

Résumé du cours de probabilités et statistiques pour l'agrégation interne.

Lisa BALSOLLIER

Ce cours est largement basé sur le poly de **Fabrice GRELA**, que je remercie beaucoup pour son aide.

Version du 25/09/2021

Quelques références :

- [1] *Probabilités pour les non-probabilistes - 2ième édition*, Walter APPEL, H&K Éditions.
- [2] *Statistique et Probabilités - 7ième édition*, Jean-Pierre LECOUTRE, Dunod.
- [3] *Exercices et problèmes de statistique et probabilités - 2ième édition*, Thérèse PHAN et Jean-Pierre ROWENCZYK, Dunod.
- [4] *Probabilités - Préparation à l'agrégation interne*, Djalil CHAFAI - Pierre-André ZITT, HAL-01374158.
- [5] *Statistique mathématiques, cours et exercices corrigés*, Benoît CADRE - Céline VIAL, Ellipse.

Introduction

Les probabilités sont à la fois une théorie mathématique et une activité de modélisation auxquelles on a recourt lorsque la science est dans l'incapacité d'apporter une réponse prédictive à un phénomène. Par exemple, les lois de la mécanique classique ne peuvent pas prédire le résultat d'un lancer de dé puisqu'il dépend de façon non continue de nombreuses conditions initiales et d'inconnues. Les premières formalisations de la notion de hasard au XVIIIe siècle répondaient essentiellement à des questions issues de la théorie des jeux et ce n'est qu'au XXe siècle que A. Kolmogorov a rigoureusement formalisé le calcul des probabilités en s'appuyant sur une théorie de l'intégration : l'intégration de Lebesgue.

Le but des probabilités est d'associer à un événement un nombre compris entre 0 et 1 qui traduit la fréquence d'apparition de l'événement si l'on répétait un grand nombre de fois les circonstances qui l'ont produit. Dans d'autres domaines des mathématiques (par exemple l'étude des équations différentielles ordinaires), on part d'une circonstance connue précisément (les conditions initiales) et on en déduit avec certitude la situation qui en résultera. La théorie des probabilités consiste à accorder une plausibilité plus ou moins grande à un événement donné : sa probabilité. Elle cherche à prédire les résultats possibles ou les issues théoriques d'une expérience à partir d'une loi de probabilité connue et issue de la modélisation du problème.

Le domaine des statistiques a une origine très ancienne, se réduisant initialement à une collecte d'observations comme le recensement des populations. Cependant, le terme statistique n'est apparu qu'au XVIIIe siècle dans un sens purement descriptif d'une collection de données chiffrées (que l'on appelle des statistiques). La statistique évoque aujourd'hui une collection de méthodes utilisées pour étendre des résultats observés sur quelques individus à l'ensemble d'une population et dégager des lois de probabilité. Ainsi, à partir de données recueillies par le statisticien (sondage, observations...), on peut traiter les données qualitativement ou quantitativement avec des tableaux, des graphiques pour interpréter ces données (c'est la statistique descriptive) et trouver des propriétés sur la loi de probabilité inconnue sous-jacente au problème (c'est la statistique inférentielle). Une étude statistique peut également conduire à des prévisions (comme avec la régression linéaire ou le machine learning) et des prises de décisions (apprentissage statistique). On peut parler de connaissance *a posteriori* puisque l'étude se fait après la collecte de données.

Table des matières

I	Probabilités	4
1	Mesure de probabilité - Variable aléatoire - Loi de probabilité	4
1.1	Modélisation d'une expérience aléatoire	4
1.2	Équiprobabilité discrète et éléments de combinatoire	6
1.3	Probabilité conditionnelle	7
1.4	Variable aléatoire	8
1.4.1	Variable aléatoire discrète et réelle	8
1.4.2	Loi de probabilité d'une variable aléatoire	9
1.4.3	Fonction de répartition	11
1.4.4	Grandeurs associées aux variables aléatoires	12
1.4.5	Inégalité de Markov et de Bienaymé-Tchebychev	14
2	Vecteur aléatoire	15
3	Indépendance	17
3.1	Indépendance d'événements	17
3.2	Indépendance de variables aléatoires	17
3.3	Somme de variables aléatoires indépendantes	18
3.3.1	Cas des variables aléatoires discrètes	18
3.3.2	Cas des variables aléatoires à densité (Hors programme)	18
4	Lois de probabilités classiques	19
4.1	Lois discrètes	19
4.2	Lois à densité	19
5	Théorèmes limites	21
5.1	Modes de convergence des variables aléatoires	21
5.2	Approximations de lois	22
5.3	Théorèmes limites	23
5.4	Intégration numérique : méthode de Monte-Carlo	25
6	Exercices	26
II	Statistiques : Estimateur et intervalle de confiance	28
1	Modèle statistique paramétrique et estimation ponctuelle	28
2	Notion d'estimateur	29
3	Notion d'intervalle de confiance et exemple de construction d'un intervalle de confiance	32
3.1	Intervalle de confiance non asymptotique avec l'inégalité de Bienaymé-Tchebychev	32
3.2	Intervalle de confiance asymptotique avec le TCL	33
3.2.1	Première approche : la variance est majorée.	33
3.2.2	Seconde approche : avec le lemme de Slutsky (Hors programme).	34
4	Intervalle de fluctuation (Hors programme)	36
4.1	Intervalle de confiance versus intervalle de fluctuations	36
4.2	Construction d'un intervalle de fluctuation	36
4.2.1	Intervalle de fluctuation non asymptotique	36
4.2.2	Intervalle de fluctuation asymptotique	37
4.3	Erreurs possibles dans la prise de décision	37
5	Exercices	39
III	Complément : Fonction génératrice et processus de Galton-Watson	41

1	Modélisation mathématique du problème	41
2	Calcul de la probabilité d'extinction	42
3	Évolution de la taille de la population	45
IV	Éléments de correction des exercices	46
1	Probabilités	46
2	Statistiques	50

Première partie

Probabilités

1 Mesure de probabilité - Variable aléatoire - Loi de probabilité

1.1 Modélisation d'une expérience aléatoire

Définition 1.1 : Espace probabilisé, Tribu, Probabilité

Un espace probabilisé est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où

1. Ω est un ensemble appelé **univers** ;
2. \mathcal{F} est une **tribu** sur l'ensemble Ω , ie un ensemble de parties de Ω qui vérifie :
 - ▶ $\emptyset \in \mathcal{F}$,
 - ▶ Si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$,
 - ▶ Si une famille finie ou dénombrable de parties de Ω appartient à \mathcal{F} alors leur réunion aussi.
3. Une **probabilité** ou une **mesure de probabilité** \mathbb{P} sur un ensemble Ω muni d'une tribu \mathcal{F} est une application de \mathcal{F} dans \mathbb{R}^+ qui vérifie :
 - ▶ $\mathbb{P}(\Omega) = 1$,
 - ▶ Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments 2 à 2 disjoints de \mathcal{F} ,

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

Exemples (Univers) :

- a) Un lancer de dé à 6 faces :
- b) Nombre de pannes d'une photocopieuse :
- c) Durée de vie d'une ampoule :

La même expérience peut être modélisée par des univers différents :

- d) Un lancer de 2 pièces :

Remarque 1.2 : Langage probabiliste et interprétation ensembliste.

1. L'univers désigne tous les résultats possibles de l'expérience. Un élément ω de Ω est appelé **épreuve** et une partie de Ω est appelée **événement**.
2. En langage ensembliste, A ou B correspond à la réunion des parties A, B de Ω (noté $A \cup B$); A et B (simultanéité) correspond à leur intersection (noté $A \cap B$); l'événement contraire A^c (ou non- A , ou \bar{A}) au complémentaire de A et l'événement **impossible** \emptyset à la partie vide.
3. On se restreint à certaines parties de Ω dont on veut calculer la probabilité. La collection de ces parties \mathcal{F} munie de quelques propriétés de stabilité en font une tribu¹. Les éléments d'une tribu sont appelés **événements**.

1. Pourquoi ne pas considérer l'ensemble de toutes les parties de Ω ? Si on se limite aux espaces finis ou dénombrables, aucune difficulté théorique ne survient. Par contre, dans le cas non dénombrable, on peut montrer qu'on ne peut pas définir de mesure de probabilité sur $[0, 1]$ si l'on munit $[0, 1]$ de l'ensemble des parties de $[0, 1]$ (ensemble de Vitali)!

Proposition 1.3 : Propriétés d'une probabilité

1. $\mathbb{P}(\emptyset) = 0$,
2. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$,
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
4. Si $(A_n)_{n \in \mathbb{N}}$ est une famille croissante de \mathcal{F} (ie. $A_n \subset A_{n+1}$) alors $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n)$.
5. Si $(B_n)_{n \in \mathbb{N}}$ est une famille décroissante de \mathcal{F} (ie. $A_{n+1} \subset A_n$) alors $\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n)$.

Exemple : événement négligeable mais pas impossible

On dit qu'un événement N est **négligeable** si $\mathbb{P}(N) = 0$. On lance un dé non truqué à six faces jusqu'à obtenir un 6 et on considère l'événement N : "la face 6 n'est jamais obtenue". Montrer que N est négligeable.

Remarque 1.4 : Tribus usuelles

1. Pour tout ensemble Ω , la tribu la plus petite ("la plus pauvre") est la **tribu grossière** (\emptyset, Ω) ; la tribu la plus grande ("la plus riche") est la **tribu discrète** $\mathcal{P}(\Omega)$ (ie l'ensemble des parties de Ω).
2. Si $\Omega := \{\omega_1, \dots, \omega_n\}$ est un ensemble fini, on le munit de sa tribu discrète.
3. Si Ω est infini dénombrable (penser à \mathbb{N}), on le munit aussi de sa tribu discrète.
4. (*Hors-programme : cas des ensembles indénombrables*) Si $\Omega = \mathbb{R}$, on le munit de la tribu borélienne (ie la tribu engendrée par les intervalles). Il existe aussi une tribu pour les ensembles indénombrables du type $\Omega = \{0, 1\}^{\mathbb{N}}$ (expérience de pile ou face avec une infinité de lancers) appelée tribu cylindrique.

Exemple : Peut-on avoir équiprobabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$?

Pour toute la suite, on fixe un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

1.2 Équiprobabilité discrète et éléments de combinatoire

Dans cette section, on considère un **univers fini** Ω (muni de la tribu discrète $\mathcal{P}(\Omega)$ et d'une probabilité \mathbb{P}). On suppose que l'on a *équiprobabilité* sur Ω , c'est-à-dire que pour tout $\omega \in \Omega$,

$$\mathbb{P}(\{\omega\}) = \frac{1}{\text{Card}(\Omega)}.$$

N.B. Pour toute partie $A \subset \Omega$, on a

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Quatre exemples fondamentaux de cardinaux.

Rappel : Si Ω est un ensemble à n éléments, il y a $n!$ *permutations* possibles des éléments de Ω (ie $n!$ listes ordonnées qui contiennent tous les éléments de Ω sans répétition).

Exemple : Une urne contient n boules étiquetées de 1 à n . On fait p tirages. Il y a alors 4 possibilités :

	Pas de répétition (tirage sans remise)	Avec répétition (tirage avec remise)
L'ordre d'apparition compte		
L'ordre d'apparition ne compte pas		

Exemple : comment retrouver ces formules ?

Dans cet exemple, les tiroirs sont toujours différenciés les uns des autres (ils ne sont donc pas interchangeables).

- Combien y a-t-il de façons de ranger 5 objets dans 8 tiroirs de sorte qu'il n'y ait pas plus d'un objet par tiroir,
 - si les objets sont discernables (ie tous différents) ?

(b) si les objets sont indiscernables (ie tous identiques) ?

- Combien y a-t-il de façons de ranger 8 objets dans 3 tiroirs,
 - si les objets sont discernables (ie tous différents) ?

(b) si les objets sont indiscernables (ie tous identiques) ?

1.3 Probabilité conditionnelle

Définition 1.5 : Probabilité conditionnelle

Si B est un événement de probabilité non nulle et A un événement alors on définit la probabilité conditionnelle de A sachant B par

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

Proposition 1.6 : La probabilité conditionnelle est une probabilité

Si \mathbb{P} est une probabilité sur (Ω, \mathcal{F}) , $\mathbb{P}(\cdot|B)$ définit une nouvelle probabilité sur (Ω, \mathcal{F}) .

Preuve :

Proposition 1.7 : Formule des probabilités totales

Soit $(A_i)_{i \in I}$ une partition (ie. tous disjoints 2 à 2 et leur réunion fait l'univers) finie ou dénombrable d'événements de Ω de probabilité non nulle. Alors

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

Proposition 1.8 : Formule de Bayes

Soit A un événement et B un événement de probabilité non nulle. Alors :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A) \times \mathbb{P}(B|A)}{\mathbb{P}(B)}.$$

Remarque 1.9 : On utilise souvent les deux formules en même temps.

Sous les mêmes hypothèses que la proposition 1.7,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i) \times \mathbb{P}(B|A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i) \times \mathbb{P}(B|A_i)}{\sum_{i \in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Exemple : (Formule des probabilités totales - Formule de Bayes)

Une personne sur 1000 est atteinte d'une maladie. Un test déclare positif 99 % des malades qu'on lui soumet et il déclare malade 2 % des personnes saines. On choisit une personne au hasard et on la soumet au test.

On considère les événements : A : "le test est positif" et B : "la personne est malade".

Quelle est la probabilité d'être malade sachant que le test est positif? A première vue, le résultat obtenu est étonnant. L'expliquer.

1.4 Variable aléatoire

Avec la notion d'événement, on s'intéresse aux conséquences qualitatives d'une expérience aléatoire (un lancer de dé a-t-il donné un résultat pair? le nombre de pannes de la photocopieuse a-t-il dépassé 10?...). Pour étudier des conséquences quantitative d'une expérience aléatoire (quel est le résultat du lancer de dé? quel est le nombre de pannes de la photocopieuse?...), on introduit la notion de *variable aléatoire*.

1.4.1 Variable aléatoire discrète et réelle

Définition 1.10 : Variable aléatoire discrète

Soit E un ensemble au plus dénombrable. On dit qu'une application $X : \Omega \rightarrow E$ est une **variable aléatoire discrète** si

$$\forall x_i \in E, X^{-1}(x_i) := \{\omega \in \Omega : X(\omega) = x_i\} \in \mathcal{F}.$$

On notera $X(\Omega) = \{x_i, i \in I\}$ où I est fini ou dénombrable (penser à \mathbb{N}).

Remarque 1.11

En pratique, si Ω est dénombrable, on prend $\mathcal{F} = \mathcal{P}(\Omega)$. Ainsi, la seule condition pour qu'une variable aléatoire soit une variable aléatoire discrète est qu'elle soit à valeurs dans un ensemble dénombrable.

Définition 1.12 : Variable aléatoire réelle

Une **variable aléatoire réelle** est une fonction de Ω dans \mathbb{R} telle que pour tout intervalle J de \mathbb{R} , l'image réciproque $X^{-1}(J)$ appartienne à \mathcal{F} .

On rappelle que $X^{-1}(J) = \{\omega \in \Omega : X(\omega) \in J\}$.

Remarque 1.13 : Pourquoi cette définition de variable aléatoire ?

Supposer qu'un événement appartient à la tribu \mathcal{F} assure de pouvoir calculer sa probabilité. Lorsque X prend ses valeurs dans un ensemble non dénombrable (comme \mathbb{R}), il n'est pas intéressant de se restreindre à des événements du type " X prend la valeur x " comme dans le cas discret, car ces événements sont de probabilité nulle (cf Remarque 1.20) ! Pouvoir calculer la probabilité de n'importe quel intervalle est plus pertinent car les intervalles engendrent une famille très riche de sous-ensembles de \mathbb{R} : la *tribu borélienne*.

