

Leçons de probabilités.

Thierry MEYRE, Université Paris-Diderot

Préparation à l'agrégation de Mathématiques. IPEST

Ce document est téléchargeable en ligne à l'adresse

<http://www.proba.jussieu.fr/pageperso/meyre>

BIBLIOGRAPHIE

- BIL Probability and Measure. Billingsley. Wiley.
- BAR** Probabilité. Barbe, Ledoux. Belin.
- BC** Modélisation stochastique et simulation. Bercu, Chafaï. Dunod.
- BJ** Probabilités. Brancovan, Jeulin. Ellipses.
- COT** Exercices de Probabilités. Cottrell et al. Cassini.
- DAC¹ cours Probabilités et Statistiques 1. Problèmes à temps fixe.
Dacunha-Castelle, Duflo. 2e édition. Editions Masson.
- DAC¹ exo Exercices de Probabilités et Statistiques 1. Problèmes à temps fixe.
Dacunha-Castelle, Duflo. 3e tirage corrigé. Masson.
- DUR Probability Theory and Examples. Durrett. Wadsworth.
- FOA Calcul des Probabilités. Foata, Fuchs. Dunod.
- FEL¹ An Introduction to Probability Theory and its Applications. Feller.
Wiley 3rd edition. Volume I
- FEL² An Introduction to Probability Theory and its Applications. Feller.
Wiley 2nd edition. Volume II
- JP** L'essentiel en théorie des probabilités. Jacod et Protter.
Cassini.
- LEJ Statistique, la théorie et ses applications. Lejeune. Springer.
- MAZ Calcul de Probabilités. Exercices et problèmes corrigés.
Mazliak. Éditions Hermann.
- MET Notions fondamentales de la théorie des probabilités. Métivier. Dunod.
- NEV Bases mathématiques du calcul des probabilités. Neveu. Masson.
- OUV** Probabilités (2 tomes). Ouvrard. Cassini
- REV** Probabilités. Revuz. Éditions Hermann. Collection Méthodes.
- REV int Mesure et Intégration. Revuz. Éditions Hermann. Collection Méthodes.
- RS Statistique en action. Rivoirard et Stoltz. Vuibert.
- ROS Initiation aux Probabilités. Ross. Presses polytechniques et universitaires romandes. Troisième édition.

RUD Analyse réelle et complexe. Rudin. Masson.

SHI Probability. Shiryaev. Springer.

TOU Thèmes de probabilité. Toulouse.

Avertissement

Ce document a pour objectif de vous aider à préparer les leçons de probabilité qui font partie de la liste générale des leçons posées à l'oral d'analyse. Il ne s'agit en aucun cas de plans «tout prêts» : vous constaterez que, sur certains titres, la liste des suggestions est beaucoup trop longue pour constituer un exposé d'un quart d'heure !

Il s'agit plutôt d'un recueil d'idées, organisées par thématiques, que vous pouvez utiliser pour réfléchir à votre propre plan et à votre choix de développement.

Chapitre 1

Loi binomiale, loi de Poisson. Applications.

1.1 Lois binomiales

1.1.1 Généralités

Loi de Bernoulli $B(p)$.

La loi $B(n, p)$ est la loi de $\sum_{i=1}^n X_i$, où les variables X_i , $1 \leq i \leq n$ sont indépendantes et de même loi $B(p)$. Exemple du pile ou face.

Si $X \sim B(n, p)$, $E[X] = np$, $\sigma^2(X) = np(1-p)$ et la fonction génératrice de X vaut $E[u^X] = (1-p+pu)^n$.

1.1.2 Application au théorème de Weierstrass

En utilisant l'inégalité de Bienaymé-Tchebychev, on démontre :

Théorème 1.1.1 *Si $f \in C^0([0, 1], \mathbb{R})$, alors elle est limite uniforme sur $[0, 1]$ de la suite polynomiale $(f_n)_{n \in \mathbb{N}^*}$ (polynômes de Bernstein) définie par :*

$$f_n(x) = \sum_{k=0}^n C_n^k x^k (1-x)^{n-k} f\left(\frac{k}{n}\right) = E\left[f\left(\frac{X_1 + \dots + X_n}{n}\right)\right],$$

avec $X_i \sim B(x)$ indépendantes.

