1(\mu,\sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$: Pour tout $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$, la vraisemblance est donnée par

$$p_n(x,(\mu,\sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

On pose $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ et $v_n = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2$. Comme $\sum_{i=1}^n (x_i - \bar{x}_n) = 0$,

$$\sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - \mu)^2 = \sum_{i=1}^n \left((x_i - \bar{x}_n)^2 + (\bar{x}_n - \mu)^2\right).$$

Donc la vraisemblance et la log-vraisemblance s'écrivent

$$p_n(x,(\mu,\sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-n\frac{(\bar{x}_n-\mu)^2+v_n}{2\sigma^2}}$$
$$l_n(x,(\mu,\sigma^2)) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - n\frac{(\bar{x}_n-\mu)^2+v_n}{2\sigma^2}.$$

À $\sigma^2 > 0$ fixé, on voit que la log-vraisemblance est maximale pour $\mu = \bar{x}_n$ et vaut alors $-\frac{n}{2}(\ln(2\pi) + f(\sigma^2))$ où

$$f(\lambda) = \ln\lambda + \frac{v_n}{\lambda}.$$

On cherche donc maintenant à minimiser $f(\lambda)$ pour $\lambda \in \mathbb{R}_+^*$. Comme la dérivée $f'(\lambda) = \frac{1}{\lambda}(1 - \frac{v_n}{\lambda})$ est négative sur $]0, v_n]$ et positive sur $[v_n, +\infty[$, la fonction f atteint son minimum en v_n. On conclut donc que la log-vraisemblance est maximale en $(\bar{x}_n, v_n)$. Ainsi l'EMV de (μ,σ^2) est le couple moyenne empirique, variance empirique $(\bar{X}_n, V_n)$. Notons que l'on obtient également l'EMV en résolvant le système

$$\begin{cases} \frac{\partial}{\partial\mu} l_n(x,(\mu,\sigma^2)) = 0 \\ \frac{\partial}{\partial\sigma^2} l_n(x,(\mu,\sigma^2)) = 0 \end{cases} \Leftrightarrow \begin{cases} n\frac{\bar{x}_n-\mu}{\sigma^2} = 0 \\ -\frac{n}{2}\left(\frac{1}{\sigma^2} - \frac{(\bar{x}_n-\mu)^2+v_n}{\sigma^4}\right) = 0 \end{cases}.$$

Comme $\mathbb{E}_{(\mu,\sigma^2)}((\bar{X}_n, V_n)) = (\mu, \frac{n-1}{n}\sigma^2)$, l'EMV est un estimateur biaisé. Il est fortement convergent d'après la loi forte des grands nombres.

Pour démontrer qu'il est asymptotiquement normal, on remarque que d'après le corollaire 6.2.4, $(\bar{X}_n, V_n) \overset{\mathcal{L}}{=} \left(\frac{1}{n}\sum_{i=1}^n X_i, \frac{\sigma^2}{n}\sum_{i=2}^n G_i^2\right)$ où $(G_i)_{i\geq 1}$ est une suite de variables I.I.D. suivant la loi normale centrée réduite indépendante de $(X_i)_{i\geq 1}$. On en déduit que

$$\sqrt{n}\left((\bar{X}_n, V_n) - (\mu, \sigma^2)\right) \overset{\mathcal{L}}{=} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n (X_i, \sigma^2 G_i^2) - (\mu, \sigma^2)\right) - \left(0, \frac{\sigma^2 G_1^2}{\sqrt{n}}\right).$$

D'après le théorème de la limite centrale multidimensionnel 6.2.8, le premier terme du second membre converge en loi vers la loi gaussienne centrée de matrice de covariance égale à celle du vecteur $(X_1, \sigma^2 G_1^2)$ c'est-à-dire

$$\Sigma(\sigma^2) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}.$$

Le second terme du second membre converge presque sûrement vers $(0, 0)$. Avec le second cas particulier du théorème de Slutsky 5.3.12, on conclut que l'Estimateur du Maximum de Vraisemblance $(\bar{X}_n, V_n)$ est asymptotiquement normal de matrice de covariance asymptotique $\Sigma(\sigma^2)$.

Exercice 7.2.13. Dans le modèle de Bernoulli $\mathcal{P} = \{\mathcal{B}(p), p \in [0,1]\}$, vérifier que l'EMV de p est la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Dans les deux exemples précédents, l'Estimateur du Maximum de Vraisemblance est fortement convergent et asymptotiquement normal. En fait, ces propriétés sont assez générales :

Théorème 7.2.14. *Sous de bonnes propriétés de régularité sur le modèle que nous ne préciserons pas, l'Estimateur du Maximum de Vraisemblance de θ est fortement convergent et asymptotiquement normal de variance asymptotique $I^{-1}(\theta)$ où la matrice $\boxed{I(\theta) = \mathbb{E}_\theta\left(\nabla_\theta l(X_1, \theta)\nabla_\theta^* l(X_1, \theta)\right)}$ s'appelle information de Fisher ($\forall 1 \leq i, j \leq d$, $I_{ij}(\theta) = \mathbb{E}_\theta\left(\frac{\partial l}{\partial \theta_i}(X_1, \theta)\frac{\partial l}{\partial \theta_j}(X_1, \theta)\right)$).*

Nous renvoyons au paragraphe II.16 de [1] pour une étude précise et rigoureuse des propriétés de l'EMV.

Remarque 7.2.15. La précision asymptotique de l'estimation de θ par maximum de vraisemblance est donnée par l'inverse de l'information de Fisher. Conformément à l'intuition, plus l'information est grande et meilleure est la précision de l'estimateur du maximum de vraisemblance.
L'information de Fisher a une autre expression qui permet souvent de simplifier les calculs. Elle s'exprime comme l'opposé de l'espérance de la matrice hessienne $\nabla_\theta^2 l(X_1, \theta)$: $\boxed{I(\theta) = -\mathbb{E}_\theta\left(\nabla_\theta^2 l(X_1, \theta)\right)}$. En effet, dans le cas d'un modèle à densité (sinon il suffit de remplacer les intégrales par des sommes), comme pour tout $\theta \in \Theta$, $1 = \int p(x_1, \theta)dx_1$ en dérivant cette égalité par rapport à θ_i où $i \in \{1, \ldots, d\}$, on obtient

$$0 = \int \frac{\partial p}{\partial \theta_i}(x_1, \theta)dx_1 = \int \frac{\partial l}{\partial \theta_i}(x_1, \theta)p(x_1, \theta)dx_1. \tag{7.1}$$

En dérivant maintenant par rapport à θ_j pour $j \in \{1, \ldots, d\}$, il vient

$$0 = \int \frac{\partial^2 l}{\partial \theta_i \partial \theta_j}(x_1, \theta)p(x_1, \theta)dx_1 + \int \frac{\partial l}{\partial \theta_i}(x_1, \theta)\frac{\partial l}{\partial \theta_j}(x_1, \theta)p(x_1, \theta)dx_1.$$

On conclut donc que

$$\mathbb{E}_\theta\left(\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}(X_1,\theta)\right) = -\mathbb{E}_\theta\left(\frac{\partial l}{\partial\theta_i}(X_1,\theta)\frac{\partial l}{\partial\theta_j}(X_1,\theta)\right).$$

Exercice 7.2.16. Calculer l'information de Fisher dans le modèle de Bernoulli et le modèle $\mathcal{P} = \{\mathcal{E}(\theta), \theta > 0\}$.
Vérifier également que dans le modèle gaussien $\mathcal{P} = \{\mathcal{N}_1(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$,

$$I(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

La plupart des modèles qui portent sur les familles de lois usuelles paramétrées par $\theta \in \Theta \subset \mathbb{R}^d$ sont dits exponentiels au sens où

$$\boxed{p(x_1, \theta) = h(x_1)\exp\left(\sum_{j=1}^d \gamma_j(\theta)T_j(x_1) - \varphi(\gamma(\theta))\right)}$$

avec, dans le cas d'un modèle à densité, $h : \mathbb{R}^k \to \mathbb{R}_+$, $\gamma = (\gamma_1, \ldots, \gamma_d) : \mathbb{R}^d \to \mathbb{R}^d$ une fonction injective et $T = (T_1, \ldots, T_d)$ une fonction telle que pour tout $\lambda \in \mathbb{R}^d$, la fonction $x_1 \in \mathbb{R}^k \to \lambda.T(x_1)$ est non constante sur l'ensemble $\{x_1 \in \mathbb{R}^k : \ h(x_1) > 0\}$. Dans le cas discret, $\mathbb{R}^k$ est remplacé par l'ensemble dénombrable F dans lequel les X_i prennent leurs valeurs.
La fonction φ assure la normalisation : pour $\lambda = (\lambda_1, \ldots, \lambda_d)$,

$$\varphi(\lambda) = \begin{cases} \ln\left(\sum_{x_1\in F} h(x_1)\exp\left(\sum_{j=1}^d \lambda_j T_j(x_1)\right)\right) & \text{dans le cas discret,} \\ \ln\left(\int_{\mathbb{R}^k} h(x_1)\exp\left(\sum_{j=1}^d \lambda_j T_j(x_1)\right)dx_1\right) & \text{dans le cas à densité.} \end{cases}$$

Exemple 7.2.17.

- Le modèle gamma $\mathcal{P} = \{\Gamma(a, \lambda), (a, \lambda) \in (\mathbb{R}_+^*)^2\}$ est exponentiel puisque

$$p(x_1, (a, \lambda)) = \frac{1_{\{x_1>0\}}}{\Gamma(a)}\lambda^a x_1^{a-1}e^{-\lambda x_1} = \frac{1_{\{x_1>0\}}}{x_1}\exp\left(a\ln x_1 - \lambda x_1 + a\ln\lambda - \ln\Gamma(a)\right).$$

- Le modèle de Bernoulli est exponentiel puisque

$$\forall x_1 \in \{0, 1\},\ p^{x_1}(1-p)^{1-x_1} = e^{x_1\ln(\frac{p}{1-p})+\ln(1-p)}$$

et la fonction $\gamma(p) = \ln(\frac{p}{1-p})$ est strictement croissante sur $]0, 1[$.

Exercice 7.2.18. Montrer que les modèles gaussien, beta $\mathcal{P} = \{\beta(a, b), (a, b) \in (\mathbb{R}_+^*)^2\}$, de Poisson $\mathcal{P} = \{\mathcal{P}(\theta), \theta > 0\}$, géométrique $\mathcal{P} = \{\mathcal{G}eo(p), p \in]0, 1]\}$ sont exponentiels mais pas le modèle de Cauchy $\mathcal{P} = \{\mathcal{C}(\theta), \theta > 0\}$.

Nous allons donner une idée de la preuve du théorème 7.2.14 dans le cas d'un modèle exponentiel avec un paramètre de dimension $d = 1$ (pour simplifier l'analyse) qui prend ses valeurs dans un ouvert Θ de $\mathbb{R}$ et une fonction γ régulière. Comme γ est injective, γ est alors strictement monotone et on la supposera strictement croissante afin de fixer les idées.

Enfin nous supposerons le modèle à densité (ce n'est pas restrictif : dans le cas discret, on remplace les intégrales par des sommes).
Démonstration : On a alors

$$p(x_1,\theta) = h(x_1)e^{\gamma(\theta)T(x_1)-\varphi(\gamma(\theta))} \text{ avec } \varphi(\lambda) = \ln\left(\int h(x_1)e^{\lambda T(x_1)}dx_1\right).$$

En dérivant sous le signe intégral par rapport à λ (ce que l'on peut justifier à l'aide du théorème de convergence dominée à l'intérieur du domaine où φ est finie), il vient

$$\begin{aligned}
\varphi'(\lambda) &= \frac{\int T(x_1)h(x_1)e^{\lambda T(x_1)}dx_1}{\int h(y)e^{\lambda T(y)}dy} = \int T(x_1)h(x_1)e^{\lambda T(x_1)-\varphi(\lambda)}dx_1 \\
\varphi''(\lambda) &= \int T(x_1)h(x_1)e^{\lambda T(x_1)-\varphi(\lambda)}(T(x_1)-\varphi'(\lambda))dx_1 \\
&= \int T^2(x_1)h(x_1)e^{\lambda T(x_1)-\varphi(\lambda)}dx_1 - \left(\int T(x_1)h(x_1)e^{\lambda T(x_1)-\varphi(\lambda)}dx_1\right)^2,
\end{aligned}$$

ce qui entraîne que $\varphi'(\gamma(\theta)) = \mathbb{E}_\theta(T(X_1))$ et $\varphi''(\gamma(\theta)) = \mathrm{Var}_\theta(T(X_1))$. Comme, d'après la définition des modèles exponentiels, la fonction T est non constante sur $\{x_1 : h(x_1) > 0\}$, la fonction $\varphi''(\lambda)$ est strictement positive par l'inégalité de Cauchy-Schwarz. Avec la croissance stricte de γ, on en déduit que la fonction $\varphi' \circ \gamma$ est également strictement croissante. Sa continuité entraîne qu'elle est inversible d'inverse ψ continu. En outre, l'image $\varphi'(\gamma(\Theta))$ de Θ par l'application $\varphi' \circ \gamma$ est un ouvert.
La log-vraisemblance $l_n(x,\theta)$ est, à une fonction ne dépendant pas de θ près, égale à $\gamma(\theta)\sum_{i=1}^n T(x_i) - n\varphi(\gamma(\theta))$ si bien que l'équation d'optimalité du premier ordre s'écrit

$$n\gamma'(\theta)\left(\frac{1}{n}\sum_{i=1}^n T(x_i) - \varphi'(\gamma(\theta))\right) = 0.$$

Cette équation admet comme unique solution $\hat\theta_n(x) = \psi(\frac{1}{n}\sum_{i=1}^n T(x_i))$ dès que $\frac{1}{n}\sum_{i=1}^n T(x_i) \in \varphi'(\gamma(\Theta))$. Par la loi forte des grands nombres, $\mathbb{P}_\theta$ presque sûrement $\frac{1}{n}\sum_{i=1}^n T(X_i)$ converge vers $\mathbb{E}_\theta(T(X_1)) = \varphi'(\gamma(\theta))$. On en déduit que $\mathbb{P}_\theta$ presque sûrement l'EMV $\hat\theta_n = \psi(\frac{1}{n}\sum_{i=1}^n T(X_i))$ existe à partir d'un certain rang n et converge vers $\psi(\varphi'(\gamma(\theta))) = \theta$ par continuité de ψ.
La fonction g définie par

$$g(t) = \begin{cases} \frac{t-\theta}{\varphi'(\gamma(t))-\varphi'(\gamma(\theta))} & \text{si } t \neq \theta \\ \frac{1}{\varphi''(\gamma(\theta))\gamma'(\theta)} & \text{si } t = \theta \end{cases},$$

est continue. Des égalités $\varphi'(\gamma(\hat\theta_n)) - \varphi'(\gamma(\theta)) = \frac{1}{n}\sum_{i=1}^n T(X_i) - \mathbb{E}_\theta(T(X_1))$ et $\varphi''(\gamma(\theta)) = \mathrm{Var}_\theta(T(X_1))$, on déduit que

$$\sqrt{n}(\hat\theta_n - \theta) = g(\hat\theta_n)\sqrt{\varphi''(\gamma(\theta))} \times \sqrt{\frac{n}{\mathrm{Var}_\theta(T(X_1))}}\left(\frac{1}{n}\sum_{i=1}^n T(X_i) - \mathbb{E}_\theta(T(X_1))\right).$$

Sous $\mathbb{P}_\theta$, la convergence presque sûre de $\hat\theta_n$ vers θ entraîne que le premier terme du produit converge presque sûrement vers $g(\theta)\sqrt{\varphi''(\gamma(\theta))} = 1/(\gamma'(\theta)\sqrt{\varphi''(\gamma(\theta))})$. Par le théorème de la limite centrale 5.4.1, le second converge en loi vers la loi normale centrée

réduite. On déduit alors du théorème de Slutsky 5.3.12 que sous $\mathbb{P}_\theta$, $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers la loi $\mathcal{N}_1(0, \frac{1}{(\gamma')^2(\theta)\varphi''(\gamma(\theta))})$.
Comme $\frac{\partial}{\partial\theta} l(x_1, \theta) = \gamma'(\theta)(T(x_1) - \varphi'(\gamma(\theta)))$, l'information de Fisher est égale à

$$I(\theta) = (\gamma')^2(\theta)\mathbb{E}_\theta((T(X_1) - \varphi'(\gamma(\theta)))^2) = (\gamma')^2(\theta)\mathrm{Var}_\theta(T(X_1)) = (\gamma')^2(\theta)\varphi''(\gamma(\theta)),$$

ce qui achève la démonstration. □

Remarque 7.2.19. Soit $Z = z(X_1, \ldots, X_n)$ un estimateur sans biais d'un paramètre θ de dimension 1 dans un modèle à densité (sinon, on remplace les intégrales par des sommes). D'après (7.1),

$$0 = \int \frac{\partial l}{\partial\theta}(x_1, \theta)p(x_1, \theta)dx_1 = \mathbb{E}_\theta\left(\frac{\partial l}{\partial\theta}(X_i, \theta)\right). \tag{7.2}$$

De manière analogue, comme $\theta = \mathbb{E}_\theta(Z) = \int z(x)p_n(x, \theta)dx$, en dérivant par rapport à θ puis en utilisant (7.2), l'inégalité de Cauchy-Schwarz et l'indépendance des variables X_i sous $\mathbb{P}_\theta$, on obtient

$$\begin{aligned}
1 &= \int z(x)\frac{\partial l_n}{\partial\theta}(x, \theta)p_n(x, \theta)dx = \mathbb{E}_\theta\left(Z\sum_{i=1}^n \frac{\partial l}{\partial\theta}(X_i, \theta)\right) \\
&= \mathbb{E}_\theta\left((Z - \theta)\sum_{i=1}^n \frac{\partial l}{\partial\theta}(X_i, \theta)\right) \\
&\leq \sqrt{\mathrm{Var}_\theta(Z)} \times \sqrt{\mathbb{E}_\theta\left(\left(\sum_{i=1}^n \frac{\partial l}{\partial\theta}(X_i, \theta)\right)^2\right)} = \sqrt{\mathrm{Var}_\theta(Z)}\sqrt{nI(\theta)}.
\end{aligned}$$

Avec la propriété 7.2.5, on en déduit la minoration de Fréchet-Darmois-Cramer-Rao du risque quadratique de l'estimateur : $R(Z, \theta) = \mathrm{Var}_\theta(Z) \geq \frac{1}{nI(\theta)}$.
Si, dans la convergence en loi énoncée dans le théorème 7.2.14, il y a convergence du moment d'ordre 2, alors le risque quadratique de l'Estimateur du Maximum de Vraisemblance vérifie $R(\hat{\theta}_n, \theta) = \frac{1}{n}\mathbb{E}_\theta((\sqrt{n}(\hat{\theta}_n - \theta))^2) \sim \frac{1}{nI(\theta)}$ pour $n \to +\infty$. Dans ces conditions, l'EMV (qui n'est pas forcément sans biais) atteint asymptotiquement la borne. Cela montre la qualité de cet estimateur.

L'exercice suivant est consacré au modèle uniforme, dans lequel l'EMV est bien fortement convergent mais pas asymptotiquement normal.

Exercice 7.2.20. On se place dans le modèle uniforme $\mathcal{P} = \{\mathcal{U}[0, \theta], \theta > 0\}$.

1. Ce modèle est-il exponentiel ?
2. Vérifier que l'EMV de θ est $\hat{\theta}_n = \max(X_1, \ldots, X_n)$.
3. Pour $x \in [0, \theta]$, calculer $\mathbb{P}_\theta(\hat{\theta}_n \leq x)$ et en déduire que $\hat{\theta}_n$ est un estimateur convergent de θ. Avec la monotonie de la suite $(\hat{\theta}_n)_n$ conclure qu'il est même fortement convergent.
4. Vérifier que $\hat{\theta}_n$ a même loi que $\theta U^{1/n}$ où $U \sim \mathcal{U}[0, 1]$. En remarquant que pour tout $u \in [0, 1]$, $n(1 - u^{1/n})$ converge vers $-\ln(u)$ lorsque n tend vers $+\infty$, conclure que sous $\mathbb{P}_\theta$, $n(\theta - \hat{\theta}_n)$ converge en loi vers $-\theta\ln(U)$ lorsque n tend vers $+\infty$. Ainsi, dans ce modèle, l'EMV converge à la vitesse $1/n$ et non $1/\sqrt{n}$.

7.2.3 Estimateurs de Moments

Dans le cas par exemple du modèle béta $\mathcal{P} = \{\beta(a,b), (a,b) \in (\mathbb{R}_+^*)^2\}$, il n'est pas possible d'obtenir une expression analytique de l'estimateur du maximum de vraisemblance de (a,b). Cependant,

$$\begin{cases} \mathbb{E}_{(a,b)}(X_1) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1} dx = \frac{\Gamma(a+b)}{\Gamma(a+b+1)} \frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{a+b} \\ \mathbb{E}_{(a,b)}(X_1(1-X_1)) = \frac{\Gamma(a+b)}{\Gamma(a+b+2)} \frac{\Gamma(a+1)}{\Gamma(a)} \frac{\Gamma(b+1)}{\Gamma(b)} = \frac{ab}{(a+b)(a+b+1)} \end{cases}.$$

En posant $c = \mathbb{E}_{(a,b)}(X_1)$ et $d = \mathbb{E}_{(a,b)}(X_1(1-X_1))$ et en inversant le système, il vient $a = \frac{cd}{c-c^2-d}$ et $b = \frac{(1-c)d}{c-c^2-d}$ où $c - c^2 - d = \frac{ab}{(a+b)^2} - \frac{ab}{(a+b)(a+b+1)} = \frac{ab}{(a+b)^2(a+b+1)} > 0$. Par la loi forte des grands nombres,

$$\hat{c}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{d}_n = \frac{1}{n} \sum_{i=1}^n X_i(1-X_i)$$

sont respectivement des estimateurs fortement convergents de c et d. Comme l'application $(y,z) \in \mathbb{R}^2 \to \frac{1}{y-y^2-z}(yz, (1-y)z)$ est définie et continue en tout point de l'ouvert $\{(y,z) \in \mathbb{R}^2 : y - y^2 - z \neq 0\}$, on en déduit que le couple

$$(\hat{a}_n, \hat{b}_n) = \frac{1}{\hat{c}_n - \hat{c}_n^2 - \hat{d}_n} (\hat{c}_n \hat{d}_n, (1-\hat{c}_n)\hat{d}_n),$$

défini au moins pour n assez grand, est un estimateur fortement convergent de (a,b).

De façon générale, pour estimer θ, la méthode des moments consiste à trouver une fonction φ telle que pour tout $\theta \in \Theta$, $\mathbb{E}_\theta|\varphi(X_1)| < +\infty$ et que la fonction $m(\theta) = \mathbb{E}_\theta(\varphi(X_1))$ soit inversible d'inverse continu. Alors l'estimateur de θ

$$m^{-1}\left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i)\right)$$

est fortement convergent par la loi forte des grands nombres (et même asymptotiquement normal sous des hypothèses un peu plus restrictives sur φ).

Exercice 7.2.21. Dans le modèle uniforme $\mathcal{P} = \{\mathcal{U}[0,\theta], \theta > 0\}$, calculer $\mathbb{E}_\theta(X_1)$. En déduire un estimateur de moments pour θ. Calculer son risque quadratique et vérifier que l'estimateur du maximum de vraisemblance lui est préférable.

Exercice 7.2.22. Pour estimer la taille l d'une population d'oiseaux, on en capture respectivement $X_1, \ldots, X_n$ pendant n journées successives en les relachant le soir. On suppose que chaque oiseau a la même probabilité inconnue p de se faire capturer chaque jour et ce indépendamment des autres oiseaux.

1. Quel modèle proposez-vous ?
2. Calculer $\mathbb{E}_{(l,p)}(X_1)$ et $\mathbb{E}_{(l,p)}(X_1^2)$ et en déduire un estimateur de moments pour (l,p).

7.2.4 Amélioration d'estimateurs

Dans le modèle de Bernoulli, pour une observation $x = (x_1, \ldots, x_n) \in \{0,1\}^n$, il semble que seul le nombre $(x_1 + \ldots + x_n)$ de 1 apporte de l'information sur le paramètre p, la manière dont les 0 et les 1 se succèdent n'ayant aucune importance. On peut formaliser cette intuition en s'intéressant à la loi conditionnelle de $X = (X_1, \ldots, X_n)$ sachant $S = s$ où $S = X_1 + \ldots + X_n$. Soit $s \in \{0, \ldots, n\}$. Comme $S \sim \mathcal{B}(n, p)$ sous $\mathbb{P}_p$, $\mathbb{P}_p(S = s) = \binom{n}{s} p^s (1-p)^{n-s}$. Donc

$$\mathbb{P}_p(X = x | S = s) = \frac{\mathbb{P}_p(X = x, S = s)}{\mathbb{P}_p(S = s)} = \begin{cases} 0 \text{ si } x_1 + \ldots + x_n \neq s, \\ \frac{\mathbb{P}_p(X=x)}{\mathbb{P}_p(S=s)} = \frac{p^s(1-p)^{n-s}}{\binom{n}{s} p^s (1-p)^{n-s}} = \frac{1}{\binom{n}{s}} \text{ sinon.} \end{cases}$$

Donc la loi conditionnelle de X sachant $S = s$ est la loi uniforme sur $\{x \in \{0, 1\}^n : x_1 + \ldots + x_n = s\}$. Elle ne dépend pas du paramètre p, ce qui signifie que toute l'information sur p dans l'échantillon X est contenue dans la statistique S.

Définition 7.2.23. *Une statistique S est dite exhaustive si pour tout s, la loi conditionnelle de $X = (X_1, \ldots, X_n)$ sachant $S = s$ ne dépend pas du paramètre θ.*

On peut caractériser sur la vraisemblance le caractère exhaustif d'une statistique :

Théorème 7.2.24 (de Halmos-Savage ou de factorisation)**.** *La statistique $S = \mathcal{S}(X_1, \ldots, X_n)$ est exhaustive si et seulement si la vraisemblance $p_n(x, \theta)$ se factorise sous la forme*

$$\boxed{\forall x = (x_1, \ldots, x_n),\ \forall \theta \in \Theta,\ p_n(x, \theta) = \psi(\mathcal{S}(x), \theta) \varphi(x).}$$

Démonstration : On effectue la démonstration dans le seul cas d'un modèle discret.
Condition nécessaire : On a

$$p_n(x, \theta) = \mathbb{P}_\theta(X = x) = \mathbb{P}_\theta(X = x, S = \mathcal{S}(x)) = \mathbb{P}_\theta(X = x | S = \mathcal{S}(x)) \mathbb{P}_\theta(S = \mathcal{S}(x)).$$

Le premier terme du produit ne dépend pas de θ et on le note $\varphi(x)$. Le second terme est de la forme $\psi(\mathcal{S}(x), \theta)$.
Condition suffisante : On a alors

$$\mathbb{P}_\theta(S = s) = \mathbb{P}_\theta(\mathcal{S}(X) = s) = \sum_{y : \mathcal{S}(y) = s} \psi(\mathcal{S}(y), \theta) \varphi(y) = \psi(s, \theta) \sum_{y : \mathcal{S}(y) = s} \varphi(y).$$

On conclut alors par un calcul généralisant celui effectué dans le modèle de Bernoulli :

$$\mathbb{P}_\theta(X = x | S = s) = \begin{cases} 0 \text{ si } \mathcal{S}(x) \neq s, \\ \frac{\mathbb{P}_\theta(X=x)}{\mathbb{P}_\theta(S=s)} = \frac{\psi(s,\theta)\varphi(x)}{\psi(s,\theta) \sum_{y:\mathcal{S}(y)=s} \varphi(y)} = \frac{\varphi(x)}{\sum_{y:\mathcal{S}(y)=s} \varphi(y)} \text{ sinon.} \end{cases}$$

□

Exercice 7.2.25. Dans le modèle gaussien, montrer que $(\bar{X}_n, V_n)$ est une statistique exhaustive.

Soit S est une statistique exhaustive et Z un estimateur de $g(\theta)$. Comme Z est fonction de $X = (X_1, \ldots, X_n)$ et comme la loi conditionnelle de X sachant $S = s$ ne dépend pas de θ, l'espérance conditionnelle de Z sachant S ne dépend pas de θ. On la note $Z_S = \mathbb{E}(Z|S)$. Ainsi Z_S est un estimateur de $g(\theta)$. Il faut noter que lorsque la statistique S n'est pas exhaustive, l'espérance conditionnelle de Z sachant S dépend en général de θ et ne constitue alors pas un estimateur.

Théorème 7.2.26 (de Rao Blackwell)**.** *Soit Z un estimateur de $g(\theta)$ et S une statistique exhaustive. Lorsque g est à valeurs réelles, alors $Z_S = \mathbb{E}(Z|S)$ est un estimateur préférable à Z.*

Démonstration : D'après le corollaire 2.5.7 pour le cas discret et la proposition 3.3.20 pour le cas à densité,

$$\mathbb{E}_\theta(Z_S) = \mathbb{E}_\theta(Z) \text{ et } \mathrm{Var}_\theta(Z_S) \leq \mathrm{Var}_\theta(Z).$$

Avec la propriété 7.2.5, on en déduit que

$$R(Z_S, \theta) = (g(\theta) - \mathbb{E}_\theta(Z_S))^2 + \mathrm{Var}_\theta(Z_S) \leq (g(\theta) - \mathbb{E}_\theta(Z))^2 + \mathrm{Var}_\theta(Z) = R(Z, \theta).$$

□

Remarque 7.2.27.
- Nous retiendrons du théorème de Rao-Blackwell qu'il faut s'efforcer de construire des estimateurs qui sont des fonctions des statistiques exhaustives.
- Notons que si $p_n(x, \theta) = \psi(\mathcal{S}(x), \theta)\varphi(x)$ alors pour maximiser la vraisemblance en θ, il suffit de maximiser la fonction $\theta \to \psi(\mathcal{S}(x), \theta)$ qui ne dépend de x qu'au travers de $\mathcal{S}(x)$. On en déduit que l'Estimateur du Maximum de Vraisemblance ne dépend de X qu'au travers de la statistique exhaustive S. Il est donc égal à son espérance conditionnelle sachant S et ne peut être amélioré par le procédé précédent. Cela souligne une nouvelle fois la bonne qualité de l'Estimateur du Maximum de Vraisemblance.
- Lorsque g est à valeurs dans $\mathbb{R}^q$ où $q \geq 1$, on peut définir le risque quadratique de l'estimateur Z de $g(\theta)$ comme la matrice symétrique positive $q \times q$:

  $$R(Z, \theta) = \mathbb{E}_\theta\left((Z - g(\theta))(Z - g(\theta))^*\right) = (\mathbb{E}_\theta(Z) - g(\theta))(\mathbb{E}_\theta(Z) - g(\theta))^* + \mathrm{Cov}_\theta(Z)$$

  où $\mathrm{Cov}_\theta(Z)$ désigne la matrice de covariance de Z sous $\mathbb{P}_\theta$. Si S est une statistique exhaustive, alors

  $$\forall u \in \mathbb{R}^q,\ u.\mathrm{Cov}_\theta(Z_S)u = \mathrm{Var}_\theta((u.Z)_S) \leq \mathrm{Var}_\theta(u.Z) = u.\mathrm{Cov}_\theta(Z)u.$$

  Comme $\mathbb{E}_\theta(Z_S) = \mathbb{E}_\theta(Z)$, on en déduit que $R(Z_S, \theta) \leq R(Z, \theta)$ au sens où $\forall u \in \mathbb{R}^q$, $u.R(Z_S, \theta)u \leq u.R(Z, \theta)u$. Ainsi Z_S est préférable à Z.

Exemple 7.2.28. Dans le modèle de Bernoulli, nous avons vu que $S = X_1 + \ldots + X_n$ est une statistique exhaustive. Si on s'intéresse à l'estimateur sans biais $Z = X_1$ de p, d'après l'exercice 2.5.8, $Z_S = \frac{S}{n} = \bar{X}_n$. La moyenne empirique $\bar{X}_n$ est bien sûr préférable à X_1.

Exercice 7.2.29. On se place dans le modèle de Poisson $\mathcal{P} = \{\mathcal{P}(\theta), \theta > 0\}$.

1. Vérifier que $Z = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i=0\}}$ est un estimateur sans biais de $e^{-\theta}$ fortement convergent et asymptotiquement normal.
2. Donner une statistique exhaustive.
3. On pose $S = \sum_{i=1}^n X_i$. Vérifier que $S \sim \mathcal{P}(n\theta)$ puis que pour tout $i \in \{1, \ldots, n\}$, la loi conditionnelle de X_i sachant $S = s$ est la loi binomiale $\mathcal{B}(s, \frac{1}{n})$. En déduire que $\mathbb{E}_\theta(Z|S) = (1 - \frac{1}{n})^S$. Vérifier que $\mathrm{Var}_\theta((1 - \frac{1}{n})^S) = e^{-2\theta}\left(e^{\frac{\theta}{n}} - 1\right) \leq \mathrm{Var}_\theta(Z)$.
4. Donner un estimateur de $e^{-\theta}$ préférable à Z.

7.3 Intervalles de confiance

Lorsque l'on estime un paramètre $g(\theta)$, **supposé de dimension** 1 **dans tout ce paragraphe**, on veut avoir une idée de la précision de l'estimation effectuée. C'est le rôle des intervalles de confiance.