Proposition 1.14 : Admis

Soient X, Y deux variables aléatoires définies sur un même espace de probabilité et $\lambda \in \mathbb{R}$. Alors $X + Y, \lambda X, XY, \max(X, Y), \min(X, Y)$ sont des variables aléatoires. Si de plus, g est une fonction continue par morceaux de \mathbb{R} dans \mathbb{R} alors $g \circ X$ définit aussi une variable aléatoire.

Remarque 1.15 : Abus de notations

Dans la suite, on écrira souvent des quantités comme $\mathbb{P}(X = x)$, $\mathbb{P}(X \leq x)$, $\mathbb{P}(X \in J)$ (où J est un intervalle)...

Il s'agit d'abus de notations. Par exemple,

$$\mathbb{P}(X \in J) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in J\}) = \mathbb{P}(X^{-1}(J)).$$

De même, la variable aléatoire $g(X)$ est une notation pour l'application composée $g \circ X$.

Exemple : (Somme aléatoire de variables aléatoires discrètes)

Soient $(X_n)_n$ une suite de v.a. et N une v.a. toutes à valeurs dans \mathbb{N} . Montrer que $S_N := \sum_{k=0}^N X_k$ définit une variable aléatoire discrète.

1.4.2 Loi de probabilité d'une variable aléatoire

N.B. La loi de probabilité d'une variable aléatoire X est une mesure de probabilité P_X sur l'ensemble $\Omega' := X(\Omega) \subset \mathbb{R}$.

Définition 1.16 : Loi d'une variable aléatoire discrète

Soit X une variable aléatoire discrète à valeurs dans $X(\Omega) = \{x_i : i \in I\}$ avec I fini ou dénombrable. La loi de X est la fonction $P_X : X(\Omega) \rightarrow [0, 1]$ définie par

$$P_X(x_i) := \mathbb{P}(X = x_i).$$

Remarque 1.17

Comme \mathbb{P} est une probabilité, on a

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

Définition 1.18 : Densité de probabilité

On dit qu'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est une **densité de probabilité** si

► $f \geq 0$ et f est intégrable sur \mathbb{R} ;

► $\int_{\mathbb{R}} f(t)dt = 1$.

Définition 1.19 : Loi d'une variable aléatoire à densité

On dit que X possède une **loi de densité** f_X si pour tout intervalle J de \mathbb{R} ,

$$P_X(J) := \mathbb{P}(X \in J) = \int_J f_X(t)dt.$$

Remarque 1.20

Une variable aléatoire à densité est dite *sans atome*, ie que les événements du type " $X = x_0$ " pour tout $x_0 \in \mathbb{R}$ sont de mesure nulle :

$$\mathbb{P}(X = x_0) = \int_{x_0}^{x_0} f_X(t)dt = 0.$$

En particulier, pour une variable aléatoire X à densité, on a :

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in]a, b[) = \mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in [a, b]).$$

Remarque 1.21 : Hors-programme

Il existe des variables aléatoires qui ne sont ni à densité, ni discrète.

Exemples : (Lois de probabilité)

a) Loi binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}, p \in [0, 1]$

b) Loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda > 0$

1.4.3 Fonction de répartition

Définition 1.22 : Fonction de répartition

La **fonction de répartition** de la variable aléatoire X est définie par

$$F_X(x) = \mathbb{P}(X \leq x) \quad \forall x \in \mathbb{R}.$$

Proposition 1.23 : Propriétés de la fonction de répartition

Si F est une fonction de répartition alors

1. F est croissante ;
2. F est continue à droite et admet une limite à gauche ;
3. $\lim_{x \rightarrow +\infty} F(x) = 1$;
4. $\lim_{x \rightarrow -\infty} F(x) = 0$.
5. $\mathbb{P}(a \leq X \leq b) = F(b) - \lim_{x \rightarrow a^-} F(x)$ pour tout $a, b \in \mathbb{R}$.

Réciproquement, toute fonction définie sur \mathbb{R} qui satisfait les propriétés 1 à 4 est la fonction de répartition d'une variable aléatoire.

Remarque 1.24 : Atomes d'une loi.

En un point de discontinuité x , la hauteur du saut de discontinuité est égale à la probabilité que la variable aléatoire prenne la valeur x . Plus formellement, on dit que les points de discontinuité de la fonction de répartition sont les *atomes* de la loi.

Remarque 1.25 : Hors-programme - La fonction de répartition caractérise la loi

La loi d'une variable aléatoire est entièrement déterminée par sa fonction de répartition.

Allure des fonctions de répartition d'une variable aléatoire discrète et continue :

a) Cas où X est discrète

b) Cas où X est à densité

1.4.4 Grandeurs associées aux variables aléatoires

	V.A. DISCRÈTE	V.A. A DENSITÉ
Fonction de répartition	$F_X(x) = \sum_{i \in I; x_i \leq x} \mathbb{P}(X = x_i).$ <p><i>Rq</i> : La fonction de répartition d'une va discrète est constante par morceaux.</p>	$F_X(x) = \int_{-\infty}^x f_X(t) dt.$ <p><i>Rq</i> : La fonction de répartition d'une va à densité est continue et dérivable (de dérivée f_X en tout point x où f_X est continue).</p>
Espérance	<p>Si $\sum_{i \in I} x_i \mathbb{P}(X = x_i)$ converge alors X possède une espérance finie définie par</p> $\mathbb{E}(X) = \sum_{i \in I} x_i \mathbb{P}(X = x_i).$	<p>Si la fonction définie sur \mathbb{R} par $x \mapsto x f_X(x)$ est intégrable sur \mathbb{R} alors X possède une espérance définie par</p> $\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx.$
Moments	<p>Pour $\alpha \in \mathbb{N}^*$, si $\sum_{i \in I} x_i ^\alpha \mathbb{P}(X = x_i)$ cv alors X possède un moment d'ordre α.</p>	<p>Si la fonction définie sur \mathbb{R} par $x \mapsto x ^k f_X(x)$ est intégrable sur \mathbb{R} alors X possède un moment d'ordre k défini par $\int_{\mathbb{R}} x^k f_X(x) dx$.</p>
Variance	<p>Si X possède un moment d'ordre 2, on définit la variance de X par</p> $\text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2).$ <p><i>Prop</i> : $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.</p>	idem.
Lemme de transfert	<p>Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue par morceaux. $g(X)$ admet une espérance ssi la série $\sum_{i \in I} g(x_i) \mathbb{P}(X = x_i)$ cv absolument. Dans ce cas,</p> $\mathbb{E}(g(X)) = \sum_{i \in I} g(x_i) \mathbb{P}(X = x_i).$	<p>Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue par morceaux. $g(X)$ admet une espérance ssi $x \mapsto g(x) f_X(x)$ est intégrable sur \mathbb{R}. Dans ce cas,</p> $\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) dx.$
Autres	<p>Fonction génératrice : Soit X une va telle que $X(\Omega) \subset \mathbb{N}$. La fonction génératrice des moments (qui caractérise la loi de X) est définie pour tout $s \in [0, 1]$ par :</p> $G_X(s) := \mathbb{E}(s^X) = \sum_k s^k \mathbb{P}(X = k).$ <p><i>Prop</i> : X possède une espérance ssi G_X est dérivable en 1. On a :</p>	<p>Fonction caractéristique : Hors programme.</p>

Remarque 1.26 : Attention !

► Si X et Y possèdent une espérance alors $X + Y$ et λX ont une espérance pour tout $\lambda \in \mathbb{R}$. De plus, on a $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ et $\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$.

► Contrairement à l'espérance, la variance n'est pas linéaire :

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y) \quad \text{et} \quad \text{Var}(\lambda X) \neq \lambda \text{Var}(X).$$

On a par contre, pour tout $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$\text{Var}(\lambda_1 X + \lambda_2) = \lambda_1^2 \text{Var}(X).$$

► **Inégalité de Jensen** : Soit φ une fonction convexe. Alors $\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$.

Exemples : (Calculs d'espérance)

a) Espérance de la loi géométrique

b) La loi de Cauchy n'admet pas d'espérance

c) Espérance d'une indicatrice

Méthode 1.27 : Trouver la loi d'une variable aléatoire

1. Identifier $X(\Omega)$ afin de savoir si X est discrète ou à densité.
2. Si X est discrète, on peut chercher à calculer $\mathbb{P}(X = x_i)$ pour tout $i \in I$.
3. Si X est à densité, on peut utiliser le lemme de transfert et calculer $\mathbb{E}(f(X))$ où f est une fonction continue par morceaux pour trouver la loi de la variable aléatoire.
4. Dans tous les cas, comme la fonction de répartition caractérise la loi, on peut aussi calculer la fonction de répartition de X .

Exemple : Calculer la loi de $X := \frac{-1}{\lambda} \ln(U)$ où $U \sim \mathcal{U}([0, 1])$.

1.4.5 Inégalité de Markov et de Bienaymé-Tchebychev

Proposition 1.28 : *Inégalité de Markov*

Soit X une variable aléatoire positive admettant une espérance, et soit $t > 0$, alors

$$\mathbb{P}(X \geq t) \leq \frac{1}{t} \mathbb{E}(X).$$

Démonstration :

Remarque 1.29

L'inégalité de Markov quantifie la tendance naturelle d'une variable aléatoire positive à rester concentrée autour de 0.

Théorème 1.30 : *Inégalité de Bienaymé-Tchebychev*

Soit X une variable aléatoire admettant une variance. Alors, pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

Démonstration :

Remarque 1.31

L'interprétation de cette inégalité est la suivante : si une variable aléatoire admet une variance, alors elle ne peut pas "trop" s'écarter de sa moyenne.

2 Vecteur aléatoire

Définition 2.1 : Vecteur aléatoire

Un vecteur aléatoire est une application de Ω dans \mathbb{R}^p dont chacune des composantes est une variable aléatoire réelle.

	VECTEUR ALÉATOIRE DISCRET	VECTEUR ALÉATOIRE A DENSITÉ
Définition	Vecteur aléatoire dont chacune des composantes est une variable aléatoire discrète.	Une fonction de \mathbb{R}^p dans \mathbb{R}^+ est une densité sur \mathbb{R}^p si elle est intégrable sur \mathbb{R}^p d'intégrale égale à 1.
Loi d'un vecteur aléatoire	C'est la fonction $P : X_1(\Omega) \times \dots \times X_n(\Omega) \rightarrow \mathbb{R}$ telle que $P(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$.	Un vecteur aléatoire possède la loi de densité f si pour tous intervalles I_1, \dots, I_p , $\mathbb{P}((X_1 \in I_1) \cap \dots \cap (X_p \in I_p)) = \int_{I_1} \dots \int_{I_p} f(x_1, \dots, x_p) dx_1 \dots dx_p.$
Covariance	La covariance de X et Y , si elle existe, est définie par $\text{cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$ <i>Prop</i> : $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. <i>Rq</i> : $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y)$.	idem.
Coefficient de corrélation	Lorsque X et Y ont des variances non nulles, le coeff. de corrélation est défini par $\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$.	idem.

Remarque 2.2 : Interprétation de la covariance et du coefficient de corrélation.

- On peut interpréter la covariance comme la mesure du défaut d'égalité entre $\mathbb{E}(XY)$ et $\mathbb{E}(X)\mathbb{E}(Y)$.
- Un coefficient de corrélation est compris entre -1 et 1 . Un coefficient égal à ± 1 indique que X et Y sont *presque sûrement* affinement liées : il existe des scalaires a et b et une partie $N \subset \Omega$ négligeable telle que

$$\forall \omega \in \Omega \setminus N, \quad X(\omega) = aY(\omega) + b.$$

Remarque 2.3 : *Vocabulaire (loi marginale)*

La première marginale du couple (X, Y) est la loi de X , la seconde marginale est celle de Y . Les lois de probabilité de X et Y sont appelées **lois marginales** du vecteur (X, Y) .

Exemple : (Lois marginales d'un couple de variables aléatoires (X, Y))

a) Cas où X et Y sont discrètes :

b) Si le couple (X, Y) est à densité alors X et Y sont à densité et on a

Contre-exemple (Si X et Y admettent une densité, le couple (X, Y) n'admet pas forcément de densité) :

3 Indépendance

3.1 Indépendance d'événements

Définition 3.1 : Indépendance d'un ensemble fini d'événements

Soient A_1, \dots, A_m m événements.

1. A_1, \dots, A_m sont **mutuellement indépendants** si pour toute famille finie I de $\{1, \dots, m\}$, on a :

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

2. A_1, \dots, A_m sont **deux à deux indépendants** si, pour tous $i, j \in \{1, \dots, m\}$, on a :

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Remarque 3.2 : Attention !

Les notions d'événements mutuellement indépendants et d'événements deux à deux indépendants ne sont pas équivalentes.

3.2 Indépendance de variables aléatoires

	V.A. DISCRÈTE	V.A. A DENSITÉ
Définition	X_1, \dots, X_n sont indépendantes si pour tout (x_1, \dots, x_n) , les événements $(X_i = x_i)$ sont indépendants.	X_1, \dots, X_n sont indépendantes si pour tous intervalles I_1, \dots, I_p , les événements $(X_i \in I_i)$ sont indépendants.
Proposition	Si X_1, \dots, X_n sont à valeurs dans \mathbb{N} et indépendantes alors la somme $S_n = X_1 + \dots + X_n$ a pour fonction génératrice $G_{S_n}(t) = \prod_{k=1}^n G_{X_k}(t).$	Si X_1, \dots, X_n sont des variables aléatoires à densité, elles sont indépendantes ssi la fonction définie par $f(x_1, \dots, x_n) = \prod_{k=1}^n f_{X_k}(x_k)$ est la densité du vecteur (X_1, \dots, X_n) .

Proposition 3.3 : Lemme des coalitions

Si X_1, \dots, X_n sont des variables aléatoires indépendantes et si $\varphi_1, \dots, \varphi_n$ sont des fonctions de \mathbb{R} dans \mathbb{R} , alors les variables aléatoires $\varphi_1(X_1), \dots, \varphi_n(X_n)$ sont indépendantes.

Proposition 3.4 : Quelques conséquences de l'indépendance

Soient X, Y deux variables aléatoires indépendantes (discrètes ou à densité). Si les quantités suivantes existent, on a :

1. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (attention ! Réciproque fausse...);
2. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ (attention ! Réciproque fausse...);
3. $\text{cov}(X, Y) = 0$ (attention ! Réciproque fausse...).

Contre-exemple :

Théorème 3.5 : Caractérisation de l'indépendance (Hors-programme)

X_1, \dots, X_n sont des v.a. indépendantes ssi pour toutes fonctions $\varphi_1, \dots, \varphi_n : \mathbb{R} \rightarrow \mathbb{R}$ continues, on a (sous réserve que les espérances aient un sens),

$$\mathbb{E} \left[\prod_{i=1}^n \varphi_i(X_i) \right] = \prod_{i=1}^n \mathbb{E}[\varphi_i(X_i)].$$

3.3 Somme de variables aléatoires indépendantes**3.3.1 Cas des variables aléatoires discrètes**

Pour calculer la somme de deux variables aléatoires discrètes indépendantes, on peut utiliser la formule des probabilités totales puis la définition de l'indépendance pour des v.a. discrètes.

Exemple : (Somme de deux v.a. discrètes indépendantes)

Soient X et Y deux v.a. indépendantes suivant une loi géométrique de paramètre p . Déterminer la loi de $X + Y$.