7.3.1 Approche non asymptotique

Définition 7.3.1. *Soit $\alpha \in]0,1[$. On dit qu'un intervalle $I(X_1,\ldots,X_n)$ qui s'exprime en fonction de l'échantillon est un intervalle de confiance pour $g(\theta)$ de niveau $1-\alpha$ si*

$$\boxed{\forall \theta \in \Theta,\ \mathbb{P}_\theta(g(\theta) \in I(X_1,\ldots,X_n)) = 1-\alpha.}$$

Lorsque $\forall \theta \in \Theta,\ \mathbb{P}_\theta(g(\theta) \in I(X_1,\ldots,X_n)) \geq 1-\alpha$, on parle d'intervalle de confiance de niveau $1-\alpha$ par excès.

Remarque 7.3.2.

- Les niveaux usuels sont 90%, 95% et 99% et correspondent respectivement à $\alpha = 10\%$, $\alpha = 5\%$ et $\alpha = 1\%$.
- Pour obtenir le maximum d'information, il faut s'efforcer de construire l'intervalle de confiance le moins large possible qui satisfait la condition de minoration donnée dans la définition.

Pour construire des intervalles de confiance, il est très utile d'introduire la notion de quantile.

Définition 7.3.3. *Soit Y une variable aléatoire réelle de fonction de répartition $F(y) = \mathbb{P}(Y \leq y)$. Pour $r \in]0,1[$, on appelle quantile (ou fractile) d'ordre r de la loi de Y le nombre*

$$q_r = \inf\{y \in \mathbb{R},\ F(y) \geq r\}.$$

Lorsque la fonction de répartition F est continue et strictement croissante (par exemple quand Y possède une densité strictement positive), elle est inversible d'inverse F^{-1} et pour tout $r \in]0,1[$, on a $q_r = F^{-1}(r)$.

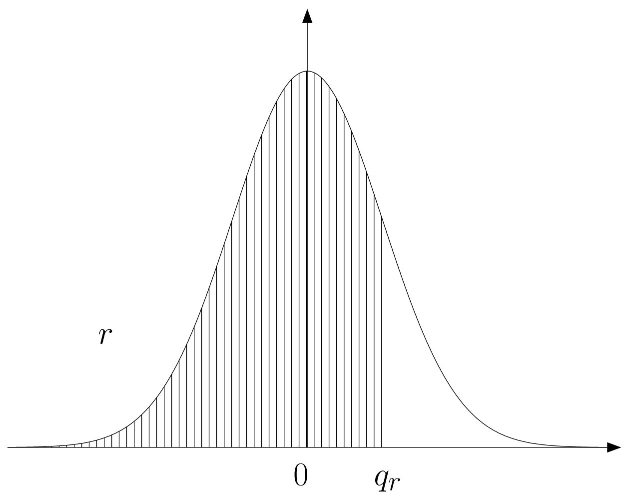


FIGURE 7.2 – Représentation d'une densité et du quantile q_r d'ordre r de cette densité : l'aire hachurée est égale à r.

La fonction de répartition est toujours croissante, ce qui entraîne la croissance de $r \to q_r$. Pour construire des intervalles de confiance et des tests, nous utiliserons les propriétés suivantes :

Propriétés 7.3.4. *On suppose que Y possède une densité.*

1. $\forall r \in]0,1[,\ F(q_r) = r$.
2. $\forall \alpha \in]0,1[,\ \mathbb{P}(Y \notin [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]) = \mathbb{P}(Y < q_\alpha) = \mathbb{P}(Y > q_{1-\alpha}) = \alpha$.
3. $\forall \alpha \in]0,1[,\ \mathbb{P}(Y \in [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]) = \mathbb{P}(Y \geq q_\alpha) = \mathbb{P}(Y \leq q_{1-\alpha}) = 1-\alpha$.
4. *Si la loi de Y est symétrique (i.e. la densité est une fonction paire), $\forall \alpha \in]0,1[,\ \mathbb{P}(|Y| > q_{1-\frac{\alpha}{2}}) = \alpha$ et $\mathbb{P}(|Y| \leq q_{1-\frac{\alpha}{2}}) = 1-\alpha$.*

Démonstration : 1. On a $\forall y < q_r$, $F(y) < r$ et par croissance de F, $\forall y > q_r$, $F(y) \geq r$. Comme F est continue, on en déduit que $F(q_r) = r$.
2. Le résultat se déduit des égalités $\mathbb{P}(Y < q_r) = \mathbb{P}(Y \leq q_r) = F(q_r) = r$ et $\mathbb{P}(Y > q_r) = 1 - F(q_r) = 1 - r$. Et le point **3.** s'obtient par passage au complémentaire.
4. Lorsque la densité de Y est une fonction paire, la variable aléatoire $-Y$ a même loi que Y. En outre $F(0) = 1/2$, ce qui entraîne que $q_{1-\frac{\alpha}{2}} > 0$. Donc

$$\begin{aligned}\mathbb{P}(|Y| > q_{1-\frac{\alpha}{2}}) &= \mathbb{P}(Y < -q_{1-\frac{\alpha}{2}}) + \mathbb{P}(Y > q_{1-\frac{\alpha}{2}}) = \mathbb{P}(-Y > q_{1-\frac{\alpha}{2}}) + \mathbb{P}(Y > q_{1-\frac{\alpha}{2}}) \\ &= 2\mathbb{P}(Y > q_{1-\frac{\alpha}{2}}) = \alpha\end{aligned}$$

et la dernière propriété s'en déduit par passage au complémentaire. □

Exemple 7.3.5. Le cas du modèle gaussien $\mathcal{P} = \{\mathcal{N}_1(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$:
Intervalles de confiance pour la moyenne μ :
D'après le corollaire 6.2.4, si $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ désignent respectivement la moyenne empirique et l'estimateur sans biais de la variance, le rapport

$\sqrt{n}\frac{\bar{X}_n-\mu}{S_n}$ suit la loi de Student $t(n-1)$ sous $\mathbb{P}_{(\mu,\sigma^2)}$. Notons $t_{n-1,r}$ le quantile d'ordre r de cette loi qui est symétrique. D'après les propriétés 7.3.4, pour $\alpha \in]0,1[$,

$$\begin{aligned}1-\alpha &= \mathbb{P}_{(\mu,\sigma^2)}\left(\sqrt{n}\left|\frac{\bar{X}_n-\mu}{S_n}\right| \leq t_{n-1,1-\frac{\alpha}{2}}\right)\\ &= \mathbb{P}_{(\mu,\sigma^2)}\left(\bar{X}_n - \frac{t_{n-1,1-\frac{\alpha}{2}}S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{t_{n-1,1-\frac{\alpha}{2}}S_n}{\sqrt{n}}\right).\end{aligned}$$

Donc $\boxed{\left[\bar{X}_n - \frac{t_{n-1,1-\frac{\alpha}{2}}S_n}{\sqrt{n}}, \bar{X}_n + \frac{t_{n-1,1-\frac{\alpha}{2}}S_n}{\sqrt{n}}\right]}$ est un intervalle de confiance de niveau $1-\alpha$ pour μ.

Exemple 7.3.6. Sur les dix dernières années, on a observé les températures moyennes suivantes en degrés Celsius au mois d'août à Paris :

$x = (x_1, \ldots, x_{10}) = (22, 19, 21, 23, 20, 22, 24, 18, 20, 25)$ telles que $\bar{x}_{10} = 21.4$ et $s_{10} = 2.22$.

D'après la table du paragraphe 11.4, $t_{9,0.975} = 2.262$ et $\frac{t_{9,0.975}\times s_{10}}{\sqrt{10}} = 1.59$. On en déduit que dans le modèle gaussien, $[19.81, 22.99]$ est un intervalle de confiance à 95% pour l'espérance de la température moyenne au mois d'août.

Exercice 7.3.7. On suppose que la variance σ^2 est connue et égale à σ_0^2.

1. Quelle est la loi de $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma_0}$ sous $\mathbb{P}_{(\mu,\sigma_0^2)}$?
2. En déduire un intervalle de confiance de niveau $1-\alpha$ pour μ (on note ϕ_r le quantile d'ordre $r \in]0,1[$ de la loi normale centrée réduite).
3. À l'aide des tables données à la fin du polycopié, comparer la largeur de cet intervalle de confiance avec celle de l'intervalle obtenu dans le modèle à variance inconnue si $S_n^2 = \sigma_0^2$.
4. Soit $Y = (Y_1, \ldots, Y_m)$ un autre échantillon de la loi $\mathcal{N}_1(\mu, \sigma_0^2)$ indépendant de X. Vérifier que les intervalles de confiance de niveau $1-\alpha$ pour μ construits à partir de X et Y sont disjoints avec probabilité $\mathbb{P}\left(|G| \geq \frac{\sqrt{n}+\sqrt{m}}{\sqrt{n+m}}\phi_{1-\frac{\alpha}{2}}\right)$ où $G \sim \mathcal{N}_1(0,1)$. Remarquer que $\left(\frac{\sqrt{n}+\sqrt{m}}{\sqrt{n+m}}\right)^2 \leq 2$ avec égalité lorsque $m = n$, ce qui assure que cette probabilité est minorée uniformément en (n, m).

Intervalles de confiance pour la variance σ^2 :

D'après le corollaire 6.2.4, $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ sous $\mathbb{P}_{(\mu,\sigma^2)}$. Notons $\chi^2_{n-1,r}$ le quantile d'ordre r de cette loi qui ne charge que $\mathbb{R}_+$. D'après les propriétés 7.3.4,

$$\begin{aligned}1-\alpha &= \mathbb{P}_{(\mu,\sigma^2)}\left(\chi^2_{n-1,\frac{\alpha}{2}} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}\right)\\ &= \mathbb{P}_{(\mu,\sigma^2)}\left(\frac{(n-1)S_n^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right).\end{aligned}$$

Ainsi $\boxed{\left[\frac{(n-1)S_n^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right]}$ est un intervalle de confiance de niveau $1-\alpha$ pour σ^2.

En regardant l'exemple du modèle gaussien, on voit qu'une étape clef pour construire un intervalle de confiance pour $g(\theta)$ consiste à trouver une fonction f telle que la loi de $f((X_1,\ldots,X_n),g(\theta))$ sous $\mathbb{P}_\theta$ ne dépend pas de θ. Une telle fonction s'appelle **fonction pivotale**. On utilise également une fonction pivotale dans l'exercice suivant.

Exercice 7.3.8. On se place dans le modèle gamma $\mathcal{P}=\{\Gamma(a,\theta),\theta>0\}$ où $a>0$ est supposé connu.

1. Donner une statistique exhaustive et déterminer l'Estimateur du Maximum de Vraisemblance $\hat{\theta}_n$ de θ.
2. Déterminer la loi de $\theta(X_1+\ldots+X_n)$ sous $\mathbb{P}_\theta$ et vérifier que cette loi ne dépend pas de θ. En déduire un intervalle de confiance de niveau $1-\alpha$ pour θ.

Lorsqu'il n'est pas possible d'exhiber une fonction pivotale, on peut se tourner vers l'approche asymptotique pour construire des intervalles de confiance.

7.3.2 Approche asymptotique

Définition 7.3.9. *Soit $\alpha\in]0,1[$. On appelle intervalle de confiance asymptotique pour $g(\theta)$ de niveau $1-\alpha$ une suite d'intervalles $(I_n(X_1,\ldots,X_n))_n$ telle que*

$$\boxed{\forall\theta\in\Theta,\ \lim_{n\to+\infty}\mathbb{P}_\theta(g(\theta)\in I_n(X_1,\ldots,X_n))=1-\alpha.}$$

Lorsqu'un estimateur est asymptotiquement normal, on peut tirer parti de cette propriété pour construire un intervalle de confiance asymptotique, comme le montre l'exemple suivant.

Exemple 7.3.10. Dans le modèle de Bernoulli $\mathcal{P}=\{\mathcal{B}(p),p\in]0,1[\}$, la moyenne empirique $\bar{X}_n$ est un estimateur de p fortement convergent. Par continuité de $y\to y(1-y)$, $\bar{X}_n(1-\bar{X}_n)$ est un estimateur fortement convergent de la variance $p(1-p)$. On a

$$\sqrt{\frac{n}{\bar{X}_n(1-\bar{X}_n)}}(\bar{X}_n-p)=\sqrt{\frac{p(1-p)}{\bar{X}_n(1-\bar{X}_n)}}\times\sqrt{\frac{n}{p(1-p)}}(\bar{X}_n-p).$$

Sous $\mathbb{P}_p$ le premier terme du produit converge presque sûrement vers 1 tandis que d'après le théorème de la limite centrale 5.4.1 le second converge en loi vers $G\sim\mathcal{N}_1(0,1)$. On déduit du théorème de Slutsky 5.3.12 que le membre de gauche converge en loi vers G. Donc pour $a\geq 0$, d'après la remarque 5.3.11,

$$\lim_{n\to+\infty}\mathbb{P}_p\left(\sqrt{\frac{n}{\bar{X}_n(1-\bar{X}_n)}}|\bar{X}_n-p|\leq a\right)=\mathbb{P}(|G|\leq a).$$

Si on note ϕ_r le quantile d'ordre r de la loi $\mathcal{N}_1(0,1)$ qui est symétrique, on a $\mathbb{P}(|G|\leq\phi_{1-\frac{\alpha}{2}})=1-\alpha$ d'après les propriétés 7.3.4. On conclut que $\left[\bar{X}_n-\phi_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}},\bar{X}_n+\phi_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]$ est un intervalle de confiance asymptotique pour p de niveau $1-\alpha$.

Exercice 7.3.11. Un sondage sur la popularité du premier ministre indique que 43% des personnes interrogées lui font confiance. Donner un intervalle de confiance asymptotique de niveau 95% pour la proportion p des Français qui lui font confiance lorsque le nombre n de personnes interrogées est 100. Même question lorsque $n=1000$.

7.4 Exercices

Exercice 7.4.1. On observe un échantillon $(X_1, \ldots, X_n)$ I.I.D. suivant la loi exponentielle $\mathcal{E}(\theta)$ où $\theta > 0$.

1. Vérifier que $S = X_1 + \ldots + X_n$ est une statistique exhaustive. Quelle est sa loi ?
2. Vérifier que sous $\mathbb{P}_\theta$, $2\theta S \sim \chi^2(2n)$. En déduire, pour $\alpha \in]0,1[$, la construction d'un intervalle de confiance de niveau $1-\alpha$ pour θ. À l'aide de la table de quantiles des lois du χ^2 donnée au paragraphe 11.3, préciser la mise-en-œuvre de cet intervalle de confiance pour $n = 10$ et $\alpha = 5\%$.

Exercice corrigé 7.4.2. Soit $(X_1, \ldots, X_n)$ un échantillon de variables aléatoires I.I.D. suivant une loi d'espérance μ et de variance $\sigma^2 \in]0, +\infty[$. Parmi les statistiques linéaires c'est-à-dire de la forme $\sum_{i=1}^n a_i X_i$ avec $(a_1, \ldots, a_n) \in \mathbb{R}^n$, quel est l'estimateur sans biais de μ de variance minimale ?

Exercice 7.4.3. On modélise la hauteur maximale annuelle d'un fleuve par une variable de Rayleigh de densité $f(x,a) = 1_{\{x>0\}} \frac{x}{a} e^{-\frac{x^2}{2a}}$ où $a > 0$ est un paramètre inconnu.

1. On suppose que l'on observe un échantillon $(X_1, \ldots, X_n)$ I.I.D. suivant cette loi. Donner l'Estimateur du Maximum de Vraisemblance $\hat{a}_n$ de a. Cet estimateur est-il sans biais ? fortement convergent ? Vérifier qu'il est asymptotiquement normal de variance asymptotique égale à l'inverse de l'information de Fisher.
2. Pendant une période de huit ans, on a observé les hauteurs maximales en mètres suivantes pour le fleuve $(x_1, \ldots, x_8) = (2.5, 1.8, 2.9, 0.9, 2.1, 1.7, 2.2, 2.8)$. On a $\sum_{i=1}^8 x_i^2 = 38.69$. Une compagnie d'assurance estime qu'une crue catastrophique avec une hauteur de 6 mètres au moins n'arrive au plus qu'une fois tous les mille ans. Est-ce justifié ?

Exercice 7.4.4. Une machine produit N pièces par jour. Chaque pièce a une probabilité θ d'être défectueuse et ce indépendamment des autres pièces. On souhaite estimer la probabilité q d'avoir au plus k pièces défectueuses sur un jour. Pour cela on dispose des nombres de pièces défecteuses observés chaque jour sur une période de n jours.

1. Préciser le modèle statistique. Est-ce un modèle exponentiel ? Donner une statistique S exhaustive.
2. Proposer un estimateur sans biais Z de q. Déterminer un estimateur Z_S qui lui est préférable.

Exercice 7.4.5. Dans l'industrie agro-alimentaire, on s'intéresse à la contamination du lait par un micro-organisme : les spores de clostridia. On souhaite estimer le nombre de spores dans 1ml de lait alors que l'on sait seulement tester si un tel échantillon contient ou non des spores. On note X_k le nombre de spores dans le k-ième échantillon, $Y_k = 1_{\{X_k=0\}}$ et $S = \sum_{k=1}^n Y_k$ le nombre d'échantillons stériles. On suppose que les variables $(X_1, \ldots, X_n)$ sont I.I.D. suivant la loi de Poisson $\mathcal{P}(\theta)$ où $\theta > 0$.

1. Donner les lois de Y_k et de S ?
2. On note $q = \mathbb{P}_\theta(Y_k = 1)$. Quel est l'Estimateur du Maximum de Vraisemblance de q ? En déduire un estimateur de θ.
3. Dans le cas où on observe 6 tubes stériles sur 10, calculer les estimateurs précédents et donner des intervalles de confiance asymptotiques de niveau 95% pour q et θ.

Exercice 7.4.6. On rappelle que dans le modèle uniforme $\mathcal{P} = \{\mathcal{U}[0,\theta], \theta > 0\}$, l'Estimateur du Maximum de Vraisemblance de θ est $\hat{\theta}_n = \max(X_1, \ldots, X_n)$.

1. Pour $x \in \mathbb{R}$, calculer $\mathbb{P}_\theta\left(\frac{\hat{\theta}_n}{\theta} \leq x\right)$ et en déduire que la loi de $\dfrac{\hat{\theta}_n}{\theta}$ ne dépend pas de θ.
2. Construire un intervalle de confiance de niveau $1-\alpha$ pour θ.

Exercice 7.4.7. La loi de Pareto de paramètre de forme $\alpha > 0$ et de paramètre d'échelle $\beta > 0$ est donnée par sa densité

$$f_{\alpha,\beta}(x) = 1_{\{x>\beta\}}\frac{\alpha\beta^\alpha}{x^{\alpha+1}}.$$

On observe un échantillon $(X_1, \ldots, X_n)$ de variables I.I.D. suivant cette loi.

1. Donnez une statistique exhaustive de dimension 2 puis déterminez l'Estimateur du Maximum de Vraisemblance du couple (α, β).
2. Le paramètre d'échelle β est maintenant supposé connu égal à 1. Vérifier que si X suit la loi de Pareto de paramètres α et 1, $\ln(X)$ suit la loi exponentielle de paramètre α. Construire un intervalle de confiance asymptotique de niveau $1-\eta$ pour α.

Problème corrigé 7.4.8. On s'intéresse à la durée de vie de deux composants électroniques se trouvant sur un système solidaire. Si l'un des deux composants tombe en panne, le système tout entier doit être changé. Les durées de vie de ces deux composants sont modélisées par l'intermédiaire de variables aléatoires exponentielles de paramètres λ et μ indépendantes. Formellement, on considère un n-échantillon I.I.D. $(X_1, \ldots, X_n)$ de loi $\mathcal{E}(\lambda)$ et un n-échantillon I.I.D. $(Y_1, \ldots, Y_n)$, indépendant du précédent, de loi $\mathcal{E}(\mu)$ où $\lambda > 0$ et $\mu > 0$.

On observe seulement la durée de vie du composant qui est tombé en panne et donc le n-échantillon I.I.D. $(Z_1, W_1), \ldots, (Z_n, W_n)$ où $Z_i = \min(X_i, Y_i)$ et $W_i = \begin{cases} 1 & \text{si} \quad Z_i = X_i \\ 0 & \text{si} \quad Z_i = Y_i \end{cases}$.

1. Donner les lois de Z_i et W_i.
2. Montrer que les variables Z_i et W_i sont indépendantes (on pourra calculer $\mathbb{E}_{(\lambda,\mu)}(f(Z_i)g(W_i))$ pour $f, g : \mathbb{R} \to \mathbb{R}$ bornées).

Les variables aléatoires Z_i et W_i étant indépendantes, la vraisemblance du n-échantillon $(Z_1, W_1), \ldots, (Z_n, W_n)$ est définie comme étant le produit de la vraisemblance du n-échantillon $Z_1, \ldots, Z_n$ par la vraisemblance du n-échantillon $W_1, \ldots, W_n$. Dans les questions 3 à 7, on suppose que la loi de la durée de vie du second composant est connue i.e. que μ est connu.

3. Donner l'Estimateur du Maximum de Vraisemblance $\hat{\lambda}_n$ de λ.
4. Montrer que $\hat{\lambda}_n$ est fortement convergent.
5. Calculer $\mathbb{E}_\lambda(\hat{\lambda}_n)$ et en déduire que $\mathbb{E}_\lambda(\hat{\lambda}_n)$ tend vers λ lorsque n tend vers l'infini.
6. Vérifier que $\hat{\lambda}_n$ est asymptotiquement normal et que sa variance asymptotique est égale à l'inverse de l'information de Fisher.
7. Donner un intervalle de confiance bilatéral symétrique asymptotique de niveau $1-\alpha$ pour λ.
8. Donner l'Estimateur du Maximum de Vraisemblance $(\hat{\lambda}_n, \hat{\mu}_n)$ de (λ, μ).

7.5 Résumé

- **Modèle paramétrique :** observation d'un **échantillon** $X = (X_1, \dots, X_n)$ où les X_i sont indépendantes de loi commune P_θ avec $\theta \in \Theta \subset \mathbb{R}^d$. Probabilité et espérance sont alors notées $\mathbb{P}_\theta$ et $\mathbb{E}_\theta$.

- **Vraisemblance :** $\boxed{\theta \in \Theta \to p_n(x, \theta) = p(x_1, \theta) \times \dots \times p(x_n, \theta)}$ où
  - lorsque les X_i sont discrètes à valeurs F, $\forall x_1 \in F$, $p(x_1, \theta) = \mathbb{P}_\theta(X_1 = x_1)$,
  - $x_1 \in \mathbb{R}^k \to p(x_1, \theta)$ densité des X_i sinon.

- **Log-vraisemblance :** $\theta \in \Theta \to l_n(x, \theta) = \sum_{i=1}^n l(x_i, \theta)$ où $l(x_1, \theta) = \ln(p(x_1, \theta))$.

- **Statistique :** variable aléatoire S fonction de X (mais pas de θ).

- **Estimateur de $g(\theta)$ où $g : \Theta \to \mathbb{R}^q$:** statistique Z à valeurs dans $g(\Theta)$
  - *sans biais :* $\forall \theta \in \Theta$, $\mathbb{E}_\theta(Z) = g(\theta)$.
  - *risque quadratique :* $R(Z, \theta) = \mathbb{E}_\theta((g(\theta) - Z)^2) = (g(\theta) - \mathbb{E}_\theta(Z))^2 + \mathrm{Var}_\theta(Z)$ si $q = 1$.
  - *fortement convergent :* suite Z_n d'estimateurs construits à partir de $(X_1, \dots, X_n)$ telle que $\forall \theta \in \Theta$, $\mathbb{P}_\theta$ presque sûrement, $\lim_{n\to+\infty} Z_n = g(\theta)$.
  - *asymptotiquement normal :* $\forall \theta \in \Theta$, sous $\mathbb{P}_\theta$, $\sqrt{n}(Z_n - g(\theta))$ converge en loi vers $Y \sim \mathcal{N}_q(0, \Sigma(\theta))$.
  - *Estimateur du Maximum de Vraisemblance :* $\hat{\theta}_n = \theta_n(X)$ où

    $$\boxed{p_n(x, \theta_n(x)) = \max_{\theta\in\Theta} p_n(x, \theta).}$$

    En général fortement convergent et asymptotiquement normal avec $\Sigma(\theta) = I^{-1}(\theta)$ où l'information de Fisher $I(\theta)$ est définie par

    $$\forall 1 \leq i, j \leq d, \ \boxed{I_{ij}(\theta) = \mathbb{E}_\theta\left(\frac{\partial l}{\partial \theta_i}(X_1, \theta)\frac{\partial l}{\partial \theta_j}(X_1, \theta)\right) = -\mathbb{E}_\theta\left(\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}(X_1, \theta)\right).}$$

- **Statistique exhaustive :** statistique $S = \mathcal{S}(X)$ telle que pour tout s la loi conditionnelle de X sachant $S = s$ ne dépend pas de θ

  $$\Leftrightarrow \boxed{\forall x, \theta, \ p(x, \theta) = \psi(\mathcal{S}(x), \theta)\varphi(x).}$$

  Alors pour tout estimateur Z de $g(\theta)$, $Z_S = \mathbb{E}(Z|S)$ est un estimateur de $g(\theta)$ tel que $\forall \theta \in \Theta$, $R(Z_S, \theta) \leq R(Z, \theta)$.

- **Intervalle de confiance pour $g(\theta)$ (g à valeurs réelles) :**
  - *de niveau $1 - \alpha$:* intervalle $I(X)$ tel que $\forall \theta \in \Theta$, $\mathbb{P}_\theta(g(\theta) \in I(X)) = 1 - \alpha$.
  - *asymptotique de niveau $1 - \alpha$:* suite d'intervalles $I_n(X_1, \dots, X_n)$ telle que $\forall \theta \in \Theta$, $\lim_{n\to+\infty} \mathbb{P}_\theta(g(\theta) \in I_n(X_1, \dots, X_n)) = 1 - \alpha$

- **Quantile** d'ordre $r \in]0, 1[$ de la loi de la variable aléatoire réelle Y : $q_r = F^{-1}(r)$ si $F(y) = \mathbb{P}(Y \leq y)$ inversible, $q_r = \inf\{y \in \mathbb{R} : F(y) \geq r\}$ sinon.

Chapitre 8

Tests d'hypothèses

L'objectif d'un test d'hypothèse est de répondre à une question que l'on formalise de la manière suivante : au vu de l'observation d'un échantillon $(X_1, \ldots, X_n)$, le paramètre θ du modèle est-il ou non dans un sous-ensemble de Θ appelé hypothèse nulle et noté H_0 ? Si on s'intéresse au changement climatique, on peut par exemple travailler sur les données de température moyenne au mois d'août à Paris. Sur l'ensemble du vingtième siècle, ces températures moyennes en degrés Celsius sont distribuées suivant une loi gaussienne de moyenne 20 et de variance 1 : $\mathcal{N}_1(20, 1)$. Sur les dix dernières années, on a observé les températures moyennes suivantes :

$$x = (x_1, \ldots, x_{10}) = (22, 19, 21, 23, 20, 22, 24, 18, 20, 25) \text{ telles que } \bar{x}_{10} = 21.4 \text{ et } s_{10} = 2.22.$$

À partir de ces éléments, on peut construire un test d'hypothèse pour savoir si la température moyenne a augmenté ces dix dernières années par rapport à l'ensemble du vingtième siècle. Bien sûr le fait que la moyenne empirique sur les dix dernières années dépasse 20 va plutôt dans le sens d'un réchauffement mais il faut procéder de manière plus fine pour pouvoir contrôler la probabilité d'affirmer à tort qu'il y a eu réchauffement.

8.1 Tests

8.1.1 Définitions

On considère $X = (X_1, \ldots, X_n)$ un échantillon de variables aléatoires I.I.D. dans le modèle $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. Soit (H_0, H_1) une partition de l'ensemble Θ des paramètres.
• On appelle **test d'hypothèse** une règle de décision qui au vu de l'observation X permet de décider si θ est dans l'ensemble H_0 appelé **hypothèse nulle** ou si θ est dans l'ensemble H_1 appelé **hypothèse alternative**.
• Un test est déterminé par sa **région critique** W qui constitue un sous-ensemble des valeurs possibles de $X = (X_1, \ldots, X_n)$. Lorsque l'on observe $x = (x_1, \ldots, x_n)$,
 – si $x \in W$, alors on rejette H_0 et on accepte H_1 i.e. on décide que $\theta \in H_1$,
 – si $x \notin W$, alors on accepte H_0 et on rejette H_1 i.e. on décide que $\theta \in H_0$.
• On appelle **erreur de première espèce** le rejet d'H_0 à tort. Cette erreur est mesurée par le **risque de première espèce** : $\theta \in H_0 \to \mathbb{P}_\theta(W)$ où on note de façon légèrement abusive W l'événement $\{X \in W\}$.

• On appelle **erreur de seconde espèce** le rejet d'H_1 à tort. Cette erreur est mesurée par le **risque de seconde espèce** : $\theta \in H_1 \to \mathbb{P}_\theta(W^c) = 1 - \rho_W(\theta)$ où la fonction $\rho_W : \theta \in H_1 \to \mathbb{P}_\theta(W)$ s'appelle **puissance** du test.
• Par convention, on minimise en priorité le risque de première espèce. On appelle **niveau du test** le nombre $\boxed{\alpha_W = \sup_{\theta \in H_0} \mathbb{P}_\theta(W)}$. Parmi les tests de niveau inférieur à un seuil α fixé, on souhaite minimiser le risque de seconde espèce ou de manière équivalente maximiser la puissance. En général, on choisit $\alpha = 10\%$, $\alpha = 5\%$ ou $\alpha = 1\%$.

Exemple 8.1.1. Dans le modèle $\mathcal{P} = \{\mathcal{N}_1(\mu, \sigma^2),\ \mu \in \{\mu_0, \mu_1\}\}$ avec $\sigma^2 > 0$ connu et $\mu_0 > \mu_1$, on souhaite tester $H_0 = \{\mu = \mu_0\}$ contre $H_1 = \{\mu = \mu_1\}$. On va bien sûr accepter H_0 (resp. H_1) si la moyenne empirique $\bar{X}_n$ est grande (resp. petite), c'est-à-dire choisir la région critique de la forme $W = \{\bar{X}_n < a\}$. Le choix $a = \frac{\mu_0+\mu_1}{2}$, qui peut sembler naturel,

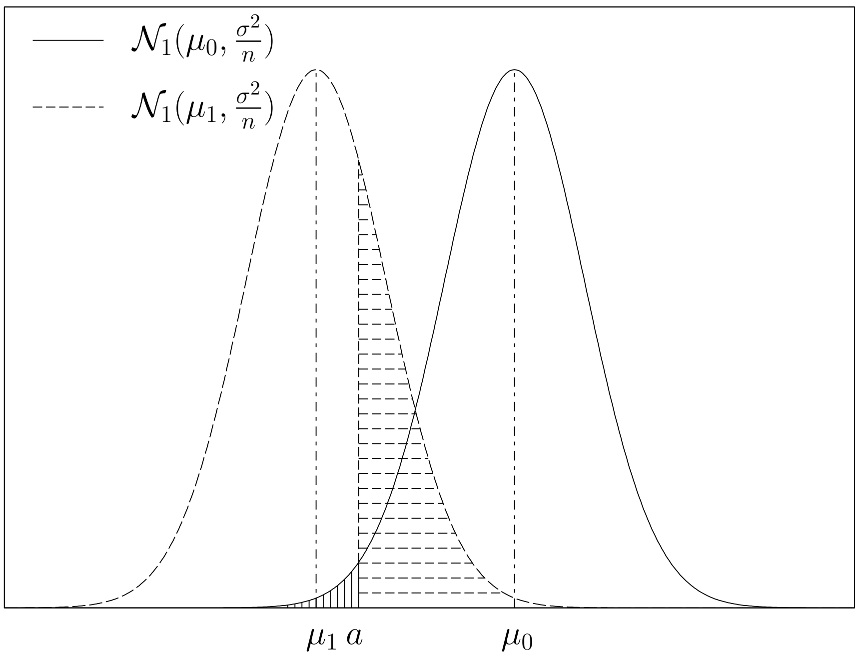


FIGURE 8.1 – Densités de $\bar{X}_n$ sous H_0 et H_1, risques de 1[ere] et de 2[e] espèce.

ne permet pas de contrôler le risque de première espèce. Pour obtenir ce contrôle de la probabilité de rejeter H_0 à tort, on utilise le fait que sous H_0, la statistique de test $\bar{X}_n$ suit la loi $\mathcal{N}_1\left(\mu_0, \frac{\sigma^2}{n}\right)$. Donc si $Z \sim \mathcal{N}_1(0,1)$,

$$\mathbb{P}_{(\mu_0,\sigma^2)}(W) = \mathbb{P}_{(\mu_0,\sigma^2)}(\bar{X}_n < a) = \mathbb{P}\left(\mu_0 + \frac{\sigma Z}{\sqrt{n}} < a\right) = \mathbb{P}\left(Z < \sqrt{n}\frac{a-\mu_0}{\sigma}\right).$$

Si ϕ_r désigne le quantile d'ordre r de la loi normale centrée réduite, le choix $a = \mu_0 + \frac{\sigma\phi_\alpha}{\sqrt{n}}$, assure que le niveau du test est α.

Pour ce choix, la probabilité de rejeter H_1 à tort est

$$\begin{aligned}\mathbb{P}_{(\mu_1,\sigma^2)}(W^c) &= \mathbb{P}_{(\mu_1,\sigma^2)}(\bar{X}_n \geq a) = \mathbb{P}\left(\mu_1 + \frac{\sigma Z}{\sqrt{n}} \geq \mu_0 + \frac{\sigma\phi_\alpha}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(Z \geq \phi_\alpha + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma}\right) \leq \mathbb{P}\left(Z \geq \phi_\alpha\right) = 1-\alpha.\end{aligned}$$

À n et σ^2 fixés, pour μ_1 assez proche de μ_0, cette probabilité peut être arbitrairement proche de $1-\alpha$ avec $1-\alpha = 95\%$ pour $\alpha = 5\%$. Ainsi le risque de seconde espèce qui correspond à l'aire hachurée horizontalement sur la figure 8.1 est beaucoup moins bien contrôlé que le risque de première espèce qui correspond à l'aire hachurée verticalement. Néanmoins pour σ^2, μ_0 et μ_1 fixés, $\mathbb{P}_{(\mu_1,\sigma^2)}(W^c) = \mathbb{E}\left(1_{\{Z\geq\phi_\alpha+\sqrt{n}\frac{\mu_0-\mu_1}{\sigma}\}}\right)$ converge vers 0 lorsque n tend vers l'infini d'après le théorème de convergence dominée. Cette propriété que l'on peut récrire $\lim_{n\to+\infty}\mathbb{P}_{(\mu_1,\sigma^2)}(W) = 1$ et qui est souhaitable pour un test porte le nom de convergence (voir définition 8.1.4 ci-après).

Définition 8.1.2. *Lorsque l'on observe $x = (x_1, \ldots, x_n)$, on appelle* p**-valeur** *du test la probabilité sous H_0 que la statistique de test, notée ζ_n, prenne des valeurs plus défavorables pour l'acceptation de H_0 que celle ζ_n^{obs} observée.*

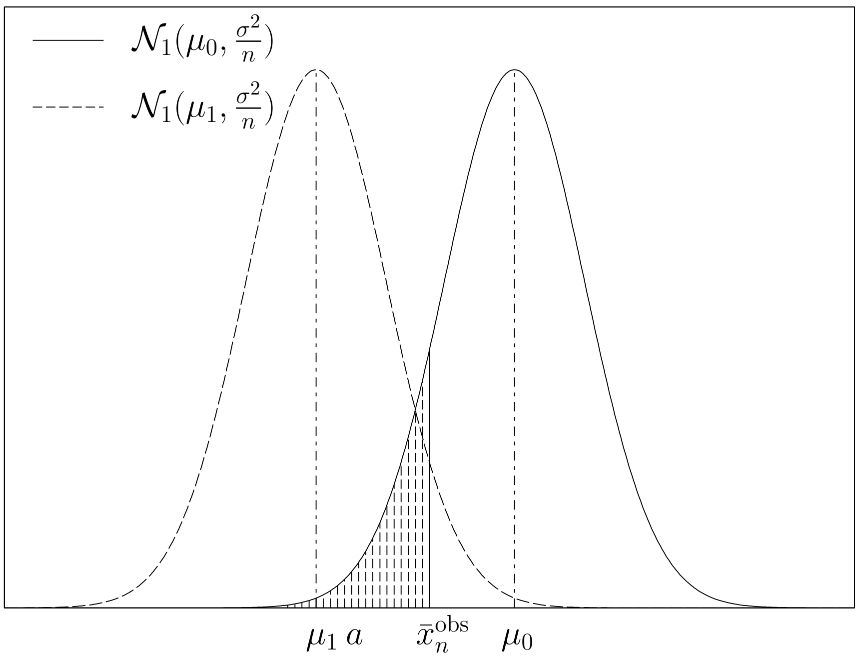


FIGURE 8.2 – p-valeur.

Dans l'exemple 8.1.1 qui précède, la statistique de test est $\zeta_n = \bar{X}_n$. La p-valeur

$$p - \text{valeur} = \mathbb{P}_{(\mu_0,\sigma^2)}(\bar{X}_n < \bar{x}_n^{\text{obs}}) = \mathbb{P}\left(Z < \sqrt{n}\frac{\bar{x}_n^{\text{obs}}-\mu_0}{\sigma}\right) \text{ où } Z \sim \mathcal{N}_1(0,1)$$

est représentée sur la figure 8.2 comme l'aire hachurée verticalement. Notons que l'on a l'équivalence suivante :

$$p - \text{valeur} < \alpha \Leftrightarrow \bar{x}_n^{\text{obs}} < \mu_0 + \frac{\sigma\phi_\alpha}{\sqrt{n}} \Leftrightarrow \bar{x}_n^{\text{obs}} \in W.$$

La connaissance de la p-valeur permet donc de décider si on accepte ou non H_0.
Règle d'acceptation : Si la p-valeur est supérieure au niveau α du test, on accepte H_0. Sinon, on rejette H_0.

Mais l'intérêt de la p-valeur est qu'elle donne plus d'information que cela : son écart avec α quantifie la marge avec laquelle on accepte H_0 dans le cas où elle est plus grande que α et la marge avec laquelle on rejette H_0 dans le cas contraire.

Remarque 8.1.3.
- La règle d'acceptation qui précède montre que la p-valeur est également définie comme le niveau de test à partir duquel on rejette H_0.
- Comme on minimise en priorité le risque de première espèce, les rôles de l'hypothèse nulle H_0 et de l'hypothèse alternative H_1 ne sont pas symétriques. Le choix de H_0 parmi deux ensembles constituant une partition de Θ dépend donc du problème considéré : on choisit comme hypothèse nulle l'ensemble que l'on ne souhaite surtout pas voir rejeté à tort (hypothèse à laquelle on tient, hypothèse de prudence, hypothèse solidement établie,...). Par exemple, dans le test de dépistage d'une maladie, on souhaite surtout éviter de dire à une personne qu'elle est en bonne santé alors qu'elle est en fait malade. On choisit comme hypothèse nulle le fait d'être malade. Dans le cas du réchauffement climatique, un homme politique qui veut éviter de prendre des mesures si le réchauffement n'est pas avéré choisira comme hypothèse nulle "il n'y a pas réchauffement". Un écologiste choisira plutôt "il y a réchauffement".
- Utiliser l'Estimateur du Maximum de Vraisemblance de θ est en général une bonne idée pour construire un test qui porte sur θ.

Définition 8.1.4. *Soit $(W_n)_n$ une suite de régions critiques où n désigne la taille de l'échantillon. La suite de tests $(W_n)_n$ est dite*
- *convergente si*
$$\forall\theta \in H_1, \lim_{n\to+\infty} \mathbb{P}_\theta(W_n) = 1.$$
- *de niveau asymptotique α si*
$$\lim_{n\to+\infty} \sup_{\theta\in H_0} \mathbb{P}_\theta(W_n) = \alpha.$$

Regarder le comportement asymptotique de la statistique que l'on utilise pour construire le test sous H_1 permet en général de guider le choix de la région critique.

8.1.2 Le cas du modèle gaussien $\mathcal{P} = \{\mathcal{N}_1(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$:

Tests pour la moyenne μ :

Soit $\mu_0 \in \mathbb{R}$. On désire tester $H_0 = \{\mu = \mu_0\}$ contre $H_1 = \{\mu \neq \mu_0\}$ au niveau $\alpha \in]0, 1[$.
D'après le corollaire 6.2.4, si $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ désignent respectivement la moyenne empirique et l'estimateur sans biais de la variance, le rapport $\zeta_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{S_n}$ suit la loi de Student $t(n-1)$ sous H_0 (i.e. si $\mu = \mu_0$).

Sous H_1, par la loi forte des grands nombres, $\bar{X}_n - \mu_0$ converge presque sûrement vers $\mu - \mu_0 \neq 0$ et S_n converge presque sûrement vers σ. Donc ζ_n tend presque sûrement vers $+\infty$ ou $-\infty$ lorsque $n \to +\infty$, suivant que $\mu > \mu_0$ ou $\mu < \mu_0$.

On choisit donc $W_n = \{|\zeta_n| > a\}$ pour la région critique. On note $t_{n-1,r}$ le quantile d'ordre r de la loi de Student $t(n-1)$. D'après les propriétés 7.3.4, si $a \geq t_{n-1,1-\frac{\alpha}{2}}$,

$$\forall \sigma^2 > 0,\ \mathbb{P}_{(\mu_0,\sigma^2)}(W_n) = \mathbb{P}_{(\mu_0,\sigma^2)}(|\zeta_n| > a) \leq \mathbb{P}_{(\mu_0,\sigma^2)}(|\zeta_n| > t_{n-1,1-\frac{\alpha}{2}}) = \alpha,$$

et le niveau du test est inférieur à α. Comme on souhaite ensuite minimiser le risque de seconde espèce $\mathbb{P}_{(\mu,\sigma^2)}(|\zeta_n| \leq a)$ pour $\mu \neq \mu_0$, on choisit $a = t_{n-1,1-\frac{\alpha}{2}}$. En conclusion, on choisit la région critique $W_n = \{|\zeta_n| > t_{n-1,1-\frac{\alpha}{2}}\}$ et on a alors $\mathbb{P}_{(\mu_0,\sigma^2)}(W_n) = \alpha$ pour tout $\sigma^2 > 0$.
La p-valeur est

$$p-\text{valeur} = \mathbb{P}(|T| > |\zeta_n^{\text{obs}}|) \ \text{ où } \ T \sim t(n-1).$$

Notons que lorsque $n \to +\infty$, $t_{n-1,1-\frac{\alpha}{2}}$ converge vers $\phi_{1-\frac{\alpha}{2}}$, le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite (voir les tables à la fin du polycopié). Donc, par le théorème de convergence dominée, pour $\mu \neq \mu_0$, $\mathbb{P}_{(\mu,\sigma^2)}(W_n) = \mathbb{E}_{(\mu,\sigma^2)}\left(1_{\{|\zeta_n| > t_{n-1,1-\frac{\alpha}{2}}\}}\right)$ tend vers 1 lorsque n tend vers l'infini. Ainsi le test est convergent.

Exercice résolu 8.1.5. *On suppose maintenant la variance σ^2 connue et égale à σ_0^2.*

1. *Quel est la loi de la statistique $\zeta_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma_0}$ sous $\mathbb{P}_{(\mu,\sigma_0^2)}$?*
2. *Construire un test de niveau α pour $H_0 = \{\mu \leq \mu_0\}$ et $H_1 = \{\mu > \mu_0\}$.*
3. *Mettre en œuvre ce test avec $\mu_0 = 20$ et $\sigma_0^2 = 1$ sur les températures moyennes des dernières années au mois d'août à Paris pour $\alpha = 0.01$.*

1. Sous $\mathbb{P}_{(\mu,\sigma_0^2)}$, $\zeta_n \sim \mathcal{N}_1\left(\sqrt{n}\frac{\mu-\mu_0}{\sigma_0}, 1\right)$.
2. La variable aléatoire $\zeta_n - \sqrt{n}\frac{\mu-\mu_0}{\sigma_0}$ suit la loi normale centrée réduite sous $\mathbb{P}_{(\mu,\sigma_0^2)}$. Ainsi ζ_n croît avec μ, ce qui conduit à choisir une région critique de la forme $\{\zeta_n \geq a\}$. Par ailleurs, si $Z \sim \mathcal{N}_1(0,1)$,
$$\sup_{H_0} \mathbb{P}_{(\mu,\sigma_0^2)}(\zeta_n \geq a) = \sup_{\mu \leq \mu_0} \mathbb{P}\left(Z + \sqrt{n}\frac{\mu - \mu_0}{\sigma_0} \geq a\right) = \mathbb{P}(Z \geq a).$$
Le choix $a = \phi_{1-\alpha}$ où ϕ_r désigne le quantile d'ordre r de la loi normale centrée réduite assure que le niveau du test est α. La p-valeur est donnée par $\mathbb{P}(Z \geq \zeta_n^{\text{obs}})$.
3. Dans le cas des données de température moyenne, on a $\zeta_n^{\text{obs}} = \sqrt{10}(21.4-20) = 4.43$ et $\phi_{0.99} = 2.33$. Comme $\zeta_n^{\text{obs}} \geq \phi_{0.99}$, on rejette H_0. La p-valeur est égale à $\mathbb{P}(Z \geq 4.43) = 4.7 \times 10^{-6}$. Cette p-valeur est très faible et on rejette donc H_0 à tous les niveaux α supérieurs à 4.7×10^{-6} c'est-à-dire à tous les niveaux usuels. Ainsi on peut conclure à l'augmentation des températures sur les dix dernières années.

Tests pour la variance σ^2 :

Soit $\sigma_0^2 > 0$. On désire tester $H_0 = \{\sigma^2 \geq \sigma_0^2\}$ contre $H_1 = \{\sigma^2 < \sigma_0^2\}$. On utilise la statistique

$$\zeta_n = \frac{(n-1)S_n^2}{\sigma_0^2} = \frac{\sigma^2}{\sigma_0^2} \times \frac{(n-1)S_n^2}{\sigma^2}.$$

Comme d'après le corollaire 6.2.4, $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ sous $\mathbb{P}_{(\mu,\sigma^2)}$, ζ_n prend des valeurs de plus en plus grandes lorsque σ^2 croît. En particulier ζ_n va avoir tendance à prendre des valeurs plus grandes sous H_0 que sous H_1. C'est pourquoi on acceptera H_0 si la statistique ζ_n est grande et on rejettera H_0 si elle est petite.
On choisit donc une région critique de la forme $W_n = \{\zeta_n \leq a\}$. En outre, si $Z \sim \chi^2(n-1)$

$$\begin{aligned}\sup_{H_0} \mathbb{P}_{(\mu,\sigma^2)}(\zeta_n \leq a) &= \sup_{\sigma^2 \geq \sigma_0^2} \mathbb{P}_{(\mu,\sigma^2)}\left(\frac{(n-1)S_n^2}{\sigma^2} \leq a\frac{\sigma_0^2}{\sigma^2}\right) \\ &= \sup_{\sigma^2 \geq \sigma_0^2} \mathbb{P}\left(Z \leq a\frac{\sigma_0^2}{\sigma^2}\right) = \mathbb{P}\left(Z \leq a\right).\end{aligned}$$

Le choix $a = \chi^2_{n-1,\alpha}$ où $\chi^2_{n-1,r}$ désigne le quantile d'ordre r de la loi $\chi^2(n-1)$ assure que le niveau du test est α.
La p-valeur du test est donnée par $\mathbb{P}(Z \leq \zeta_n^{\text{obs}})$.

Remarque 8.1.6.

$$\begin{aligned}\sup_{H_1} \mathbb{P}_{(\mu,\sigma^2)}(W_n^c) &= \sup_{\sigma^2 < \sigma_0^2} \mathbb{P}_{(\mu,\sigma^2)}\left(\frac{(n-1)S_n^2}{\sigma^2} > \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha}\right) = \sup_{\sigma^2 < \sigma_0^2} \mathbb{P}\left(Z > \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha}\right) \\ &= \mathbb{P}(Z > \chi^2_{n-1,\alpha}) = 1-\alpha.\end{aligned}$$

Notons que cela n'empêche pas la convergence du test. Soit en effet $\sigma^2 < \sigma_0^2$ et $\varepsilon > 0$ t.q. $(1-\varepsilon)\frac{\sigma_0^2}{\sigma^2} > 1$. Si $(G_i)_{i\geq 1}$ est une suite de variables I.I.D. suivant la loi normale centrée réduite $\mathcal{N}_1(0,1)$, la loi forte des grands nombres entraîne que

$$\mathbb{P}\left(\sum_{i=1}^{n-1} G_i^2 \leq (n-1)(1-\varepsilon)\right) = \mathbb{E}\left(1_{\{\frac{1}{n-1}\sum_{i=1}^{n-1} G_i^2 \leq 1-\varepsilon\}}\right) \longrightarrow_{n\to+\infty} 0.$$

Comme $\sum_{i=1}^{n-1} G_i^2 \sim \chi^2(n-1)$, on en déduit que $\chi^2_{n-1,\alpha} \geq (n-1)(1-\varepsilon)$ pour n grand. Alors

$$\begin{aligned}\mathbb{P}_{(\mu,\sigma^2)}(W_n) &= \mathbb{P}_{(\mu,\sigma^2)}\left(\frac{(n-1)S_n^2}{\sigma^2} \leq \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha}\right) = \mathbb{P}\left(\sum_{i=1}^{n-1} G_i^2 \leq \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha}\right) \\ &\geq \mathbb{P}\left(\frac{1}{n-1}\sum_{i=1}^{n-1} G_i^2 \leq (1-\varepsilon)\frac{\sigma_0^2}{\sigma^2}\right).\end{aligned}$$

Et la loi forte des grands nombres entraîne que le membre de droite tend vers 1 lorsque n tend vers l'infini.
Ce résultat se généralise en $\forall n \in \mathbb{N}^*$, $\sup_{\theta\in H_1} \mathbb{P}_\theta(W_n^c) = 1-\alpha$, propriété vérifiée pour la plupart des tests et qui illustre le fait que l'on contrôle beaucoup moins bien l'erreur de seconde espèce que l'erreur de première espèce.

Exercice 8.1.7. Dans le cas du modèle gaussien à moyenne et variance inconnues,

1. Donner le principe du test de l'hypothèse nulle $H_0 = \{\mu \leq \mu_0\}$ où μ_0 est fixé contre l'hypothèse alternative $H_1 = \{\mu > \mu_0\}$. Mettre en œuvre ce test au niveau $\alpha = 0.01$ pour $\mu_0 = 20$ sur les données de température moyenne à Paris au mois d'août présentées en introduction.
2. Donner le principe du test de $H_0 = \{\sigma^2 = \sigma_0^2\}$ contre $H_1 = \{\sigma^2 \neq \sigma_0^2\}$.

8.2 Le test du χ^2

Le test du χ^2 permet de répondre à des questions comme :

– Un dé à six faces est-il pipé ? Pour cela on observe les fréquences d'apparition des faces lors de n lancers de ce dé et on les compare au vecteur $(\frac{1}{6}, \ldots, \frac{1}{6})$.
– La répartition entre accouchements avec et sans césarienne dépend-elle du jour de la semaine ? Pour cela, on introduit $\hat{q}_j$ les proportions d'accouchements qui ont lieu le j-ème jour de la semaine, $\hat{r}_A$ et $\hat{r}_S$ les proportions globales d'accouchements avec et sans césarienne et enfin $\hat{p}_{j,l}$, $(j,l) \in \{1, \ldots, 7\} \times \{A, S\}$ la proportions d'accouchements qui ont lieu le j-ème jour de la semaine avec césarienne si $l = A$ et sans si $l = S$. Bien sûr, on va comparer la matrice $(\hat{p}_{j,l})$ à la matrice $(\hat{q}_j \hat{r}_l)$.

8.2.1 Test d'adéquation à une loi

On observe un échantillon $(X_1, \ldots, X_n)$ de variables aléatoires I.I.D. à valeurs dans un espace fini $A = \{a_1, \ldots, a_k\}$. La loi commune est paramétrisée par $p = (p_1, \ldots, p_k)$: $\mathbb{P}_p(X_1 = a_j) = p_j$ pour $j \in \{1, \ldots, k\}$. Le paramètre p appartient à $\Theta = \{(p_1, \ldots, p_k) \in \mathbb{R}_+^k : p_1 + \ldots + p_k = 1\}$.

Pour $p^0 = (p_1^0, \ldots, p_k^0) \in \Theta \cap]0,1[^k$, on souhaite tester l'hypothèse nulle $H_0 = \{p = p^0\}$ contre l'hypothèse alternative $H_1 = \{p \neq p^0\}$.

Exemple 8.2.1. Dans le cas du dé à six faces évoqué plus haut, $A = \{1, \ldots, 6\}$ et $p^0 = (\frac{1}{6}, \ldots, \frac{1}{6})$.

Pour $j \in \{1, \ldots, k\}$ on note $\hat{p}_j = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i = a_j\}}$ la fréquence empirique de a_j. Le vecteur des fréquences empiriques est $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_k)$.

Remarque 8.2.2. Notons que $\hat{p}$ est l'EMV de p. En effet, la vraisemblance et la log-vraisemblance s'écrivent en posant $\rho_j = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i = a_j\}}$

$$p_n(x,p) = \prod_{i=1}^n \left(\prod_{j=1}^k p_j^{1_{\{x_i = a_j\}}} \right) = \prod_{j=1}^k p_j^{n\rho_j} \text{ et } l_n(x,p) = n \sum_{j:\rho_j>0} \rho_j \ln p_j.$$

Or, pour tout $y > 0$, $\ln(y) = y - 1 + \int_1^y \left(\frac{1}{z} - 1\right) dz \leq y - 1$. On en déduit que si $p_j > 0$ dès que $\rho_j > 0$,

$$\frac{1}{n}\left(l_n(x,p) - l_n(x,\rho)\right) = \sum_{j:\rho_j>0} \rho_j \ln\left(\frac{p_j}{\rho_j}\right) \leq \sum_{j:\rho_j>0} \rho_j \left(\frac{p_j}{\rho_j} - 1\right) \leq \sum_{j=1}^k p_j - \sum_{j=1}^k \rho_j = 0.$$

Comme s'il existe $j \in \{1, \ldots, k\}$ tel que $p_j = 0$ et $\rho_j > 0$, $l_n(x,p) = -\infty$, on conclut que la log-vraisemblance est maximale pour $p = \rho$.

L'idée qui est à la base du test est bien sûr que le vecteur $\hat{p}$ est plus proche de p^0 sous H_0 que sous H_1. Afin de quantifier la "proximité", on utilise la pseudo-distance du χ^2 : $\sum_{j=1}^k \frac{(\hat{p}_j - p_j^0)^2}{p_j^0}$. En multipliant cette distance par n, on obtient le comportement asymptotique suivant :

Proposition 8.2.3. *Soit* $\zeta_n = n\sum_{j=1}^{k} \frac{(\hat{p}_j - p_j^0)^2}{p_j^0}$. *Sous* H_0, ζ_n *converge en loi vers* $Z \sim \chi^2(k-1)$. *Sous* H_1, ζ_n *tend presque sûrement vers* $+\infty$.

D'après les propriétés 7.3.4, en notant $\chi^2_{k-1,r}$ le quantile d'ordre r de la loi $\chi^2(k-1)$, $\mathbb{P}(Z \geq \chi^2_{k-1,1-\alpha}) = \alpha$. On en déduit le corollaire suivant :

Corollaire 8.2.4. *Le test de région critique* $W_n = \{\zeta_n > \chi^2_{k-1,1-\alpha}\}$ *est convergent de niveau asymptotique* α. *La p-valeur de ce test est égale à* $\mathbb{P}(Z \geq \zeta_n^{\text{obs}})$ *où* $Z \sim \chi^2(k-1)$.

Démonstration du corollaire : Sous H_1, on déduit de la proposition 8.2.3 que $1_{\{\zeta_n \geq \chi^2_{k-1,1-\alpha}\}}$ converge $\mathbb{P}_p$ presque sûrement vers 1. Le théorème de convergence dominée entraîne alors que

$$\mathbb{P}_p\left(\zeta_n > \chi^2_{k-1,1-\alpha}\right) = \mathbb{E}_p\left(1_{\{\zeta_n > \chi^2_{k-1,1-\alpha}\}}\right) \overset{n\to+\infty}{\longrightarrow} \mathbb{E}_p(1) = 1,$$

ce qui assure que le test est convergent.
Pour vérifier que son niveau asymptotique est α, il suffit de vérifier que $\mathbb{E}_{p^0}\left(1_{\{\zeta_n > \chi^2_{k-1,1-\alpha}\}}\right)$ converge vers $\mathbb{E}\left(1_{\{Z > \chi^2_{k-1,1-\alpha}\}}\right) = \alpha$ lorsque n tend vers l'infini. La fonction bornée $z \in \mathbb{R} \to 1_{\{z > \chi^2_{k-1,1-\alpha}\}}$ est continue en dehors du point $\chi^2_{k-1,1-\alpha}$. Comme Z, qui suit la loi du χ^2 à $k-1$ degrés de liberté, possède une densité, $\mathbb{P}(Z = \chi^2_{k-1,1-\alpha}) = 0$. On conclut alors avec la remarque 5.3.11. □

Remarque 8.2.5. En pratique, on considère que l'approximation en loi par $\chi^2(k-1)$ est valide sous H_0 si $n \times \min_{1\leq j\leq k} p_j^0 \geq 5$. Si cette condition n'est pas satisfaite, on peut regrouper les valeurs de a_j pour lesquelles p_j^0 est trop faible et augmenter ainsi le minimum.

L'exercice suivant a pour objet de démontrer la convergence en loi énoncée dans la proposition 8.2.3 dans le cas particulier $k = 2$.

Exercice 8.2.6. On suppose $k = 2$.

1. Vérifier que $\zeta_n = Y_n^2$ où $Y_n = \sqrt{\frac{n}{p_1^0(1-p_1^0)}}\left(\frac{1}{n}\sum_{i=1}^n 1_{\{X_i = a_1\}} - p_1^0\right)$.
2. Sous H_0, montrer que Y_n converge en loi vers une limite à préciser et conclure.

Nous allons maintenant démontrer la proposition 8.2.3 dans le cas général.
Démonstration de la proposition 8.2.3 : Tout d'abord, sous H_1, $p \neq p^0$. Donc il existe $j \in \{1,\ldots,k\}$ t.q. $p_j \neq p_j^0$. Par la loi forte des grands nombres, $\mathbb{P}_p$ presque sûrement, $\hat{p}_j = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i = a_j\}}$ converge vers p_j et $n\frac{(\hat{p}_j - p_j^0)^2}{p_j^0}$ tend vers $+\infty$.
Plaçons nous maintenant sous H_0.
Le vecteur $\left(\frac{1_{\{X_1 = a_1\}}}{\sqrt{p_1^0}}, \ldots, \frac{1_{\{X_1 = a_k\}}}{\sqrt{p_k^0}}\right)$ a pour matrice de covariance

$$\Gamma_{jl} = \frac{1}{\sqrt{p_j^0 p_l^0}}\left(\mathbb{P}_{p^0}(X_1 = a_j, X_1 = a_l) - \mathbb{P}_{p^0}(X_1 = a_j)\mathbb{P}_{p^0}(X_1 = a_l)\right) = \begin{cases} 1 - p_j^0 & \text{si } j = l \\ -\sqrt{p_j^0 p_l^0} & \text{sinon.} \end{cases}$$

Le théorème de la limite centrale multidimensionnel 6.2.8 assure que $Y_n = \sqrt{n}\left(\frac{\hat{p}_1 - p_1^0}{\sqrt{p_1^0}}, \ldots, \frac{\hat{p}_k - p_k^0}{\sqrt{p_k^0}}\right)$ converge en loi vers $Y \sim \mathcal{N}_k(0, \Gamma)$. Par continuité de l'application $y \in \mathbb{R}^k \to |y|^2$, $\zeta_n = |Y_n|^2$ converge en loi vers $|Y|^2$. Tout le travail consiste maintenant à identifier la loi de cette limite.
Le vecteur $e = (\sqrt{p_1^0}, \ldots, \sqrt{p_k^0})$ est unitaire dans $\mathbb{R}^k$ et $\Gamma = I_k - ee^*$ est la matrice représentant la projection orthogonale sur $e^\perp$ dans la base canonique. Pour $V = (V_1, \ldots, V_k) \sim \mathcal{N}_k(0, I_k)$, $V^{e^\perp} = (I_k - ee^*)V \sim \mathcal{N}_k(0, (I_k - ee^*)I_k(I_k - ee^*)^*) = \mathcal{N}_k(0, (I_k - ee^*))$. Donc $Y \stackrel{\mathcal{L}}{=} V^{e^\perp}$ et $|Y|^2 \stackrel{\mathcal{L}}{=} |V^{e^\perp}|^2$. Le théorème de Cochran 6.2.3 assure que $|V^{e^\perp}|^2 \sim \chi^2(k-1)$. $\square$

Exercice résolu 8.2.7. *Lors de 100 lancers du dé à 6 faces on observe les résultats suivants*

x	*1*	*2*	*3*	*4*	*5*	*6*
$N(x)$	*20*	*13*	*17*	*12*	*23*	*15*

Tester au niveau 5% si le dé n'est pas pipé.
Les fréquences observées sont

x	1	2	3	4	5	6
$\hat{p}_x$	0.2	0.13	0.17	0.12	0.23	0.15

On applique le test d'adéquation à la loi uniforme $p^0 = (\frac{1}{6}, \ldots, \frac{1}{6})$. La statistique de test vaut

$$\zeta_{100}^{\text{obs}} = 600 \sum_{x=1}^{6} (\hat{p}_x - 1/6)^2 = 5.36.$$

D'après la table du paragraphe 11.3, $\chi^2_{5,0.95} = 11.07$. On rejette donc, au niveau 0.05, l'hypothèse que le dé est pipé.

8.2.2 Test d'adéquation à une famille de lois

Cette généralisation du test précédent permet entre autres de répondre à une question du type : la répartition entre accouchements avec et sans césarienne dépend-elle du jour de la semaine ?
On observe toujours un échantillon $(X_1, \ldots, X_n)$ de variables aléatoires I.I.D. à valeurs dans $A = \{a_1, \ldots, a_k\}$. Mais on souhaite maintenant tester l'hypothèse nulle $H_0 = \{p \in \{p^\theta, \theta \in \Theta_0\}\}$ contre $H_1 = \{p \notin \{p^\theta, \theta \in \Theta_0\}\}$ où Θ_0 est une partie de $\mathbb{R}^h$ d'intérieur non vide avec $h < k - 1$.
L'idée consiste alors à utiliser un estimateur $\hat{\theta}_n$ de θ à valeurs dans Θ_0 et à comparer les vecteurs $\hat{p}$ et $p^{\hat{\theta}_n}$.

Proposition 8.2.8. *Soit $\zeta_n = n \sum_{j=1}^{k} \frac{(\hat{p}_j - p_j^{\hat{\theta}_n})^2}{p_j^{\hat{\theta}_n}}$. Sous des hypothèses de régularité non précisées (vérifiées en général lorsque $\hat{\theta}_n$ est l'EMV de θ), sous H_0, ζ_n converge en loi vers $Z \sim \chi^2(k - h - 1)$. Sous H_1, ζ_n tend presque sûrement vers $+\infty$.*

Il est essentiel de noter que le nombre de degrés de liberté dans la limite en loi sous H_0 est $\boxed{k-h-1}$ et non plus $k-1$ comme dans le test d'adéquation à une loi. Nous ne démontrerons pas cette proposition qui assure que le test de région critique $W_n = \{\zeta_n \geq \chi^2_{k-h-1,1-\alpha}\}$ est convergent de niveau asymptotique α. La p-valeur associée est $\mathbb{P}(Z \geq \zeta_n^{\text{obs}})$ où $Z \sim \chi^2(k-h-1)$.

Exemple 8.2.9. Test d'indépendance : On se demande si l'échantillon $((Y_1, Z_1), \ldots, (Y_n, Z_n))$ avec les Y_i à valeurs dans $\{b_1, \ldots, b_d\}$ et les Z_i à valeurs dans $\{c_1, \ldots, c_m\}$ est tel que Y_1 est indépendante de Z_1. On pose bien sûr $X_i = (Y_i, Z_i)$ variable à valeurs dans l'ensemble $\{b_1, \ldots, b_d\} \times \{c_1, \ldots, c_m\}$ de cardinal $k = dm$. On note $p = (p_{jl}, 1 \leq j \leq d, 1 \leq l \leq m)$ où $\mathbb{P}_p(X_i = (b_j, c_l)) = \mathbb{P}_p(Y_i = b_j, Z_i = c_l) = p_{jl}$.
La famille des lois p de forme produit qui correspondent à l'indépendance de Y_i est Z_i est paramétrée par :

$$\forall \theta = (q_1, \ldots, q_{d-1}, r_1, \ldots, r_{m-1}) \in \Theta_0 = \left\{ \lambda \in]0,1[^{d+m-2} : \sum_{j=1}^{d-1} \lambda_j < 1 \text{ et } \sum_{l=d}^{d+m-2} \lambda_l < 1 \right\},$$

$$p^\theta = (q_j r_l, 1 \leq j \leq d, 1 \leq l \leq m) \text{ où } q_d = 1 - \sum_{j=1}^{d-1} q_j \text{ et } r_m = 1 - \sum_{l=1}^{m-1} r_l.$$

On pose $\hat{p}_{jl} = \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i = b_j, Z_i = c_l\}}$, $\hat{q}_j = \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i = b_j\}}$ et $\hat{r}_l = \frac{1}{n}\sum_{i=1}^n 1_{\{Z_i = c_l\}}$. Notons que d'après la remarque 8.2.2, l'EMV de $(q_1, \ldots, q_{d-1}, r_1, \ldots, r_{m-1})$ est $(\hat{q}_1, \ldots, \hat{q}_{d-1}, \hat{r}_1, \ldots, \hat{r}_{m-1})$. La statistique de test

$$\zeta_n = n \sum_{j=1}^d \sum_{l=1}^m \frac{(\hat{p}_{jl} - \hat{q}_j \hat{r}_l)^2}{\hat{q}_j \hat{r}_l}$$

mesure la distance entre la matrice $\hat{p}$ des fréquences des couples (b_j, c_l) et la matrice $\hat{q}\hat{r}^*$ produit des fréquences marginales.
Comme la dimension h de Θ_0 est $d+m-2$, on a $k-h-1 = dm-d-m+1 = (d-1)(m-1)$. On rejette l'hypothèse H_0 d'indépendance si ζ_n dépasse $\chi^2_{(d-1)(m-1),1-\alpha}$ et on l'accepte sinon.

8.3 Exercices

Exercice corrigé 8.3.1. On se place dans le modèle exponentiel $\mathcal{P} = \{\mathcal{E}(\theta), \theta > 0\}$.

1. Rappeler $\mathbb{E}_\theta(X_1)$.
2. En remarquant que $2\theta(X_1 + \ldots + X_n) \sim \chi^2(2n)$ construire un test de niveau α pour $H_0 = \{\theta = \theta_0\}$ et $H_1 = \{\theta \neq \theta_0\}$.
   A.N. : $n = 15$, $\bar{x}_n = 1.47$, $\theta_0 = 1$, $\alpha = 5\%$.
3. Tester maintenant $H_0 = \{\theta \geq \theta_0\}$ contre $H_1 = \{\theta < \theta_0\}$.