3.3.2 Cas des variables aléatoires à densité (Hors programme)**Théorème 3.6**

Soient X et Y deux v.a. réelles indépendantes. On suppose que X et Y admettent des densités f_X et f_Y . Alors la v.a. $X + Y$ est à densité et

$$\forall z \in \mathbb{R}, f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx.$$

Démonstration :

4 Lois de probabilités classiques

4.1 Lois discrètes

Loi uniforme (équiprobabilité) $\mathcal{U}(\{1, \dots, n\})$

Cette loi modélise une expérience à n issues possibles, chacune ayant la même probabilité $\frac{1}{n}$. Par exemple, X suit une loi uniforme si elle représente la valeur obtenue par un lancer de dé équilibré, la valeur d'une carte tirée dans un jeu usuel (et non truqué!). . .

$$\forall k \in \{1, \dots, n\}, \mathbb{P}(X = k) = \frac{1}{n}, \mathbb{E}[X] = \frac{n+1}{2}, \text{Var}(X) = \frac{n^2-1}{12}.$$

Loi de Bernoulli $\mathcal{B}(p)$

Cette loi modélise une expérience à deux issues possibles (succès et échec), dont la probabilité de succès vaut p . Par exemple, X suit cette loi si elle représente le résultat d'un jeu du *Pile ou Face* où l'on n'a pas nécessairement la même chance de tomber sur l'un ou l'autre des côtés.

$$\mathbb{P}(X = 1) = p \text{ et } \mathbb{P}(X = 0) = 1 - p, \mathbb{E}[X] = p, \text{Var}(X) = p(1 - p).$$

Loi binomiale $\mathcal{B}(n, p)$

Cette loi modélise le nombre de succès obtenus au cours de n expériences indépendantes, chacune ayant une probabilité p de succès. Par exemple, X suit cette loi si elle représente le nombre de pile obtenus sur n lancers avec une pièce qui a une probabilité p de tomber sur pile.

$$\forall k \in \{0, \dots, n\}, \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \mathbb{E}[X] = np, \text{Var}(X) = np(1-p).$$

Loi géométrique $\mathcal{G}(p)$

Cette loi modélise le rang du premier succès dans une suite d'expériences indépendantes, chacune ayant une probabilité p de succès. Par exemple, X suit cette loi si elle représente la première fois qu'on obtient pile dans une suite de *Pile ou Face* indépendantes avec probabilité p de tomber sur pile.

$$\forall k \in \mathbb{N}^*, \mathbb{P}(X = k) = (1-p)^{k-1} p, \mathbb{E}[X] = \frac{1}{p}, \text{Var}(X) = \frac{1-p}{p^2}.$$

Loi de Poisson $\mathcal{P}(\lambda)$

Elle modélise de manière relativement satisfaisante le nombre X de clients dans certains types de file d'attente. Elle sert aussi à modéliser les occurrences d'événements peu probables. Le paramètre λ est strictement positif.

$$\forall k \in \mathbb{N}, \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \mathbb{E}[X] = \lambda, \text{Var}(X) = \lambda.$$

4.2 Lois à densité

Loi uniforme $\mathcal{U}([a, b])$

Cette loi modélise des phénomènes dont les valeurs sont uniformément réparties sur un intervalle $[a, b]$ (avec ou sans les bornes), tel que $-\infty < a < b < +\infty$.

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x), F_X(x) = \frac{x-a}{b-a} \mathbb{1}_{[a,b]}(x) + \mathbb{1}_{]b,+\infty[}(x), \mathbb{E}[X] = \frac{a+b}{2}, \text{Var}(X) = \frac{(b-a)^2}{12}.$$

Loi exponentielle $\mathcal{E}(\lambda)$

Cette loi modélise généralement des temps d'attente ou des durées de vie. Le paramètre λ est strictement positif.

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x), F_X(x) = (1 - e^{-\lambda x}) \mathbb{1}_{\mathbb{R}_+}(x), \mathbb{E}[X] = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}.$$

Loi normale $\mathcal{N}(m, \sigma^2)$

Cette loi apparaît notamment dans le cadre du théorème central limite : elle permet donc parfois d'approcher le comportement d'une grande somme de variables aléatoires indépendantes et identiquement distribuées.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \mathbb{E}[X] = m, \text{Var}(X) = \sigma^2.$$

Remarque 4.1 : Quelques propriétés de la loi normale1. *Loi normale centrée réduite* $\mathcal{N}(0, 1)$

➤ Densité d'une loi normale centrée réduite :

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Rq : f est paire et admet un maximum en 0 égal à $\frac{1}{\sqrt{2\pi}}$.

➤ Conséquences :

- (a) $\forall x \in \mathbb{R}, F_X(x) = 1 - F_X(-x)$;
- (b) $\forall x > 0, \mathbb{P}(|X| \leq x) = 2F_X(x) - 1$.

En effet,

2. *Centrer - Réduire*

A partir d'une loi normale quelconque, on peut se ramener à une loi centrée réduite.

En effet,

$$Y \sim \mathcal{N}(m, \sigma^2) \text{ ssi } \frac{Y - m}{\sigma} \sim \mathcal{N}(0, 1).$$

3. *Somme de lois normales indépendantes*

Si $Y_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $Y_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ sont deux v.a. indépendantes alors $Y_1 + Y_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ (cf Théorème 3.6).

Idée de preuve :

5 Théorèmes limites

5.1 Modes de convergence des variables aléatoires

Définition 5.1 : Convergence presque sûre - en probabilité - en loi

1. On dit qu'une suite $(X_n)_n$ de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ converge **presque sûrement** vers X également définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ si

$$\mathbb{P}(\{\omega \in \Omega; \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}) = 1.$$

Notation : $X_n \rightarrow X$ p.s.

2. On dit qu'une suite $(X_n)_n$ de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ converge **en probabilité** vers X également définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ si pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\{\omega \in \Omega; |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

Notation : $X_n \xrightarrow{\mathbb{P}} X$.

3. **Convergence en loi**

► *Une première définition*

Soit $(X_n)_n$ une suite de variables aléatoires et soit X une variable aléatoire. On dit que $(X_n)_n$ converge en loi vers X si, en notant F_n la fonction de répartition de X_n et F celle de X , en tout point où F est continue, on a :

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

► *Une deuxième définition*

Soit $(X_n)_n$ une suite de variables aléatoires et soit X une variable aléatoire. On suppose que X et les X_n sont toutes à valeurs dans le même espace $E = \mathbb{R}^d$. On dit que $(X_n)_n$ converge en loi vers X si, pour toute fonction $f : E \rightarrow \mathbb{R}$ continue et bornée, on a

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)).$$

Notation : $X_n \xrightarrow{\mathcal{L}} X$.

Remarque 5.2 : Convergence en loi vers une constante

Dire que la suite $(X_n)_n$ converge en loi vers $c \in \mathbb{R}$ a bien un sens : d'après la définition de la convergence en loi, cela revient à dire que

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x),$$

où, pour tout $n \in \mathbb{N}^*$, F_n est la fonction de répartition de X_n et F est définie, pour tout $x \in \mathbb{R}$ par

$$F(x) = \mathbb{P}(c \leq x) = \mathbf{1}_{c \leq x}.$$

Autrement dit, on peut dire que X_n converge en loi vers X où X est constante égale à c . La loi d'une telle variable aléatoire X est appelée loi de Dirac en c , c'est-à-dire que $\mathbb{P}(X = c) = 1$.

Remarque 5.3 : Attention !

La convergence en loi n'implique aucune relation de "proximité" entre les variables aléatoires X_n et la variable X . Pour être rigoureux, on devrait dire, la loi de X_n converge vers la loi de X . La convergence en loi n'est pas une convergence de variables aléatoires contrairement à la convergence presque sûre ou en probabilité; contrairement à ces autres modes de convergence, les énoncés suivants ne sont pas équivalents :

$$X_n \xrightarrow{\mathcal{L}} X \text{ et } X_n - X \xrightarrow{\mathcal{L}} 0.$$

	V.A. DISCRÈTE	V.A. A DENSITÉ
Propriétés de la cv en loi	$X_n \xrightarrow{\mathcal{L}} X$ ssi $\mathbb{P}(X_n = x_i) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(X = x_i),$ pour tout $x_i \in X(\Omega)$ ssi $\forall t \in [0, 1[, G_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} G_X(t).$	Lemme de Scheffé : On note f_n et f les fonctions de densité des X_n et de X . Si $f_n \xrightarrow[n \rightarrow \infty]{} f$ presque partout alors $X_n \xrightarrow{\mathcal{L}} X.$

Proposition 5.4 : *Convergence et image de variables aléatoires*

Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue.

1. Si $X_n \xrightarrow[n \rightarrow +\infty]{} X$ p.s. alors $g(X_n) \xrightarrow[n \rightarrow +\infty]{} g(X)$ p.s.
2. Idem pour les convergences en probabilité et en loi.

Théorème 5.5 : *Lien entre les modes de convergence - Admis*

Convergence presque sûr \implies Convergence en probabilité \implies Convergence en loi.

5.2 Approximations de lois

Théorème 5.6 : *Approximation de la loi de Poisson par la loi binomiale*

Soit $(p_n)_{n \geq 1}$ une suite de réels de $]0, 1[$ telle que $np_n \xrightarrow[n \rightarrow \infty]{} \lambda > 0$. Soit X_n des variables aléatoires indépendantes de loi $\mathcal{B}(n, p_n)$. Alors $X_n \xrightarrow{\mathcal{L}} X$ où $X \sim \mathcal{P}(\lambda)$, ie pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \xrightarrow[n \rightarrow \infty]{} e^{-\lambda} \frac{\lambda^k}{k!}$.

Une conséquence de ce théorème est l'approximation suivante :

"Corollaire" 5.7 : *Approximation de la loi binomiale par la loi de Poisson*

On peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$ lorsque n est assez grand et p prend des petites valeurs (si $n \geq 20$ et $p \leq 0,05$ ou bien si $n \geq 100$ et $np \leq 10$).

Une autre approximation de la loi binomiale peut être faite avec des lois normales.

Théorème 5.8 : *Théorème de Moivre-Laplace*

Soit $p \in]0, 1[$ et soit (X_n) une suite de variables aléatoires suivant une loi binomiale $B(n, p)$. Alors la variable aléatoire

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

converge en loi vers une variable aléatoire de loi normale $\mathcal{N}(0, 1)$.

Le théorème de Moivre-Laplace est un cas particulier du théorème central limite (donné dans le prochain paragraphe).

Une conséquence de ce théorème est l'approximation suivante :

"Corollaire" 5.9 : *Approximation de la loi binomiale par la loi normale*

On peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, np(1-p))$ lorsque n est assez grand ($n \geq 30$) et p pas trop proche de 0 ou 1 ($np \geq 5$ et $n(1-p) \geq 5$).

5.3 Théorèmes limites

Rappels : On a :

$$\limsup_{n \rightarrow +\infty} A_n := \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k,$$

c'est-à-dire :

$$\omega \in \limsup_{n \rightarrow +\infty} A_n \iff \forall n \geq 0, \exists k \geq n \text{ tel que } \omega \in A_k$$

Proposition 5.10 : *Lemmes de Borel-Cantelli*

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.

1. Si la série $\sum \mathbb{P}(A_n)$ converge alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$, ie presque sûrement, seul un nombre fini d'événements A_n se réalisent.
2. Supposons que les événements A_n soient deux à deux indépendants. Si la série $\sum \mathbb{P}(A_n)$ diverge alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$ ie, presque sûrement, une infinité d'événements A_n se réalisent.

Remarque 5.11

Les lemmes de Borel-Cantelli donnent un critère univoque pour la réalisation d'une infinité d'événements A_n .

Exemple : (Pile ou Face)

Proposition 5.12 : Loi faible des grands nombres

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes, de même espérance m et de même variance σ^2 . Alors la suite de variables aléatoires de terme général

$$Y_n = \frac{X_1 + \dots + X_n}{n}$$

converge en probabilité vers (la variable aléatoire constante égale à) m :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(|Y_n - m| \geq \varepsilon) = 0.$$

Démonstration :

Théorème 5.13 : Loi forte des grands nombres - Admis

Soit X une variable aléatoire. Soit $(X_n)_n$ une suite de variables aléatoires indépendantes, admettant une espérance et de même loi que X . Soit m un réel. Alors la suite de variables aléatoires de terme général

$$Y_n = \frac{X_1 + \dots + X_n}{n}$$

converge presque sûrement vers m ssi X est intégrable et $\mathbb{E}(X) = m$.

Théorème 5.14 : Théorème central limite

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes, de même loi, et admettant une variance $\sigma^2 > 0$. Notons $m = \mathbb{E}(X_1)$ leur espérance. Posons enfin, $S_n = X_1 + \dots + X_n$. Alors

$$\sqrt{n} \left(\frac{S_n}{n} - m \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Remarque 5.15

La loi des grands nombres décrit l'évolution moyenne d'une expérience répétée une infinité de fois. En pratique, on ne peut pas réitérer une infinité de fois une expérience. La question naturelle qui se pose est alors de quantifier la vitesse de cette convergence. Le théorème central limite répond à cette question. Plus précisément, il quantifie le comportement, qui est alors gaussien, des fluctuations de la moyenne pour des v.a. indépendantes et identiquement distribuées.

5.4 Intégration numérique : méthode de Monte-Carlo

Soit $\varphi : [a; b] \rightarrow \mathbb{R}$ une fonction définie sur un segment, intégrable. L'objectif est d'évaluer numériquement l'intégrale $I = \int_a^b \varphi(x)dx$. Il existe plusieurs méthodes déterministes (par exemple la méthode des rectangles, les méthodes de quadratures...). Nous allons présenter ici une méthode probabiliste basée sur la loi des grands nombres.

Si X_1 est une variable aléatoire uniforme sur $[a; b]$, alors, par le théorème de transfert, $\varphi(X_1)$ admet une espérance et

$$\mathbb{E}(\varphi(X_1)) = \int_a^b \varphi(x)f(x)dx = \frac{1}{b-a} \int_a^b \varphi(x)dx,$$

car la densité f de X_1 est constante et égale à $1/(b-a)$ sur $[a; b]$, nulle en dehors. On obtient alors :

$$\int_a^b \varphi(x)dx = (b-a)\mathbb{E}(\varphi(X_1)).$$

Considérons à présent $(X_n)_n$ une suite de variables aléatoires indépendantes et uniformément distribuées sur $[a; b]$. D'après la loi forte des grands nombres,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \varphi(X_k) = \mathbb{E}(\varphi(X_1)) \text{ p.s.}$$

On a donc presque sûrement

$$\int_a^b \varphi(x)dx = (b-a)\mathbb{E}(\varphi(X_1)) = \lim_{n \rightarrow +\infty} \frac{b-a}{n} \sum_{k=1}^n \varphi(X_k).$$

Autrement dit, la loi des grands nombres permet d'obtenir une estimation de $\mathbb{E}(\varphi(X_1))$ par l'évaluation d'une moyenne empirique. Lorsque n est assez grand, on a en pratique

$$\mathbb{E}(\varphi(X_1)) \approx \frac{1}{n} \sum_{k=1}^n \varphi(X_k)$$

et donc

$$\int_a^b \varphi(x)dx \approx \frac{b-a}{n} \sum_{k=1}^n \varphi(X_k).$$

Remarque 5.16 : Effet de la dimension

La vitesse de convergence de cette méthode peut être donnée par le théorème central limite. Le point remarquable dans cette méthode est la vitesse de convergence en $1/\sqrt{n}$ quelque soit la dimension dans laquelle on travaille ! Si cette méthode n'est pas très rapide en faible dimension (la méthode des rectangles pour l'intégrale I précédente est plus performante car sa vitesse de convergence est en $1/n$!), elle devient intéressante en dimension supérieure (plus grande que 4).

6 Exercices

Exercice 1 : Modéliser une expérience

Deux individus A et B jouent au jeu suivant.

Le joueur A dispose de trois boîtes et met une pièce de monnaie dans l'une des boîtes. Le joueur B n'ayant pas vu cette opération essaie de deviner où se trouve la pièce en proposant à A l'une des boîtes.

Le joueur A indique alors à B l'une des deux boîtes que B n'a pas choisies et où il est sûr que la pièce ne se trouve pas.

B doit-il maintenir son premier choix ou se reporter sur l'autre boîte ?

Après avoir défini un espace probabilisé qui décrit cette expérience, répondre à la question.