Exercice 8.3.2. On désire comparer les moyennes de deux échantillons indépendants $(X_{1,1}, \ldots, X_{1,n_1})$ et $(X_{2,1}, \ldots, X_{2,n_2})$ distribués respectivement suivant $\mathcal{N}_1(\mu_1, \sigma^2)$ et $\mathcal{N}_1(\mu_2, \sigma^2)$ où $\mu_1, \mu_2 \in \mathbb{R}$ et $\sigma^2 > 0$. Pour $i \in \{1, 2\}$, on pose $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ et $S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{i,j} - \bar{X}_i)^2$.

1. Quelle est la loi de $\left(\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sigma}, \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}\right)$? En déduire que $\sqrt{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}} \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}} \sim t(n_1+n_2-2)$.
2. On s'interroge sur les tailles des garçons et des filles de douze ans dans la population. Dans une classe de cinquième, on a observé
   – 16 garçons : moyenne $\bar{x}_G = 148$ cm, écart-type $s_G = 6.9$ cm,
   – 15 filles : moyenne $\bar{x}_F = 153$ cm, écart-type $s_F = 5.8$ cm.
   À l'aide de la question précédente, tester au niveau 5% l'hypothèse nulle "la taille moyenne des filles dépasse celle des garçons d'au moins 2 cm".

Exercice 8.3.3. On se place dans le modèle gaussien à espérance et variance égales $\mathcal{P} = \{\mathcal{N}(\theta,\theta), \theta > 0\}$.

1. Donner une statistique exhaustive et déterminer l'Estimateur du Maximum de Vraisemblance $\hat{\theta}_n$ de θ. Cet estimateur est-il convergent ?
2. Calculer l'information de Fisher $I(\theta)$. Vérifier que
$$\sqrt{n}(\hat{\theta}_n - \theta) = f\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) \times \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - (\theta+\theta^2)\right)$$
   où $f :]-\frac{1}{4}, +\infty[\to \mathbb{R}$ est définie par
$$f(y) = \begin{cases} \frac{\sqrt{\frac{1}{4}+y}-\frac{1}{2}-\theta}{y-(\theta+\theta^2)} & \text{si } \ y \neq \theta+\theta^2, \\ \frac{1}{1+2\theta} & \text{sinon.} \end{cases}$$
   En déduire que $\hat{\theta}_n$ est asymptotiquement normal de variance asymptotique égale à $1/I(\theta)$.
3. On note $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $V_n = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ la moyenne et la variance empiriques et pour $\lambda \in [0,1]$, on pose $T_n^\lambda = \lambda \bar{X}_n + (1-\lambda)V_n$. Montrer que T_n^λ est un estimateur convergent et asymptotiquement normal de θ dont on calculera la variance asymptotique $V(\lambda,\theta)$ (*on pourra utiliser les résultats donnés dans l'exemple 7.2.12*).
   Vérifier que $\min_{\lambda\in[0,1]} V(\lambda,\theta) = 1/I(\theta)$. Dans la classe des estimateurs $(T_n^\lambda)_{\lambda\in[0,1]}$, existe-il un estimateur aussi performant que $\hat{\theta}_n$ pour tout $\theta > 0$ en termes de variance asymptotique ?

Pour $\theta_0 > 0$ fixé, on souhaite tester l'hypothèse nulle $H_0 = \{\theta \leq \theta_0\}$ contre l'hypothèse alternative $H_1 = \{\theta > \theta_0\}$.

4. Construire un test asymptotique de niveau α à l'aide de $\hat{\theta}_n$.
   A.N. : $\theta_0 = 3.8$, $n = 100$, $\bar{x}_n = 4.18$ et $v_n = 3.84$.
5. Construire un test asymptotique de niveau α à l'aide de T_n^λ.
   A.N. : $\lambda \in \{0, 0.5, 1, \lambda_0\}$ où λ_0 minimise $\lambda \in [0,1] \to V(\lambda,\theta_0)$.

Exercice 8.3.4. On souhaite vérifier la qualité du générateur de nombres aléatoires d'une calculatrice scientifique. Pour cela, on procède à 250 tirages dans l'ensemble $\{0,\ldots,9\}$ et on obtient les résultats suivants :

x	0	1	2	3	4	5	6	7	8	9
$N(x)$	28	32	23	26	23	31	18	19	19	31

Tester au niveau $\alpha = 0.1$ si le générateur produit des entiers uniformément répartis sur $\{0, \ldots, 9\}$.

Exercice corrigé 8.3.5. On désire étudier la répartition des naissances suivant le type du jour dans la semaine (jours ouvrables ou week-end) et suivant le mode d'accouchement (naturel ou par césarienne). Les données proviennent du "National Vital Statistics Report" et concernent les naissances aux USA en 1997.

Naissances	Naturelles	César.	Total
J.O.	2331536	663540	2995076
W.E.	715085	135493	850578
Total	3046621	799033	3845654

Naissances	Naturelles	César.	Total
J.O.	60.6 %	17.3 %	77.9%
W.E.	18.6 %	3.5 %	22.1%
Total	79.2 %	20.8 %	100.0%

1. Tester au niveau 0.001 l'hypothèse d'indépendance entre le type du jour de naissance (jour ouvrable ou week-end) et le mode d'accouchement (naturel ou césarienne).
2. On désire savoir s'il existe une évolution significative dans la répartition des naissances par rapport à 1996. Tester au niveau 0.01 si $p = p^0$, où p^0 correspond aux données de 1996 :

Naissances	Naturelles	Césariennes
J.O.	60.5 %	17.0 %
W.E.	18.9 %	3.6 %

.

Exercice 8.3.6. On se propose de comparer les réactions produites par deux vaccins B.C.G. désignés par A et B . Un groupe de 348 enfants a été divisé par tirage au sort en deux séries qui ont été vaccinées, l'une par A, l'autre par B. La réaction a été ensuite lue par une personne ignorant le vaccin utilisé. Les résultats figurent dans le tableau suivant :

Vaccin	Réaction légère	Réaction moyenne	Ulcération	Abcès	Total
A	12	156	8	1	177
B	29	135	6	1	171
Total	41	291	14	2	348

On désire tester l'hypothèse selon laquelle les réactions aux deux vaccins ont même loi.

1. Expliquer pourquoi cette situation relève d'un test du χ^2 d'indépendance.
2. Les effectifs observés permettent-ils d'effectuer le test ? Sinon, procéder aux opérations nécessaires sur ces données, puis effectuer le test au niveau $\alpha = 0.05$.

8.4 Résumé

- **Tests d'hypothèses :** $X = (X_1, \ldots, X_n)$ échantillon d'un modèle paramétrique $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$.
  - L'*hypothèse nulle* H_0 et l'*hypothèse alternative* H_1 forment une partition de Θ.
  - *Région critique* W *:* sous-ensemble de l'ensemble des valeurs possibles de X. Si l'observation x est dans W on rejette H_0 et on accepte H_1. Sinon on accepte H_0. Par abus de notation, on note aussi W l'événement $\{X \in W\}$. Le plus souvent, $W = \{S \in A\}$ où S est une statistique à valeurs réelles et $A \subset \mathbb{R}$.
  - *Niveau du test :* $\boxed{\alpha_W = \sup_{\theta \in H_0} \mathbb{P}_\theta(W)}$ contrôle l'erreur de première espèce, c'est-à-dire le rejet de H_0 à tort.
  - *Test asymptotique :* Soit $(W_n)_n$ suite de régions critiques dépendant de la taille n de l'échantillon. Le test est dit
    – *de niveau asymptotique* α si $\lim_{n\to+\infty} \sup_{\theta\in H_0} \mathbb{P}_\theta(W_n) = \alpha$,
    – *convergent* si $\forall\theta \in H_1$, $\lim_{n\to+\infty} \mathbb{P}_\theta(W_n) = 1$.
  - *p-valeur :* probabilité sous H_0 que la statistique de test S prenne des valeurs plus défavorables pour l'acceptation de H_0 que la valeur observée. Si la p-valeur est supérieure au niveau α, on accepte H_0. Sinon, on accepte H_1.

- **Test d'adéquation à une loi discrète : le test du χ^2**
  - *Modèle :* $(X_1, \ldots, X_n)$ échantillon de variables à valeurs dans $A = \{a_1, \ldots, a_k\}$ indépendantes et de loi commune paramétrisée par $p = (p_1, \ldots, p_k)$: $\forall j \in \{1, \ldots, k\}$, $\mathbb{P}_p(X_i = a_j) = p_j$.
  - *Hypothèses :* $H_0 = \{p = p^0\}$ et $H_1 = \{p \neq p^0\}$.
  - *Statistique de test :* $\boxed{\zeta_n = n \sum_{j=1}^{k} \frac{(\hat{p}_j - p_j^0)^2}{p_j^0}}$ où $\hat{p}_j = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i = a_j\}}$ désigne la fréquence empirique de a_j.
  - *Comportement asymptotique :* Lorsque $n \to +\infty$, sous H_0, ζ_n converge en loi vers $Z \sim \chi^2(k-1)$. Sous H_1, ζ_n tend p.s. vers $+\infty$.
  - *Région critique pour un niveau asymptotique* α *:* $W_n = \{\zeta_n \geq \chi^2_{k-1,1-\alpha}\}$ où $\chi^2_{k-1,r}$ quantile d'ordre r de la loi $\chi^2(k-1)$. Le test est convergent.
  - *p-valeur asymptotique :* $\mathbb{P}(Z \geq \zeta_n^{\text{obs}})$ où $Z \sim \chi^2(k-1)$ et ζ_n^{obs} est la valeur observée pour la statistique de test.
  - *Critère d'utilisation :* $n \times \min_{1\leq j\leq k} p_j^0 \geq 5$ pour assurer que la loi de ζ_n sous H_0 est suffisamment proche de sa limite $\chi^2(k-1)$.

Chapitre 9

Régression Linéaire

L'objectif de la régression linéaire est d'étudier l'effet de certains facteurs explicatifs appelés **régresseurs** sur un phénomène observé ou sur des résultats expérimentaux. On peut par exemple s'intéresser à l'influence de l'âge, du poids, de la concentration de bactéries dans le sang sur la durée d'hospitalisation des patients. Pour $i \in \{1, \ldots, n\}$, on note X_i la i-ème observation (durée d'hospitalisation du i-ème patient par exemple) et $(r_i^j)_{1 \leq j \leq p}$ les p variables explicatives correspondantes (âge, poids et concentration de bactéries du i-ème patient). On suppose que chaque X_i s'écrit comme fonction affine des r_i^j plus un résidu aléatoire :

$$\forall i \in \{1, \ldots, n\},\ X_i = \beta_0 + \sum_{j=1}^{p} \beta_j r_i^j + \varepsilon_i$$

où les β_j sont des constantes et les résidus ε_i sont I.I.D. suivant la loi normale $\mathcal{N}_1(0, \sigma^2)$ centrée de variance σ^2. Bien sûr, pour que cette étude puisse être significative, il faut que le nombre n d'observations soit grand devant le nombre p de variables explicatives. En notant X et ε les vecteurs colonnes dont les coordonnées sont respectivement les X_i et les ε_i, $\gamma = (\beta_0, \beta_1, \ldots, \beta_p)^*$ et

$$M = \begin{pmatrix} 1 & r_1^1 & \ldots & r_1^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_n^1 & \ldots & r_n^p \end{pmatrix},$$

le modèle se récrit

$$\boxed{X = M\gamma + \varepsilon}. \tag{9.1}$$

Dans la suite, nous supposerons que M est une matrice de rang $p+1$, c'est-à-dire de rang égal à son nombre de colonnes. Lorsque ce n'est pas vrai, on peut s'y ramener en supprimant des colonnes qui sont combinaisons linéaires des autres colonnes.

Après avoir vu comment estimer le paramètre $\boxed{\theta = (\gamma, \sigma^2)}$ et construire des intervalles de confiance pour ses coordonnées, nous verrons comment tester si le j-ème régresseur r_j est utile i.e. si $\beta_j \neq 0$.

9.1 Estimation

La densité du vecteur ε est $\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2}$. On en déduit par translation que la vraisemblance et la log-vraisemblance du modèle sont

$$p_n(x,(\gamma,\sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-(M\gamma)_i)^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{|x-M\gamma|^2}{2\sigma^2}}$$
$$l_n(x,(\gamma,\sigma^2)) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{|x-M\gamma|^2}{2\sigma^2}.$$

À $\sigma^2 > 0$ fixé, la log-vraisemblance est maximale pour $\gamma \in \mathbb{R}^p$ qui minimise $J(\gamma) = |x-M\gamma|^2 = |x|^2 - 2(M^*x).\gamma + (M^*M\gamma).\gamma$. Le point de l'image $\boxed{E = \{Mu, u \in \mathbb{R}^{p+1}\}}$ de $\mathbb{R}^{p+1}$ par M à distance minimale de x est la projection orthogonale x_E de x sur E. On en déduit que J atteint son minimum pour γ tel que $M\gamma = x_E$, équation qui admet une unique solution puisque M est de rang $p+1$. On a

$$\nabla_\gamma J(\gamma) = -2M^*x + 2M^*M\gamma$$

et la condition d'optimalité du premier ordre s'écrit donc $M^*M\gamma = M^*x$. Comme la matrice $M^*M \in \mathbb{R}^{(p+1)\times(p+1)}$ est de rang $p+1$, elle est inversible et on conclut qu'à σ^2 fixé la log-vraisemblance est maximale pour $\gamma = (M^*M)^{-1}M^*x$ et vaut alors $f(\sigma^2)$ où

$$f(\lambda) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\lambda) - \frac{|x-x_E|^2}{2\lambda} \quad \text{avec} \quad x_E = M(M^*M)^{-1}M^*x.$$

Comme la dérivée $f'(\lambda) = \frac{n}{2\lambda}(\frac{|x-x_E|^2}{n\lambda} - 1)$ est positive sur $]0, \frac{|x-x_E|^2}{n}]$ et négative sur $[\frac{|x-x_E|^2}{n}, +\infty[$, la fonction f atteint son maximum en $\frac{|x-x_E|^2}{n}$. On conclut donc que la log-vraisemblance est maximale en $((M^*M)^{-1}M^*x, \frac{|x-x_E|^2}{n})$. Ainsi l'Estimateur du Maximum de Vraisemblance de (γ,σ^2) est $((M^*M)^{-1}M^*X, \frac{|X-X_E|^2}{n})$ où $\boxed{X_E = M(M^*M)^{-1}M^*X}$.

Remarque 9.1.1. Nous venons de voir que l'EMV $(M^*M)^{-1}M^*X$ de γ est le point $u \in \mathbb{R}^{p+1}$ qui minimise $|X-Mu|^2$, c'est-à-dire la solution au sens des moindres carrés de l'équation $X = Mu$. Ainsi, $(M^*M)^{-1}M^*X$ conserve une interprétation naturelle même lorsque l'on ne suppose pas que les résidus ε_i sont aléatoires. L'intérêt de l'hypothèse que nous avons faite sur les résidus est ailleurs : elle va nous permettre de construire des intervalles de confiance et des tests d'hypothèses.

Proposition 9.1.2. *Le vecteur aléatoire* $\boxed{\hat\gamma = (M^*M)^{-1}M^*X}$ *est un estimateur sans biais de* γ. *Plus précisément,* $\hat\gamma \sim \mathcal{N}_{p+1}(\gamma, \sigma^2(M^*M)^{-1})$ *et ce vecteur est indépendant de la variable aléatoire* $\frac{|X-X_E|^2}{\sigma^2} \sim \chi^2(n-(p+1))$.

On déduit de cette proposition que $\frac{|X-X_E|^2}{n}$ est un estimateur biaisé de σ^2 indépendant de $\hat\gamma$. On lui préférera l'estimateur sans biais $\frac{|X-X_E|^2}{n-(p+1)}$.

Démonstration : D'après (9.1), $\hat\gamma = (M^*M)^{-1}M^*(M\gamma+\varepsilon) = \gamma + (M^*M)^{-1}M^*\varepsilon$ où $\varepsilon \sim \mathcal{N}_n(0,\sigma^2 I_n)$. Comme

$$(M^*M)^{-1}M^*\left((M^*M)^{-1}M^*\right)^* = (M^*M)^{-1}M^*\left(M(M^*M)^{-1}\right) = (M^*M)^{-1},$$

on en déduit que $\hat\gamma \sim \mathcal{N}_{p+1}(\gamma, \sigma^2(M^*M)^{-1})$.

Comme $M\gamma \in E$, $X_E = M\gamma + \varepsilon_E$ où ε_E désigne la projection orthogonale de ε sur E. Donc $\frac{|X-X_E|^2}{\sigma^2} = \left|\frac{\varepsilon-\varepsilon_E}{\sigma}\right|^2$ où $\frac{\varepsilon-\varepsilon_E}{\sigma}$ est la projection orthogonale du vecteur aléatoire $\frac{\varepsilon}{\sigma} \sim \mathcal{N}_n(0, I_n)$ sur l'orthogonal de E. Notons que comme M est de rang $p+1$, E est de dimension $p+1$ et son orthogonal de dimension $n-(p+1)$. On déduit alors du théorème de Cochran 6.2.3 que la variable aléatoire $\frac{|X-X_E|^2}{\sigma^2}$ suit la loi $\chi^2(n-(p+1))$ et est indépendante de ε_E. Comme $M\hat{\gamma} = X_E$, on a $M\hat{\gamma} = M\gamma + \varepsilon_E$. Après multiplication de cette égalité par $(M^*M)^{-1}M^*$, on obtient que $\hat{\gamma} = \gamma + (M^*M)^{-1}M^*\varepsilon_E$ d'où l'on conclut que $\hat{\gamma}$ est indépendant de $\frac{|X-X_E|^2}{\sigma^2}$. □

À l'aide de la proposition on peut construire des intervalles de confiance pour les coefficients β_j. En effet pour $k \in \{1, \dots, p+1\}$, la variable aléatoire $\frac{\hat{\gamma}_k-\gamma_k}{\sigma\sqrt{(M^*M)^{-1}_{kk}}} \sim \mathcal{N}_1(0,1)$ est indépendante de $\frac{|X-X_E|^2}{\sigma^2} \sim \chi^2(n-(p+1))$. Donc

$$\frac{\hat{\gamma}_k-\gamma_k}{\sqrt{(M^*M)^{-1}_{kk}|X-X_E|^2/(n-(p+1))}} \sim t(n-(p+1)).$$

Si $t_{n-(p+1),r}$ désigne le quantile d'ordre r de la loi de Student $t(n-(p+1))$, on en déduit que $\boxed{\left[\hat{\gamma}_k - t_{n-(p+1),1-\frac{\alpha}{2}}\sqrt{\frac{(M^*M)^{-1}_{kk}|X-X_E|^2}{n-(p+1)}}, \hat{\gamma}_k + t_{n-(p+1),1-\frac{\alpha}{2}}\sqrt{\frac{(M^*M)^{-1}_{kk}|X-X_E|^2}{n-(p+1)}}\right]}$ est un intervalle de confiance de niveau $1-\alpha$ pour β_{k-1}.

Exercice 9.1.3. On suppose que la variance σ^2 est connue et égale à σ_0^2.

1. Quelle est la loi de $\frac{\hat{\gamma}_k-\gamma_k}{\sigma_0\sqrt{(M^*M)^{-1}_{kk}}}$?
2. En déduire un intervalle de confiance de niveau $1-\alpha$ pour β_{k-1} ?

En procédant comme dans l'exemple 7.3.5, on obtient l'intervalle de confiance de niveau $1-\alpha$ suivant pour la variance σ^2 : $\boxed{\left[\frac{|X-X_E|^2}{\chi^2_{n-(p+1),1-\frac{\alpha}{2}}}, \frac{|X-X_E|^2}{\chi^2_{n-(p+1),\frac{\alpha}{2}}}\right]}$ où $\chi^2_{n-(p+1),r}$ désigne le quantile d'ordre r de la loi $\chi^2(n-(p+1))$.

9.2 Test de l'utilité des régresseurs

On souhaite tester l'utilité de $p-q$ régresseurs où $q \in \{0, \dots, p-1\}$. Quitte à permuter les régresseurs, nous supposerons que ce sont les $p-q$ derniers. Ainsi, l'objectif est de tester l'hypothèse nulle $H_0 = \{\beta_j = 0, q+1 \leq j \leq p\}$ contre l'hypothèse alternative $H_1 = \{\exists j \in \{q+1, \dots, p\} : \beta_j \neq 0\}$.
Sous H_0, pour $i \in \{1, \dots, n\}$, $X_i = \beta_0 + \sum_{j=1}^q \beta_j r_i^j + \varepsilon_i$ et on a $X = M_0\tilde{\gamma} + \varepsilon$ où $\tilde{\gamma} = (\beta_0, \dots, \beta_q)^*$ et

$$M_0 = \begin{pmatrix} 1 & r_1^1 & \dots & r_1^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_n^1 & \dots & r_n^q \end{pmatrix}.$$

Comme les colonnes de M forment une famille libre, la matrice M_0 est de rang $q+1$. L'ensemble $H = \{M_0 u, u \in \mathbb{R}^{q+1}\}$ est un sous-espace vectoriel de dimension $q+1$ de E.

Pour tout vecteur v de $\mathbb{R}^n$, on note $v_H = M_0(M_0^* M_0)^{-1} M_0^* v$ la projection orthogonale de v sur H.

Les vecteurs aléatoires $\varepsilon - \varepsilon_E$ et $\varepsilon_E - \varepsilon_H$ sont respectivement les projections orthogonales de ε sur l'orthogonal de E de dimension $n-(p+1)$ et sur l'orthogonal de H dans E de dimension $p-q$. On déduit alors du théorème de Cochran 6.2.3 que $|\varepsilon - \varepsilon_E|/\sigma^2$ et $|\varepsilon_E - \varepsilon_H|^2/\sigma^2$ sont des variables aléatoires indépendantes respectivement distribuées suivant les lois du χ^2 à $n-(p+1)$ degrés de liberté et à $p-q$ degrés de liberté.

Sous H_0, $X_E = M_0\tilde{\gamma} + \varepsilon_E$ et $X_H = M_0\tilde{\gamma} + \varepsilon_H$ ce qui entraîne que $X - X_E = \varepsilon - \varepsilon_E$ et $X_E - X_H = \varepsilon_E - \varepsilon_H$. En conséquence, le rapport $F = \dfrac{|X_E - X_H|^2/(p-q)}{|X - X_E|^2/(n-(p+1))}$ suit la loi de Fisher $\mathcal{F}_{p-q,n-(p+1)}$.

Sous H_1, $X_E = M\gamma + \varepsilon_E$ et $X_H = (M\gamma)_H + \varepsilon_H$ ce qui entraîne que $X - X_E = \varepsilon - \varepsilon_E$ et $X_E - X_H = M\gamma - (M\gamma)_H + \varepsilon_E - \varepsilon_H$. Le numérateur et le dénominateur du rapport $F = \dfrac{|X_E - X_H|^2/(p-q)}{|X - X_E|^2/(n-(p+1))}$ restent indépendants avec le dénominateur de loi inchangée. Mais comme $(M\gamma)_H \neq M\gamma$, le numérateur et donc le rapport va avoir tendance à prendre des valeurs plus grandes que sous H_0 (voir la remarque 9.2.1 qui suit).

On choisit donc comme région critique $W = \{F \geq \mathcal{F}_{p-q,n-(p+1),1-\alpha}\}$ où $\mathcal{F}_{p-q,n-(p+1),r}$ désigne le quantile d'ordre r de la loi de Fisher $\mathcal{F}_{p-q,n-(p+1)}$. La p-valeur du test est $\mathbb{P}(Z \geq F^{\text{obs}})$ où $Z \sim \mathcal{F}_{p-q,n-(p+1)}$ et F^{obs} désigne la statistique de test observée.

Remarque 9.2.1. On se place sous H_1. Pour vérifier que $|X_E - X_H|^2$ prend des valeurs plus grandes que sous H_0, on se donne $e_1, \ldots, e_{p-q}$ une base orthornormée de l'orthogonal de H dans E telle que $e_1 = \frac{M\gamma - (M\gamma)_H}{|M\gamma - (M\gamma)_H|}$. D'après la preuve du théorème de Cochran 6.2.3, les variables aléatoires $(G_i = \frac{e_i.\varepsilon}{\sigma})_{1\leq i\leq p-q}$ sont I.I.D. suivant la loi $\mathcal{N}_1(0,1)$. En outre, $\varepsilon_E - \varepsilon_H = \sigma \sum_{i=1}^{p-q} G_i e_i$ et $X_E - X_H = \sigma\left(\left(G_1 + \frac{|M\gamma - (M\gamma)_H|}{\sigma}\right) e_1 + \sum_{i=2}^{p-q} G_i e_i\right)$. Donc $|\varepsilon_E - \varepsilon_H|^2 = \sigma^2 \sum_{i=1}^{p-q} G_i^2$ et $|X_E - X_H|^2 = \sigma^2\left(\left(G_1 + \frac{|M\gamma - (M\gamma)_H|}{\sigma}\right)^2 + \sum_{i=2}^{p-q} |G_i|^2\right)$ ne diffèrent que par le premier terme. Comme sous H_0, $|X_E - X_H|^2 = |\varepsilon_E - \varepsilon_H|^2$, il suffit de vérifier que pour $a = \frac{|M\gamma - (M\gamma)_H|}{\sigma}$, $(G_1 + a)^2$ prend des valeurs plus grandes que G_1^2. Pour $x \geq 0$,

$$\mathbb{P}((G_1 + y)^2 \leq x) = \mathbb{P}(-y - \sqrt{x} \leq G_1 \leq -y + \sqrt{x}) = \int_{-y-\sqrt{x}}^{-y+\sqrt{x}} e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}}.$$

Pour $y, x > 0$, comme $(y - \sqrt{x})^2 = y^2 + x - 2y\sqrt{x} < y^2 + x + 2y\sqrt{x} = (y + \sqrt{x})^2$,

$$\frac{\partial}{\partial y}\mathbb{P}((G_1 + y)^2 \leq x) = \frac{e^{-\frac{(y+\sqrt{x})^2}{2}} - e^{-\frac{(y-\sqrt{x})^2}{2}}}{\sqrt{2\pi}} < 0.$$

Ainsi $\forall x > 0$, $\mathbb{P}((G_1 + a)^2 > x) = 1 - \mathbb{P}((G_1 + a)^2 \leq x) > 1 - \mathbb{P}(G_1^2 \leq x) = \mathbb{P}(G_1^2 > x)$, ce qui montre bien que $(G_1 + a)^2$ prend des valeurs plus grandes que G_1^2.

L'exercice suivant a pour objectif de vérifier que si on s'intéresse à l'utilité d'un seul régresseur ($q = p - 1$), on peut effectuer un test de Student à la place du test de Fisher que nous venons de présenter.

Exercice 9.2.2. On se place dans le cas $q = p - 1$ et on souhaite tester l'hypothèse nulle $H_0 = \{\gamma_{p+1} = 0\}$ contre l'hypothèse alternative $H_1 = \{\gamma_{p+1} \neq 0\}$.

1. Quelle est la loi de $T = \frac{\hat{\gamma}_{p+1}}{\sqrt{(M^*M)^{-1}_{p+1,p+1}|X-X_E|^2/(n-(p+1))}}$ sous H_0 ?
2. Expliquez pourquoi T prend des valeurs plus grandes en valeur absolue sous H_1 que sous H_0.
3. En déduire la région critique du test.

Coefficient de détermination : Les logiciels de statistiques, en plus de l'estimation des paramètres (γ, σ^2), du test de Fisher de l'utilité de l'ensemble des régresseurs ($q = 0$) et des tests de Student d'utilité de chacun des régresseurs pris individuellement fournissent le coefficient de détermination $R^2 = \frac{|X_E - X_{\mathbf{1}_n}|^2}{|X - X_{\mathbf{1}_n}|^2}$ où $X_{\mathbf{1}_n} = (\frac{1}{n}\sum_{i=1}^n X_i, \ldots, \frac{1}{n}\sum_{i=1}^n X_i)^*$ est la projection orthogonale de X sur le sous-espace vectoriel engendré par le vecteur $\mathbf{1}_n$ dont toutes les composantes sont égales à 1 (première colonne de M). Comme $X_{\mathbf{1}_n} \in E$, le vecteur $X_E - X_{\mathbf{1}_n} \in E$ est orthogonal à $X - X_E \in E^\perp$ et $|X - X_{\mathbf{1}_n}|^2 = |X - X_E|^2 + |X_E - X_{\mathbf{1}_n}|^2$. Le coefficient R^2 est relié au test de l'utilité de l'ensemble des régresseurs puiqu'alors la statistique de test

$$\begin{aligned} F &= \frac{|X_E - X_{\mathbf{1}_n}|^2/p}{|X - X_E|^2/(n-(p+1))} = \frac{|X_E - X_{\mathbf{1}_n}|^2/p}{(|X - X_{\mathbf{1}_n}|^2 - |X_E - X_{\mathbf{1}_n}|^2)/(n-(p+1))} \\ &= \frac{R^2}{1-R^2} \times \frac{n-(p+1)}{p} \end{aligned}$$

s'exprime en fonction de R^2. Lorsque $R^2 \to 1$, $F \to +\infty$ et on accepte H_1, c'est-à-dire que l'on conclut qu'au moins un régresseur est utile. En revanche, lorsque $R^2 \to 0$, $F \to 0$ et on accepte H_0, c'est-à-dire que l'on conclut que tous les régresseurs sont inutiles. Ainsi plus le coefficient de détermination est faible, moins l'utilité de l'ensemble des régresseurs est probante.

9.3 Exercices

Exercice 9.3.1. La durée d'une maladie semble liée au nombre de bactéries dans l'organisme et à la température du patient lors de son admission à l'hôpital. On détermine pour $n = 10$ malades leur décompte en milliers de bactéries, r^1, et leur température r^2, et on observe la durée Y de persistance des symptômes de la maladie en jours :

r^1	r^2	Y
8	37.6	29
7	39.2	29
4	38.5	19
6	37.4	23
9	38.1	32
7	39.1	28
8	39.0	30
3	37.8	18
8	38.2	30
7	39.1	31

1. On propose tout d'abord un modèle (noté vectoriellement)

$$Y = \beta_0 \mathbf{1}_n + \beta_1 r^1 + \varepsilon,$$

où $\mathbf{1}_n$ est le vecteur de taille n dont toutes les coordonnées sont égales à 1 et ε est un n-échantillon distribué suivant la loi $\mathcal{N}_1(0, \sigma^2)$. Donner des estimateurs sans biais de $\beta_0, \beta_1, \sigma^2$. Tester la significativité du régresseur au niveau 5% ; qu'en concluez-vous ?

2. On propose ensuite le modèle

$$Y = \beta_0 \mathbf{1}_n + \beta_1 r^1 + \beta_2 r^2 + \varepsilon.$$

Donner des estimateurs sans biais de $\beta_0, \beta_1, \beta_2, \sigma^2$. Tester, au niveau 5%, l'hypothèse "$\beta_2 = 0$" contre "$\beta_2 \neq 0$" *(attention, ce modèle est assez lourd à traiter numériquement ; il peut être plus pratique d'effectuer les calculs sous* `Scilab`*).*

3. Quel modèle conseillez-vous ?

Pour faciliter les calculs numériques, on donne la matrice des sommes de produits croisés ; on a par exemple $\sum_{i=1}^{10} r_i^2 r_i^1 = 2574.9$.

	$\mathbf{1}_{10}$	r^1	r^2	Y
r^1	67	481		
r^2	384	2574.9	14749.72	
Y	269	1884	10341.4	7465

Exercice corrigé 9.3.2. On souhaite tester l'égalité des moyennes de k échantillons gaussiens. Pour $i \in \{1, \ldots, k\}$ et $j \in \{1, \ldots, n_i\}$, on note $X_{i,j}$ la j-ème valeur du i-ème échantillon et on suppose que

$$X_{i,j} = \mu_i + \varepsilon_{i,j}$$

où les variables $\varepsilon_{i,j}$ sont I.I.D. suivant la loi normale $\mathcal{N}_1(0, \sigma^2)$ centrée de variance σ^2. Par exemple $X_{i,j}$ peut représenter le rendement dans la j-ème parcelle ensemencée avec l'espèce de blé i. La question de l'égalité des moyennes μ_i des échantillons revient alors à savoir si les différentes espèces de blé sont équivalentes en termes de rendement. On pose

$$\begin{cases} X = (X_{1,1}, \ldots, X_{1,n_1}, X_{2,1}, \ldots, X_{2,n_2}, \ldots, X_{k,1}, \ldots, X_{k,n_k})^* \\ \varepsilon = (\varepsilon_{1,1}, \ldots, \varepsilon_{1,n_1}, \varepsilon_{2,1}, \ldots, \varepsilon_{2,n_2}, \ldots, \varepsilon_{k,1}, \ldots, \varepsilon_{k,n_k})^* \\ \mu = (\mu_1, \ldots, \mu_1, \mu_2, \ldots, \mu_2, \ldots, \mu_k, \ldots, \mu_k)^* \end{cases},$$

où μ_i apparaît n_i fois si bien que le modèle se récrit $\boxed{X = \mu + \varepsilon}$. Enfin, on note $n = \sum_{i=1}^k n_k$, $E = \{(y_1, \ldots, y_1, y_2, \ldots, y_2, \ldots, y_k, \ldots, y_k)^* : (y_1, \ldots, y_k) \in \mathbb{R}^k\}$ où y_i apparaît n_i fois et $H = \{(z, \ldots, z)^* : z \in \mathbb{R}\}$ où z apparaît n fois.