Exercice 2 : Variable aléatoire discrète sans mémoire et loi géométrique

1. Si X suit une loi géométrique, montrer que X est sans mémoire ie :

$$\forall k, l \in \mathbb{N}^*, \mathbb{P}(X > k + l \mid X > k) = \mathbb{P}(X > l).$$

2. Si X est à valeurs entières strictement positives telles que pour tout $k \in \mathbb{N}^*$, $\mathbb{P}(X = k) > 0$ et telle que pour tout $k, l \in \mathbb{N}^*$, $\mathbb{P}(X > k + l \mid X > k) = \mathbb{P}(X > l)$, montrer que X suit une loi géométrique.

Exercice 3 : Variable aléatoire à densité sans mémoire et loi exponentielle

Soit X une variable à densité, à valeurs positives, et telle que pour tout $x > 0$, $\mathbb{P}(X > x) > 0$. Montrer que la loi de X est une variable sans mémoire ssi X suit une loi exponentielle.

Si X modélise la durée de vie d'un individu (et suit une loi exponentielle), comment interprétez-vous le fait que X soit sans mémoire ?

Exercice 4 : Théorème de Weierstrass par les polynômes de Bernstein

Soit $f : [0, 1] \rightarrow \mathbb{R}$ une fonction continue. L'objectif de cet exercice est de montrer le théorème d'approximation de Weierstrass : il existe une suite $(B_n)_n$ de polynômes qui converge uniformément vers f .

Pour tout $n \in \mathbb{N}^*$, on définit le n -ième polynôme de Bernstein de f :

$$B_n(X) = \sum_{k=1}^n \binom{n}{k} X^k (1-X)^{n-k} f\left(\frac{k}{n}\right).$$

Soit $x \in [0, 1]$. On considère $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées de

loi de Bernoulli de paramètre x . On note $S_n = \sum_{k=1}^n X_k$.

1. Calculer l'espérance de $f(S_n/n)$.
2. En déduire une majoration uniforme de $|f(x) - B_n(x)|$.
3. Conclure.

Exercice 5 : Caractérisation de l'existence d'un moment d'ordre 1 pour une v.a. discrète

Soit X une v.a. à valeurs dans \mathbb{N} . Montrer que X admet une espérance finie ssi $\sum \mathbb{P}(X > n)$ converge.

Exercice 6 : Quelques calculs autour de la loi de Poisson

Soit X une v.a. de loi de Poisson de paramètre $\lambda > 0$.

1. Trouver la loi de

$$Y = \begin{cases} X/2 & \text{si } X \text{ est pair} \\ \frac{(1-X)}{2} & \text{si } X \text{ est impair.} \end{cases}$$

2. Calculer les moments factoriels pour tout $n \geq 0$,

$$m_n := \mathbb{E}[X(X-1)\dots(X-n+1)].$$

Exercice 7 : *Quelques calculs autour de lois continues*

1. Soit X une v.a. de loi exponentielle de paramètre $\theta > 0$. Déterminer la loi de $Y = \sqrt{X}$.
2. Soit X une variable aléatoire de loi uniforme sur $[0, 1]$.
 - (a) Déterminer la loi et l'espérance de $Y = \min(X, 1 - X)$, ainsi que $Z = \max(X, 1 - X)$.
 - (b) Calculer, si elles existent, les espérances de Y/Z et de Z/Y .

Exercice 8 : *Couple de variables aléatoires discrètes*

Soient X, Y deux v.a. à valeurs dans \mathbb{N} . On suppose que la loi conjointe de X et Y est donnée par $\mathbb{P}(X = j, Y = k) = \frac{a}{j!k!}$ pour tout $j, k \in \mathbb{N}$ et avec $a \in \mathbb{R}$.

1. Déterminer a .
2. Calculer la loi marginale de X .
3. X et Y sont-elles indépendantes?

Exercice 9 : *Couple de variables aléatoires à densité*

Soient X et Y deux variables aléatoires indépendantes de lois exponentielles de paramètre respectif $\lambda > 0$ et $\mu > 0$. Calculer

$$\mathbb{P}(X > Y).$$

Exercice 10 : *Identité de Wald*

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires discrètes, identiquement distribuées, et N une variable aléatoire à valeurs entières positives. On suppose que N et les X_k sont mutuellement indépendantes et qu'elles admettent une espérance. On pose

$$S_N := \sum_{k=1}^N X_k,$$

où si $N(\omega) = 0$ alors $S_N(\omega) = 0$. Montrer que $\mathbb{E}(S_N) = \mathbb{E}(N)\mathbb{E}(X_1)$.

Exercice 11 : *Truquer des dés*

On lance deux dés à 6 faces. On note S la variable aléatoire égale à la somme des deux faces. On va montrer qu'il n'est pas possible de truquer les deux dés (pas nécessairement de la même façon) de telle sorte que S suive une loi uniforme sur $\{2, 3, \dots, 12\}$.

1. Soit S une v.a. uniforme sur $\{2, 3, \dots, 12\}$. Calculer la fonction génératrice G_S de S . Puis montrer qu'il existe un polynôme R de degré 10 et sans racine réelle telle que $G_S(z) = z^2 R(z)$ pour tout $z \in [0, 1]$.
2. En raisonnant par l'absurde, conclure.

Exercice 12 : *Somme de variables aléatoires indépendantes*

Soit X et Y deux v.a. indépendantes suivant une loi de Poisson de paramètres respectifs $\lambda > 0$ et $\mu > 0$. Quelle est loi de $X + Y$?

Exercice 13 : *Lemmes de Borel-Cantelli et convergence presque sûr*

Soit $(X_n)_n$ une suite de variables aléatoires telle que pour tout $\varepsilon > 0$, la série $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon)$ converge. Montrer que la suite $(X_n)_n$ converge p.s. vers 0. La réciproque est-elle vraie?

Exercice 14 : *Convergence en loi*

Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes, de même loi uniforme $\mathcal{U}([0, 1])$. On note $M_n := \max(X_1, \dots, X_n)$ et $Y_n := n(1 - M_n)$ pour tout $n \geq 1$. Calculer la fonction de répartition de Y_n et montrer que la suite $(Y_n)_{n \geq 1}$ converge en loi vers une loi à préciser.

Deuxième partie

Statistiques : Estimateur et intervalle de confiance

1 Modèle statistique paramétrique et estimation ponctuelle

La démarche statistique est la suivante : on dispose de n réels x_1, \dots, x_n qui sont des valeurs observées de variables aléatoires X_1, \dots, X_n , de lois inconnues. En se basant sur ces observations, on veut "trouver" leurs lois de probabilité ou un paramètre qui intervient dans leurs lois de probabilité.

Définition 1.1 : *Modèle statistique paramétrique*

Un **modèle statistique** pour n observations est un triplet $(\chi^n, \mathcal{A}, \mathcal{P})$ où

1. χ est l'espace des résultats ou des observations ;
2. \mathcal{A} est une tribu sur χ^n ;
3. \mathcal{P} est une famille de lois de probabilités sur l'espace χ^n muni de la tribu \mathcal{A} .

On parle de modèle statistique **paramétrique** lorsque \mathcal{P} est paramétré par un ensemble Θ (appelé espace des paramètres) inclus dans un espace vectoriel de dimension finie :

$$\mathcal{P} = \{\mathbb{P}_\theta; \theta \in \Theta \subset \mathbb{R}^p\}.$$

Remarque 1.2 : *Observations indépendantes*

Si on dispose de n observations indépendantes, alors la famille \mathcal{P} a la forme $\mathcal{P} = \{\mathbb{P}^{\otimes n}, \mathbb{P} \in \mathcal{Q}\}$, où \mathcal{Q} est une famille de lois sur χ .

Si de plus, le modèle est paramétrique, alors la famille \mathcal{P} aura la forme $\mathcal{P} = \{\mathbb{P}_\theta^{\otimes n}; \theta \in \Theta \subset \mathbb{R}^p\}$.

Un cas important, que nous rencontrons fréquemment, est celui où les observations (x_1, \dots, x_n) sont issues de n répétitions indépendantes d'une même expérience. Les variables aléatoires X_1, \dots, X_n dont elles proviennent forment ce qu'on appelle un n -échantillon.

Définition 1.3 : *n -échantillon*

On appelle **n -échantillon** un n -uplet (X_1, \dots, X_n) de variables aléatoires indépendantes et identiquement distribuées dont on observe une réalisation.

Dans la plupart des cas, en plus du fait que les observations soient indépendantes, le modèle est paramétrique. Autrement dit, à partir des observations (x_1, \dots, x_n) issues d'une réalisation d'un n -échantillon (X_1, \dots, X_n) de variables aléatoires (indépendantes, de même loi et) à valeurs dans χ , on cherche à estimer la valeur de $\theta \in \Theta$ et donc la loi $\mathbb{P}_\theta \in \mathcal{P}$ qui s'ajuste le mieux aux observations. De manière plus générale, on peut aussi chercher à estimer une fonction $g(\theta)$ du paramètre $\theta \in \Theta$.

Remarque 1.4 : *Identifiabilité*

Si θ_1 et θ_2 sont deux paramètres de Θ tels que $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$, on ne peut pas distinguer θ_1 et θ_2 au vu du résultat de l'expérience uniquement. Pour éviter cela, on suppose que le modèle statistique est **identifiable**, ie que l'application $\theta \mapsto \mathbb{P}_\theta$ est injective.

Exemple 1.5

Une urne contient des boules rouges en proportion p et des boules noires en proportion $1 - p$. On tire n boules avec remise. Les observations sont ici les tirages de boules dans l'urne. On note :

$$X_i = \begin{cases} 1 & \text{si la boule tirée au } i^{\text{ème}} \text{ tirage est rouge} \\ 0 & \text{sinon.} \end{cases}$$

Le paramètre qui nous intéresse est p . Le modèle statistique de cette expérience est donc :

$$(\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), \{\mathcal{B}er(p)^{\otimes n}, p \in [0, 1]\})$$

2 Notion d'estimateur**Définition 2.1 : Estimateur, Biais**

1. Un **estimateur** \hat{T}_n de $g(\theta)$ est une fonction de X_1, \dots, X_n , qui s'exprime indépendamment de θ .
2. Le **biais** de \hat{T}_n est $b(\hat{T}_n) := \mathbb{E}(\hat{T}_n) - g(\theta)$.
3. Un estimateur est dit **sans biais** si son biais est nul.
4. Un estimateur \hat{T}_n est dit **asymptotiquement sans biais** si $\lim_{n \rightarrow +\infty} \mathbb{E}(\hat{T}_n) = g(\theta)$.

Exemple 2.2

On reprend l'exemple 1.5 ci-dessus. $\hat{T}_n = \frac{X_1 + \dots + X_n}{n}$ est un estimateur de p . De plus, $\mathbb{E}(\hat{T}_n) = p$ (par indépendance des X_i) donc \hat{T}_n est un estimateur sans biais de p .

Remarque 2.3

Le biais fait intervenir l'espérance de l'estimateur. Vouloir un estimateur sans biais, c'est demander, "qu'en moyenne", l'estimateur soit égale à la quantité estimée. On veut cependant davantage : quand n tend vers l'infini, on voudrait que \hat{T}_n "converge" vers $g(\theta)$.

Définition 2.4 : Estimateur convergent

On dit que \hat{T}_n est un **estimateur convergent (ou consistant)** de $g(\theta)$ si $\hat{T}_n \xrightarrow{\mathbb{P}} g(\theta)$ ie

$$\mathbb{P}(|\hat{T}_n - g(\theta)| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0, \quad \forall \varepsilon > 0.$$

Si de plus, $\hat{T}_n \xrightarrow{\text{P.s.}} g(\theta)$, alors on dit que \hat{T}_n est un **estimateur fortement convergent (ou fortement consistant)**.

Remarque 2.5

Pour prouver la convergence d'un estimateur, on pourra utiliser l'inégalité de Bienaymé-Tchebychev ou prouver que \hat{T}_n est fortement convergent.

Définition 2.6 : *Risque quadratique*

Si l'estimateur \hat{T}_n admet un moment d'ordre 2, on appelle **risque quadratique** de \hat{T}_n la quantité

$$R(T_n, \theta) := \mathbb{E}[(T_n - g(\theta))^2].$$

Remarque 2.7

► Le risque quadratique quantifie les fluctuations que peut présenter un estimateur autour de sa valeur moyenne. Plus le risque quadratique est petit, plus l'estimateur est "performant". Le risque quadratique est donc un critère permettant de comparer deux estimateurs d'un même paramètre.

► On peut montrer que

$$R(\hat{T}_n, \theta) = \text{Var}(\hat{T}_n) + b(\hat{T}_n)^2.$$

Cette formule s'appelle la *décomposition biais-variance*.

Proposition 2.8

Si le risque quadratique d'un estimateur \hat{T}_n tend vers 0 alors \hat{T}_n est un estimateur convergent de $g(\theta)$.

Preuve : C'est une conséquence de l'inégalité de Markov :

Exemple 2.9

On reprend l'exemple 1.5. On cherche le risque quadratique de \hat{T}_n . On a vu que $b(\hat{T}_n) = 0$. De plus, $\text{Var}(\hat{T}_n) = \frac{\text{Var}(X_1) + \dots + \text{Var}(X_n)}{n^2} = \frac{p(1-p)}{n}$. Ainsi,

$$R(\hat{T}_n, p) = \frac{p(1-p)}{n}.$$

De plus, $R(\hat{T}_n, p) \xrightarrow{n \rightarrow +\infty} 0$ donc \hat{T}_n est convergent.

L'estimateur que nous avons utilisé dans cet exemple est en fait un estimateur bien connu appelé **moyenne empirique**.

Exemple 2.10 : *Moyenne empirique*

Soit X une variable aléatoire réelle de loi donnée, admettant une espérance $m := \mathbb{E}(X)$, qui est le paramètre que l'on cherche à estimer. Soit un n -échantillon (X_1, \dots, X_n) de même loi que X . Alors :

$$\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$$

est un estimateur de m appelé **moyenne empirique**.

On en déduit facilement les résultats suivants :

- \bar{X}_n est un estimateur sans biais de m ;
- \bar{X}_n est un estimateur fortement convergent de m (conséquence de la Loi forte des Grands Nombres).

Exemple 2.11 : Variance empirique

Si cette fois le paramètre que l'on cherche à estimer n'est plus l'espérance de la loi commune du n -échantillon (X_1, \dots, X_n) considéré mais sa variance σ^2 , alors une estimation "naturelle" de ce paramètre σ^2 est le calcul de la moyenne des écarts quadratiques à la moyenne observée :

$$\hat{V}_n := \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \stackrel{(1)}{=} \left(\frac{1}{n} \sum_{k=1}^n X_k^2 \right) - \bar{X}_n^2.$$

Cet estimateur s'appelle la **variance empirique**. On a le résultat suivant :

$$\mathbb{E}(\hat{V}_n) = \frac{n-1}{n} \sigma^2. \quad (2)$$

L'estimateur \hat{V}_n est donc biaisé mais asymptotiquement sans biais. De plus, il est convergent (3). On en déduit que l'estimateur :

$$\hat{S}_n := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

appelé **variance empirique corrigée**, est un estimateur sans biais et convergent de la variance σ^2 .

Preuve :

3 Notion d'intervalle de confiance et exemple de construction d'un intervalle de confiance

Définition 3.1 : Intervalle de confiance

Soit un nombre $\alpha \in (0, 1)$ et un n -échantillon $X = (X_1, \dots, X_n)$ dont la loi dépend d'un paramètre inconnu θ . Un **intervalle de confiance** pour le paramètre θ de niveau $1 - \alpha$ (mais il faut comprendre "de probabilité de confiance $1 - \alpha$ ") est un intervalle qui dépend de l'échantillon mais pas du paramètre inconnu θ tel que la probabilité que cet intervalle contienne θ soit égale à $1 - \alpha$.

Autrement dit, un intervalle de confiance au niveau $1 - \alpha$ est un intervalle $I_n(X)$ telle que

$$\mathbb{P}_\theta(\theta \in I_n(X)) = 1 - \alpha.$$

On parle d'**intervalle de confiance asymptotique** de niveau $1 - \alpha$, si

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta(\theta \in I_n(X)) = 1 - \alpha.$$

Remarque 3.2

► L'intervalle de confiance est aléatoire car il dépend de l'observation.

► Les notations ci-dessus sont des abus de langage : l'interprétation " θ a une probabilité $1 - \alpha$ d'appartenir à $I_n(X)$ " est fautive : le paramètre θ est déterministe et l'aléa porte sur l'observation $X = (X_1, \dots, X_n)$. Il faudrait écrire :

$$\mathbb{P}_\theta(I_n(X) \ni \theta) = 1 - \alpha,$$

et comprendre "la probabilité que l'intervalle aléatoire $I_n(X)$ contienne θ est égale à $1 - \alpha$ ".

Problème : Une pièce a une probabilité inconnue $\theta_0 \in]0, 1[$ de tomber sur pile.

Observations : On note (x_1, \dots, x_n) les résultats de n lancers de pièce (en notant $x_i = 1$ si le i -ième lancer a donné pile, 0 dans le cas contraire).

L'observation (x_1, \dots, x_n) est une réalisation d'un n -uplet de variables aléatoires (X_1, \dots, X_n) à valeurs dans l'espace $\{0, 1\}^n$.

Les conditions de l'expérience apportent des informations sur ce n -uplet : les variables aléatoires X_1, \dots, X_n sont indépendantes (car les lancers de pièce le sont) et sont de même loi de Bernoulli de paramètre inconnu θ_0 . Comme on ne connaît pas θ_0 , la seule affirmation raisonnable est que (X_1, \dots, X_n) est un n -échantillon de variables aléatoires de loi de Bernoulli $\mathcal{B}er(\theta)$ pour un certain $\theta \in]0, 1[$.

Le modèle statistique de l'expérience est donc :

$$(\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), \{\mathcal{B}er(\theta), \theta \in]0, 1[\}).$$

Objectif : Construire un intervalle de confiance de niveau 95 % pour le paramètre θ .

Données numériques de l'expérience : Sur $n = 1000$ lancers, on a obtenu 520 piles.

3.1 Intervalle de confiance non asymptotique avec l'inégalité de Bienaymé-Tchebychev

On dispose d'un n -échantillon (X_1, \dots, X_n) de variables aléatoires indépendantes et identiquement distribuées de loi de Bernoulli $\mathcal{B}er(\theta)$ et l'on souhaite estimer $\theta \in]0, 1[$.

On va commencer par construire un estimateur de θ . On sait que, si $X \sim \mathcal{B}er(\theta)$, alors :

$$\mathbb{E}(X) = \theta.$$

On en déduit que la moyenne empirique

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

est un estimateur (sans biais) de θ . Les X_i ayant un moment d'ordre 2, d'après l'inégalité de Bienaymé-Tchebychev,

$$\begin{aligned} \forall \varepsilon > 0, \quad \mathbb{P}(|\bar{X}_n - \theta| \geq \varepsilon) &\leq \frac{\theta(1-\theta)}{n\varepsilon^2} \quad \text{car } \text{Var}(\bar{X}_n) = \frac{n\text{Var}(X_1)}{n^2} = \frac{\theta(1-\theta)}{n} \\ &\leq \frac{1}{4n\varepsilon^2} \quad \text{car } x \mapsto x(1-x) \text{ admet sur } [0, 1] \text{ un maximum en } 1/2 \text{ valant } 1/4. \end{aligned}$$

On en déduit ainsi un intervalle de confiance de niveau $1 - \alpha$ en considérant,

$$I_1 = \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}} \right].$$

En effet,

Application numérique :

Pour les arrondis dans les applications numériques : il faut choisir $I_{\text{théo}} \subset I_{\text{approché}}$ car dans ce cas, on a $\{\theta \in I_{\text{théo}}\} \subset \{\theta \in I_{\text{approché}}\}$ et donc $1 - \alpha \leq \mathbb{P}(\theta \in I_{\text{théo}}) \leq \mathbb{P}(\theta \in I_{\text{approché}})$.

Ici, on obtient :

$$\theta_0 \in I_1 = [0.44; 0.60]$$

avec un niveau de confiance au moins égal à 95 %. On parle ici d'**intervalle de confiance par excès** (car $\mathbb{P}_\theta(\theta \in I_1) \geq 1 - \alpha$.)

3.2 Intervalle de confiance asymptotique avec le TCL

On suppose ici que n est grand, ie que l'on a réalisé un grand nombre d'observations ($n \geq 30$ en pratique).

D'après le TCL, on sait que

$$\sqrt{\frac{n}{\text{Var}(X_1)}} (\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

avec

$$\text{Var}(X_1) = \theta(1 - \theta).$$

3.2.1 Première approche : la variance est majorée.

On note F la fonction de répartition d'une loi normale $Z \sim \mathcal{N}(0, 1)$. Pour tout $q \in \mathbb{R}$, le TCL (+la composition par la fonction valeur absolue qui conserve la convergence en loi) fournit :

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{\frac{n}{\theta(1-\theta)}} |\bar{X}_n - \theta| \leq q \right) &= \mathbb{P}(|Z| \leq q) \\ &= F(q) - F(-q) \\ &= 2F(q) - 1. \end{aligned}$$

On choisit $q := q_\alpha \in \mathbb{R}$ de sorte que

$$F(q_\alpha) = \mathbb{P}(Z \leq q_\alpha) = 1 - \frac{\alpha}{2}.$$

Remarque 3.3 : Quantile

- ▶ q_α est appelé **quantile** d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$.
- ▶ Pour les applications numériques, si $\alpha = 5\%$ alors $q_\alpha = 1,96$.

Comme $\theta \in]0, 1[$, on remarque que

$$\text{Var}(X_1) = \theta(1 - \theta) \leq 1/4. \quad (1)$$

On a alors

$$\begin{aligned} 1 - \alpha &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{\frac{n}{\theta(1 - \theta)}} |\bar{X}_n - \theta| \leq q_\alpha \right) \\ &\leq \lim_{n \rightarrow +\infty} \mathbb{P} (2\sqrt{n} |\bar{X}_n - \theta| \leq q_\alpha) \quad \text{avec (1)} \\ &= \lim_{n \rightarrow +\infty} \mathbb{P} (-q_\alpha \leq 2\sqrt{n}(\bar{X}_n - \theta) \leq q_\alpha) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\bar{X}_n - \frac{q_\alpha}{2\sqrt{n}} \leq \theta \leq \bar{X}_n + \frac{q_\alpha}{2\sqrt{n}} \right) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\theta \in \left[\bar{X}_n - \frac{q_\alpha}{2\sqrt{n}}; \bar{X}_n + \frac{q_\alpha}{2\sqrt{n}} \right] \right) \end{aligned}$$

On en déduit alors l'intervalle de confiance asymptotique par excès de niveau $1 - \alpha$ pour θ :

$$I_2 = \left[\bar{X}_n - \frac{q_\alpha}{2\sqrt{n}}, \bar{X}_n + \frac{q_\alpha}{2\sqrt{n}} \right].$$

Application numérique :

On obtient avec un niveau asymptotique de niveau de 95 % :

$$\theta_0 \in I_2 = [0.48; 0.56].$$

Remarque : On a $I_2 \subset I_1$ donc on privilégie I_2 , qui est un intervalle plus précis (car n est grand).

3.2.2 Seconde approche : avec le lemme de Slutsky (Hors programme).

La variance de X_1 est inconnue car elle dépend du paramètre θ . On dispose néanmoins d'un estimateur de θ : \bar{X}_n . On en déduit donc un estimateur de $\text{Var}(X_1)$: $\bar{X}_n(1 - \bar{X}_n)$.

On utilise alors le lemme de Slutsky :

Lemme 3.4 : Lemme de Slutsky (hors-programme)

Si X_n converge en loi vers une variable aléatoire X et Y_n converge en loi vers une constante c alors le couple (X_n, Y_n) converge en loi vers (X, c) .

On en déduit alors (d'après la Proposition 5.3) que $X_n + Y_n$ converge en loi vers $X + c$, que $X_n Y_n$ converge en loi vers cX ...

Le lemme de Slutsky assure donc que

$$\sqrt{\frac{n}{\bar{X}_n(1 - \bar{X}_n)}} (\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

On note toujours F la fonction de répartition d'une loi normale $Z \sim \mathcal{N}(0, 1)$. On a donc, pour q_α le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{\frac{n}{\bar{X}_n(1 - \bar{X}_n)}} |\bar{X}_n - \theta| \leq q_\alpha \right) &= \mathbb{P}(|Z| \leq q_\alpha) \\ &= 1 - \alpha. \end{aligned}$$

On en déduit enfin l'intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ :

$$I_3 = \left[\bar{X}_n - \frac{q_\alpha}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)}, \bar{X}_n + \frac{q_\alpha}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)} \right].$$

Application numérique :

On obtient avec un niveau asymptotique de niveau de 95 % :

$$\theta_0 \in [0.48; 0.56].$$

Remarque : On aurait pu utiliser la variance empirique pour estimer la variance (au lieu de $\bar{X}_n(1 - \bar{X}_n)$) et utiliser le Lemme de Slutsky.

4 Intervalle de fluctuation (Hors programme)

4.1 Intervalle de confiance versus intervalle de fluctuations

Remarque 4.1 : *Intervalle de confiance vs intervalle de fluctuation*

Reprenons le premier exemple (l'intervalle I_1) ci-dessus.

On a

$$\mathbb{P}\left(\theta \in \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]\right) = 1 - \alpha.$$

De manière équivalente, on peut aussi écrire que

$$\mathbb{P}\left(\bar{X}_n \in \left[\theta - \frac{1}{2\sqrt{n\alpha}}, \theta + \frac{1}{2\sqrt{n\alpha}}\right]\right) = 1 - \alpha.$$

Dans cette exemple,

- ▶ $\left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]$ est un intervalle de confiance : il est aléatoire ;
- ▶ $\left[\theta - \frac{1}{2\sqrt{n\alpha}}, \theta + \frac{1}{2\sqrt{n\alpha}}\right]$ est un intervalle de fluctuation : il est *déterministe*, ie qu'il ne dépend pas de l'aléa ω .

L'intervalle de fluctuation est un intervalle fixe dans lequel se trouve la moyenne empirique \bar{X}_n avec une probabilité égale à $1 - \alpha$. Cet intervalle a un intérêt plus théorique puisqu'en général, on cherche à estimer θ .

Formalisons ce qui a été dit dans la remarque précédente.

On considère toujours un problème de pile ou face. On suppose maintenant que le paramètre θ est **connu**. Autrement dit, la probabilité de succès θ est connue. Comme précédemment, on répète n fois l'expérience de façon indépendante et on se demande où se situe la fréquence de succès en fonction du nombre de réalisations n .

Définition 4.2 : *Intervalle de fluctuation*

Soit X_1, \dots, X_n des variables aléatoires indépendantes de loi de Bernoulli $\mathcal{B}er(\theta)$ avec θ connu. Un **intervalle de fluctuation** de la fréquence de succès au niveau $1 - \alpha$ est un intervalle déterministe I_f tel que

$$\mathbb{P}(\bar{X}_n \in I_f) = 1 - \alpha.$$

La quantité α est l'erreur que l'on s'autorise et est appelée **niveau de risque**.

4.2 Construction d'un intervalle de fluctuation

4.2.1 Intervalle de fluctuation non asymptotique

Dans ce cadre théorique, on connaît la loi de $S_n := \sum_{i=1}^n X_i$: il s'agit de la loi Binomiale $\mathcal{B}(n, \theta)$ (θ étant connu!).

Pour trouver un intervalle de fluctuations $I_f = [a; b]$ de la fréquence de succès, on détermine des réels a et b tels que :

$$\mathbb{P}(a \leq \bar{X}_n \leq b) = \mathbb{P}(na \leq S_n \leq nb) = \mathbb{P}(S_n \leq nb) - \mathbb{P}(S_n < na) = 1 - \alpha.$$

Les valeurs de a et b vont dépendre de θ , n et α . On peut utiliser un outil numérique pour déterminer ces réels.

Remarque 4.3

Il n'y a pas unicité des valeurs a et b satisfaisant les conditions de l'intervalle de fluctuation.

4.2.2 Intervalle de fluctuation asymptotique

On suppose ici que n est grand.

D'après le TCL, on sait que

$$\sqrt{\frac{n}{\theta(1-\theta)}} (\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Pour trouver un intervalle de fluctuation $I_f = [a; b]$ de la fréquence de succès, on détermine des réels a et b tels que, si $Z \sim \mathcal{N}(0, 1)$,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P} \left(\theta + \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}} a \leq \bar{X}_n \leq \theta + \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}} b \right) &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{\frac{n}{\theta(1-\theta)}} (\bar{X}_n - \theta) \in [a; b] \right) \\ &= \mathbb{P}(Z \in [a; b]) \\ &= 1 - \alpha. \end{aligned}$$

Là encore, il n'y a pas unicité de a et b .

Par exemple, si on choisit $a = -b$ alors,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\theta - \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}} b \leq \bar{X}_n \leq \theta + \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}} b \right) = 2F(b) - 1.$$

On prend pour b la valeur du quantile d'ordre $1 - \alpha/2$ de la loi normale $\mathcal{N}(0, 1)$.

4.3 Erreurs possibles dans la prise de décision

Les intervalles de fluctuation sont un outil pour la prise de décision. En effet, lorsqu'on ne connaît pas la probabilité de succès d'une épreuve de Bernoulli (ie la valeur de θ), on peut émettre une hypothèse sur sa valeur. On réalise alors plusieurs réalisations de l'expérience et au vu de la valeur de la moyenne empirique, on considère la *procédure de décision* suivante :

- ▶ si la moyenne empirique n'est pas dans l'intervalle de fluctuation, *on rejette l'hypothèse* ;
- ▶ si la moyenne empirique est dans l'intervalle de fluctuation, *on ne rejette pas l'hypothèse*.

Une telle procédure peut nous amener à faire deux erreurs possibles :

- ▶ l'*erreur de première espèce* : c'est rejeter l'hypothèse alors qu'elle est vraie ;
- ▶ l'*erreur de seconde espèce* : c'est accepter l'hypothèse alors qu'elle n'est pas vraie.

Remarque 4.4 : Niveau de risque et erreurs.

On a vu qu'un intervalle de fluctuation est construit avec un certain niveau de confiance $1 - \alpha$ fixé à l'avance. La quantité α correspond à la probabilité de rejeter à tort l'hypothèse (ie l'erreur de première espèce).

Attention ! Si l'erreur de première espèce est contrôlée par α , on ne contrôle, a priori, pas du tout l'erreur de seconde espèce. Elle peut donc se produire avec une grande probabilité...

Remarque 4.5 : Abus de langage.

La procédure de décision permet de rejeter ou de ne pas rejeter l'hypothèse. Attention, ne pas rejeter l'hypothèse ne signifie pas qu'elle est vraie !

Exemple 4.6

On dispose d'une pièce pour jouer à pile ou face.

On note θ_0 la probabilité d'obtenir Pile.

On se demande si la pièce est équilibrée afin que le jeu ne soit pas truqué.

Hypothèse : La pièce est équilibrée, ie $\theta_0 = 0.5$.

On suppose désormais que la pièce est équilibrée, ie que $\theta_0 = 0.5$. Le but est à présent d'"accepter" ou de "réfuter" cette hypothèse.

On réalise 1000 lancers de pièce et on obtient 520 piles. Nous allons construire un intervalle de fluctuation asymptotique à l'aide du TCL de la forme $I_f = [-a; a]$ (cf paragraphe précédent).

Application numérique :

En reprenant les notations précédentes, on a : $\alpha = 0.05$; $n = 1000$; $\overline{X}_n = 0.52$; $\theta = \theta_0 = 0.50$ et $b = 1,96$. On obtient :

$$I_f = [0.47; 0.53].$$

$\overline{X}_n \in I_f$ donc notre hypothèse n'est pas rejetée au niveau 95%.