1. Donner l'Estimateur du Maximum de Vraisemblance de $\theta = (\mu, \sigma^2)$.
2. Vérifier que $X_E = (X_{1,.}, \ldots, X_{1,.}, X_{2,.}, \ldots, X_{2,.}, \ldots, X_{k,.}, \ldots, X_{k,.})^*$ où pour $i \in \{1, \ldots, k\}$, $X_{i,.} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}$ apparaît n_i fois. Déterminer X_H.
3. Quelle est la loi de $(|X - X_E|^2/\sigma^2, |X_E - X_H|^2/\sigma^2)$ sous $H_0 = \{\mu \in H\}$? En deduire que sous H_0 le rapport $F = \dfrac{|X_E - X_H|^2/(k-1)}{|X - X_E|^2/(n-k)}$ suit la loi de Fisher $\mathcal{F}_{k-1,n-k}$.
4. Pourquoi les variables aléatoires $|X_E - X_H|^2$ et F prennent-elles des valeurs plus grandes sous $H_1 = \{\mu \notin H\}$ que sous H_0 ?
5. En déduire la région critique du test.

6. **Application :** on souhaite tester, pour une chaîne de magasins, les politiques de publicité suivantes :

A : aucune publicité
B : tracts distribués dans le voisinage
C : tracts distribués et annonces dans les journaux.

On sélectionne 18 magasins divisés au hasard en 3 groupes de 6, et chaque groupe applique l'une des politiques de publicité. On enregistre ensuite les ventes cumulées sur un mois pour chaque magasin, et on obtient les moyennes et écart-types empiriques suivants (en milliers d'euros) :

	A	B	C
$\bar{X}$	130.17	139.5	169.17
S	8.57	14.71	18.23

où, par exemple, pour le groupe A d'effectif $n_A = 6$,

$$\bar{X}_A = \frac{1}{n_A}\sum_{j=1}^{n_A} X_{A,j}, \quad S_A = \sqrt{\frac{1}{n_A - 1}\sum_{j=1}^{n_A}(X_{A,j} - \bar{X}_A)^2}.$$

On suppose que les observations pour chaque groupe sont gaussiennes, de moyennes respectives μ_A, μ_B, μ_C et de même variance σ^2. Tester l'hypothèse nulle "il n'existe aucune différence entre les politiques de publicité" au niveau 5%. Évaluer approximativement la p-valeur.

9.4 Résumé

Modèle $\boxed{X = M\gamma + \varepsilon}$ où

- les colonnes de la matrice

$$M = \begin{pmatrix} 1 & r_1^1 & \dots & r_1^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_n^1 & \dots & r_n^p \end{pmatrix} \quad \text{de rang } p+1,$$

sont un vecteur de coordonnées toutes égales à 1 et les régresseurs,

- $\gamma \in \mathbb{R}^{p+1}$ est le vecteur des paramètres,
- $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$.

- **Projection orthogonale de X sur $E = \{Mu : u \in \mathbb{R}^{p+1}\}$:**

$$\boxed{X_E = M(M^*M)^{-1}M^*X}.$$

- **Estimation de γ :** $\boxed{\hat{\gamma} = (M^*M)^{-1}M^*X}$ estimateur sans biais de γ indépendant de $\frac{|X-X_E|^2}{\sigma^2} \sim \chi^2(n-(p+1))$.

- **Intervalle de confiance de niveau α pour γ_k :**

$$\left[\hat{\gamma}_k \pm t_{n-(p+1),1-\frac{\alpha}{2}} \sqrt{\frac{(M^*M)^{-1}_{kk}|X - X_E|^2}{n-(p+1)}}\right].$$

- **Test de l'utilité des $p-q$ derniers régresseurs où $q \in \{0, \dots, p-1\}$:**

$$\text{Soit } M_0 = \begin{pmatrix} 1 & r_1^1 & \dots & r_1^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_n^1 & \dots & r_n^q \end{pmatrix} \quad \text{et} \quad H = \{M_0 u : u \in \mathbb{R}^{q+1}\}.$$

- *Hypothèses :* $H_0 = \{M\gamma \in H\}$ et $H_1 = \{M\gamma \notin H\}$.
- *Projection orthogonale de X sur H :* $X_H = M_0(M_0^*M_0)^{-1}M_0^*X$.
- *Statistique de test :*

$$\boxed{F = \frac{|X_E - X_H|^2/(p-q)}{|X - X_E|^2/(n-(p+1))}}$$

qui suit la loi de Fisher $\mathcal{F}_{p-q,n-(p+1)}$ sous H_0 et prend des valeurs plus grandes sous H_1.

- *Région critique :* $W = \{F \geq \mathcal{F}_{p-q,n-(p+1),1-\alpha}\}$ où $\mathcal{F}_{p-q,n-(p+1),r}$ quantile d'ordre r de la loi de Fisher.
- *p-valeur :* $\mathbb{P}(Z \geq F^{\text{obs}})$ où $Z \sim \mathcal{F}_{p-q,n-(p+1)}$ et F^{obs} désigne la statistique de test observée.

Chapitre 10

Corrigés d'une sélection d'exercices et problèmes

10.1 Probabilité sur un espace fini

Corrigé de l'exercice 1.3.4. On note de 1 à 6 les positions des emplacements qui sont percutés successivement à chaque fois que l'on presse la détente. On modélise le jeu par $\omega \in \Omega = \{1, \ldots, 6\}$ tel que les deux balles sont sur les emplacements ω et $(\omega + 1)$ mod 6. Comme le barillet est positionné au hasard, Ω est muni de la probabilité uniforme. L'événement "votre adversaire reste en vie" s'écrit $A = \{2, 3, 4, 5\}$. Si vous choisissez de ne pas faire tourner à nouveau le barillet au hasard, l'événement "vous restez en vie" s'écrit $B = \{3, 4, 5, 6\}$. On a

$$\mathbb{P}(A) = \frac{\text{Card }(A)}{\text{Card }(\Omega)} = \frac{2}{3} \text{ et } \mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\text{Card }(B \cap A)}{\text{Card }(A)} = \frac{3}{4}.$$

Vous avez donc trois chances sur quatre de rester en vie si vous ne faites pas tourner à nouveau le barillet. Sinon, vous vous mettez dans la situation initiale de votre adversaire et votre probabilité de rester en vie est $\mathbb{P}(A) = \frac{2}{3}$.

Corrigé de l'exercice 1.3.8. On note $C = \{\text{suspect coupable}\}$, $I = \{\text{suspect innocent}\}$ et $G = \{\text{suspect gaucher}\}$. D'après l'énoncé, $\mathbb{P}(C) = 1 - \mathbb{P}(I) = 0.6$, $\mathbb{P}(G|C) = 1$ et $\mathbb{P}(G|I) = 0.07$. D'après la formule de Bayes, la probabilité pour que le suspect soit coupable sachant qu'il est gaucher est

$$\mathbb{P}(C|G) = \frac{\mathbb{P}(G|C)\mathbb{P}(C)}{\mathbb{P}(G|C)\mathbb{P}(C) + \mathbb{P}(G|I)\mathbb{P}(I)} = 0.955.$$

Ainsi, s'il se trouve que le suspect est gaucher, l'inspecteur peut tabler sur sa culpabilité avec moins de 5% de chances de se tromper.

10.2 Variables aléatoires discrètes

Corrigé de l'exercice 2.6.4. Un composant électronique a une durée de vie X qu'on mesure en nombre entier d'unités de temps. On fait l'hypothèse que, à chaque unité de temps, ce composant a une probabilité $p \in]0, 1[$ de tomber en panne, de sorte que

$X \sim \mathcal{G}eo(p)$. On considère un autre composant dont la durée de vie Y est indépendante de X et de même loi. On pose

$$S = \min(X, Y) \text{ et } T = |X - Y|.$$

1. **Que représentent S et T ?**
   S représente le premier temps de panne et T la durée qui sépare les deux temps de panne.
2. **Calculer $\mathbb{P}(S = s$ et $T = t)$ pour $s \geq 1$ et $t \geq 0$ (distinguer $t = 0$ et $t \geq 1$).**
   Pour $s \geq 1$ et $t \geq 0$, on a toujours

$$\{S = s, T = t\} = \{X = s, Y = s + t\} \cup \{X = s + t, Y = s\}$$

   mais l'union n'est disjointe que si $t \geq 1$ (sinon les deux événements sont égaux à $\{X = Y = s\}$). Donc pour $s, t \geq 1$, on a, en utilisant l'indépendance de X et Y,

$$\begin{aligned}
\mathbb{P}(S = s \text{ et } T = t) &= \mathbb{P}(X = s \text{ et } Y = s + t) + \mathbb{P}(Y = s \text{ et } X = s + t) \\
&= \mathbb{P}(X = s)\mathbb{P}(Y = s + t) + \mathbb{P}(Y = s)\mathbb{P}(X = s + t) \\
&= 2p^2(1 - p)^{2s+t-2}.
\end{aligned}$$

   Pour $s \geq 1$ et $t = 0$, on a, toujours par indépendance de X et Y,

$$\begin{aligned}
\mathbb{P}(S = s \text{ et } T = 0) &= \mathbb{P}(X = s \text{ et } Y = s) \\
&= \mathbb{P}(X = s)\mathbb{P}(Y = s) = p^2(1 - p)^{2s-2}.
\end{aligned}$$

   On conclut que

$$\forall s \geq 1,\ \forall t \geq 0,\ \mathbb{P}(S = s, T = t) = p^2 \left(1 + 1_{\{t>0\}}\right) (1 - p)^t (1 - p)^{2(s-1)}. \tag{10.1}$$

3. **En déduire les lois de S et T puis $\mathbb{E}(T)$. Quel est le nom de la loi de S ?**
   Pour $s \geq 1$, on a par la formule des lois marginales

$$\begin{aligned}
\mathbb{P}(S = s) &= \sum_{t=0}^{\infty} \mathbb{P}(S = s \text{ et } T = t) \\
&= p^2(1 - p)^{2s-2} + \sum_{t=1}^{\infty} 2p^2(1 - p)^{2s+t-2} \\
&= (1 - p)^{2(s-1)} \left(p^2 + 2p^2(1 - p)\sum_{u=0}^{\infty}(1 - p)^u\right) \text{ où } u = t - 1 \\
&= \left(p^2 + 2p^2(1 - p)\frac{1}{p}\right) (1 - p(2 - p))^{s-1} = p(2 - p)(1 - p(2 - p))^{s-1}.
\end{aligned}$$

   On reconnaît la loi géométrique de paramètre $p(2 - p)$.
   Maintenant, toujours par la formule des lois marginales, pour $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(T = t) &= \sum_{s=1}^{\infty} \mathbb{P}(S = s \text{ et } T = t) = \sum_{s=1}^{\infty} \left(1 + 1_{\{t>0\}}\right) p^2(1 - p)^{2(s-1)+t} \\
&= \left(1 + 1_{\{t>0\}}\right) p^2(1 - p)^t \sum_{u=0}^{\infty}(1 - p(2 - p))^u = \left(1 + 1_{\{t>0\}}\right) \frac{p(1 - p)^t}{2 - p}.
\end{aligned}$$

D'où

$$\begin{aligned}\mathbb{E}(T) &= \sum_{t=0}^{\infty} t\mathbb{P}(T=t) = \sum_{t=1}^{\infty} t\frac{2p(1-p)^t}{2-p} \\ &= \frac{2p(1-p)}{2-p}\sum_{t=1}^{\infty} t(1-p)^{t-1} = \frac{2p(1-p)}{2-p} f'(1-p),\end{aligned}$$

où $f(x) = \sum_{t=0}^{\infty} x^t = \frac{1}{1-x}$. Comme $f'(x) = \frac{1}{(1-x)^2}$, $f'(1-p) = 1/p^2$ et $\mathbb{E}(T) = \frac{2(1-p)}{p(2-p)}$.

4. **Les variables aléatoires S et T sont-elles indépendantes ?**
   On peut montrer l'indépendance en utilisant la définition. En effet, on a pour $s \geq 1$ et $t \geq 0$,

$$\begin{aligned}\mathbb{P}(S=s)\mathbb{P}(T=t) &= p(2-p)[(1-p)^2]^{s-1}\left(1+1_{\{t>0\}}\right)\frac{p(1-p)^t}{2-p} \\ &= p^2\left(1+1_{\{t>0\}}\right)(1-p)^{t+2(s-1)} = \mathbb{P}(S=s \text{ et } T=t),\end{aligned}$$

   d'après (10.1). On peut aussi remarquer sur (10.1) que la loi du couple (S,T) se met sous forme produit $c\mu(s)\nu(t)$ et conclure que ces variables sont indépendantes en utilisant la remarque 2.2.12.

Corrigé du problème 2.6.12. Soit $A_1, \cdots, A_n$ une suite d'événements.

1. **Montrer que $1_{A_1\cup\ldots\cup A_n} = 1 - \prod_{i=1}^n (1-1_{A_i})$.**

$$1_{A_1\cup\ldots\cup A_n} = 1 - 1_{(A_1\cup\ldots\cup A_n)^c} = 1 - 1_{A_1^c\cap\ldots\cap A_n^c} = 1 - \prod_{i=1}^n 1_{A_i^c} = 1 - \prod_{i=1}^n (1-1_{A_i}).$$

2. **En déduire la formule de Poincaré :**

$$\mathbb{P}(A_1\cup\ldots\cup A_n) = \sum_{k=1}^n (-1)^{k+1} \sum_{1\leq i_1<\ldots<i_k\leq n} \mathbb{P}(A_{i_1}\cap\ldots\cap A_{i_k}).$$

   En écrivant,

$$\prod_{i=1}^n (1-1_{A_i}) = \sum_{k=0}^n (-1)^k \sum_{1\leq i_1<\ldots<i_k\leq n} 1_{A_{i_1}}\ldots 1_{A_{i_k}},$$

   et en prenant l'espérance, on obtient la formule de Poincaré.

   Application : Une personne écrit à n correspondants des lettres personnelles, met chaque lettre dans une enveloppe, ferme les enveloppe puis écrit les adresses au hasard. On s'intéresse au nombre X_n de lettres qui parviennent à leur destinataire. Pour $1 \leq i \leq n$, on note A_i l'événement : la lettre i arrive à son destinataire.

3. **Préciser l'espace fini Ω choisi pour modéliser le problème ainsi que la probabilité $\mathbb{P}$ dont il est muni.** On suppose que pour $i \in \{1,\ldots,n\}$, le i-ième correspondant est celui pour lequel i-ième lettre a été écrite. On considère $\Omega = \mathcal{S}_n$, l'espace des permutations de $\{1,\cdots,n\}$. Pour $\sigma \in \Omega$ et $i \in \{1,\ldots,n\}$, $\sigma(i)$ désigne le numéro de la lettre reçue par le i-ième correspondant. On munit Ω de la probabilité uniforme. On a Card $(\Omega) = n!$.
   Pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, calculer $\mathbb{P}(A_{i_1}\cap\ldots\cap A_{i_k})$.
   $A_{i_1}\cap\ldots\cap A_{i_k}$ représente l'ensemble des permutations σ telles que pour $1 \leq j \leq k$, $\sigma(i_j) = i_j$. On a Card $(A_{i_1}\cap\ldots\cap A_{i_k}) = (n-k)!$ et $\mathbb{P}(A_{i_1}\cap\ldots\cap A_{i_k}) = \dfrac{(n-k)!}{n!}$.

4. **En déduire $\mathbb{P}(X_n > 0)$ puis $\mathbb{P}(X_n = 0)$. Quel est le nombre de permutations σ de $\{1, \cdots, n\}$ sans point fixe i.e. telles que $i \in \{1, \cdots, n\}$, $\sigma(i) \neq i$?**

L'événement au moins une lettre parvient à son destinataire est $\{X_n > 0\} = A_1 \cup \ldots \cup A_n$. En utilisant la formule de Poincaré, on obtient

$$\begin{aligned}\mathbb{P}(X_n > 0) = \mathbb{P}(A_1 \cup \ldots \cup A_n) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \\ &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!},\end{aligned}$$

puisque $\mathrm{Card}\left(\{(i_1, \ldots, i_k) : 1 \leq i_1 < \ldots < i_k \leq n\}\right) = \binom{n}{k}$. Donc

$$\mathbb{P}(X_n = 0) = 1 - \mathbb{P}(X_n > 0) = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Le nombre de permutations σ de $\{1, \cdots, n\}$ sans point fixe est égal à

$$\mathrm{Card}\left(\{X_n = 0\}\right) = n!\mathbb{P}(X_n = 0) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

5. **En déduire que pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,**
$\mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \dfrac{(n-k)!}{n!} \displaystyle\sum_{l=0}^{n-k} \frac{(-1)^l}{l!}$ **puis donner la loi de X_n.**

L'application qui à $\sigma \in \mathcal{S}_n$ associe sa restriction à $\mathcal{T} = \{1, \ldots, n\} \backslash \{i_1, \ldots, i_k\}$ établit une bijection entre $A_{i_1} \cap \ldots \cap A_{i_k}$ et l'ensemble des permutations de $\mathcal{T}$. L'image de $\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}$ par cette restriction est l'ensemble des permutations sans point fixe de $\mathcal{T}$ dont le cardinal est $(n-k)! \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}$ d'après la question précédente. Donc

$$\mathrm{Card}\left(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}\right) = (n-k)! \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}$$

$$\text{puis} \quad \mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}.$$

Comme $\{X_n = k\} = \displaystyle\bigcup_{1 \leq i_1 < \ldots < i_k \leq n} \{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}$ où l'union est disjointe,

$$\begin{aligned}\mathbb{P}(X_n = k) &= \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) \\ &= \binom{n}{k} \frac{(n-k)!}{n!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!} = \frac{1}{k!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}.\end{aligned}$$

6. **Vérifier que $k\mathbb{P}(X_n = k) = \mathbb{P}(X_n = k-1) - \dfrac{(-1)^{n-(k-1)}}{(k-1)!(n-(k-1))!}$ pour k dans $\{1, \ldots, n\}$. En déduire que $\mathbb{E}(X_n) = 1$.**

Il est facile de vérifier que

$$\begin{aligned} k\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k-1) &= \frac{1}{(k-1)!}\sum_{l=0}^{n-k}\frac{(-1)^l}{l!} - \frac{1}{(k-1)!}\sum_{l=0}^{n-k+1}\frac{(-1)^l}{l!} \\ &= -\frac{(-1)^{n-k+1}}{(k-1)!(n-k+1)!}. \end{aligned}$$

Pour l'espérance de X_n, on a

$$\begin{aligned} \mathbb{E}(X_n) &= \sum_{k=0}^{n} k\mathbb{P}(X_n = k) = \sum_{k=1}^{n} k\mathbb{P}(X_n = k) \\ &= \sum_{k=1}^{n}\mathbb{P}(X_n = k-1) - \sum_{k=1}^{n}\frac{(-1)^{n-(k-1)}}{(k-1)!(n-(k-1))!} \\ &= 1 - \mathbb{P}(X_n = n) - \sum_{l=0}^{n-1}\frac{(-1)^{n-l}}{l!(n-l)!} = 1 - \frac{1}{n!}\sum_{l=0}^{n}\binom{n}{l}(-1)^{n-l} = 1. \end{aligned}$$

Comme $X_n = \sum_{i=1}^n 1_{A_i}$, par linéarité de l'espérance,

$$\mathbb{E}(X_n) = \sum_{i=1}^{n}\mathbb{P}(A_i) = n \times \frac{1}{n} = 1.$$

7. **Soit X une variable de loi de Poisson de paramètre 1. Vérifier que pour tout entier k, $\lim_{n\to\infty}\mathbb{P}(X_n = k) = \mathbb{P}(X = k)$.**
   Soit $k \in \mathbb{N}$,
   $$\lim_{n\to\infty}\mathbb{P}(X_n = k) = \lim_{n\to\infty}\frac{1}{k!}\sum_{l=0}^{n-k}\frac{(-1)^l}{l!} = \frac{e^{-1}}{k!} = \mathbb{P}(X = k).$$

Corrigé du problème 2.6.13. Pour fidéliser ses clients, une marque de chocolat place dans chaque tablette une pièce d'un puzzle. Le puzzle est composé de n morceaux distincts. Le morceau qui se trouve dans une tablette est supposé suivre une loi uniforme sur les n morceaux possibles. Le client est supposé acheter les différentes tablettes au hasard. On s'intéresse au nombre N d'achats à réaliser pour obtenir l'ensemble des morceaux du puzzle.

Événements

1. **Pour $0 \leq m \leq n-1$, que vaut $\mathbb{P}(N > m)$?**
   Il est clair qu'il faut au minimum n achats pour obtenir l'ensemble des pièces du puzzle. Ainsi, nécessairement $N \geq n$ et donc pour $m \leq n-1$, $\mathbb{P}(N > m) = 1$.
2. **On suppose maintenant $m \geq n$.**
   (a) **Pour $1 \leq k \leq n$, on désigne par A_k^m l'événement "la pièce k n'a toujours pas été obtenue au bout de m achats". Calculer pour $k_1, \ldots, k_r \in \{1, \ldots, n\}$ des entiers distincts, la probabilité $\mathbb{P}(A_{k_1}^m \cap \ldots \cap A_{k_r}^m)$.**
   On modélise l'achat des m tablettes par la suite $\omega = (\omega_1, \ldots, \omega_m) \in \{1, \ldots, n\}^m$ des numéros des tablettes obtenues. On ne dispose d'aucun élément pour affirmer qu'une configuration ω particulière est plus probable qu'une autre ; on munit donc $\Omega = \{1, \ldots, n\}^m$ de la probabilité uniforme $\mathbb{P}$. Une configuration ω

est dans $A_{k_1}^m \cap \ldots \cap A_{k_r}^m$ si et seulement si au bout de m achats, on n'a toujours pas obtenu les pièces $k_1, \ldots, k_r$.
Il est donc clair que $A_{k_1}^m \cap \ldots \cap A_{k_r}^m = (\{1, \ldots, n\} \setminus \{k_1, \ldots, k_r\})^m$. Par conséquent :

$$\begin{aligned}\mathbb{P}\left(A_{k_1}^m \cap \ldots \cap A_{k_r}^m\right) &= \frac{\text{Card }(\{1, \ldots, n\} \setminus \{k_1, \ldots, k_r\})^m}{\text{Card }(\Omega)} \\ &= \frac{(n-r)^m}{n^m} = \left(1 - \frac{r}{n}\right)^m\end{aligned}$$

(b) **Montrer que $\{N > m\} = \bigcup_{k=1}^{n} A_k^m$.**
Soit maintenant $\omega \in \{N > m\}$. Pour cette configuration ω particulière, toutes les pièces n'ont pas encore été obtenues au bout de m achats. Ceci signifie également qu'au bout de m achats, il manque encore au moins une pièce.

(c) **En déduire que $\mathbb{P}(N > m) = \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \left(1 - \frac{r}{n}\right)^m$.**
Par application de la formule de Poincaré, on peut alors calculer :

$$\begin{aligned}\mathbb{P}(N > m) &= \mathbb{P}\left(\bigcup_{k=1}^{n} A_k^m\right) \\ &= \sum_{r=1}^{n} (-1)^{r+1} \sum_{\mathcal{K} \subset \{1,\ldots,n\}, \text{Card }(\mathcal{K}) = r} \mathbb{P}\left(\bigcap_{j \in \mathcal{K}} A_j^m\right).\end{aligned}$$

Or, d'après 2.(a), $\mathbb{P}\left(\bigcap_{j \in \mathcal{K}} A_j^m\right) = (1 - r/n)^m$ dès lors que Card $(\mathcal{K}) = r$. Comme il y a exactement $\binom{n}{r}$ parties $\mathcal{K}$ de cardinal r dans $\{1, \ldots, n\}$, la formule proposée est démontrée.

3. **En remarquant que $\forall l \in \mathbb{N}$, $l = \sum_{m=0}^{+\infty} 1_{\{l > m\}}$, m.q. $\mathbb{E}(N) = \sum_{m=0}^{+\infty} \mathbb{P}(N > m)$. En déduire une expression pour cette espérance.**
En appliquant à $N(\omega)$ la décomposition de l'énoncé, il vient :

$$N(\omega) = \sum_{m=0}^{+\infty} 1_{\{N(\omega) > m\}}$$

On calcule l'espérance des deux membres et, tout les termes étant positifs, on permute l'espérance et la somme :

$$\mathbb{E}(N) = \mathbb{E}\left(\sum_{m=0}^{+\infty} 1_{\{N > m\}}\right) = \sum_{m=0}^{+\infty} \mathbb{E}\left(1_{\{N > m\}}\right) = \sum_{m=0}^{+\infty} \mathbb{P}(N > m).$$

En utilisant les deux expressions de $\mathbb{P}(N > m)$ suivant que $m \leq n-1$ (voir 1.) ou que $m \geq n$ (voir 2.(c)), on obtient :

$$\begin{aligned}\mathbb{E}(N) = n + \sum_{m=n}^{+\infty} \left[\sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \left(1 - \frac{r}{n}\right)^m\right] &= n + \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \left[\sum_{m=n}^{+\infty} \left(1 - \frac{r}{n}\right)^m\right] \\ = n + \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \frac{n}{r} \left(1 - \frac{r}{n}\right)^n.\end{aligned}$$

4. **Donner la loi de N.**
La variable aléatoire N prend ses valeurs dans $\{n, n+1, \dots\}$. Pour $m \geq n$, on a :
$$\mathbb{P}(N = m) = \mathbb{P}(N > m-1) - \mathbb{P}(N > m).$$
On calcule cette valeur en distinguant suivant que $m \geq n+1$ ou $m = n$. Dans le premier cas, on utilise l'expression en 2.(c) pour obtenir :
$$\mathbb{P}(N = m) = \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \frac{r}{n} \left(1 - \frac{r}{n}\right)^{m-1}.$$
Dans le cas où $m = n$, on utilise 1. pour écrire que $\mathbb{P}(N > n-1) = 1$ et donc $\mathbb{P}(N = n) = 1 - \mathbb{P}(N > n)$. Ainsi,
$$\mathbb{P}(N = n) = -\sum_{r=0}^{n} (-1)^{r+1} \binom{n}{r} \left(1 - \frac{r}{n}\right)^{n}.$$

Variables aléatoires

La première pièce ayant été découverte dans la première tablette ($N_1 = 1$), on note N_2 le nombre d'achats supplémentaires nécessaires à l'obtention d'une seconde pièce, puis N_3 le nombre d'achats supplémentaires nécessaires à l'obtention d'une troisième, et ainsi de suite.

1. **Exprimer N en fonction de $N_1, N_2, \dots, N_n$.**
Il est clair que pour obtenir toutes les pièces du puzzle, il faut obtenir la première, puis la seconde, puis la troisième, etc. Ainsi, $N = N_1 + N_2 + \cdots + N_n$.
2. **On note X_i le numéro de la i-ème pièce de puzzle obtenue. Pour $m \geq 1$, montrer que $\{N_2 = m\} = \bigcup_{i=1}^{n} \{X_1 = i, X_2 = i, \dots, X_m = i, X_{m+1} \neq i\}$.**
Soit $\omega \in \{N_2 = m\}$ et soit i le numéro de la première pièce obtenue i.e. $X_1(\omega)$. Dire que $N_2(\omega) = m$, c'est dire qu'il a fallu m pièces supplémentaires pour obtenir une pièce différente de la première. C'est donc dire que $X_2(\omega) = \ldots = X_m(\omega) = i$ mais que $X_{m+1}(\omega) \neq i$. D'où le résultat.
3. **Par hypothèse, les X_i sont des variables aléatoires indépendantes, de loi uniforme sur $\{1, \dots, n\}$. En déduire la loi de N_2.**
Les différents éléments dans la réunion ci-dessus sont disjoints car ils correspondent à des valeurs distinctes de X_1. Aussi,
$$\begin{aligned} \mathbb{P}(N_2 = m) &= \sum_{i=1}^{n} \mathbb{P}(X_1 = i, X_2 = i, \dots, X_m = i, X_{m+1} \neq i) \\ &= \sum_{i=1}^{n} \mathbb{P}(X_1 = i)\, \mathbb{P}(X_2 = i) \ldots \mathbb{P}(X_m = i)\, \mathbb{P}(X_{m+1} \neq i) \end{aligned}$$
On a utilisé dans la dernière égalité l'indépendance des variables X_i. Comme elles suivent une même loi uniforme sur $\{1, \dots, n\}$, il est clair que :
$$\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = i) = \ldots = \mathbb{P}(X_m = i) = \frac{1}{n}, \qquad \mathbb{P}(X_{m+1} \neq i) = \frac{n-1}{n}.$$
Par conséquent,
$$\mathbb{P}(N_2 = m) = \sum_{i=1}^{n} \frac{1}{n^m} \times \frac{n-1}{n} = \left(1 - \frac{1}{n}\right) \times \left(\frac{1}{n}\right)^{m-1}.$$
D'après le cours, ceci signifie que N_2 suit la loi géométrique de paramètre $1 - 1/n$.

4. **Pour $m_2, m_3 \geq 1$, exprimer l'événement $\{N_2 = m_2, N_3 = m_3\}$ en fonction des X_i. Calculer sa probabilité et en déduire l'indépendance de N_2 et N_3.**
En examinant les numéros des deux premières pièces différentes obtenues comme ci-dessus, on peut écrire :

$$\{N_2 = m_2, N_3 = m_3\} = \bigcup_{1 \leq i \neq j \leq n} \{X_1 = i, X_2 = i, \ldots, X_{m_2} = i, X_{m_2+1} = j,$$
$$X_{m_2+2} \in \{i,j\}, \ldots, X_{m_2+m_3} \in \{i,j\}, X_{m_2+m_3+1} \notin \{i,j\}\}$$

On en déduit en utilisant le fait que les événements ci-dessus sont disjoints puis que les X_i sont indépendantes que :

$$\begin{aligned}\mathbb{P}(N_2 = m_2, N_3 = m_3) &= n(n-1)\frac{1}{n^{m_2+1}}\left(\frac{2}{n}\right)^{m_3-1}\left(\frac{n-2}{n}\right) \\ &= \left(1 - \frac{1}{n}\right) \times \left(\frac{1}{n}\right)^{m_2-1} \times \left(1 - \frac{2}{n}\right) \times \left(\frac{2}{n}\right)^{m_3-1}\end{aligned}$$

Ceci étant vrai pour tous $m_2, m_3 \geq 1$, cette égalité prouve d'une part l'indépendance de N_2 et N_3 (l'expression est le produit d'une fonction de m_2 par une fonction de m_3) et d'autre part que N_3 suit la loi géométrique de paramètre $1 - 2/n$.

5. **Justifier intuitivement que les N_i sont des variables de loi géométrique et donner leurs paramètres.**
Le raisonnement intuitif est le suivant : une fois que les $(k-1)$ premières pièces ont été obtenues, on achète des tablettes jusqu'à obtenir une nouvelle pièce. À chaque achat, on a une probabilité $(k-1)/n$ d'obtenir une pièce que l'on possédait déjà et une probabilité $(1-(k-1)/n)$ d'en obtenir une que l'on ne possédait pas. Les achats successifs sont indépendants. On reconnaît ainsi le temps de premier succès dans une suite d'expériences indépendantes avec probabilité de succès $(1-(k-1)/n)$, qui suit la loi géométrique de paramètre cette probabilité. Ainsi, N_k est une variable de loi géométrique de paramètre $(1 - (k-1)/n)$.

6. **En déduire une autre expression de l'espérance de N.**
On sait que l'espérance d'une variable géométrique est l'inverse de son paramètre. Ainsi, $\mathbb{E}(N_k) = \frac{n}{n-k+1}$ et donc d'après 1., il vient :

$$\mathbb{E}(N) = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{i=1}^{n} \frac{1}{i}.$$

7. **Conclure que $\sum_{r=1}^{n-1} (-1)^{r+1} \binom{n}{r} \frac{1}{r} \left(1 - \frac{r}{n}\right)^n = \sum_{i=2}^{n} \frac{1}{i}$.**
En confrontant l'expression obtenue à la question précédente avec celle obtenue au 3. de la première partie, on peut écrire :

$$n + n \sum_{r=1}^{n-1} (-1)^{r+1} \binom{n}{r} \frac{1}{r} \left(1 - \frac{1}{n}\right)^n = n \sum_{i=1}^{n} \frac{1}{i} = n + n \sum_{i=2}^{n} \frac{1}{i}.$$

D'où le résultat annoncé.

10.3 Variables aléatoires à densité

Corrigé de l'exercice 3.5.8. On coupe un bâton de longueur 1 au hasard en trois morceaux : les abscisses U et V des découpes sont supposées indépendantes et uniformément réparties sur $[0,1]$. On veut calculer la probabilité p pour que l'on puisse faire un triangle avec les trois morceaux (on peut faire un triangle avec trois segments de longueur l_1, l_2 et l_3 si et seulement si $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$).

1. **Exprimer en fonction de U et V les longueurs respectives L_1, L_2 et L_3 du morceau le plus à gauche, du morceau du milieu et du morceau le plus à droite.**

   Il est clair que $L_1 = \inf(U,V)$, $L_2 = \sup(U,V) - \inf(U,V)$ et $L_3 = 1 - \sup(U,V)$.