Conclusion : Au vu des données, on ne peut pas affirmer que la pièce soit truquée (mais est-elle pour autant parfaitement équilibrée ? Rien n'est moins sûr...)

5 Exercices

Exercice 1 : Choix d'un estimateur

Soit X une variable aléatoire uniforme sur l'intervalle $[0, 2a]$.

On considère une suite X_1, \dots, X_n de n variables aléatoires indépendantes et de même loi que X .

On pose $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ et $T = \max(X_1, \dots, X_n)$.

1. Déterminer $\mathbb{E}[X]$ et $\text{Var}(X)$.
2. Montrer que \bar{X} est un estimateur sans biais et convergent du paramètre a .
3. Déterminer la densité de la variable aléatoire T . Calculer l'espérance et la variance de T . En déduire un autre estimateur \hat{T} sans biais et convergent de a .
4. Comparer les deux estimateurs de a .

Exercice 2 : Un estimateur de la variance lorsque la moyenne est connue

Soit X une variable aléatoire qui admet un moment d'ordre 4 dont on note $m = \mathbb{E}[X]$ et $\sigma^2 = \text{Var}(X)$. On suppose que m est connue.

Soit (X_1, \dots, X_n) un n -uplet de variables aléatoires indépendantes et identiquement distribuées de même loi que X . Proposer un estimateur sans biais et convergent de la variance.

Exercice 3 : Estimation d'une proportion par intervalle de confiance

On effectue un sondage auprès de la population française dans le but de déterminer la proportion p d'individus éprouvant de la peur à l'idée d'effectuer un voyage en avion. On interroge 1000 personnes. Sur ces 1000 personnes, 253 affirment avoir peur de prendre l'avion.

1. Proposer un intervalle de confiance pour la proportion p à 90 %,
 - (a) en utilisant l'inégalité de Bienaymé-Tchebychev ;
 - (b) en utilisant le théorème central limite.
2. Comparer les résultats obtenus.

Exercice 4 : Estimation d'un paramètre d'une densité de probabilité

Soit X une variable aléatoire de densité :

$$f_{\theta}(x) = \begin{cases} \frac{1}{\theta} \exp\left(-\frac{x-\theta}{\theta}\right) & \text{si } \theta \leq x \\ 0 & \text{sinon} \end{cases}$$

où θ est un réel strictement positif.

1. Calculer $\mathbb{E}[X]$.
2. Soit X_1, \dots, X_n , n variables aléatoires indépendantes et de même loi que X . Déterminer un estimateur \hat{T} de θ . Étudier ses propriétés.
3. Déterminer la loi limite de \hat{T} quand n tend vers l'infini et en déduire un intervalle de confiance de niveau 0.95 pour θ dans le cas où l'on a observé la valeur $\sum_{i=1}^{100} x_i = 660$ sur un échantillon de taille $n = 100$.

Exercice 5 : Méthode des moments

On observe un n -échantillon X_1, \dots, X_n de v.a. indépendantes et de même loi définie sur \mathbb{R} de densité

$$f_{\theta}(x) = \frac{1}{2}(1 + \theta x)\mathbb{1}_{]-1,1[}(x),$$

où $\theta \in]-1, 1[$ est inconnu.

1. Calculer l'espérance de X_1 . En déduire un estimateur $\hat{\theta}$ de θ .
2. Calculer son biais et son risque quadratique.
3. Soit $\alpha \in]0, 1[$. Déterminer un intervalle asymptotique pour θ au niveau de confiance $(1 - \alpha)$.

Exercice 6 : *Un nouvel estimateur*

On observe x_1, \dots, x_n des réalisations de variables aléatoires X_1, \dots, X_n indépendantes et de même loi uniforme sur $[0, \theta]$ où $\theta > 0$ est inconnu.

1. Justifier (heuristiquement) que $\hat{\theta} := \max_{1 \leq i \leq n} X_i$ est un estimateur de θ .
2. Quelle est la loi de $\hat{\theta}$?
3. $\hat{\theta}$ est-il biaisé ? asymptotiquement biaisé ?

Exercice 7 : *Intervalle de fluctuations*

Deux individus jouent ensemble. Un joueur choisit au hasard une carte dans un jeu de 32 cartes. S'il obtient un as, l'autre joueur doit lui donner 1 euro. On constate que sur 50 essais, il a retourné 11 fois un as. Peut-on présumer, au risque 5 % que ce joueur est un tricheur ?

Troisième partie

Complément : Fonction génératrice et processus de Galton-Watson

1 Modélisation mathématique du problème

Cadre : De nombreux phénomènes d'évolution de population peuvent être modélisés en première approximation par un processus de branchement (réactions nucléaires en chaîne, étude des gènes, survivance des noms de famille...).

Modélisation mathématique :

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

On note, pour $n \in \mathbb{N}$, $p_n = \mathbb{P}(X = n)$ et $m = \mathbb{E}[X] = \sum_{n=0}^{\infty} np_n < \infty$.

Soit $(X_{i,j})_{i,j \in \mathbb{N}}$ une famille de variable aléatoire indépendantes et identiquement distribuées, suivant la loi \mathbb{P}_X .

On définit la suite $(Z_n)_{n \in \mathbb{N}}$ de la façon suivante :

$$\begin{cases} Z_0 = 1 \\ \forall n \in \mathbb{N}, Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n} \end{cases}$$

L'idée est de modéliser avec la suite $(Z_n)_n$ la taille d'une population. Plus précisément, Z_n symbolisera le nombre d'individus à la $n^{\text{ème}}$ génération, et pour $i \in \llbracket 1, Z_n \rrbracket$, $X_{i,n}$ représente le nombre de descendants que l'individu de la $n^{\text{ème}}$ génération portant le numéro i a engendré (les individus de la population que l'on considère génèrent des enfants tous seuls ; on pourra, par exemple penser aux végétaux). Aussi, chaque individu a la probabilité p_n d'engendrer n individus. On veut répondre à la question : quelle est la probabilité que la population considérée s'éteigne ?

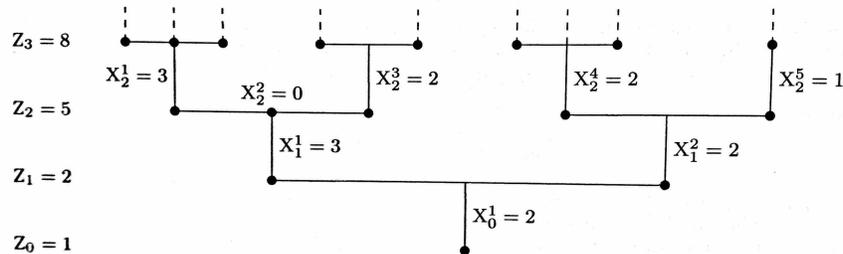


FIGURE 1 – Arbre de la descendance de Z_0 .

Lemme 1.1

Pour tout $n \in \mathbb{N}$ et pour tout $i \in \mathbb{N}^*$, la variable Z_n est indépendante de $X_{i,n}$.

Démonstration.

Soit $n \in \mathbb{N}^*$; Z_n ne dépend que de Z_{n-1} et de la famille $(X_{i,n-1})_{i \in \mathbb{N}}$.

Ainsi, par une récurrence immédiate, il vient : Z_n ne dépend que de la famille $(X_{i,j})_{i \geq 0, j \leq n-1}$.

Et, par indépendance des variables $X_{i,j}$, on obtient que $\forall i \in \mathbb{N}, Z_n \perp\!\!\!\perp X_{i,n}$. □

2 Calcul de la probabilité d'extinction

Objectif : Calculer $\mathbb{P}(\exists n \in \mathbb{N}, Z_n = 0)$.

Notations :

Pour tout $n \in \mathbb{N}$, on note :

$$\pi_n := \mathbb{P}(Z_n = 0) \text{ et } \pi_\infty := \mathbb{P}(\exists n \in \mathbb{N}, Z_n = 0) \text{ appelée probabilité d'extinction.}$$

Comme $Z_n = 0 \Rightarrow Z_{n+1} = 0$, la suite d'événements $(\{Z_n = 0\})_{n \in \mathbb{N}}$ est croissante et donc :

$$\pi_\infty = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \{Z_n = 0\}\right) = \lim_{n \rightarrow +\infty} \pi_n.$$

Si $p_0 = 0$, alors on a $\forall n \in \mathbb{N}^*, Z_n \geq 1$ ps et $\pi_\infty = 0$.

Si $p_0 = 1$, alors on a $\forall n \in \mathbb{N}^*, Z_n = 0$ ps et $\pi_\infty = 1$.

On suppose donc désormais que $p_0 \in]0, 1[$.

Proposition 2.1 : Propriétés générales sur la fonction génératrice

On définit la fonction génératrice de X par $G : s \mapsto \mathbb{E}[s^X] = \sum_{k=0}^{\infty} p_k s^k$.

On a les résultats suivants :

1. G est bien définie sur $[0, 1]$ et y est de classe \mathcal{C}^1 .
2. (a) G est strictement croissante sur $]0, 1[$.
- (b) G est convexe sur $]0, 1[$.
- (c) G est strictement convexe sur $]0, 1[\Leftrightarrow p_0 + p_1 < 1$.

Démonstration.

La série entière $\sum_{k \geq 0} p_k s^k$ ayant un rayon de convergence supérieure ou égale à 1 (car $\sum_{k \geq 0} p_k = 1$), on a que la série $\sum p_n s^n$ est \mathcal{C}^∞ sur $[0, 1]$.

1. Intéressons nous à la régularité de la série sur $[0, 1]$.

$\forall k \in \mathbb{N}, s \mapsto p_k s^k$ est de classe \mathcal{C}^1 sur $[0, 1]$, la série $\sum_{k \geq 0} p_k 1^k$ converge, et la série de fonctions $\sum_{k \geq 1} k p_k s^{k-1}$

converge normalement (car X est intégrable) donc uniformément sur $[0, 1]$.

Par conséquent, d'après le théorème de dérivation des séries, G est bien définie et est \mathcal{C}^1 sur $[0, 1]$.

2. On a, par théorème de dérivation terme à terme d'une série entière, :

$$\forall s \in [0, 1[, G'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \text{ et } G''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2}$$

Comme $p_0 < 1$, il existe $k_0 > 0$ tel que $p_{k_0} > 0$.

(a) Ainsi : $\forall s \in]0, 1[, G'(s) \geq k_0 p_{k_0} s^{k_0-1} > 0$ et G est strictement croissante sur $]0, 1[$.

(b) Aussi : $\forall s \in]0, 1[, G''(s) \geq k_0(k_0-1) p_{k_0} s^{k_0-2} \geq 0$ et G est convexe sur $]0, 1[$.

(c) Si $p_0 + p_1 = 1$, alors on a $k_0 = 1$ et G est affine donc n'est pas strictement convexe sur $]0, 1[$.

Si $p_0 + p_1 < 1$, alors il existe $k_1 > 1$ tq $p_{k_1} > 0$ et $G'' > 0$ sur $]0, 1[$ d'où la stricte convexité. □

Proposition 2.2

Pour $n \in \mathbb{N}$, on définit la série génératrice de Z_n par $G_n : s \mapsto \mathbb{E}[s^{Z_n}] = \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) s^k$.

Comme précédemment, on peut montrer que G_n est bien définie sur $[0, 1]$.

Pour $n \in \mathbb{N}^*$, on a $G_n = \underbrace{G \circ \dots \circ G}_{n \text{ fois}}$ (sur $[0, 1]$).

Démonstration.

On procède par récurrence.

Initialisation : $G_1(s) = \mathbb{E}[s^{Z_1}] = \mathbb{E}[s^{X_{1,0}}] = \mathbb{E}[s^X] = G(s)$

Supposons que $G_n = G \circ \dots \circ G$.

$$\begin{aligned}
 G_{n+1}(s) &= \mathbb{E}[s^{Z_{n+1}}] = \mathbb{E}\left[s^{\sum_{i=1}^{Z_n} X_{i,n}}\right] \\
 &= \mathbb{E}\left[\prod_{i=1}^{Z_n} s^{X_{i,n}}\right] \\
 &= \mathbb{E}\left[\sum_{k=0}^{+\infty} \left(\prod_{i=1}^k s^{X_{i,n}} \mathbb{1}_{Z_n=k}\right)\right] \quad (\text{car } Z_n < \infty \text{ ps, } X \text{ étant finie ps}) \\
 &= \sum_{k=0}^{+\infty} \mathbb{E}\left[\prod_{i=1}^k s^{X_{i,n}} \mathbb{1}_{Z_n=k}\right] \quad (\text{par Fubini Tonelli, car les termes sont tous positifs}) \\
 &= \sum_{k=0}^{+\infty} \mathbb{E}\left[\prod_{i=1}^k s^{X_{i,n}}\right] \mathbb{E}[\mathbb{1}_{Z_n=k}] \quad (\text{car } Z_n \perp\!\!\!\perp X_{i,n} \text{ par le lemme 1.1.}) \\
 &= \sum_{k=0}^{+\infty} \prod_{i=1}^k \mathbb{E}[s^{X_{i,n}}] \mathbb{P}(Z_n = k) \quad (\text{car les } X_{i,n} \text{ sont indépendants}) \\
 &= \sum_{k=0}^{+\infty} \mathbb{E}[s^X]^k \mathbb{P}(Z_n = k) \quad (\text{car les } X_{i,n} \text{ ont même loi}) \\
 &= \sum_{k=0}^{+\infty} \mathbb{P}(Z_n = k) G(s)^k = G_n(G(s)) \text{ par hypothèse de récurrence.}
 \end{aligned}$$

Ce qui achève la récurrence. □

Proposition 2.3

La probabilité d'extinction π_∞ est le plus petit point fixe de G sur l'intervalle $[0, 1]$.

Démonstration.

La proposition précédente donne : $\forall n \in \mathbb{N}, \forall s \in [0, 1], G_{n+1}(s) = G(G_n(s))$.

En évaluant en 0, on obtient la relation : $\pi_{n+1} = G(\pi_n)$, puis par continuité de G sur $[0, 1]$, on obtient que π_∞ est un point fixe de G .

Montrons que c'est le plus petit. Soit $u \in]0, 1]$ un point fixe de G . Par récurrence, montrons que $\forall n \in \mathbb{N}^*, \pi_n \leq u$.

• Initialisation : on a $\pi_1 = G(\pi_0) = G(\mathbb{P}(Z_0 = 0)) = G(0) \leq G(u) = u$, car G est croissante.

• Supposons que $\pi_n \leq u$. Alors $\pi_{n+1} = G(\pi_n) \leq G(u) = u$, par croissance de G .

Par conséquent, $\forall n \in \mathbb{N}^*, \pi_n \leq u$, donc par passage à la limite : $\pi_\infty \leq u$. □

Théorème 2.4

Si $m := \mathbb{E}(X) \leq 1$, alors $\pi_\infty = 1$.

Si $m > 1$, alors π_∞ est l'unique point fixe de G sur $]0, 1[$.

Démonstration.

Puisque $G(1) = 1$, le graphe de G coupe la droite $y = x$ sur l'intervalle $[0, 1]$ au moins au point $(1, 1)$. On a deux

cas :

- Si $p_0 + p_1 = 1$ (ie $m = 1$) alors G est une fonction affine et comme $G(0) = p_0 > 0$, G a un unique point fixe en $(1, 1)$.
- Sinon, G est strictement convexe sur $]0, 1[$ et il en va de même pour $x \mapsto G(x) - x$ qui s'annule donc au plus deux fois.¹

Rappelons que l'on a aussi : $G(0) = p_0, G'(0) = p_1$ et $G(1) = 1, G'(1) = \sum_{n=1}^{\infty} np_n = m$. Étudions la fonction $x \in [0, 1] \mapsto G(x) - x$.

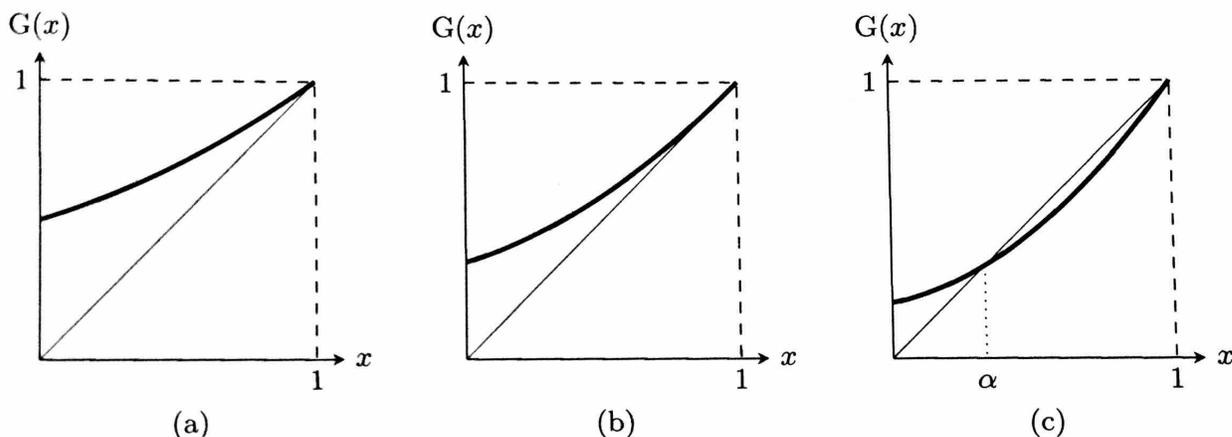


FIGURE 2 – Fonction génératrice de X , dans le cas : (a) $m < 1$, (b) $m = 1$ et (c) $m > 1$.

Supposons $m > 1$.

Alors $G' - 1$ est une fonction croissante. Sa valeur en 0 est $p_1 - 1 < 0$ (car $p_1 = 1 \Rightarrow m = 1$ ou bien plus simplement car $p_0 > 0$) et sa valeur en 1 est $m - 1 > 0$, donc elle s'annule en un point $\alpha \in]0, 1[$.

La fonction $G - \text{Id}$ est alors décroissante sur $[0, \alpha]$ puis croissante sur $[\alpha, 1]$. Comme $G(0) - 0 = p_0 > 0$ et $G(1) - 1 = 0$, il existe un point dans l'intervalle $]0, \alpha]$ où $G - \text{Id}$ s'annule.

π_∞ est donc l'unique point fixe de G sur l'intervalle $]0, 1[$ (car G en a au plus 2 car strictement convexe).

Supposons $m \leq 1$.

Alors $G' - 1$ est une fonction croissante sur $[0, 1]$, négative ou nulle en 1 ; donc négative sur $[0, 1]$.

Donc $G - \text{Id}$ est décroissante sur $[0, 1]$, et s'annule en 1. Comme cette fonction admet au plus 2 annulations, elle ne s'annule qu'en 1 (car sinon elle s'annulerait sur un intervalle non-réduit à un singleton).

Par conséquent, $\pi_\infty = 1$.

□

Remarque :

Supposons que l'équation $G(x) = x$ admette une solution $0 < p < 1$. On a alors $p = \pi_\infty$ par la proposition précédente. De plus, puisque $G(p) - p = 0$ et $G(1) - 1 = 0$, le théorème de Rolle appliqué à $x \mapsto G(x) - x$ montre qu'il existe $z \in]p, 1[$ tel que $G'(z) = 1$. Comme G est strictement convexe on a nécessairement $m = G'(1) > 1$.

1. En effet, si cette fonction s'annule en trois points distincts, par le théorème de Rolle, sa dérivée s'annulera en deux points distincts, $a < b$. Mais $x \mapsto G(x) - x$ est convexe, donc sa dérivée est croissante, donc nulle sur $[a, b]$. Cela implique que $G' - 1$ n'est pas strictement croissante, et donc que $x \mapsto G(x) - x$ n'est pas strictement convexe. Attention à ne pas dire que "strictement convexe" équivaut à "dérivée seconde strictement positive" pour les fonctions deux fois dérivables (on pourra penser par exemple à la fonction $x \mapsto x^4$).

3 Évolution de la taille de la population

On donne ici une première idée de l'évolution de la taille de la population, *i.e.* de la suite Z_n .

Proposition 3.1

Pour tout $n \in \mathbb{N}$, on a $\mathbb{E}[Z_n] = m^n$.

Démonstration.

On va raisonner par récurrence.

- $Z_0 = 1$ donc $\mathbb{E}[Z_0] = 1 = m^0$.
- Pour $n \in \mathbb{N}$, supposons que $\mathbb{E}[Z_n] = m^n$.

On peut dériver G_n (comme pour G), et on a, pour $s \in [0, 1]$,

$$G'_{n+1}(s) = G'(s)(G'_n \circ G(s))$$

Donc en évaluant en 1 :

$$G'_{n+1}(1) = \mathbb{E}[X](G'_n(G(1)) = mG'_n(1) = m^{n+1}.$$

Ce qui conclut la récurrence. □

Quatrième partie

Éléments de correction des exercices

1 Probabilités

Exercice 1 :

Considérer l'espace Ω défini par l'ensemble des couples (numéro de la boîte choisie initialement par le joueur A, numéro de la boîte gagnante). On a équiprobabilité sur cet espace probabilisé et chaque couple a donc une probabilité $1/9$ de se réaliser. Le joueur B gagne sans changer de boîte si les couples $(1, 1)$, $(2, 2)$ ou $(3, 3)$ se réalisent donc avec une probabilité de $\frac{2}{9} = \frac{1}{3}$ s'il maintient son choix. En revanche, s'il décide de changer de boîte, les couples gagnants deviennent $(1, 2)$, $(1, 3)$, $(2, 1)$, $(2, 3)$, $(3, 1)$, $(3, 2)$ donc il a une probabilité de gagner de $\frac{6}{9} = \frac{2}{3}$ s'il change de boîte!

Exercice 2 :

- Il suffit de remarquer que pour tout $k \in \mathbb{N}^*$, $\mathbb{P}(X > k) = \sum_{i=k+1}^{+\infty} \mathbb{P}(X = i)$ puis que pour tout $(k, l) \in \mathbb{N}^*$, $\mathbb{P}(X > k+l, X > k) = \mathbb{P}(X > k+l)$.
- En utilisant le fait que $\mathbb{P}(X > 1) = 1 - \mathbb{P}(X = 1)$, on montre que $(\mathbb{P}(X > k))_{k \in \mathbb{N}^*}$ est une suite géométrique de raison $1 - p$ où $p = \mathbb{P}(X = 1)$. On a donc $\mathbb{P}(X > k) = (1 - p)^{k-1} \times \mathbb{P}(X > 1) = (1 - p)^k$. On conclut en remarquant que $\mathbb{P}(X = k) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k) = (1 - p)^{k-1} - (1 - p)^k = p(1 - p)^{k-1}$.

Exercice 3 :

Si X suit une loi exponentielle, on montre que X est sans mémoire de la même manière que dans l'exercice 2. Réciproquement, supposons que la loi de X est une loi sans mémoire. Posons $r(x) = \mathbb{P}(X > x)$ pour tout $x \in \mathbb{R}$. On commence par montrer que pour tout $(x, y) \in \mathbb{R}^+$, $r(x+y) = r(x)r(y)$. En prenant $x = y = 0$, on obtient que $r(0) = 0$ ou $r(0) = 1$. Par hypothèse, on a $r(0) = \mathbb{P}(X > 0) > 0$ donc $r(0) = 1$. Par croissance de la fonction de répartition de X , noté F_X sur \mathbb{R} , on obtient alors que $F_X(x) = 0$ pour tout $x \geq 0$. De plus, on obtient par récurrence,

$$\forall (x_1, \dots, x_n) \in \mathbb{R}_+^n, r(x_1 + \dots + x_n) = r(x_1) \dots r(x_n).$$

On pose $a = r(1) > 0$. On va raisonner par densité de \mathbb{Q} dans \mathbb{R} . Soit $p, q \in \mathbb{N}^*$ et $x = p/q \in \mathbb{Q}_+^*$. On a

$$a = r(1) = r\left(\underbrace{\frac{1}{q} + \dots + \frac{1}{q}}_{q \text{ termes}}\right) = r\left(\frac{1}{q}\right)^q,$$

et donc $r(1/q) = a^{1/q}$. Ainsi,

$$r\left(\frac{p}{q}\right) = r\left(\underbrace{\frac{1}{q} + \dots + \frac{1}{q}}_{p \text{ termes}}\right) = r\left(\frac{1}{q}\right)^p = \left(a^{1/q}\right)^p = a^{p/q} = a^x.$$

A présent, soit $x \in \mathbb{R}^{+*}$. Alors x est la limite d'une suite décroissante de rationnels x_n . Comme $r = 1 - F_X$ alors r est continue à droite en tout point de \mathbb{R} et comme $x \leq x_n$ pour tout entier naturel n , la continuité à droite de F_X implique que

$$r(x) = \lim_{n \rightarrow +\infty} r(x_n) = \lim_{n \rightarrow +\infty} a^{x_n} = a^x = e^{x \ln(a)}.$$

Comme F_X est croissante alors r est décroissante et donc $a \in]0, 1[$. En posant, $\lambda = -\ln(a) > 0$, on obtient finalement

$$\forall x \in \mathbb{R}^{+*}, F_X(x) = 1 - e^{-\lambda x},$$

ce qui achève la preuve!

Exercice 4 :

- Le lemme de transfert permet de conclure que $\mathbb{E}[f(S_n/n)] = B_n(x)$.

2. On a

$$|f(x) - B_n(x)| \leq \mathbb{E}[|f(x) - f(S_n/n)|].$$

Soit $\varepsilon > 0$. Comme f est uniformément continue, il existe $\nu > 0$ tel que pour tout $x, y \in [0, 1]$ vérifiant $|x - y| \leq \nu$, on ait $|f(x) - f(y)| \leq \varepsilon$. Ainsi, on a

$$\mathbb{E}[|f(x) - f(S_n/n)|] = \mathbb{E}[|f(x) - f(S_n/n)|\mathbb{1}_{|x - S_n/n| \leq \nu}] + \mathbb{E}[|f(x) - f(S_n/n)|\mathbb{1}_{|x - S_n/n| > \nu}],$$

d'où

$$\mathbb{E}[|f(x) - f(S_n/n)|] \leq \varepsilon + 2\|f\|_\infty \mathbb{P}\left(\left|x - \frac{S_n}{n}\right| \geq \nu\right).$$

On conclut avec l'inégalité de Bienaymé-Tchebychev ou avec la loi faible des grands nombres.

Exercice 5 :

$$\begin{aligned} \mathbb{E}[X] < \infty &\iff \sum_{k=0}^{\infty} k\mathbb{P}(X = k) \text{ converge} \\ &\iff \sum_{k=1}^{\infty} \mathbb{P}(X = k) \sum_{n=0}^{k-1} 1 \text{ converge} \\ &\iff \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} \mathbb{P}(X = k) \text{ converge} \\ &\iff \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} \mathbb{P}(X = k) \text{ converge} \\ &\iff \sum_{n=0}^{\infty} \mathbb{P}(X > n) \text{ converge.} \end{aligned}$$

Exercice 6 :

1. Calculer $\mathbb{P}(Y = k)$ en distinguant les cas $k < 0$, $k > 0$ et $k = 0$.
2. En appliquant la formule de transfert, on trouve $m_n = \lambda^n$.

Exercice 7 :

1. Pour toute fonction f continue par morceaux, on trouve $\mathbb{E}[f(Y)] = \int_0^{+\infty} f(u)2u \exp(-\theta u^2) du$.
2. (a) Y est à valeurs dans $[0, 1/2]$. Soit $t \in [0, 1/2]$. On rappelle que la fonction de répartition d'une loi uniforme est donnée par $F(x) = x$ pour tout $x \in [0, 1]$. On a

$$\mathbb{P}(Y > t) = \mathbb{P}(X > t, 1 - X > t) = \mathbb{P}(t < X < 1 - t) = 1 - 2t.$$

D'autre part, $\mathbb{P}(Y > t) = 1 - F_Y(t)$ où F_Y est la fonction de répartition de Y . Ainsi, $F_Y(t) = 2t$ et $F'_Y(t) = 2$ pour tout $t \in [0, 1/2]$. Ainsi Y suit une loi uniforme sur $[0, 1/2]$. De même, Z suit une loi uniforme sur $[1/2, 1]$.

- (b) La fonction $t \mapsto \frac{\max(t, 1-t)}{\min(t, 1-t)}$ n'est pas intégrable sur $[0, 1]$ donc Z/Y n'admet pas d'espérance. La fonction $t \mapsto \frac{\min(t, 1-t)}{\max(t, 1-t)}$ est continue sur $[0, 1]$ donc Y/Z admet une espérance :

$$\mathbb{E}[Y/Z] = \int_0^1 \frac{\min(t, 1-t)}{\max(t, 1-t)} dt = \int_0^{1/2} \frac{t}{1-t} dt + \int_{1/2}^1 \frac{1-t}{t} dt.$$

Un calcul d'intégrale fournit $\mathbb{E}[Y/Z] = 2 \log(2) - 1$.

Exercice 8 :

1. $a = e^{-2}$
2. $\mathbb{P}(X = j) = a/j!$
3. $\mathbb{P}(Y = k) = a/k!$ donc X et Y sont indépendantes.

Exercice 9 :

Les va X et Y sont indépendantes donc la densité du couple (X, Y) vaut

$$f_{(X,Y)}(x, y) = \mathbb{1}_{[0, +\infty[^2}(x, y) \lambda e^{-\lambda x} \mu e^{-\mu y}.$$

On a alors

$$\mathbb{P}(X > Y) = \mathbb{E}[\mathbb{1}_{X > Y}] = \int_{y=0}^{+\infty} \int_{x=y}^{+\infty} f_{(X,Y)}(x, y) dx dy,$$

et on trouve $\mathbb{P}(X > Y) = \frac{\mu}{\lambda + \mu}$.

Exercice 10 :

S_N est à valeurs dans un ensemble dénombrable noté $\{s_k, k \in \mathbb{N}^*\}$.

$$\begin{aligned} \mathbb{P}(S_N = s_k) &= \sum_{n=1}^{\infty} \mathbb{P}(X_1 + \dots + X_n = s_k, N = n) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X_1 + \dots + X_n = s_k) \mathbb{P}(N = n) \text{ par indépendance.} \end{aligned}$$

Alors

$$\begin{aligned} \mathbb{E}[S_N] &= \sum_{k=1}^{\infty} s_k \mathbb{P}(S_N = s_k) \\ &= \sum_{k=1}^{\infty} s_k \sum_{n=1}^{\infty} \mathbb{P}(X_1 + \dots + X_n = s_k) \mathbb{P}(N = n) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(N = n) \sum_{k=1}^{\infty} s_k \mathbb{P}(X_1 + \dots + X_n = s_k) \text{ par le théorème de Fubini} \\ &= \sum_{n=1}^{\infty} \mathbb{P}(N = n) \mathbb{E}[X_1 + \dots + X_n] \\ &= \sum_{n=1}^{\infty} \mathbb{P}(N = n) n \mathbb{E}[X_1] \text{ par linéarité de l'espérance} \\ &= \mathbb{E}[X_1] \mathbb{E}[N]. \end{aligned}$$

Exercice 11 :

1. Le truquage de chaque dé étant individuel, la méthode de truquage préserve l'indépendance des deux dés. Soit S une variable aléatoire uniforme sur $\{2, 3, \dots, 12\}$. On a

$$\mathbb{P}(S = n) = \frac{1}{11} \text{ pour tout } 2 \leq n \leq 12.$$

La fonction génératrice s'écrit

$$G_S(z) = \frac{1}{11}(z^2 + z^3 + \dots + z^{12}) = \frac{z^2}{11}(1 + z + z^2 + \dots + z^{10}).$$

Les zéros de $1 + z + z^2 + \dots + z^{10}$ sont les racines onzièmes de l'unité (sauf 1), que l'on note $s_k = \exp(2ik\pi/11)$ pour tout $k \in \{1, \dots, 10\}$.