2. **Montrer que**
$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$
   En remarquant que lorsque $l_1 + l_2 + l_3 = 1$, la condition $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$ est équivalente à $l_1 \leq \frac{1}{2}$, $l_2 \leq \frac{1}{2}$ et $l_3 \leq \frac{1}{2}$, on en déduit que p est égale à la probabilité de $\left\{L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right\}$. En décomposant cet événement sur la partition $\{U < V\}$, $\{V < U\}$ et $\{U = V\}$ et en utilisant l'égalité $\mathbb{P}(U = V) = 0$, on obtient
$$p = \mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right) + \mathbb{P}\left(V \leq U, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$
   Comme les couples des variables (U,V) et (V,U) ont même loi, les deux probabilités à droite de l'égalité sont égales. Donc
$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$

3. **Calculer p.**
$$\begin{aligned}
p &= 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right)\\
&= 2\mathbb{P}\left(U \leq V, U \leq \frac{1}{2}, V - U \leq \frac{1}{2}, V \geq \frac{1}{2}\right)\\
&= 2\int\int 1_{\{x\leq y, x\leq\frac{1}{2}, y-x\leq\frac{1}{2}, y\geq\frac{1}{2}\}} 1_{[0,1]}(x) 1_{[0,1]}(y)\,dxdy\\
&= 2\int_0^{\frac{1}{2}}\left(\int_{\frac{1}{2}}^{x+\frac{1}{2}} dy\right)dx = 2\int_0^{\frac{1}{2}} x\,dx = \frac{1}{4}.
\end{aligned}$$

Corrigé de l'exercice 3.5.9. Soient X et Y deux variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi de Pareto de paramètre 2 i.e. qui possèdent la densité $p(x) = \frac{1_{\{x>1\}}}{x^2}$. On pose
$$(Z,W) = \left(\ln(X), 1 + \frac{\ln(Y)}{\ln(X)}\right).$$

1. **Quelle est la loi de (Z, W) ? Les variables Z et W sont-elles indépendantes ?**
On se donne une fonction muette $f : \mathbb{R}^2 \to \mathbb{R}$ bornée. On a

$$\mathbb{E}(f(Z,W)) = \mathbb{E}(f(\ln(X), 1 + \ln(Y)/\ln(X))).$$

Comme X et Y sont indépendantes, (X, Y) a pour densité $p(x)p(y)$. Donc

$$\mathbb{E}(f(Z,W)) = \int_{]1,+\infty[^2} f(\ln(x), 1 + \ln(y)/\ln(x)) \frac{dx}{x^2}\frac{dy}{y^2}.$$

Pour $(x, y) \in]1, +\infty[^2$, on pose $\varphi(x, y) = (z, w) = (\ln(x), 1 + \ln(y)/\ln(x))$. Comme $\ln$ est une fonction injective et positive sur $]1, +\infty[$, φ est injective sur $]1, +\infty[^2$ et $\varphi(]1, +\infty[^2) \subset]0, +\infty[\times]1, +\infty[$. Soit $(z, w) \in]0, +\infty[\times]1, +\infty[$,

$$\begin{cases} z = \ln(x) \\ w = 1 + \ln(y)/\ln(x) \end{cases} \Leftrightarrow \begin{cases} \ln(x) = z \\ \ln(y) = z(w-1) \end{cases} \Leftrightarrow \begin{cases} x = e^z \\ y = e^{z(w-1)} \end{cases}.$$

Comme $(e^z, e^{z(w-1)}) \in]1, +\infty[^2$ on en déduit que $\varphi(]1, +\infty[^2) =]0, +\infty[\times]1, +\infty[$ et que $(x, y) = \varphi^{-1}(z, w) = (e^z, e^{z(w-1)})$. Comme

$$\frac{D(x,y)}{D(z,w)} = \begin{pmatrix} e^z & 0 \\ (w-1)e^{z(w-1)} & ze^{z(w-1)} \end{pmatrix},$$

en calculant la valeur absolue du déterminant de cette matrice, il vient $dxdy = ze^{zw}dzdw$. Donc en appliquant la formule de changement de variables, on obtient

$$\mathbb{E}(f(Z,W)) = \int_{]0,+\infty[\times]1,+\infty[} f(z,w) \frac{ze^{zw}}{e^{2z}e^{2z(w-1)}} dzdw.$$

Donc (Z, W) possède la densité $1_{\{z>0\}}1_{\{w>1\}}ze^{-zw}$. Comme cette densité ne peut pas s'écrire comme le produit d'une fonction de z et d'une fonction de w, les variables Z et W ne sont pas indépendantes.

2. **Quelle est la loi de W ?**
Par la formule des densités marginales, W a pour densité

$$\begin{aligned}\int_{\mathbb{R}} 1_{\{z>0\}}1_{\{w>1\}}ze^{-zw}dz &= 1_{\{w>1\}} \int_0^{+\infty} ze^{-zw}dz \\ &= 1_{\{w>1\}} \left[z \times \frac{-1}{w}e^{-zw} \right]_{z=0}^{+\infty} + \frac{1_{\{w>1\}}}{w} \int_0^{+\infty} e^{-zw}dz \\ &= 0 + \frac{1_{\{w>1\}}}{w} \left[\frac{-1}{w}e^{-zw} \right]_{z=0}^{+\infty} = \frac{1_{\{w>1\}}}{w^2}.\end{aligned}$$

Donc W suit la loi de Pareto de paramètre 2.

3. **Quelle est la loi de Z ?**
Toujours par la formule des densités marginales, Z a pour densité

$$\int_{\mathbb{R}} 1_{\{z>0\}}1_{\{w>1\}}ze^{-zw}dw = 1_{\{z>0\}} \left[-e^{-zw} \right]_{w=1}^{+\infty} = 1_{\{z>0\}}e^{-z}.$$

Donc $Z \sim \mathcal{E}(1)$.

Corrigé de l'exercice 3.5.10. Soient X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\Gamma(p,1)$ et $\Gamma(q,1)$, $p>0$ et $q>0$.

1. **Déterminer l'espérance et la variance de X.**

$$\mathbb{E}(X)=\frac{1}{\Gamma(p)}\int_0^\infty x^p\exp(-x)dx=\frac{\Gamma(p+1)}{\Gamma(p)}=p.$$

$$\mathbb{E}(X^2)=\frac{1}{\Gamma(p)}\int_0^\infty x^{p+1}\exp(-x)dx=\frac{\Gamma(p+2)}{\Gamma(p)}=p(p+1).$$

Donc $\mathrm{Var}(X)=\mathbb{E}(X^2)-(\mathbb{E}(X))^2=p$.

2. On veut calculer la loi de la variable aléatoire $Z=\dfrac{X}{Y}$. On dit que Z suit la loi bêta de seconde espèce de paramètres p et q.

(a) **On pose $W=X-Y$. Exprimer en fonction de X et Y l'événement $\{Z<1, W>0\}$. En déduire sans calcul sa probabilité. Les variables Z et W sont-elles indépendantes ?**

Il est clair que $\{Z<1,W>0\}=\{X<Y,X>Y\}=\emptyset$. La probabilité de cet événement est nulle. Comme il est facile de vérifier que $\mathbb{P}(Z<1)=\mathbb{P}(X<Y)$ et $\mathbb{P}(W>0)=\mathbb{P}(Y<X)$ sont strictement positives, on en déduit que

$$\mathbb{P}(Z<1,W>0)\neq\mathbb{P}(Z<1)\mathbb{P}(W>0).$$

Ceci montre que les deux variables ne sont pas indépendantes.

(b) **Déterminer la loi du vecteur (Z,W).**

Soit $U=\mathbb{R}_+^*\times\mathbb{R}_+^*\setminus\{(x,y)\text{ t.q. }x=y\}$ et $V=]0,1[\times]-\infty,0[\cup]1,+\infty[\times]0,+\infty[$. U et V sont deux ouverts de $\mathbb{R}^2$. On note par φ la l'application de U dans V définie par $\varphi(x,y)=(z,w)=\left(\dfrac{x}{y},x-y\right)$. En résolvant le système d'équations

$$\left\{\begin{array}{rcl} z&=&\frac{x}{y}\\ w&=&x-y\end{array}\right.\Leftrightarrow\left\{\begin{array}{rcl} x&=&yz\\ w&=&y(z-1)\end{array}\right.\Leftrightarrow\left\{\begin{array}{rcl} x&=&\frac{wz}{z-1}\\ y&=&\frac{w}{z-1}\end{array}\right.$$

on déduit que φ est bijective et que $\varphi^{-1}(z,w)=(x,y)=\left(\dfrac{wz}{z-1},\dfrac{w}{z-1}\right)$. On remarque que φ et φ^{-1} sont de classe C^1. On a

$$\frac{D(z,w)}{D(x,y)}=\begin{pmatrix}\frac{1}{y} & -\frac{x}{y^2}\\ 1 & -1\end{pmatrix},$$

ce qui entraîne que $\frac{dzdw}{dxdy}=\frac{|y-x|}{y^2}=\frac{(z-1)^2}{|w|}$. Pour appliquer la méthode de la fonction muette, on considère h une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ bornée,

$$\mathbb{E}\left(h(Z,W)\right)=\mathbb{E}\left(h\left(\frac{X}{Y},X-Y\right)\right)=\int\int_{\mathbb{R}^2}h\left(\frac{x}{y},x-y\right)p_{X,Y}(x,y)dxdy,$$

avec $p_{X,Y}$ la densité du couple (X,Y). Comme les variables X et Y sont indépendantes, on a $p_{X,Y}(x,y)=p_X(x)p_Y(y)$, avec p_X la densité de X et p_Y la densité de Y. Maintenant on écrit

$$\begin{aligned}
\mathbb{E}\left(h(Z,W)\right) &= \int\int_U h\circ\varphi(x,y)p_{X,Y}(x,y)dxdy\\
&= \int\int_V h(z,w)p_{X,Y}\left(\frac{wz}{z-1},\frac{w}{z-1}\right)\frac{|w|}{(z-1)^2}dzdw\\
&= \frac{1}{\Gamma(p)\Gamma(q)}\int\int_V h(z,w)\left(\frac{wz}{z-1}\right)^{p-1}\left(\frac{w}{z-1}\right)^{q-1}e^{-w\frac{z+1}{z-1}}\frac{|w|}{(z-1)^2}dzdw\\
&= \frac{1}{\Gamma(p)\Gamma(q)}\int\int_V h(z,w)\frac{|w|^{p+q-1}\,z^{p-1}}{|z-1|^{p+q}}e^{-|w|\frac{z+1}{|z-1|}}dzdw.
\end{aligned}$$

Ainsi (Z,W) a pour densité

$$p_{Z,W}(z,w)=\frac{1}{\Gamma(p)\Gamma(q)}\frac{|w|^{p+q-1}\,z^{p-1}}{|z-1|^{p+q}}e^{-|w|\frac{z+1}{|z-1|}}1_V(z,w).$$

(c) **En déduire la loi de la variable aléatoire Z.**
La densité marginale de Z est

$$\begin{aligned}
p_Z(z) &= \int_{\mathbb{R}} f_{Z,W}(z,w)dw\\
&= 1_{]0,1[}(z)\frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{|z-1|^{p+q}}\int_{-\infty}^{0}|w|^{p+q-1}\,e^{-|w|\frac{z+1}{|z-1|}}dw\\
&+ 1_{]1,+\infty[}(z)\frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{|z-1|^{p+q}}\int_{0}^{+\infty}|w|^{p+q-1}\,e^{-|w|\frac{z+1}{|z-1|}}dw\\
&= \frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{|z-1|^{p+q}}\int_0^{+\infty}w^{p+q-1}e^{-w\frac{z+1}{|z-1|}}dw \quad \left(\text{on pose } a=w\frac{z+1}{|z-1|}\right)\\
&= \frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{(z+1)^{p+q}}\int_0^{+\infty}a^{p+q-1}e^{-a}da=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{(z+1)^{p+q}}.
\end{aligned}$$

Donc Z possède la densité $p_Z(z)=\dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\dfrac{z^{p-1}}{(z+1)^{p+q}}1_{]0,+\infty[}(z)$.

3. **Déterminer la loi de $T=\dfrac{Z}{1+Z}$.**
Soit $h:\mathbb{R}\to\mathbb{R}$ bornée.

$$\mathbb{E}\left(h(T)\right) \;=\; \mathbb{E}\left(h\left(\frac{Z}{1+Z}\right)\right)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^{+\infty}h\left(\frac{z}{1+z}\right)\frac{z^{p-1}}{(z+1)^{p+q}}dz.$$

On pose $t=\frac{z}{1+z}$, ce qui donne $z=\frac{t}{1-t}=-1+\frac{1}{1-t}$ et $dz=\frac{dt}{(1-t)^2}$. Donc $\mathbb{E}\left(h(T)\right)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^1 h(t)t^{p-1}(1-t)^{q-1}dt$ et T suit la loi béta de paramètres p et q. Comme $T=\frac{X}{X+Y}$, ce résultat est aussi une conséquence de la proposition 3.4.4.

Corrigé du problème 3.5.16. Soient $\Theta\sim\mathcal{U}[-\pi,\pi]$ et R une variable aléatoire réelle indépendante de Θ possédant une densité p nulle sur $\mathbb{R}_-$. On pose

$$(X,Y)=(R\cos\Theta,R\sin\Theta).$$

1. **Déterminer la loi du vecteur (X,Y). Que peut-on dire des lois des vecteurs (X,Y) et (Y,X) ? En déduire que X et Y ont la même densité.**
Comme R et Θ sont indépendantes, le couple (R,Θ) a pour densité $\frac{1}{2\pi}p(r)1_{[-\pi,\pi]}(\theta)$ fonction qui est nulle pour $r \leq 0$. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ bornée.
$$\mathbb{E}(f(X,Y)) = \mathbb{E}(f(R\cos\Theta, R\sin\Theta)) = \int_{]0,+\infty[\times]-\pi,\pi[} f(r\cos\theta, r\sin\theta)\frac{p(r)}{2\pi}drd\theta.$$
A l'aide de l'exercice résolu 3.1.5, on vérifie que le changement de variables $(x,y) = \varphi(r,\theta) = (r\cos\theta, r\sin\theta)$ est une bijection de $]0,+\infty[\times]-\pi,\pi[$ sur $\mathbb{R}^2 \setminus \{\mathbb{R}_- \times \{0\}\}$ de classe C^1 ainsi que son inverse
$$(r,\theta) = \varphi^{-1}(x,y) = \left(\sqrt{x^2+y^2}, 2\arctan\left(\frac{x}{x+\sqrt{x^2+y^2}}\right)\right).$$
$$\text{Comme } \frac{D(x,y)}{D(r,\theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$
$dxdy = r(\cos^2\theta + \sin^2\theta)drd\theta$, ce qui s'écrit aussi $drd\theta = \frac{dxdy}{\sqrt{x^2+y^2}}$. Donc par la formule de changement de variable,
$$\mathbb{E}(f(X,Y)) = \int_{\mathbb{R}^2\setminus\{\mathbb{R}_-\times\{0\}\}} f(x,y)\frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}dxdy.$$
Comme on ne change pas la valeur du membre de droite en rajoutant la demi-droite $\mathbb{R}_- \times \{0\}$ au domaine d'intégration, on conclut que (X,Y) a pour densité $p_{(X,Y)}(x,y) = p(\sqrt{x^2+y^2})/2\pi\sqrt{x^2+y^2}$.
Puisque cette densité est symétrique, pour f de $\mathbb{R}^2$ dans $\mathbb{R}$ bornée,
$$\mathbb{E}(f(Y,X)) = \int_{\mathbb{R}^2} f(y,x)p_{X,Y}(x,y)dxdy = \int_{\mathbb{R}^2} f(y,x)p_{X,Y}(y,x)dxdy.$$
Donc (Y,X) admet $p_{X,Y}$ pour densité et $(Y,X) \stackrel{\mathcal{L}}{=} (X,Y)$. Les images respectives X et Y de ces deux vecteurs par la projection $(z,w) \in \mathbb{R}^2 \to z \in \mathbb{R}$ ont également même loi. Comme d'après la formule des densités marginales X et Y ont une densité, leurs densités sont égales.
2. **Pour $a > 0$, quelle est la loi de $Z = aX/Y$ (on pourra exprimer Z en fonction de Θ) ? En déduire sans calcul la loi de l'inverse d'une variable de Cauchy de paramètre a.**
Comme $Z = a\cot\Theta$, pour f de $\mathbb{R}$ dans $\mathbb{R}$ bornée,
$$\mathbb{E}(f(Z)) = \mathbb{E}(f(a\cot\Theta)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(a\cot\theta)d\theta = \frac{1}{\pi}\int_0^{\pi} f(a\cot\theta)d\theta,$$
en utilisant la périodicité de la fonction cotangente. Si on pose $z = a\cot\theta$, $\frac{dz}{d\theta} = -a(1+\cot^2\theta)$ ce qui implique que $d\theta = \frac{dz}{-a(1+z^2/a^2)}$. On en déduit que
$$\mathbb{E}(f(Z)) = \int_{+\infty}^{-\infty} f(z)\frac{dz}{-a\pi(1+z^2/a^2)} = \int_{\mathbb{R}} f(z)\frac{adz}{\pi(a^2+z^2)}.$$
Donc la variable $Z = aX/Y$ suit la loi de Cauchy de paramètre a. Comme $(X,Y) \stackrel{\mathcal{L}}{=} (Y,X)$, $Y/aX \stackrel{\mathcal{L}}{=} X/aY$. Donc $1/Z = Y/aX \stackrel{\mathcal{L}}{=} \frac{1}{a}X/Y$. D'après ce qui précède, $\frac{1}{a}X/Y \sim \mathcal{C}(1/a)$. Donc la loi de l'inverse d'une variable de Cauchy de paramètre a est la loi de Cauchy de paramètre $1/a$.

3. **Quelle doit être la loi de R pour que le couple (X,Y) soit uniformément distribué sur le disque $D(0,\rho)$ où $\rho > 0$ (i.e. possède une densité constante sur ce disque et nulle en dehors) ?**
Pour que (X,Y) soit uniformément réparti sur le disque $D(0,\rho)$ i.e. possède la densité $1_{\{\sqrt{x^2+y^2}<\rho\}}/\pi\rho^2$, il faut et il suffit que
$$\frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}} = 1_{\{\sqrt{x^2+y^2}<\rho\}}/\pi\rho^2$$
c'est-à-dire que $p(r) = 1_{]0,\rho[}(r)2r/\rho^2$.

4. **On suppose la densité p de R continue sur $\mathbb{R}_+$ et strictement positive sur $\mathbb{R}_+^*$. On souhaite trouver à quelle condition sur p les variables X et Y sont indépendantes.**
   (a) **Dans le cas où X et Y sont indépendantes, si q désigne leur densité commune, exprimer $q(x)q(y)$ en fonction de p.**
   En commençant par choisir $y = 0$ dans cette relation, établir que pour $x, y \in \mathbb{R}^*$,
$$\frac{p(|x|)}{2\pi q(0)|x|} \times \frac{p(|y|)}{2\pi q(0)|y|} = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}. \tag{10.2}$$
   Si X et Y sont indépendantes, $p_{X,Y}(x,y) = q(x)q(y)$, ce qui s'écrit aussi
$$q(x)q(y) = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}.$$
   Le choix $y = 0$ implique alors que $q(x) = p(|x|)/2\pi q(0)|x|$ et on obtient (10.2) en reportant cette expression dans l'égalité précédente.

   (b) **Pour $s > 0$, on pose $f(s) = \ln\left(\frac{p(\sqrt{s})}{2\pi q^2(0)\sqrt{s}}\right)$. Vérifier que f est une fonction continue sur $\mathbb{R}_+^*$ telle que pour $s,t > 0$, $f(s) + f(t) = f(s+t)$.**
   Pour $s,t > 0$, en posant $x = \sqrt{s}$, $y = \sqrt{t}$ et en divisant (10.2) par $q^2(0)$, on obtient
$$\frac{p(\sqrt{s})}{2\pi q^2(0)\sqrt{s}} \times \frac{p(\sqrt{t})}{2\pi q(0)^2\sqrt{t}} = \frac{p(\sqrt{t+s})}{2\pi q^2(0)\sqrt{t+s}}.$$
   Par passage au logarithme, il vient $f(s) + f(t) = f(s+t)$, la continuité de f se déduisant de celle de p.
   En déduire qu'il existe une constante $\sigma > 0$ telle que $p(r) = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}1_{\{r>0\}}$ i.e. que R suit la loi de Rayleigh de paramètre σ^2.
   Par récurrence, on en déduit que pour $k \in \mathbb{N}^*$ et $s_1, \ldots, s_k > 0$, $\sum_{j=1}^k f(s_j) = f(\sum_{j=1}^k s_j)$. Soit $n \in \mathbb{N}^*$. Pour le choix $s_j = \frac{1}{n}$ pour tout $j \geq 1$, on en déduit que pour tout $k \in \mathbb{N}^*$, $kf(1/n) = f(k/n)$. Comme en particulier pour $k = n$, cette relation s'écrit $f(1/n) = f(1)/n$, on en déduit que pour tout $k, n \in \mathbb{N}^*$, $f(k/n) = kf(1)/n$. La densité des rationnels dans $\mathbb{R}_+^*$ et la continuité de f impliquent que pour tout $s > 0$, $f(s) = f(1)s$.
   En utilisant la définition de f, on obtient qu'il existe deux constantes c et a telles que $p(r) = cre^{ar^2}1_{\{r>0\}}$. La condition de sommabilité de p impose que

$a < 0$ i.e. qu'il existe une constante $\sigma > 0$ telle que $a = -1/2\sigma^2$. Puis la condition de normalisation de p impose que $c = 1/\sigma^2$. Ainsi R suit la loi de Rayleigh de paramètre σ^2.

Quelle est alors la loi de (X, Y) ? On a

$$p_{X,Y}(x,y) = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}} = \frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \times \frac{e^{-y^2/2\sigma^2}}{\sigma\sqrt{2\pi}},$$

c'est-à-dire que X et Y sont indépendantes et suivent la loi normale $\mathcal{N}_1(0, \sigma^2)$.

10.4 Simulation

Corrigé de l'exercice 4.3.5. Soit $((X_i, Y_i))_{i\geq 1}$ une suite I.I.D. avec X_1 et Y_1 variables exponentielles de paramètre 1 indépendantes. On se donne également indépendamment de cette suite une variable ε t.q. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$ et on pose

$$N = \inf\{i \geq 1 : \ 2Y_i \geq (1 - X_i)^2\} \ \text{ et } \ Z = \varepsilon X_N.$$

1. **Quelle est la loi de N ? Donner $\mathbb{E}(N)$.**
   Comme temps de premier succès, N suit une loi géométrique de paramètre

$$p = \mathbb{P}(2Y_1 \geq (1-X_1)^2) = \mathbb{E}\left(1_{\{Y_1 \geq \frac{(1-X_1)^2}{2}\}}\right) = \int_0^{+\infty}\int_{(1-x)^2/2}^{+\infty} e^{-y}dy e^{-x}dx$$
$$= \int_0^{+\infty} e^{-\frac{(1-x)^2+2x}{2}}dx = \frac{1}{2\sqrt{e}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}dx = \sqrt{\frac{\pi}{2e}}.$$

2. **Quelle est la loi de Z ?**
   Soit $f : \mathbb{R} \to \mathbb{R}$ bornée. En décomposant sur les valeurs prises par N et ε, puis en utilisant les propriétés d'indépendance, on obtient

$$\mathbb{E}(f(Z)) = \sum_{n\geq 1}\mathbb{E}\left(\left(1_{\{\varepsilon=1\}}f(X_n) + 1_{\{\varepsilon=-1\}}f(-X_n)\right)1_{\{N=n\}}\right) = \sum_{n\geq 1}\mathbb{E}\left(g(X_n)1_{\{N=n\}}\right)$$

   où $g(x) = (f(x) + f(-x))/2$. Or

$$\begin{aligned}
\mathbb{E}\left(g(X_n)1_{\{N=n\}}\right) &= \mathbb{E}\left(\prod_{k=1}^{n-1} 1_{\{Y_k < \frac{(1-X_k)^2}{2}\}} g(X_n) 1_{\{Y_n \geq \frac{(1-X_n)^2}{2}\}}\right) \\
&= (1-p)^{n-1}\int_0^{+\infty}\int_{(1-x)^2/2}^{+\infty} e^{-y}dy\, g(x)e^{-x}dx \\
&= \frac{(1-p)^{n-1}}{2}\int_0^{+\infty}(f(x)+f(-x))e^{-\frac{1+x^2}{2}}dx \\
&= \frac{(1-p)^{n-1}}{2\sqrt{e}}\int_{\mathbb{R}} f(x)e^{-\frac{x^2}{2}}dx = p(1-p)^{n-1}\int_{\mathbb{R}} f(x)e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}
\end{aligned}$$

   On en déduit que

$$\mathbb{E}(f(Z)) = p\int_{\mathbb{R}} f(x)e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}\sum_{n\geq 1}(1-p)^{n-1} = \int_{\mathbb{R}} f(x)e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}$$

   et $Z \sim \mathcal{N}_1(0, 1)$.

3. **En déduire une méthode pour simuler une variable distribuée suivant la loi normale centrée réduite.**
   À partir d'une suite $(U_i)_{i\geq 1}$ de variables I.I.D. suivant la loi uniforme sur $[0,1]$, on pose $\varepsilon = 1_{\{U_1\leq 1/2\}} - 1_{\{U_1>1/2\}}$, $X_i = -\ln(U_{2i})$ et $Y_i = -\ln(U_{2i+1})$ pour $i \geq 1$ et on applique la technique de type rejet qui vient d'être étudiée.

10.5 Convergence et théorèmes limites

Corrigé de l'exercice 5.5.1. On considère une machine qui tourne à une cadence de n cycles par unité de temps et qui a une probabilité $p \in]0,1[$ de tomber en panne lors de chaque cycle et ce indépendamment des autres cycles. Soit N le numéro du cycle où se produit la première panne.

1. **Quelle est la loi de N ? Que représente $T = N/n$?**
   N suit la loi géomètrique de paramètre p. La variable T représente le temps avant la première panne.
   Soit τ l'espérance de T. Exprimer τ en fonction de p et n.
   $\tau = \mathbb{E}(N)/n = 1/pn$.
2. Dans le cas où n est grand, on souhaite approcher la variable discrète T par une variable aléatoire à densité. Pour cela on va effectuer le passage à la limite $n \to +\infty$ en choisissant $p(n)$ de façon à ce que l'espérance de $T(n)$ reste égale à τ.
   (a) **Que vaut $p(n)$?**
   $p(n) = 1/n\tau$.
   (b) **Calculer la fonction caractéristique de $T(n)$.**
   $$\Phi_{T(n)}(u) = \sum_{k\geq 1} p(n)(1-p(n))^{k-1}\exp(iuk/n) = \frac{p(n)\exp(iu/n)}{1-(1-p(n))\exp(iu/n)}.$$
   En remplaçant $p(n)$ par sa valeur $1/n\tau$ on obtient
   $$\Phi_{T(n)}(u) = \frac{1}{1+n\tau(\exp(-iu/n)-1)}.$$
   (c) **En déduire que cette variable converge en loi vers une limite que l'on précisera.**
   Comme $\lim_{n\to+\infty} n(\exp(-iu/n)-1) = -iu$, on a
   $$\forall u \in \mathbb{R},\ \lim_{n\to+\infty} \Phi_{T(n)}(u) = \frac{1}{1-iu\tau} = \frac{1/\tau}{1/\tau - iu}.$$
   On reconnaît la fonction caractéristique de la loi exponentielle de paramètre $1/\tau$. Donc les variables $T(n)$ convergent en loi vers la loi exponentielle de paramètre $1/\tau$.

Corrigé de l'exercice 5.5.4. Chaque jour, dans une certaine ville, 100 personnes ont besoin d'un examen radioscopique. Pour préserver le libre choix, n centres d'imagerie sont installés dans cette ville. On admet que les patients choisissent indifféremment l'un ou l'autre centre d'imagerie.
Soit N le nombre de clients journaliers dans un centre d'imagerie.

1. **Quelle est la probabilité qu'un client choisisse le centre d'imagerie considéré ?**
   Comme on suppose que les patients choisissent indifféremment l'un ou l'autre des n centres d'imagerie, la probabilité qu'un client choisisse le centre d'imagerie considéré est $\frac{1}{n}$.
2. **Montrer que N peut s'écrire $N = \sum_{i=1}^{i=100} X_i$ où les $(X_i)_{1 \leq i \leq n}$ sont n variables indépendantes distribuées suivant la loi de Bernoulli de même paramètre p que l'on déterminera.**
   Soit X_i la variable aléatoire discrète réelle qui vaut 1 si le client i va au centre d'imagerie étudié et 0 sinon. Alors X_i suit la loi de Bernoulli de paramètre p tel que
   $$p = \mathbb{P}(X_i = 1) = \frac{1}{n}.$$
3. **Quelle est la loi de N ?**
   N suit la loi binomiale de paramètres 100 et $\frac{1}{n}$. Son espérance est donc $\mathbb{E}(N) = \frac{100}{n}$, et sa variance $\mathrm{Var}(N) = 100\frac{1}{n}(1 - \frac{1}{n})$.

4. **On donne que si Y suit la loi normale centrée réduite, alors $\mathbb{P}(Y \leq 2) = 98\%$. En utilisant le théorème de la limite centrale, déterminer quelle capacité $c(n)$ chaque centre d'imagerie doit avoir pour être capable de répondre à la demande avec une probabilité de 98% ? Cas où $n = 2$, $n = 3$, $n = 4$?**
   Le Théorème Central Limite dit que
   $$X = \frac{N - \mathbb{E}(N)}{\sqrt{\mathrm{Var}(N)}}$$
   peut être considérée comme une variable normale centrée réduite.
   Or
   $$\{N \leq c\} = \left\{ X \leq \frac{c - \mathbb{E}(N)}{\sqrt{\mathrm{Var}(N)}} \right\}.$$
   et les tables de la loi normale centrée réduite donnent que $\mathbb{P}(X \leq 2) = 98\%$. Donc pour pouvoir faire face à demande avec une probalité de 98%, le centre doit avoir la capacité $c(n)$ donnée par
   $$\frac{c(n) - \mathbb{E}(N)}{\sqrt{\mathrm{Var}(N)}} = 2 \text{ i.e. } c(n) = \mathbb{E}(N) + 2\sqrt{\mathrm{Var}(N)} = \frac{100}{n} + 20\sqrt{\frac{1}{n}\left(1 - \frac{1}{n}\right)}.$$
   On a les résultats suivants :
   $$c(1) = 100 \text{ (heureusement !)}, \quad c(2) = 60, \quad c(3) = 42.8, \quad c(4) = 33.7.$$

5. **Quel est le coût de la concurrence : quelle surcapacité $s(n)$ la concurrence entraîne-t-elle par rapport à une situation où chaque centre se verrait affecter un même nombre de clients ? Cas où $n = 2$, $n = 3$, $n = 4$?**
   La surcapacité est $s(n) = n \times c(n) - 100 = 20 \times \sqrt{n-1}$ soit
   $$s(1) = 0, \quad s(2) = 20, \quad s(3) = 28.3, \quad s(4) = 34.6.$$

Corrigé de l'exercice 5.5.11.

1. **Calculer la fonction caractéristique d'une variable X suivant la loi de Poisson $\mathcal{P}(\lambda)$. Calculer $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.**
Soit $u \in \mathbb{R}$. Par définition, la fonction caractéristique Φ_X de X vaut
$$\Phi_X(u) = \mathbb{E}\left(e^{iuX}\right) = \sum_{k=0}^{\infty} e^{iuk}\mathbb{P}\{X = k\} = \sum_{k=0}^{\infty} e^{iuk} e^{-\lambda}\frac{\lambda^k}{k!} = \exp(\lambda(e^{iu} - 1)).$$
Développons cette expression au voisinage de $u = 0$:
$$e^{iu} - 1 = iu - \frac{1}{2}u^2 + o(u^2).$$
Par conséquent,
$$\begin{aligned}\Phi_X(u) = \exp\{\lambda(e^{iu} - 1)\} &= \exp\{iu\lambda - \frac{\lambda}{2}u^2 + o(u^2)\}\\ &= 1 + \left(iu\lambda - \frac{\lambda}{2}u^2\right) - \frac{\lambda^2}{2}u^2 + o(u^2)\\ &= 1 + iu\lambda - \frac{\lambda + \lambda^2}{2}u^2 + o(u^2).\end{aligned}$$
En identifiant avec le développement limité en $u = 0$ de Φ_X donné par le lemme 5.3.5 ($\Phi_X(u) = 1 + i\mathbb{E}(X)u - \frac{\mathbb{E}(X^2)}{2}u^2 + o(u^2)$), il vient : $\mathbb{E}(X) = \lambda$ et $\mathbb{E}(X^2) = \lambda + \lambda^2$ donc $\mathrm{Var}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$.
2. **Soit $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ deux variables de Poisson indépendantes. Quelle est la loi de leur somme ?**
Les deux variables étant indépendantes, la fonction caractéristique de leur somme est le produit de leurs fonctions caractéristiques. Ainsi, pour tout $u \in \mathbb{R}$, on a
$$\begin{aligned}\Phi_{X+Y}(u) = \Phi_X(u)\Phi_Y(u) &= \exp\{\lambda(e^{iu} - 1)\}\exp\{\mu(e^{iu} - 1)\}\\ &= \exp\{(\lambda + \mu)(e^{iu} - 1)\}.\end{aligned}$$
On reconnaît la fonction caractéristique d'une variable de Poisson de paramètre $\lambda + \mu$. Comme la fonction caractéristique caractérise la loi, on conclut que $X + Y \sim \mathcal{P}(\lambda + \mu)$.
3. **Soit $X_n \sim \mathcal{P}(n)$. Montrer que pour $n \to +\infty$, $Y_n = (X_n - n)/\sqrt{n}$ converge en loi vers une limite que l'on précisera. Commenter ce résultat à la lumière du théorème de la limite centrale.**
Soit $u \in \mathbb{R}$.
$$\begin{aligned}\Phi_{Y_n}(u) &= \exp\left(-iu\sqrt{n}\right)\Phi_{X_n}\left(\frac{u}{\sqrt{n}}\right)\\ &= \exp\left(-iu\sqrt{n}\right)\exp\{n(e^{i\frac{u}{\sqrt{n}}} - 1)\}\\ &= \exp\left(-iu\sqrt{n}\right)\exp\left[n\left(i\frac{u}{\sqrt{n}} - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right)\right] = \exp\left(-\frac{u^2}{2} + o(1)\right).\end{aligned}$$
Ainsi, $\forall u \in \mathbb{R}$, $\lim_{n\to\infty}\Phi_{Y_n}(u) = \exp(-\frac{u^2}{2}) = \Phi_G(u)$ où $G \sim \mathcal{N}_1(0,1)$. Par conséquent, Y_n converge en loi vers G qui suit la loi normale centrée réduite.