Aucun des s_k n'est réel.

En effet, si l'un d'entre eux était réel, il vaudrait soit 1 soit -1 .

Par exemple, si on suppose qu'il existe $l \in \mathbb{Z}$ tel que $\frac{2k\pi}{11} = 2l\pi$ alors $k = 11l$ ce qui est impossible (car $k \in \{1, \dots, 10\}$).

De même, si on suppose qu'il existe $l \in \mathbb{Z}$ tel que $\frac{2k\pi}{11} = \pi + 2l\pi$ alors $2k = 11(2l + 1)$ ce qui est impossible (car $11(2l + 1)$ est impair).

On a alors

$$G_S(z) = z^2 R(z)$$

où $R(z) = \frac{1}{11}(1 + z + z^2 + \dots + z^{10})$ est un polynôme de degré 10 sans racine réelle.

2. Supposons que l'on puisse réaliser un truquage de chaque dé de telle sorte que la somme de chaque face suive une loi uniforme sur $\{2, \dots, 12\}$.

On définit deux variables aléatoires X et Y indépendantes à valeurs dans $\{1, \dots, 6\}$ de telle sorte que $S = X + Y$.

On a

$$G_X(z) = \sum_{n=1}^6 \mathbb{P}(X = n)z^n = zQ_X(z)$$

où $Q_X(z)$ est un polynôme à coefficients réels de degré au plus 5.

De même

$$G_Y(z) = zQ_Y(z)$$

où $Q_Y(z)$ est un polynôme à coefficients réels de degré au plus 5.

Ainsi, par indépendance de X et Y ,

$$G_S(z) = G_X(z)G_Y(z) = z^2 Q_X(z)Q_Y(z) = z^2 R(z).$$

Comme R est de degré 10, Q_X et Q_Y sont de degré 5. Or tout polynôme à coefficients réels et de degré impair a une racine réelle. Donc R aurait au moins une racine réelle. On arrive à une contradiction car on vient de montrer en 1. que R n'a aucune racine réelle...

Il est donc impossible de réaliser un trucage de deux dés tel que la somme des points suive un loi uniforme.

Exercice 12 :

$X + Y$ est à valeurs dans \mathbb{N} . Pour tout $k \in \mathbb{N}$, on a $\mathbb{P}(X + Y = k) = \sum_{j=0}^k \mathbb{P}(X + Y = k, Y = j)$ et donc par indépendance, $\mathbb{P}(X + Y = k) = \sum_{j=0}^k \mathbb{P}(X = k - j)\mathbb{P}(Y = j)$. On trouve alors que $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.

Exercice 13 :

On suppose que pour tout $\varepsilon > 0$, $\sum_{n=0}^{\infty} \mathbb{P}(|X_n| > \varepsilon) < +\infty$. Montrons que X_n converge p.s. vers 0.

$$\begin{aligned} X_n \xrightarrow{ps} 0 &\Leftrightarrow \mathbb{P}(\{\omega \in \Omega : \forall \varepsilon > 0, \exists N(\omega), n \geq N(\omega) \Rightarrow |X_n(\omega)| \leq \varepsilon\}) = 1 \\ &\Leftrightarrow \mathbb{P}(\{\omega \in \Omega : \forall k \geq 1, \exists N(\omega), n \geq N(\omega) \Rightarrow |X_n(\omega)| \leq 1/k\}) = 1 \\ &\Leftrightarrow \mathbb{P}\left(\bigcap_{k \geq 1} \bigcup_{N \geq 1} \bigcap_{n \geq N} \{|X_n| \leq 1/k\}\right) = 1 \\ &\Leftrightarrow 1 - \mathbb{P}\left(\bigcup_{k \geq 1} \bigcap_{N \geq 1} \bigcup_{n \geq N} \{|X_n| > 1/k\}\right) = 1 \\ &\Leftrightarrow 1 - \mathbb{P}\left(\bigcup_{k \geq 1} \limsup_n \{|X_n| > 1/k\}\right) = 1. \end{aligned}$$

D'après le lemme de Borel-Cantelli, on a pour tout $k \geq 1$, $\mathbb{P}(\limsup_n \{|X_n| > 1/k\}) = 0$. Une réunion dénombrable d'événements de probabilité nulle étant encore de probabilité nulle (exercice), on en déduit que $\mathbb{P}(X_n \xrightarrow{ps} 0) = 1$. La réciproque est vraie si les X_n sont indépendants pour pouvoir utiliser le point 2. du Lemme de Borel-Cantelli.

Exercice 14 :

Notons tout d'abord que pour tout $x \in [0, 1]$,

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leq x) \\ &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &= \mathbb{P}(X_1 \leq x)^n \text{ car } X_1, \dots, X_n \text{ sont i.i.d.} \\ &= x^n. \end{aligned}$$

Déterminons la fonction de répartition de Y_n . Y_n est à valeurs dans $[0, n]$. Pour tout $t \in [0, n]$,

$$F_{Y_n}(t) = \mathbb{P}(1 - M_n \leq t/n) = \mathbb{P}(M_n \geq 1 - t/n) = 1 - \mathbb{P}(M_n \leq 1 - t/n) = 1 - (1 - t/n)^n.$$

Montrons que $(Y_n)_n$ converge en loi vers une variable aléatoire Y de loi exponentielle de paramètre 1. Soit $t \geq 0$. Pour n suffisamment grand, on a $0 \leq t \leq n$ et donc

$$F_{Y_n}(t) \xrightarrow[n \rightarrow +\infty]{} F(t),$$

où $F(t) = 1 - e^{-t}$ est la fonction de répartition d'une loi exponentielle de paramètre 1.

2 Statistiques

Exercice 1 :

1. $\mathbb{E}(X) = a$ et $\text{Var}(X) = \frac{4a^2}{3}$.
2. cf cours.
3. T est à valeurs dans $[0, 2a]$. On calcule d'abord la fonction de répartition de T : pour tout $t \in [0, 2a]$, on a :

$$\begin{aligned} F_T(t) &= \mathbb{P}(\max(X_1, \dots, X_n) \leq t) \\ &= \mathbb{P}(X_1 \leq t, \dots, X_n \leq t) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq t) \text{ par indépendance} \\ &= \mathbb{P}(X_1 \leq t)^n \end{aligned}$$

Or, par calcul, $F_{X_1}(t) = \frac{t}{2a}$ donc $F_T(t) = \left(\frac{t}{2a}\right)^n$.

Donc F_T est continue sur \mathbb{R} , donc $f_T(t) = F'_T(t) = \frac{n}{2a} \left(\frac{t}{2a}\right)^{n-1}$.

On a alors :

$$\mathbb{E}(T) = \int_0^{2a} t f_T(t) dt = \frac{2n}{n+1} a.$$

On en déduit que $\hat{T} = \frac{n+1}{2n} T$ est un estimateur sans biais de a . On a :

$$\mathbb{E}(T^2) = \frac{4n}{n+2} a^2 \quad \text{et} \quad \text{Var}(T) = \frac{4na^2}{(n+2)(n+1)^2}.$$

Donc $\text{Var}(\hat{T}) = \left(\frac{n+1}{2n}\right)^2 \text{Var}(T) = \frac{4a^2}{n(n+2)} \xrightarrow[n \rightarrow +\infty]{} 0$ donc \hat{T} est consistant.

4. Pour comparer deux estimateurs, on calcule leurs risques quadratiques.

$$R(\hat{T}, a) = \text{Var}(\hat{T}) + b(\hat{T})^2 = \text{Var}(\hat{T}).$$

Or $\text{Var}(\hat{T}) < \text{Var}(\bar{X})$ pour $n \geq 10$ donc \hat{T} est un meilleur estimateur que \bar{X} .

Exercice 2 :

On définit l'estimateur

$$S_n := \frac{1}{n} \sum_{k=1}^n (X_k - m)^2.$$

On montre que S_n est sans biais : $\mathbb{E}[S_n] = \sigma^2$.

Montrons que S_n est un estimateur convergent.

$$\begin{aligned} \text{Var}(S_n) &= \frac{1}{n^2} \text{Var} \left(\sum_{k=1}^n (X_k - m)^2 \right) \\ &= \frac{1}{n^2} \sum_{k=1}^n \text{Var}((X_k - m)^2) \quad \text{car les } (X_k - m)^2 \text{ sont indépendants} \\ &= \frac{1}{n} \text{Var}((X_1 - m)^2) \quad \text{car les } X_k \text{ ont la même loi} \\ &= \frac{1}{n} (\mathbb{E}[(X_1 - m)^4] - \mathbb{E}[(X_1 - m)^2]^2) \\ &= \frac{1}{n} (\mathbb{E}[(X_1 - m)^4] - \sigma^4). \end{aligned}$$

D'après l'inégalité de Bienaymé-Tchebychev, on montre que pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|S_n - \sigma^2| \leq \varepsilon) = 0.$$

Exercice 3 :

Même démarche que dans le cours.

Exercice 4 :

1. Par intégration par parties, on trouve :

$$\mathbb{E}(X) = \int_{\theta}^{+\infty} x f_{\theta}(x) dx = 2\theta.$$

2. Comme $\mathbb{E}(X) = 2\theta$, et que $\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$ est un estimateur de $\mathbb{E}(X)$, on en déduit que $\hat{T} = \frac{\bar{X}_n}{2}$ est un estimateur de θ .

\hat{T} est sans biais par construction. En effet,

$$\mathbb{E}(\hat{T}) = \frac{1}{2} \mathbb{E}(\bar{X}_n) = \theta.$$

De plus, d'après la loi des grands nombres, $\bar{X}_n \xrightarrow{ps} \mathbb{E}(X) = 2\theta$ donc $\hat{T} \xrightarrow{ps} \theta$, ie. \hat{T} est consistant.

3. D'après le TCL, on sait que

$$\sqrt{\frac{n}{\text{Var}(X)}} (\bar{X}_n - 2\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Or par intégration par parties, on a $\text{Var}(X) = \theta^2$, donc :

$$\begin{aligned}
 1 - \alpha &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{\frac{n}{\theta^2}} |\bar{X}_n - 2\theta| \leq q_\alpha \right) \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{n} \left| \frac{\bar{X}_n - 2\theta}{\theta} \right| \leq q_\alpha \right) \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\sqrt{n} \left| \frac{\hat{T} - \theta}{\theta/2} \right| \leq q_\alpha \right) \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(-q_\alpha \leq \sqrt{n} \frac{\hat{T} - \theta/2}{\theta/2} \leq q_\alpha \right) \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(-\theta \frac{q_\alpha}{2\sqrt{n}} \leq \hat{T} - \theta/2 \leq \theta \frac{q_\alpha}{2\sqrt{n}} \right) \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{\hat{T}}{1 - q_\alpha/2\sqrt{n}} \leq \theta \leq \frac{\hat{T}}{1 + q_\alpha/2\sqrt{n}} \right) \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P} \left(\theta \in \left[\frac{\hat{T}}{1 + q_\alpha/2\sqrt{n}}; \frac{\hat{T}}{1 - q_\alpha/2\sqrt{n}} \right] \right)
 \end{aligned}$$

où q_α est le quantile d'ordre $1 - \alpha/2$ de la loi normale. Pour $1 - \alpha = 0,95$, on a $q_\alpha = 1,96$. De plus, avec $n = 100$, on a $\hat{T} = 3,3$ donc l'intervalle est :

$$I = [3; 3,67].$$

Exercice 5 :

1. Comme $\mathbb{E}[X_1] = \theta/3$, on en déduit que $\hat{\theta} := 3\bar{X}_n$ est un estimateur de θ .
2. Par construction, $\hat{\theta}$ est sans biais. On a donc

$$R(\hat{\theta}, \theta) = \text{Var}(\hat{\theta}) = \frac{9}{n} \text{Var}(X_1) = \frac{9}{n} \left(\frac{1}{3} - \frac{\theta^2}{9} \right),$$

car $\mathbb{E}[X_1^2] = 1/3$.

3. Le TCL assure que

$$\lim_{n \rightarrow +\infty} \sqrt{n} \left(\bar{X}_n - \frac{\theta}{3} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(\frac{1}{3} - \frac{\theta^2}{9} \right) \right),$$

d'où

$$\lim_{n \rightarrow +\infty} \sqrt{n} (\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N} (0, 3 - \theta^2).$$

D'après le Lemme de Slutsky,

$$\lim_{n \rightarrow +\infty} \sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{3 - \hat{\theta}^2}} \xrightarrow{\mathcal{L}} \mathcal{N} (0, 1).$$

Ainsi, pour $\alpha \in]0, 1[$, si on note q_α le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{N} (0, 1)$ et F sa fonction de répartition, on obtient

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta \left(\sqrt{n} \left| \frac{\hat{\theta} - \theta}{\sqrt{3 - \hat{\theta}^2}} \right| \leq q_\alpha \right) = 2F(q_\alpha) - 1 = 1 - \alpha,$$

d'où l'intervalle de confiance asymptotique pour θ au niveau (1_α) :

$$\left[\hat{\theta} - q_\alpha \sqrt{\frac{3 - \hat{\theta}^2}{n}}; \hat{\theta} + q_\alpha \sqrt{\frac{3 - \hat{\theta}^2}{n}} \right].$$

Exercice 6 :

1. La plus grande valeur des X_i doit "s'approcher" de la borne supérieure de l'intervalle $[0, \theta]$, ie de θ .
2. Déterminons la loi de $\hat{\theta}$. Soit $t \in \mathbb{R}^+$, la fonction de répartition de $\hat{\theta}$ vaut

$$F_{\hat{\theta}}(t) = \mathbb{P}_{\theta}(\hat{\theta} \leq t) = \mathbb{P}_{\theta}(\max(X_1, \dots, X_n) \leq t) = \mathbb{P}_{\theta}(X_1 \leq t)^n = \begin{cases} \left(\frac{t}{\theta}\right)^n & \text{si } t \leq \theta \\ 1 & \text{sinon.} \end{cases}$$

On en déduit que la densité de $\hat{\theta}$ vaut :

$$f_{\hat{\theta}}(t) = \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} \mathbb{1}_{[0, \theta]}(t).$$

3. On calcule alors

$$\mathbb{E}_{\theta}[\hat{\theta}] = \int_{\mathbb{R}^+} t f_{\hat{\theta}}(t) dt = \frac{n}{\theta^n} \int_0^{\theta} t^n dt = \frac{n}{\theta^n} \int_0^{\theta} t^n dt = \frac{n}{n+1} \theta.$$

Donc $\hat{\theta}$ est un estimateur biaisé mais est asymptotiquement sans biais car $\lim_{n \rightarrow +\infty} \mathbb{E}_{\theta}[\hat{\theta}] = \theta$.

Exercice 7 :

La probabilité de tirer un as est $= 1/8$. On note X_i la variable aléatoire qui modélise le résultat du i -ième tirage : $X_i = 1$ si le joueur obtient un as et $X_i = 0$ sinon. Les X_i sont i.i.d. et de même loi de Bernoulli de paramètre $\theta = 1/8$. D'après le cours, si le joueur ne triche pas au niveau de confiance 95%, la fréquence de succès doit être dans l'intervalle

$$\left[\theta - \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}} q; \theta + \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}} q \right],$$

où q est le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{N}(0, 1)$. L'application numérique fournit $[0, 033; 0, 217]$.

Le joueur a obtenu une fréquence de succès de $11/50 = 0,22$ ce qui est hors de l'intervalle de fluctuation. Par conséquent, au niveau de confiance 95%, on peut mettre en doute l'honnêteté du joueur !