Pour expliquer ce résultat, il suffit de considérer une suite $(Z_j)_{j\geq 1}$ de variables I.I.D. de loi $\mathcal{P}(1)$. Il résulte de la deuxième question que $Z_1 + \ldots + Z_n$ suit la loi $\mathcal{P}(n)$ et par conséquent Y_n a même loi que $W_n = \sqrt{n}(\frac{Z_1+\cdots+Z_n}{n} - 1)$. Comme $\mathbb{E}(Z_1) = 1$ et $\mathrm{Var}(Z_1) = 1$, le théorème de la limite centrale implique que la suite $(W_n)_{n\geq 1}$ converge en loi vers G.

4. **Quelle est la limite de la suite $u_n = e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$ lorsque $n \to +\infty$?**
On a

$$u_n = \sum_{k=0}^{n} \mathbb{P}(X_n = k) = \mathbb{P}(X_n \leq n) = \mathbb{P}(Y_n \leq 0) = \mathbb{E}\left(1_{\{Y_n\leq 0\}}\right).$$

La fonction $y \to 1_{\{y\leq 0\}}$ est bornée et continue sur $\mathbb{R}^*$. Comme la limite en loi G de la suite $(Y_n)_n$ est telle que $\mathbb{P}(G = 0)$, on déduit de la remarque 5.3.11 que $\lim_{n\to+\infty} u_n = \mathbb{E}\left(1_{\{G\leq 0\}}\right) = \mathbb{P}(G \leq 0) = \frac{1}{2}$.

Corrigé du problème 5.5.16. Un automobiliste emprunte tous les jours le même trajet qui comporte un feu tricolore pour se rendre à son travail. Comme le trajet est peu encombré, lorsque le feu est rouge, l'automobiliste peut redémarrer dès que le feu passe au vert. Mais pour faire passer le temps, il se demande quelle est la durée θ pendant laquelle le feu reste rouge. On note $(X_i)_{i\geq 1}$ ses durées d'attente successives au feu lorsque celui-ci est rouge et on suppose ces variables I.I.D. suivant la loi uniforme $\mathcal{U}[0, \theta]$. Pour $n \geq 1$, on note $Z_n = \max_{1\leq i\leq n} X_i$.

1. **Montrer que $\mathbb{P}(X_1 = X_2) = 0$.**
Les variables X_1 et X_2 étant indépendantes et à densité, le couple (X_1, X_2) admet une densité qui est le produit des densités de X_1 et de X_2. Ainsi,

$$\mathbb{P}(X_1 = X_2) = \frac{1}{\theta^2}\int_{[0,\theta]^2} 1_{\{x_1=x_2\}}dx_1dx_2 = 0$$

car la droite d'équation $x_1 = x_2$ est de mesure nulle dans le plan.

2. **Montrer que plus généralement $\mathbb{P}(\exists 1 \leq i \neq j \leq n, X_i = X_j) = 0$. En déduire que pour toute fonction f bornée sur $\mathbb{R}$, on a :**

$$\mathbb{E}(f(Z_n)) = \sum_{i=1}^{n} \mathbb{E}\left(f(X_i)1_{\{\forall j\neq i,\ X_j<X_i\}}\right) = n\mathbb{E}\left(f(X_1)1_{\{X_2<X_1,\ldots,X_n<X_1\}}\right).$$

Il suffit de remarquer que comme pour $i \neq j$, $\mathbb{P}(X_i = X_j) = \mathbb{P}(X_1 = X_2) = 0$,

$$\mathbb{P}(\exists 1 \leq i \neq j \leq n, X_i = X_j) = \mathbb{P}\left(\cup_{1\leq i\neq j\leq n}\{X_i = X_j\}\right) \leq \sum_{1\leq i\neq j\leq n} \mathbb{P}(X_i = X_j) = 0.$$

Par suite, comme $\{\exists 1 \leq i \neq j \leq n, X_i = X_j\}^c = \{\forall 1 \leq i \neq j \leq n, X_i \neq X_j\}$ est un événement de probabilité 1, on peut écrire :

$$\mathbb{E}(f(Z_n)) = \mathbb{E}(f(Z_n)1_{\{\forall 1\leq i\neq j\leq n, X_i\neq X_j\}}) = \sum_{k=1}^{n} \mathbb{E}\left(f(Z_n)1_{\{\forall 1\leq i\neq j\leq n, X_i\neq X_j\}}1_{\{Z_n=X_k\}}\right)$$

où on a décomposé suivant l'indice k tel que X_k est le maximum des valeurs distinctes $X_1, \ldots, X_n$. Examinons l'un des termes de cette somme :

$$\mathbb{E}\left(f(Z_n)1_{\{\forall 1\leq i\neq j\leq n, X_i\neq X_j\}}1_{\{Z_n=X_k\}}\right) = \mathbb{E}\left(f(X_k)1_{\{\forall i\neq k, X_i<X_k\}}\right)$$

car l'événement $\{\exists i \neq j \in \{1,\dots,k-1,k+1,\dots,n\} : X_i = X_j\}$ que l'on rajoute dans l'indicatrice est de probabilité nulle. Il ne reste plus qu'à remarquer que l'expression ne dépend pas de l'indice k choisi car pour toute permutation σ de $\{1,\dots,n\}$, le vecteur $(X_{\sigma(1)},\dots,X_{\sigma(n)})$ est constitué de n variables indépendantes de loi uniforme sur $[0,\theta]$ et a même loi que le vecteur $(X_1,\dots,X_n)$. Ainsi,

$$\mathbb{E}\left(f(X_k)1_{\{\forall i\neq k, X_i<X_k\}}\right) = \mathbb{E}\left(f(X_1)1_{\{\forall i\neq 1, X_i<X_1\}}\right).$$

Le résultat en découle immédiatement.

3. **En déduire que Z_n suit la densité $\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1}1_{\{0\leq z\leq\theta\}}$.**
Les variables $X_1,\dots,X_n$ étant indépendantes et à densité, le n-uplet $(X_1,\dots,X_n)$ admet une densité, produit des densités marginales. Ainsi,

$$\begin{aligned}\mathbb{E}\left(f(X_1)1_{\{\forall i\neq 1, X_i<X_1\}}\right) &= \int_{[0,\theta]^n} 1_{\{\forall i\neq 1, x_i<x_1\}}\frac{dx_1\dots dx_n}{\theta^n}\\ &= \int_{x_1=0}^{\theta} f(x_1)\left[\int_{x_2=0}^{x_1}\frac{dx_2}{\theta}\right]\cdots\left[\int_{x_n=0}^{x_1}\frac{dx_n}{\theta}\right]\frac{dx_1}{\theta}\\ &= \int_0^{\theta} f(x_1)\left(\frac{x_1}{\theta}\right)^{n-1}\frac{dx_1}{\theta}.\end{aligned}$$

Il en résulte que $\mathbb{E}(f(Z_n)) = n\int_0^{\theta} f(x_1)\left(\frac{x_1}{\theta}\right)^{n-1}\frac{dx_1}{\theta}$. D'où le résultat.

4. **Calculer $\mathbb{E}(Z_n)$ et $\mathrm{Var}(Z_n)$. En conclure que Z_n converge presque sûrement vers θ (*indication* : on rappelle (voir preuve du théorème 5.2.2) que si $(\tilde{Z}_n)_{n\geq 1}$ est une suite de variables aléatoires telle que $\sum_{n\geq 1}\mathbb{E}(\tilde{Z}_n^2) < +\infty$, alors la suite $(\tilde{Z}_n)_{n\geq 1}$ converge presque sûrement vers 0).**
On a :
$$\mathbb{E}(Z_n) = \int_0^{\theta} z\frac{n}{\theta^n}z^{n-1}\,dz = \frac{n}{\theta^n}\times\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta.$$
De même,
$$\mathbb{E}(Z_n^2) = \frac{n}{n+2}\theta^2$$
puis
$$\mathrm{Var}(Z_n) = \mathbb{E}(Z_n^2) - \mathbb{E}(Z_n)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2.$$
Comme $\mathrm{Var}(Z_n) = O(1/n^2)$, la série de terme général $\mathbb{E}((Z_n-\mathbb{E}(Z_n))^2)$ est convergente et donc $(\tilde{Z}_n = Z_n - \mathbb{E}(Z_n))_{n\geq 1}$ converge presque sûrement vers 0. Or la suite (déterministe) $(\mathbb{E}(Z_n))_{n\geq 1}$ converge vers θ. Par conséquent, $(Z_n)_{n\geq 1}$ converge presque sûrement vers θ.

5. **Soit f une fonction bornée sur $\mathbb{R}$. Montrer que :**

$$\mathbb{E}\left(f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right) = \int_0^n f(u)\left(1-\frac{u}{n}\right)^{n-1}\,du.$$

On rappelle que pour $0<v<1$, $\ln(1-v)\leq -v$. En déduire que pour $n\geq 2$, $\left(1-\frac{u}{n}\right)^{n-1}1_{\{0<u\leq n\}} \leq e^{-u/2}1_{\{u>0\}}$ et conclure que :

$$\lim_{n\to\infty}\mathbb{E}\left(f\left(\frac{n}{\theta}(\theta-Z_n)\right)\right) = \int_0^{\infty} f(u)e^{-u}\,du$$

En utilisant la densité de Z_n, il est clair que :

$$\mathbb{E}\left(f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right) = \int_0^\theta f\left(\frac{n}{\theta}(\theta - z)\right)\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1} dz.$$

Le changement de variable $u = \frac{n}{\theta}(\theta - z)$ donne immédiatement le résultat souhaité. Soit $u \in]0, n[$. En appliquant l'inégalité donnée avec $v = u/n$, il vient

$$\left(1 - \frac{u}{n}\right)^{n-1} = \exp\left((n-1)\ln\left(1 - \frac{u}{n}\right)\right) \leq \exp\left(-\frac{n-1}{n}u\right).$$

Or, dès lors que $n \geq 2$, on a $(n-1)/n \geq 1/2$ et $\left(1 - \frac{u}{n}\right)^{n-1} 1_{\{0<u\leq n\}} \leq e^{-u/2}1_{\{u>0\}}$. Ainsi, pour $n \geq 2$,

$$\forall u \in \mathbb{R},\ |f(u)|\left(1 - \frac{u}{n}\right)^{n-1} 1_{\{0<u\leq n\}} \leq \sup_x |f(x)| e^{-u/2}1_{\{u>0\}},$$

avec $\int_0^\infty e^{-u/2}1_{\{u>0\}} < +\infty$ et

$$\forall u \in \mathbb{R},\ \lim_{n\to\infty} f(u)\left(1 - \frac{u}{n}\right)^{n-1} 1_{\{0<u\leq n\}} = f(u)e^{-u}1_{\{u>0\}}$$

Le résultat annoncé découle du théorème de convergence dominée.

6. **En déduire que $\frac{n}{\theta}(\theta - Z_n)$ converge en loi et identifier la limite.**
Ainsi pour toute fonction f bornée, si X suit la loi exponentielle de paramètre 1,

$$\lim_{n\to\infty} \mathbb{E}\left(f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right) = \mathbb{E}(f(X)).$$

Ce résultat est a fortiori vrai pour les fonctions f continues bornées, ce qui signifie par définition que $\frac{n}{\theta}(\theta - Z_n)$ converge en loi vers $X \sim \mathcal{E}(1)$.

7. **Montrer que $M_n = \frac{2}{n}(X_1 + \cdots + X_n)$ converge presque sûrement vers θ. À quelle vitesse à lieu cette convergence ?**
Par application de la loi forte des grands nombres aux variables X_i I.I.D. et intégrables (car bornées) :

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.}_{n\to\infty} \mathbb{E}(X_1) = \int_0^\theta x\frac{dx}{\theta} = \frac{\theta}{2}$$

Ainsi M_n converge presque sûrement vers θ. Pour la vitesse de convergence, on utilise le théorème de la limite centrale. Les variables X_i étant de carré intégrable et de variance commune $\theta^2/12$,

$$\frac{\sqrt{12n}}{\theta}\left(\frac{X_1 + \cdots + X_n}{n} - \frac{\theta}{2}\right) \xrightarrow{\mathcal{L}}_{n\to\infty} G \sim \mathcal{N}_1(0, 1).$$

Comme la fonction $\varphi(x) = \theta x/\sqrt{3}$ est continue, on en déduit que

$$\sqrt{n}(M_n - \theta) = \varphi\left(\frac{\sqrt{12n}}{\theta}\left(\frac{X_1 + \cdots + X_n}{n} - \frac{\theta}{2}\right)\right) \xrightarrow{\mathcal{L}}_{n\to\infty} \varphi(G) = \frac{\theta G}{\sqrt{3}} \sim \mathcal{N}_1(0, \frac{\theta^2}{3}).$$

La convergence de M_n vers θ a donc lieu à la vitesse $1/\sqrt{n}$.

8. **L'automobiliste doit-il préférer Z_n ou M_n pour estimer rapidement θ ?**
La convergence de Z_n étant en $1/n$ tandis que celle de M_n est en $1/\sqrt{n}$, l'utilisation de Z_n est préférable.

10.6 Vecteurs gaussiens

Corrigé de l'exercice 6.3.1. Comme X et Y sont des gaussiennes réelles indépendantes, le vecteur (X,Y) est gaussien de même que le vecteur $(X+Y, X-Y)$ qui s'en déduit par transformation linéaire. Les coordonnées $X+Y$ et $X-Y$ sont dont indépendantes si et seulement si

$$0 = \mathrm{Cov}(X+Y, X-Y) = \mathrm{Var}(X) - \mathrm{Var}(Y),$$

c'est-à-dire si X et Y ont même variance.

Corrigé de l'exercice 6.3.5. Soit $X_1, \dots, X_n$ des variables aléatoires réelles I.I.D. et de carré intégrable telles que la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ et la variable $S_n^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X}_n)^2$ sont indépendantes.

1. **Vérifier que** $S_n^2 = \frac{1}{n}\sum_{j=1}^n X_j^2 - \frac{1}{n(n-1)}\sum_{\substack{j,k=1\\k\neq j}}^n X_j X_k$ **et en déduire** $\mathbb{E}(S_n^2)$**.**

   On a

$$\sum_{j=1}^n (X_j - \bar{X}_n)^2 = \sum_{j=1}^n X_j^2 - 2\bar{X}_n \sum_{j=1}^n X_j + n(\bar{X}_n)^2 = \sum_{j=1}^n X_j^2 - n(\bar{X}_n)^2$$
$$= \sum_{j=1}^n X_j^2 - \frac{1}{n}\sum_{j,k=1}^N X_j X_k = \frac{n-1}{n}\sum_{j=1}^n X_j^2 - \frac{1}{n}\sum_{\substack{j,k=1\\k\neq j}}^n X_j X_k.$$

   On en déduit l'expression voulue de S_n^2. Ainsi,

$$\mathbb{E}(S_n^2) = \frac{1}{n}\sum_{j=1}^n \mathbb{E}(X_1^2) - \frac{1}{n(n-1)}\sum_{\substack{j,k=1\\k\neq j}}^n (\mathbb{E}(X_1))^2 = \mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2 = \mathrm{Var}(X_1).$$

2. **En utilisant l'hypothèse d'indépendance, exprimer** $\mathbb{E}(S_n^2 e^{iun\bar{X}_n})$ **en fonction de** $\mathrm{Var}(X_1)$ **et de la fonction caractéristique commune des** X_j **que l'on notera** Φ_X**.**
   Par indépendance de $\bar{X}_n$ et S_n^2 puis par le caractère I.I.D. des X_j, on obtient

$$\mathbb{E}(S_n^2 e^{iun\bar{X}_n}) = \mathbb{E}(S_n^2)\mathbb{E}\left(\prod_{j=1}^n e^{iuX_j}\right) = \mathrm{Var}(X_1)(\Phi_X(u))^n.$$

3. **Montrer par ailleurs que**

$$\mathbb{E}(S_n^2 e^{iun\bar{X}_n}) = \mathbb{E}(X_1^2 e^{iuX_1})(\Phi_X(u))^{n-1} - (\mathbb{E}(X_1 e^{iuX_1}))^2(\Phi_X(u))^{n-2}.$$

   En utilisant l'expression de S_n^2 obtenue à la question 1 et le caractère I.I.D. des X_j, on obtient

$$\begin{aligned}
\mathbb{E}(S_n^2 e^{iun\bar{X}_n}) &= \frac{1}{n}\sum_{j=1}^n \mathbb{E}\left(X_j^2\prod_{k=1}^n e^{iuX_k}\right) - \frac{1}{n(n-1)}\sum_{\substack{j,k=1\\k\neq j}}^n \mathbb{E}\left(X_jX_k\prod_{l=1}^n e^{iuX_l}\right)\\
&= \frac{(\Phi_X(u))^{n-1}}{n}\sum_{j=1}^n \mathbb{E}\left(X_j^2 e^{iuX_j}\right)\\
&\qquad - \frac{(\Phi_X(u))^{n-2}}{n(n-1)}\sum_{\substack{j,k=1\\k\neq j}}^n \mathbb{E}\left(X_j e^{iuX_j}\right)\mathbb{E}\left(X_k e^{iuX_k}\right)\\
&= \mathbb{E}(X_1^2 e^{iuX_1})(\Phi_X(u))^{n-1} - (\mathbb{E}(X_1 e^{iuX_1}))^2(\Phi_X(u))^{n-2}.
\end{aligned}$$

4. **Remarquer que $\mathbb{E}(X_1e^{iuX_1})$ et $\mathbb{E}(X_1^2e^{iuX_1})$ s'expriment simplement à l'aide des deux premières dérivées de Φ_X et en déduire une équation différentielle satisfaite par cette fonction.**
   D'après la preuve du lemme 5.3.5, $\Phi'_X(u) = i\mathbb{E}(X_1e^{iuX_1})$ et $\Phi''_X(u) = -\mathbb{E}(X_1^2e^{iuX_1})$. Avec les deux questions précédentes, on en déduit que
   $$\mathrm{Var}(X_1)(\Phi_X(u))^2 = -\Phi''_X(u)\Phi_X(u) + (\Phi'_X(u))^2.$$
5. **Poser $f(u) = \Phi'_X(u)/\Phi_X(u)$ et calculer $f'(u)$. En déduire Φ_X puis la loi commune des X_j.**
   On a $f'(u) = \frac{\Phi''_X(u)\Phi_X(u)-(\Phi'_X(u))^2}{\Phi_X^2(u)}$. Avec l'équation différentielle de la question précédente, on en déduit que $f'(u) = -\mathrm{Var}(X_1)$. En intégrant, il vient
   $$\frac{\Phi'_X(u)}{\Phi_X(u)} = \frac{\Phi'_X(0)}{\Phi_X(0)} - \mathrm{Var}(X_1)u = \frac{i\mathbb{E}(X_1)}{1} - \mathrm{Var}(X_1)u.$$
   En intégrant une nouvelle fois, on obtient
   $$\ln(\Phi_X(u)) = \ln(\Phi_X(0)) + iu\mathbb{E}(X_1) - \frac{\mathrm{Var}(X_1)u^2}{2}.$$
   Comme $\Phi_X(0) = 1$, on conclut que $\Phi_X(u) = e^{iu\mathbb{E}(X_1)-\frac{\mathrm{Var}(X_1)u^2}{2}}$, c'est-à-dire que les X_j sont distribuées suivant la loi normale $\mathcal{N}_1(\mathbb{E}(X_1), \mathrm{Var}(X_1))$.

10.7 Estimateurs

Corrigé de l'exercice 7.4.2. Comme $\mathbb{E}(\sum_{i=1}^n a_iX_i) = \mu\sum_{i=1}^n a_i$, l'estimateur $\sum_{i=1}^n a_iX_i$ est sans biais si et seulement si $\sum_{i=1}^n a_i = 1$. L'inégalité de Cauchy-Schwarz implique alors que $\sum_{i=1}^n a_i^2 \geq \frac{1}{n}\left(\sum_{i=1}^n a_i\right)^2 = \frac{1}{n}$ avec égalité si et seulement si les a_i valent tous $\frac{1}{n}$. Comme $\mathrm{Var}\left(\sum_{i=1}^n a_iX_i\right) = \sigma^2\sum_{i=1}^n a_i^2$, on conclut que la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est l'estimateur linéaire et sans biais de μ de variance minimale.

Corrigé du problème 7.4.8. On s'intéresse à la durée de vie de deux composants électroniques se trouvant sur un système solidaire. Si l'un des deux composants tombe en panne, le système tout entier doit être changé. Les durées de vie de ces deux composants sont modélisées par l'intermédiaire de variables aléatoires exponentielles de paramètres λ

et μ indépendantes. Formellement, on considère un n-échantillon I.I.D. $(X_1, \ldots, X_n)$ de loi $\mathcal{E}(\lambda)$ et un n-échantillon I.I.D. $(Y_1, \ldots, Y_n)$, indépendant du précédent, de loi $\mathcal{E}(\mu)$ où $\lambda > 0$ et $\mu > 0$.

On observe seulement la durée de vie du composant qui est tombé en panne et donc le n-échantillon I.I.D. $(Z_1, W_1), \ldots, (Z_n, W_n)$ où $Z_i = \min(X_i, Y_i)$ et $W_i = \begin{cases} 1 & \text{si} \quad Z_i = X_i \\ 0 & \text{si} \quad Z_i = Y_i \end{cases}$.

1. **Donner les lois de Z_i et W_i.**
   Pour tout $z > 0$, $\mathbb{P}_{(\lambda,\mu)}(Z_i \le z) = 1 - \mathbb{P}_{(\lambda,\mu)}(\min(X_i, Y_i) > z) = 1 - \exp(-(\lambda+\mu)z)$. Ainsi, $Z_i \sim \mathcal{E}(\lambda + \mu)$. Par ailleurs,
   $\mathbb{P}_{(\lambda,\mu)}(W_i = 1) = \mathbb{P}_{(\lambda,\mu)}(X_i \le Y_i) = \int_0^\infty \left[\int_x^\infty \mu \exp(-\mu y) dy\right] \lambda \exp(-\lambda x) dx = \frac{\lambda}{\lambda+\mu}$,
   et $W_i \sim \mathcal{B}\left(\frac{\lambda}{\lambda+\mu}\right)$.
2. **Montrer que les variables Z_i et W_i sont indépendantes (on pourra calculer $\mathbb{E}_{(\lambda,\mu)}(f(Z_i)g(W_i))$ pour $f, g : \mathbb{R} \to \mathbb{R}$ bornées).**
   En décomposant sur les valeurs prises par W_i, il vient

$$\mathbb{E}_{(\lambda,\mu)}(f(Z_i)g(W_i)) = g(1)\mathbb{E}_{(\lambda,\mu)}\left(f(Z_i)1_{\{W_i=1\}}\right) + g(0)\mathbb{E}_{(\lambda,\mu)}\left(f(Z_i)1_{\{W_i=0\}}\right).$$

   Or

$$\begin{aligned}
\mathbb{E}_{(\lambda,\mu)}\left(f(Z_i)1_{\{W_i=1\}}\right) &= \mathbb{E}_{(\lambda,\mu)}\left(f(X_i)1_{\{X_i \le Y_i\}}\right) \\
&= \int_0^\infty f(x) \left[\int_x^\infty \mu \exp(-\mu y) dy\right] \lambda \exp(-\lambda x) dx \\
&= \frac{\lambda}{\lambda+\mu} \int_0^\infty f(x)(\lambda+\mu)\exp(-(\lambda+\mu)x) dx \\
&= \frac{\lambda}{\lambda+\mu} \mathbb{E}_{(\lambda,\mu)}(f(Z_i)).
\end{aligned}$$

   Et de même $\mathbb{E}_{(\lambda,\mu)}\left(f(Z_i)1_{\{W_i=1\}}\right) = \frac{\mu}{\lambda+\mu}\mathbb{E}_{(\lambda,\mu)}(f(Z_i))$. On conclut que

$$\mathbb{E}_{(\lambda,\mu)}(f(Z_i)g(W_i)) = \mathbb{E}_{(\lambda,\mu)}(f(Z_i)) \times \mathbb{E}_{(\lambda,\mu)}(g(W_i)),$$

   ce qui assure l'indépendance de Z_i et W_i.

Les variables aléatoires Z_i et W_i étant indépendantes, la vraisemblance du n-échantillon $(Z_1, W_1), \ldots, (Z_n, W_n)$ est définie comme étant le produit de la vraisemblance du n-échantillon $Z_1, \ldots, Z_n$ par la vraisemblance du n-échantillon $W_1, \ldots, W_n$. Dans les questions 3 à 7, on suppose que la loi de la durée de vie du second composant est connue i.e. que μ est connu.

3. **Donner l'Estimateur du Maximum de Vraisemblance $\hat{\lambda}_n$ de λ.**
   La vraisemblance et la log-vraisemblance du modèle sont respectivement

$$\begin{aligned}
p_n((z,w),(\lambda,\mu)) &= \exp\left(-(\lambda+\mu)\sum_{i=1}^n z_i\right)\lambda^{\sum_{i=1}^n w_i}\mu^{n-\sum_{i=1}^n w_i} \\
l_n((z,w),(\lambda,\mu)) &= -(\lambda+\mu)\sum_{i=1}^n z_i + \sum_{i=1}^n w_i \ln(\lambda) + \left(n - \sum_{i=1}^n w_i\right)\ln(\mu).
\end{aligned}$$

   La fonction $\frac{\partial l_n}{\partial \lambda}((z,w),(\lambda,\mu)) = -\sum_{i=1}^n z_i + \frac{1}{\lambda}\sum_{i=1}^n w_i$ est décroissante en λ et telle que

$$\frac{\partial l_n}{\partial \lambda}((z,w),(\lambda,\mu)) = 0 \Leftrightarrow \lambda = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n z_i}.$$

On en déduit que l'Estimateur du Maximum de Vraisemblance de λ est $\hat{\lambda}_n = \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i}$.

4. **Montrer que $\hat{\lambda}_n$ est fortement convergent.**
D'après la Loi Forte des Grands Nombres, sous $\mathbb{P}_\lambda$, lorsque $n \to +\infty$,
$$\frac{1}{n}\sum_{i=1}^n W_i \xrightarrow{p.s.} \frac{\lambda}{\lambda+\mu} \quad \text{et} \quad \frac{1}{n}\sum_{i=1}^n Z_i \xrightarrow{p.s.} \frac{1}{\lambda+\mu}.$$
Ainsi, $\hat{\lambda}_n \xrightarrow{p.s.} \lambda$.

5. **Calculer $\mathbb{E}_\lambda(\hat{\lambda}_n)$ et en déduire que $\mathbb{E}_\lambda(\hat{\lambda}_n)$ tend vers λ lorsque n tend vers l'infini.**
Sous $\mathbb{P}_\lambda$, $\sum_{i=1}^n Z_i \sim \Gamma(n, \lambda+\mu)$ d'après la proposition 3.4.4. Cette somme est indépendante de $\sum_{i=1}^n W_i \sim \mathcal{B}(n, \frac{\lambda}{\lambda+\mu})$. Donc
$$\begin{aligned}\mathbb{E}_\lambda(\hat{\lambda}_n) &= \mathbb{E}_\lambda\left(\sum_{i=1}^n W_i\right)\mathbb{E}_\lambda\left(\frac{1}{\sum_{i=1}^n Z_i}\right)\\ &= \frac{n\lambda}{\lambda+\mu}\int_0^{+\infty}\frac{1}{z}(\lambda+\mu)^n z^{n-1}e^{-(\lambda+\mu)z}\frac{dz}{\Gamma(n)}\\ &= \frac{n\lambda\Gamma(n-1)}{\Gamma(n)}\int_0^{+\infty}(\lambda+\mu)^{n-1}z^{n-2}e^{-(\lambda+\mu)z}\frac{dz}{\Gamma(n-1)} = \frac{n\lambda}{n-1}.\end{aligned}$$
Donc $\hat{\lambda}_n$ est un estimateur biaisé de λ et tel que $\mathbb{E}_\lambda(\hat{\lambda}_n)$ tend vers λ lorsque n tend vers l'infini.

6. **Vérifier que $\hat{\lambda}_n$ est asymptotiquement normal et que sa variance asymptotique est égale à l'inverse de l'information de Fisher.**
On a
$$\sqrt{n}(\hat{\lambda}_n - \lambda) = \frac{1}{\bar{Z}_n} \times \frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i - \lambda Z_i)$$
avec $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ qui converge presque sûrement vers $\mathbb{E}_\lambda(Z_1) = \frac{1}{\lambda+\mu}$ sous $\mathbb{P}_\lambda$ d'après la loi forte des grands nombres. Sous $\mathbb{P}_\lambda$, les variables aléatoires $W_i - \lambda Z_i$ sont I.I.D., centrées et de variance
$$\mathrm{Var}_\lambda(W_i - \lambda Z_i) = \mathrm{Var}_\lambda(W_i) + \lambda^2\mathrm{Var}_\lambda(Z_i) = \frac{\lambda\mu}{(\lambda+\mu)^2} + \frac{\lambda^2}{(\lambda+\mu)^2} = \frac{\lambda}{\lambda+\mu}.$$
D'après le théorème de la limite centrale $\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i - \lambda Z_i)$ converge en loi vers $G \sim \mathcal{N}_1(0, \frac{\lambda}{\lambda+\mu})$. On déduit du théorème de Slutsky que $\sqrt{n}(\hat{\lambda}_n - \lambda)$ converge en loi vers $(\lambda+\mu)G \sim \mathcal{N}_1(0, \lambda(\lambda+\mu))$. Donc $\hat{\lambda}_n$ est asymptotiquement normal et sa variance asymptotique est égale à $\lambda(\lambda+\mu)$.
Comme $\frac{\partial^2 l}{\partial\lambda^2}((z_1, w_1), (\lambda, \mu)) = -\frac{w_1}{\lambda^2}$, l'information de Fisher est
$$I(\lambda) = -\mathbb{E}_\lambda\left(-\frac{W_1}{\lambda^2}\right) = \frac{\lambda}{\lambda^2(\lambda+\mu)} = \frac{1}{\lambda(\lambda+\mu)}.$$

7. **Donner un intervalle de confiance bilatéral symétrique asymptotique de niveau $1-\alpha$ pour λ.**
D'après les questions précédentes, sous $\mathbb{P}_\lambda$,
$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}_1(0, \lambda(\lambda+\mu)) \text{ et } \hat{\lambda}_n(\hat{\lambda}_n + \mu) \xrightarrow{p.s.} \lambda(\lambda+\mu)$$

ce qui, avec le théorème de Slutsky, entraîne que $\sqrt{\frac{n}{\hat{\lambda}_n(\hat{\lambda}_n+\mu)}}(\hat{\lambda}_n-\lambda) \xrightarrow{\mathcal{L}} \mathcal{N}_1(0,1)$. Comme dans l'exemple 7.3.10, on en déduit que si ϕ_r désigne le quantile d'ordre r de la loi normale centrée réduite, alors

$$\left[\hat{\lambda}_n - \phi_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}_n(\hat{\lambda}_n+\mu)}{n}}, \hat{\lambda}_n + \phi_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}_n(\hat{\lambda}_n+\mu)}{n}}\right]$$

est un intervalle de confiance bilatéral symétrique asymptotique pour λ de niveau $1-\alpha$.

8. **Donner l'Estimateur du Maximum de Vraisemblance $(\hat{\lambda}_n, \hat{\mu}_n)$ de (λ, μ).**
À μ fixé, $l_n((z,w),(\lambda,\mu))$ est maximum pour $\lambda = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n z_i}$ et vaut alors $\sum_{i=1}^n w_i \left(\ln\left(\frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n z_i}\right) - 1\right) + f(\mu)$ avec $f(\mu) = -\mu\sum_{i=1}^n z_i + (n - \sum_{i=1}^n w_i)\ln(\mu)$. La fonction $f'(\mu) = -\sum_{i=1}^n z_i + \frac{1}{\mu}(n - \sum_{i=1}^n w_i)$ est décroissante et s'annule pour $\mu = \frac{n-\sum_{i=1}^n w_i}{\sum_{i=1}^n z_i}$. On conclut donc que l'EMV de (λ,μ) est $(\hat{\lambda}_n, \hat{\mu}_n) = \left(\frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i}, \frac{n-\sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i}\right)$.

10.8 Tests d'hypothèses

Corrigé de l'exercice 8.3.1. On se place dans le modèle exponentiel $\mathcal{P} = \{\mathcal{E}(\theta), \theta > 0\}$.

1. **Rappeler $\mathbb{E}_\theta(X_1)$.**

$$\mathbb{E}_\theta(X_1) = \int_0^{+\infty} \theta x e^{-\theta x}dx = [-xe^{-\theta x}]_0^{+\infty} + \int_0^{+\infty} e^{-\theta x}dx = 0 + \frac{1}{\theta} = \frac{1}{\theta}.$$

2. **En remarquant que $2\theta(X_1+\ldots+X_n) \sim \chi^2(2n)$ construire un test de niveau α pour $H_0 = \{\theta = \theta_0\}$ et $H_1 = \{\theta \neq \theta_0\}$.**
A.N. : $n = 15$, $\bar{x}_n = 1.47$, $\theta_0 = 1$, $\alpha = 5\%$.
Soit $f : \mathbb{R} \to \mathbb{R}$ bornée.

$$\mathbb{E}_\theta(f(2\theta X_1)) = \int_0^{+\infty} f(2\theta x)\theta e^{-\theta x}dx = \int_0^{+\infty} f(y)\frac{1}{2}e^{-y/2}dy.$$

Donc, sous $\mathbb{P}_\theta$, les variables $2\theta X_i$ sont I.I.D. suivant la loi $\mathcal{E}(\frac{1}{2}) = \Gamma(1,\frac{1}{2})$. D'après la proposition 3.4.4, sous $\mathbb{P}_\theta$, $2\theta(X_1+\ldots+X_n) \sim \Gamma(n,\frac{1}{2}) = \chi^2(2n)$.
On choisit comme statistique de test $\zeta_n = 2\theta_0(X_1+\ldots+X_n)$. Sous H_0, $\zeta_n \sim \chi^2(2n)$. Sous H_1, comme $\zeta_n = \frac{\theta_0}{\theta} \times 2\theta(X_1+\ldots+X_n)$, cette statistique va avoir tendance à prendre des valeurs plus petites ou plus grandes que sous H_0. On choisit donc $W = \{\zeta_n \notin [\chi^2_{2n,\frac{\alpha}{2}}, \chi^2_{2n,1-\frac{\alpha}{2}}]\}$ où $\chi^2_{2n,r}$ désigne le quantile d'ordre r de la loi $\chi^2(2n)$.
A.N. : $\zeta_{15}^{\text{obs}} = 44.1$ et d'après la table du paragraphe 11.3, $\chi^2_{30,0.025} = 16.79$ et $\chi^2_{30,0.975} = 46.98$. On accepte donc H_0 au niveau 5%.

3. **Tester maintenant $H_0 = \{\theta \geq \theta_0\}$ contre $H_1 = \{\theta < \theta_0\}$.**
On conserve la même statistique de test mais la nouvelle région critique est $W = \{\zeta_n > \chi^2_{2n,1-\alpha}\}$.
A.N. : Comme $\chi^2_{30,0.95} = 43.77$, on rejette maintenant H_0 au niveau 5%. Notons que la p-valeur du test est comprise entre 2.5% et 5%

Corrigé de l'exercice 8.3.5. On désire étudier la répartition des naissances suivant le type du jour dans la semaine (jours ouvrables ou week-end) et suivant le mode d'accouchement (naturel ou par césarienne). Les données proviennent du "National Vital Statistics Report" et concernent les naissances aux USA en 1997.

Naissances	Naturelles	César.	Total
J.O.	2331536	663540	2995076
W.E.	715085	135493	850578
Total	3046621	799033	3845654

Naissances	Naturelles	César.	Total
J.O.	60.6 %	17.3 %	77.9%
W.E.	18.6 %	3.5 %	22.1%
Total	79.2 %	20.8 %	100.0%

1. **Tester au niveau 0.001 l'hypothèse d'indépendance entre le type du jour de naissance (jour ouvrable ou week-end) et le mode d'accouchement (naturel ou césarienne).** Les fréquences observées sont $\hat{p}_J = 0.779$, $\hat{p}_W = 0.221$, $\hat{p}_N = 0.792$, $\hat{p}_C = 0.208$, $\hat{p}_{JN} = 0.606$, $\hat{p}_{JC} = 0.173$, $\hat{p}_{WN} = 0.186$ et $\hat{p}_{WC} = 0.035$ où les indices J, W, N, C signifient respectivement jour ouvrable, week-end, naissance naturelle, naissance par césarienne. On en déduit que $\hat{p}_J\hat{p}_N = 0.617$, $\hat{p}_J\hat{p}_C = 0.162$, $\hat{p}_W\hat{p}_N = 0.175$ et $\hat{p}_W\hat{p}_C = 0.046$. La statistique de test pour le test d'indépendance est $\zeta^{\text{obs}}_{3\,845\,654} = 16\,401.3$. D'après la table du paragraphe 11.3, $\chi^2_{1,0.999} = 10.83$. On rejette donc, au niveau 0.001, l'hypothèse d'indépendance entre le type du jour de naissance et le mode d'accouchement. Il y a plus de naissance par césarienne les jours ouvrables que les week-ends.
2. **On désire savoir s'il existe une évolution significative dans la répartition des naissances par rapport à 1996. Tester au niveau 0.01 si $p = p^0$, où p^0 correspond aux données de 1996 :**

Naissances	**Naturelles**	**Césariennes**
J.O.	**60.5 %**	**17.0 %**
W.E.	**18.9 %**	**3.6 %**

   La statistique pour le test d'adéquation à la loi p^0 est $\zeta^{\text{obs}}_{3\,845\,654} = 499.9$. Comme $\chi^2_{3,0.99} = 6.25$, on accepte l'hypothèse d'une évolution significative dans la répartition des naissances. Comme $\chi^2_{3,0.999} = 16.27$, la p-valeur du test est bien plus petite que 0.001. En fait la taille de l'échantillon est suffisamment importante pour qu'une modification faible des proportions soit significative.

10.9 Régression linéaire

Corrigé de l'exercice 9.3.2. On souhaite tester l'égalité des moyennes de k échantillons gaussiens. Pour $i \in \{1, \dots, k\}$ et $j \in \{1, \dots, n_i\}$, on note $X_{i,j}$ la j-ème valeur du i-ème échantillon et on suppose que

$$X_{i,j} = \mu_i + \varepsilon_{i,j}$$

où les variables $\varepsilon_{i,j}$ sont I.I.D. suivant la loi normale $\mathcal{N}_1(0, \sigma^2)$ centrée de variance σ^2. Par exemple $X_{i,j}$ peut représenter le rendement dans la j-ème parcelle ensemencée avec l'espèce de blé i. La question de l'égalité des moyennes μ_i des échantillons revient alors à savoir si les différentes espèces de blé sont équivalentes en termes de rendement. On pose

$$\begin{cases} X = (X_{1,1}, \dots, X_{1,n_1}, X_{2,1}, \dots, X_{2,n_2}, \dots, X_{k,1}, \dots, X_{k,n_k})^* \\ \varepsilon = (\varepsilon_{1,1}, \dots, \varepsilon_{1,n_1}, \varepsilon_{2,1}, \dots, \varepsilon_{2,n_2}, \dots, \varepsilon_{k,1}, \dots, \varepsilon_{k,n_k})^* \\ \mu = (\mu_1, \dots, \mu_1, \mu_2, \dots, \mu_2, \dots, \mu_k, \dots, \mu_k)^* \end{cases},$$

où μ_i apparaît n_i fois si bien que le modèle se récrit $\boxed{X = \mu + \varepsilon}$. Enfin, on note $n = \sum_{i=1}^k n_k$, $E = \{(y_1, \ldots, y_1, y_2, \ldots, y_2, \ldots, y_k, \ldots, y_k)^* : (y_1, \ldots, y_k) \in \mathbb{R}^k\}$ où y_i apparaît n_i fois et $H = \{(z, \ldots, z)^* : z \in \mathbb{R}\}$ où z apparaît n fois.

1. **Donner l'Estimateur du Maximum de Vraisemblance de $\theta = (\mu, \sigma^2)$.**
La vraisemblance et la log-vraisemblance du modèle sont

$$p_n(x, (\mu, \sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^k \sum_{j=1}^{n_i}(x_{i,j}-\mu_i)^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{|x-\mu|^2}{2\sigma^2}}$$
$$l_n(x, (\mu, \sigma^2)) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{|x-\mu|^2}{2\sigma^2}.$$

À $\sigma^2 > 0$ fixé, la log-vraisemblance est maximale pour $\mu \in E$ qui minimise $|x-\mu|^2$ c'est-à-dire pour $\mu = x_E$. Elle vaut alors $f(\sigma^2)$ où $f(\lambda) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\lambda) - \frac{|x-x_E|^2}{2\lambda}$ et comme dans le paragraphe 9.1, on conclut que l'EMV de (μ, σ^2) est $(X_E, \frac{|X-X_E|^2}{n})$.

2. **Vérifier que $X_E = (X_{1,.}, \ldots, X_{1,.}, X_{2,.}, \ldots, X_{2,.}, \ldots, X_{k,.}, \ldots, X_{k,.})^*$ où pour $i \in \{1, \ldots, k\}$, $X_{i,.} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$ apparaît n_i fois. Déterminer X_H.**
Le vecteur $Z = (X_{1,.}, \ldots, X_{1,.}, X_{2,.}, \ldots, X_{2,.}, \ldots, X_{k,.}, \ldots, X_{k,.})^*$ appartient à E. Pour $y = (y_1, \ldots, y_1, y_2, \ldots, y_2, \ldots, y_k, \ldots, y_k)^* \in E$, on a

$$\begin{aligned} y.(X - Z) &= \sum_{i=1}^k y_i \sum_{j=1}^{n_i}(X_{i,j} - X_{i,.}) = \sum_{i=1}^k y_i \sum_{j=1}^{n_i}\left(X_{i,j} - \frac{1}{n_i}\sum_{l=1}^{n_i} X_{i,l}\right) \\ &= \sum_{i=1}^k y_i \left(\sum_{j=1}^{n_i} X_{i,j} - \sum_{l=1}^{n_i} X_{i,l}\right) = 0. \end{aligned}$$

On en déduit que Z est la projection orthogonale de X sur E c'est-à-dire que $X_E = Z$. De même, on montre que $X_H = (X_{.,.}, \ldots, X_{.,.})^*$ où $X_{.,.} = \frac{1}{n}\sum_{i=1}^k \sum_{j=1}^{n_i} X_{i,j}$ apparaît n fois.

3. **Quelle est la loi de $(|X - X_E|^2/\sigma^2, |X_E - X_H|^2/\sigma^2)$ sous $H_0 = \{\mu \in H\}$? En deduire que sous H_0 le rapport $F = \dfrac{|X_E - X_H|^2/(k-1)}{|X - X_E|^2/(n-k)}$ suit la loi de Fisher $\mathcal{F}_{k-1,n-k}$.**
Les variables aléatoires $\varepsilon - \varepsilon_E$ et $\varepsilon_E - \varepsilon_H$ sont respectivement les projections orthogonales de ε sur l'orthogonal de E de dimension $n-k$ et sur l'orthogonal de H dans E de dimension $k-1$. On déduit alors du théorème de Cochran 6.2.3 que $|\varepsilon - \varepsilon_E|/\sigma^2$ et $|\varepsilon_E - \varepsilon_H|^2/\sigma^2$ sont des variables aléatoires indépendantes respectivement distribuées suivant les lois du χ^2 à $n-k$ degrés de liberté et à $k-1$ degrés de liberté. Donc le rapport $\frac{|\varepsilon_E - \varepsilon_H|^2/(k-1)}{|\varepsilon - \varepsilon_E|^2/(n-k)}$ suit la loi de Fischer $\mathcal{F}_{k-1,n-k}$. Comme $\mu \in E$, $X_E = \mu + \varepsilon_E$ et $\varepsilon - \varepsilon_E = X - X_E$. Sous H_0, $\mu \in H$ et $X_H = \mu + \varepsilon_H$, ce qui assure que $\varepsilon_E - \varepsilon_H = X_E - X_H$ et donc que le rapport précédent est égal à F.

4. **Pourquoi les variables aléatoires $|X_E - X_H|^2$ et F prennent-elles des valeurs plus grandes sous $H_1 = \{\mu \notin H\}$ que sous H_0 ?**
L'argument est le même que celui donné dans la remarque 9.2.1.

5. **En déduire la région critique du test.**
On choisit comme région critique $W = \{F \geq \mathcal{F}_{k-1,n-k,1-\alpha}\}$ où $\mathcal{F}_{k-1,n-k,r}$ désigne le quantile d'ordre r de la loi de Fisher $\mathcal{F}_{k-1,n-k}$ et α le niveau du test.

6. **Application : on souhaite tester, pour une chaîne de magasins, les politiques de publicité suivantes :**

$$\begin{array}{lcl} A & : & \textbf{aucune publicité} \\ B & : & \textbf{tracts distribués dans le voisinage} \\ C & : & \textbf{tracts distribués et annonces dans les journaux}. \end{array}$$

On sélectionne 18 magasins divisés au hasard en 3 groupes de 6, et chaque groupe applique l'une des politiques de publicité. On enregistre ensuite les ventes cumulées sur un mois pour chaque magasin, et on obtient les moyennes et écart-types empiriques suivants (en milliers d'euros) :

	A	B	C
$\bar{X}$	**130.17**	**139.5**	**169.17**
S	**8.57**	**14.71**	**18.23**

où, par exemple, pour le groupe A d'effectif $n_A = 6$,

$$\bar{X}_A = \frac{1}{n_A}\sum_{j=1}^{n_A} X_{A,j}, \quad S_A = \sqrt{\frac{1}{n_A - 1}\sum_{j=1}^{n_A}(X_{A,j} - \bar{X}_A)^2}.$$

On suppose que les observations pour chaque groupe sont gaussiennes, de moyennes respectives μ_A, μ_B, μ_C et de même variance σ^2. Tester l'hypothèse nulle "il n'existe aucune différence entre les politiques de publicité" au niveau 5%. Évaluer approximativement la p-valeur.
On a $|X - X_E|^2 = 5(S_A^2 + S_B^2 + S_C^2) = 3\,110.8$. La moyenne des ventes sur tous les magasins est $\bar{X} = \frac{1}{3}(\bar{X}_A + \bar{X}_B + \bar{X}_C) = 146.28$, ce qui entraîne que $|X_E - X_H|^2 = 6 * \big((\bar{X}_A - \bar{X})^2 + (\bar{X}_B - \bar{X})^2 + (\bar{X}_C - \bar{X})^2\big) = 4\,976.72$. Ainsi $F^{\text{obs}} = 12$. D'après la table du paragraphe 11.5, $\mathcal{F}_{2,15,0.95} = 3.68$. On rejette donc l'hypothèse nulle au niveau 5%. Comme $\mathcal{F}_{2,15,0.99} = 6.36$, la p-valeur est inférieure à 1%.

Chapitre 11

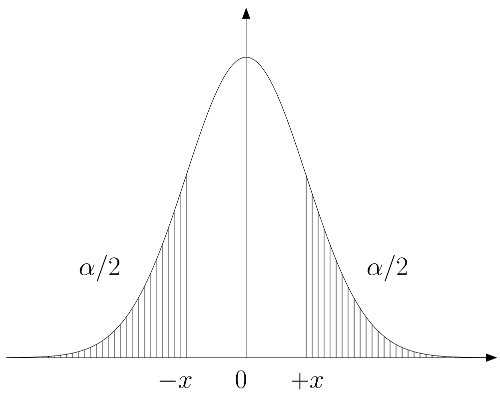
Tables statistiques

Les tables qui suivent ont été générées à l'aide du logiciel Scilab.

11.1 Quantiles de la loi $\mathcal{N}_1(0,1)$

Soit $X \sim \mathcal{N}_1(0,1)$. On pose

$$2\int_x^{\infty} \mathrm{e}^{-y^2/2}\ \frac{dy}{\sqrt{2\pi}} = \mathbb{P}(|X| \geq x) = \alpha.$$



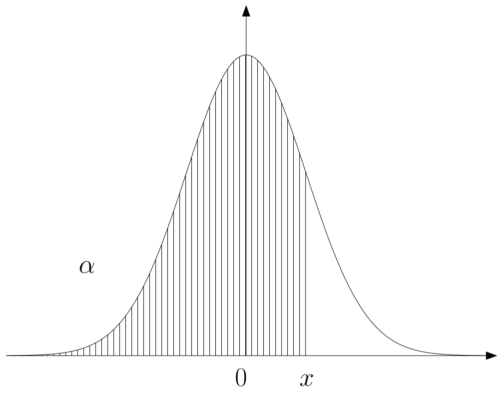
La table donne les valeurs de x en fonction de α. Par exemple $\mathbb{P}(|X| \geq 0.6280) \simeq 0.53$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	$\boxed{0.6280}$	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

11.2 Fonction de répartition de la loi $\mathcal{N}_1(0,1)$

Soit $X \sim \mathcal{N}_1(0,1)$. On pose

$$\int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\ \frac{dy}{\sqrt{2\pi}} = \mathbb{P}(X \leq x) = \alpha.$$



La table suivante donne les valeurs de $1-\alpha$ pour les grandes valeurs de x.

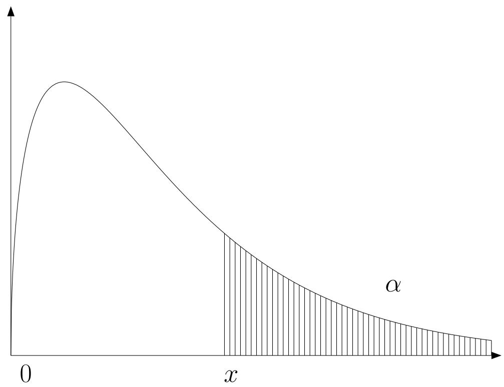
x	2	3	4	5	6	7	8	9	10
$1-\alpha$	2.28e-02	1.35e-03	3.17e-05	2.87e-07	9.87e-10	1.28e-12	6.22e-16	1.13e-19	7.62e-24

11.3 Quantiles de la loi du χ^2

Soit $X_n \sim \chi^2(n)$. On pose

$$\int_x^{\infty} \frac{1}{2^{n/2}\Gamma(n/2)}\, y^{\frac{n}{2}-1}\,\mathrm{e}^{-y/2}\; dy = \mathbb{P}(X_n \geq x) = \alpha.$$

La table donne les valeurs de x en fonction de n et α.
Par exemple $\mathbb{P}(X_8 \geq 20.09) \simeq 0.01$.
Lorsque $n > 30$, on peut utiliser l'approximation $\sqrt{2X_n} - \sqrt{2n-1} \overset{\mathcal{L}}{\simeq} G \sim \mathcal{N}_1(0,1)$ (voir l'exercice 5.5.15) qui assure que pour $x \geq 0$,



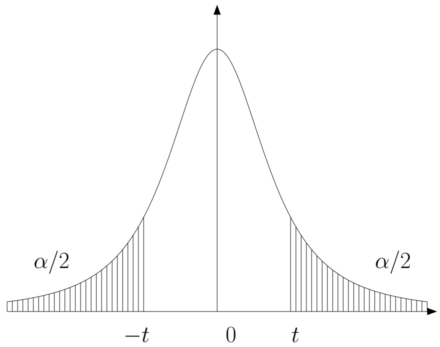
$$\mathbb{P}(X_n \geq x) = \mathbb{P}(\sqrt{2X_n}-\sqrt{2n-1} \geq \sqrt{2x}-\sqrt{2n-1}) \simeq \mathbb{P}(G \geq \sqrt{2x}-\sqrt{2n-1}).$$

$n\backslash\alpha$	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.001
1	0.0002	0.0010	0.0039	0.0158	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.52
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	$\boxed{20.09}$	26.12
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	39.25
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.31
21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	54.05
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70

11.4 Quantiles de la loi de Student

Soit $X_n \sim t(n)$. On pose

$$2\int_t^{\infty} \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)} \frac{1}{\left(1+\frac{y^2}{n}\right)^{(n+1)/2}}\, dy = \mathbb{P}(|X_n| \geq t)$$



La table donne les valeurs de t en fonction de n et α.
Par exemple $\mathbb{P}(|X_{20}| \geq 2.086) \simeq 0.05$.

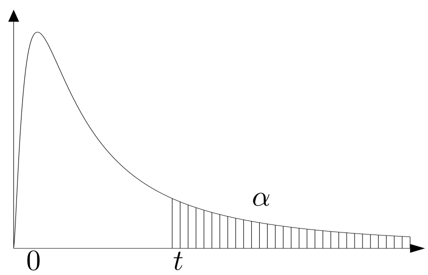
$n\backslash\alpha$	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100	0.050	0.020	0.010	0.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	$\boxed{2.086}$	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
80	0.126	0.254	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.416
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

11.5 Quantiles de la loi de Fisher (ou Fisher-Snedecor)

Soit $X_{n,m}$ une v.a. de loi de Fisher de paramètre (n, m). On pose

$$\mathbb{P}(X_{n,m} \geq f) = \alpha.$$

La table donne les valeurs de t en fonction de n, m et $\alpha \in \{0,05; 0,01\}$.
Par exemple $\mathbb{P}(X_{4,20} \geq 4.43) \simeq 0.01$.



	$n=1$		$n=2$		$n=3$		$n=4$		$n=5$	
m	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01
1	161.45	4052.18	199.50	4999.50	215.71	5403.35	224.58	5624.58	230.16	5763.65
2	18.51	98.50	19.00	99.00	19.16	99.17	19.25	99.25	19.30	99.30
3	10.13	34.12	9.55	30.82	9.28	29.46	9.12	28.71	9.01	28.24
4	7.71	21.20	6.94	18.00	6.59	16.69	6.39	15.98	6.26	15.52
5	6.61	16.26	5.79	13.27	5.41	12.06	5.19	11.39	5.05	10.97
6	5.99	13.75	5.14	10.92	4.76	9.78	4.53	9.15	4.39	8.75
7	5.59	12.25	4.74	9.55	4.35	8.45	4.12	7.85	3.97	7.46
8	5.32	11.26	4.46	8.65	4.07	7.59	3.84	7.01	3.69	6.63
9	5.12	10.56	4.26	8.02	3.86	6.99	3.63	6.42	3.48	6.06
10	4.96	10.04	4.10	7.56	3.71	6.55	3.48	5.99	3.33	5.64
11	4.84	9.65	3.98	7.21	3.59	6.22	3.36	5.67	3.20	5.32
12	4.75	9.33	3.89	6.93	3.49	5.95	3.26	5.41	3.11	5.06
13	4.67	9.07	3.81	6.70	3.41	5.74	3.18	5.21	3.03	4.86
14	4.60	8.86	3.74	6.51	3.34	5.56	3.11	5.04	2.96	4.69
15	4.54	8.68	3.68	6.36	3.29	5.42	3.06	4.89	2.90	4.56
16	4.49	8.53	3.63	6.23	3.24	5.29	3.01	4.77	2.85	4.44
17	4.45	8.40	3.59	6.11	3.20	5.18	2.96	4.67	2.81	4.34
18	4.41	8.29	3.55	6.01	3.16	5.09	2.93	4.58	2.77	4.25
19	4.38	8.18	3.52	5.93	3.13	5.01	2.90	4.50	2.74	4.17
20	4.35	8.10	3.49	5.85	3.10	4.94	2.87	$\boxed{4.43}$	2.71	4.10
21	4.32	8.02	3.47	5.78	3.07	4.87	2.84	4.37	2.68	4.04
22	4.30	7.95	3.44	5.72	3.05	4.82	2.82	4.31	2.66	3.99
23	4.28	7.88	3.42	5.66	3.03	4.76	2.80	4.26	2.64	3.94
24	4.26	7.82	3.40	5.61	3.01	4.72	2.78	4.22	2.62	3.90
25	4.24	7.77	3.39	5.57	2.99	4.68	2.76	4.18	2.60	3.85
26	4.23	7.72	3.37	5.53	2.98	4.64	2.74	4.14	2.59	3.82
27	4.21	7.68	3.35	5.49	2.96	4.60	2.73	4.11	2.57	3.78
28	4.20	7.64	3.34	5.45	2.95	4.57	2.71	4.07	2.56	3.75
29	4.18	7.60	3.33	5.42	2.93	4.54	2.70	4.04	2.55	3.73
30	4.17	7.56	3.32	5.39	2.92	4.51	2.69	4.02	2.53	3.70
40	4.08	7.31	3.23	5.18	2.84	4.31	2.61	3.83	2.45	3.51
80	3.96	6.96	3.11	4.88	2.72	4.04	2.49	3.56	2.33	3.26
120	3.92	6.85	3.07	4.79	2.68	3.95	2.45	3.48	2.29	3.17
∞	3.84	6.63	3.00	4.61	2.60	3.78	2.37	3.32	2.21	3.02

	$n=6$		$n=8$		$n=12$		$n=24$		$n=\infty$	
m	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01	α =0.05	α =0.01
1	233.99	5858.99	238.88	5981.07	243.91	6106.32	249.05	6234.63	254.31	6365.86
2	19.33	99.33	19.37	99.37	19.41	99.42	19.45	99.46	19.50	99.50
3	8.94	27.91	8.85	27.49	8.74	27.05	8.64	26.60	8.53	26.13
4	6.16	15.21	6.04	14.80	5.91	14.37	5.77	13.93	5.63	13.46
5	4.95	10.67	4.82	10.29	4.68	9.89	4.53	9.47	4.36	9.02
6	4.28	8.47	4.15	8.10	4.00	7.72	3.84	7.31	3.67	6.88
7	3.87	7.19	3.73	6.84	3.57	6.47	3.41	6.07	3.23	5.65
8	3.58	6.37	3.44	6.03	3.28	5.67	3.12	5.28	2.93	4.86
9	3.37	5.80	3.23	5.47	3.07	5.11	2.90	4.73	2.71	4.31
10	3.22	5.39	3.07	5.06	2.91	4.71	2.74	4.33	2.54	3.91
11	3.09	5.07	2.95	4.74	2.79	4.40	2.61	4.02	2.40	3.60
12	3.00	4.82	2.85	4.50	2.69	4.16	2.51	3.78	2.30	3.36
13	2.92	4.62	2.77	4.30	2.60	3.96	2.42	3.59	2.21	3.17
14	2.85	4.46	2.70	4.14	2.53	3.80	2.35	3.43	2.13	3.00
15	2.79	4.32	2.64	4.00	2.48	3.67	2.29	3.29	2.07	2.87
16	2.74	4.20	2.59	3.89	2.42	3.55	2.24	3.18	2.01	2.75
17	2.70	4.10	2.55	3.79	2.38	3.46	2.19	3.08	1.96	2.65
18	2.66	4.01	2.51	3.71	2.34	3.37	2.15	3.00	1.92	2.57
19	2.63	3.94	2.48	3.63	2.31	3.30	2.11	2.92	1.88	2.49
20	2.60	3.87	2.45	3.56	2.28	3.23	2.08	2.86	1.84	2.42
21	2.57	3.81	2.42	3.51	2.25	3.17	2.05	2.80	1.81	2.36
22	2.55	3.76	2.40	3.45	2.23	3.12	2.03	2.75	1.78	2.31
23	2.53	3.71	2.37	3.41	2.20	3.07	2.01	2.70	1.76	2.26
24	2.51	3.67	2.36	3.36	2.18	3.03	1.98	2.66	1.73	2.21
25	2.49	3.63	2.34	3.32	2.16	2.99	1.96	2.62	1.71	2.17
26	2.47	3.59	2.32	3.29	2.15	2.96	1.95	2.58	1.69	2.13
27	2.46	3.56	2.31	3.26	2.13	2.93	1.93	2.55	1.67	2.10
28	2.45	3.53	2.29	3.23	2.12	2.90	1.91	2.52	1.65	2.06
29	2.43	3.50	2.28	3.20	2.10	2.87	1.90	2.49	1.64	2.03
30	2.42	3.47	2.27	3.17	2.09	2.84	1.89	2.47	1.62	2.01
40	2.34	3.29	2.18	2.99	2.00	2.66	1.79	2.29	1.51	1.80
80	2.21	3.04	2.06	2.74	1.88	2.42	1.65	2.03	1.32	1.49
120	2.18	2.96	2.02	2.66	1.83	2.34	1.61	1.95	1.25	1.38
∞	2.10	2.80	1.94	2.51	1.75	2.18	1.52	1.79	1.00	1.00

Bibliographie

[1] A. Borovkov. *Mathematical statistics.* Gordon and Breach Science Publishers, 1998.

[2] N. Bouleau. *Probabilités de l'ingénieur. Variables aléatoires et simulation.* Hermann, 1986.

[3] L. Breiman. *Probability.* Number 7 in Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992.

[4] P. Brémaud. *Introduction aux probabilités.* Springer Verlag, 1984.

[5] D. Dacunha-Castelle et M. Duflo. *Probabilités et statistiques : Problèmes à temps fixe.* Masson, 1982.

[6] J.-F. Delmas. *Introduction au calcul des probabilités et à la statistique.* Polycopié ENSTA, 2008.

[7] W. Feller. *An introduction to probability theory and its applications*, volume 1. Wiley and Sons, Inc., 3rd edition, 1968.

[8] G. Grimmett et D. Stirzaker. *Probability and random processes.* Oxford University Press, 2nd edition, 1992.

[9] J. Jacod et P. Protter. *Probability essentials.* Springer, 2000.

[10] A. Montfort. *Cours de statistique mathématique.* Economica, 1988.

[11] A. Montfort. *Introduction à la statistique.* École Polytechnique, 1991.

[12] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson, 1964.

[13] J. Neveu. *Introduction aux probabilités.* École Polytechnique, 1990.

[14] S. Port. *Theoretical probability for applications.* Wiley & Sons, 1993.

[15] D. Revuz. *Mesure et Intégration.* Hermann, 1994.

[16] G. Saporta. *Probabilités, Statistique et Analyse des Données.* Technip, 1990.

[17] Y. Sinai. *Probability theory : an introductory course.* Springer Verlag, 1992.

[18] D. Stirzaker. *Elementary probability.* Cambridge Univ. Press, 1994.

Index

Notations et symboles

$\binom{n}{k}$	$= \frac{n!}{k!(n-k)!}$ coefficient binomial
$\lfloor x \rfloor$	partie entière du réel x
$\|x\|$	$= \sqrt{\sum_{i=1}^d x_i^2}$ norme euclidienne du vecteur $x \in \mathbb{R}^d$
M^*	transposée de la matrice $M \in \mathbb{R}^{n\times m}$
$x.y$	$= \sum_{i=1}^d x_i y_i$ produit scalaire des vecteurs x et y de $\mathbb{R}^d$
x_E	projection orthogonale du vecteur $x \in \mathbb{R}^d$ sur le sous-espace E de $\mathbb{R}^d$
I.I.D.	Indépendantes et Identiquement Distribuées
1_A	fonction indicatrice de A : $1_A(\omega)$ vaut 1 si $\omega \in A$ et 0 sinon
$\mathbb{E}(X)$	espérance de la variable aléatoire X
$\mathrm{Var}(X)$	$= \mathbb{E}(X^2) - \mathbb{E}(X)^2$ variance de la variable aléatoire X
$\Phi_X(u)$	$= \mathbb{E}(e^{iu.X})$ fonction caractéristique du vecteur aléatoire X
$F_X(x)$	$= \mathbb{P}(X \leq x)$ fonction de répartition de la variable aléatoire X
$g_X(s)$	$= \mathbb{E}(s^X)$ fonction génératrice de la variable aléatoire entière X
$\mathrm{Cov}(X,Y)$	$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ covariance des variables aléatoires X et Y
$\mathbb{E}(X\|Y)$	espérance conditionnelle de X sachant Y
$\mathbb{P}(A\|B)$	$= \mathbb{P}(A \cap B)/\mathbb{P}(B)$ probabilité conditionnelle de A sachant B
$X \stackrel{\mathcal{L}}{=} Y$	égalité en loi des variables aléatoires X et Y
$p_{X,y}(x)$	$= p_{X,Y}(x,y)/p_Y(y)$ densité conditionnelle de X sachant $Y = y$
$X_n \xrightarrow{\mathcal{L}} X$	convergence en loi de la suite $(X_n)_n$ vers X
$X_n \xrightarrow{p.s.} X$	convergence presque sûre de la suite $(X_n)_n$ vers X
$\beta(a,b)$	loi béta de paramètres a et b
$\mathcal{B}(n,p)$	loi binomiale de paramètres n et p
$\mathcal{B}(p)$	loi de Bernoulli de paramètre p
$\mathcal{C}(a)$	loi de Cauchy de paramètre a
$\mathcal{E}(\lambda)$	loi exponentielle de paramètre λ
$\chi^2(n)$	loi du chi 2 à n degrés de liberté

$\mathcal{N}_d(\mu, \Gamma)$	loi gaussienne en dimension d d'espérance μ et de matrice de covariance Γ
$\mathcal{P}(\lambda)$	loi de Poisson de paramètre λ
$\mathcal{U}[a, b]$	loi uniforme sur $[a, b]$
$\Gamma(a)$	$= \int_0^{+\infty} x^{a-1} e^{-x} dx$ fonction gamma d'Euler
$\Gamma(a, \theta)$	loi gamma de paramètres a et θ
$t(n)$	loi de Student de paramètre n
$\mathcal{F}_{n,m}$	loi de Fisher de paramètres n et m
$\mathcal{G}eo(p)$	loi géométrique de paramètre p
$\bar{X}_n$	$= \frac{1}{n} \sum_{j=1}^n X_j$ moyenne empirique des X_j
S_n^2	$= \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$ estimateur sans biais de la variance des X_j
H_0	hypothèse nulle
H_1	hypothèse alternative
W	région critique d'un test
$\hat{\theta}$	estimateur du paramètre θ
$I(\theta)$	$= \mathbb{E}_\theta \left(\nabla_\theta l(X_1, \theta) \nabla_\theta^* l(X_1, \theta) \right) = -\mathbb{E}_\theta \left(\nabla_\theta^2 l(X_1, \theta) \right)$ information de Fisher
$l_n(x, \theta)$	$= \sum_{i=1}^n \ln(p(x_i, \theta))$ log-vraisemblance de l'échantillon $X = (X_1, \ldots, X_n)$
$p_n(x, \theta)$	$= \prod_{i=1}^n p(x_i, \theta)$ vraisemblance de l'échantillon $X = (X_1, \ldots, X_n)$
EMV	Estimateur du Maximum de Vraisemblance