1.1.3 Inégalité de grandes déviations et application statistique

Considérons des variables X_i , $1 \leq i \leq n$ indépendantes et de même loi $B(p)$ et posons $S_n = \sum_{i=1}^n X_i$. [SHI I paragraphe 6] établit le résultat suivant :

Proposition 1.1.2

$$\forall \epsilon > 0 \quad P \left[\left| \frac{S_n}{n} - p \right| \geq \epsilon \right] \leq 2 \exp(-2n\epsilon^2) \quad (1.1)$$

Cette proposition admet une généralisation dans le cas de variables presque sûrement bornées : c'est l'inégalité de Hoeffding (voir chapitre 6.2.2), qui est une inégalité de type «grandes déviations». Dans les deux cas, la démonstration utilise la transformation de Laplace.

D'un point de vue statistique, l'inégalité (1.1) permet de construire des intervalles de confiance, même en dehors du cas « n grand». Il est intéressant de comparer cette inégalité à celle obtenue par Bienaymé-Tchebychev :

$$\forall \epsilon > 0 \quad P \left[\left| \frac{S_n}{n} - p \right| \geq \epsilon \right] \leq \frac{p(1-p)}{n\epsilon^2}$$

Bien sûr, pour n suffisamment grand, c'est l'inégalité (1.1) qui est la plus précise.

1.1.4 Loi binomiale négative

Voir section 2.2.2 page 13.

1.2 Loi de Poisson

1.2.1 Généralités

Définition. Si $X \sim \mathcal{P}(\lambda)$, alors $E[X] = \lambda$, $\sigma^2(X) = \lambda$.
La fonction génératrice vaut $E[u^X] = \exp(\lambda(u-1))$ et $\mathcal{P}(\lambda) \star \mathcal{P}(\mu) = \mathcal{P}(\lambda+\mu)$.

1.2.2 Théorème de Raïkov

Théorème 1.2.1 *Si M et N sont deux v.a.r. indépendantes telles que $M + N \sim \mathcal{P}(\lambda)$, alors il existe $\tau \in \mathbb{R}$ tel que $M + \tau$ et $N - \tau$ suivent des lois de Poisson*

La démonstration se trouve dans [BJ 173] et utilise les fonctions génératrices ainsi que le résultat d'analyse complexe suivant (une version du théorème de Liouville) : Soit $n_0 \in \mathbb{N}$ et F une fonction entière telle que :

$$\sup_{z \in \mathbb{C}} \frac{\operatorname{Re} F(z)}{1 + |z|^{n_0}} < +\infty.$$

Alors F est un polynôme de degré n_0 au plus.

1.2.3 Lois infiniment divisibles sur \mathbb{N}

Définition 1.2.2 Une loi de probabilité μ est dite *infiniment divisible* si pour tout entier $n \in \mathbb{N}^*$, il existe une loi ν_n telle que $\mu = (\nu_n)^{*n}$.

Définition 1.2.3 On appelle loi de Poisson composée sur \mathbb{N} toute loi de probabilité d'une variable aléatoire de la forme

$$S = \sum_{i=0}^N X_i,$$

où $(X_i)_{i \in \mathbb{N}}$ est une suite de v.a. à valeurs dans \mathbb{N} indépendantes identiquement distribuées et N est une variable de Poisson indépendante de la suite $(X_i)_{i \in \mathbb{N}}$.

Remarque: La notation précédente signifie que, pour tout $\omega \in \Omega$,

$$S(\omega) = \sum_{i=0}^{N(\omega)} X_i(\omega).$$

On vérifie que S est bien une variable aléatoire en l'écrivant sous la forme :

$$S = \sum_{n=0}^{+\infty} \mathbf{1}_{\{N=n\}} \sum_{i=0}^n X_i.$$

Théorème 1.2.4 Les lois infiniment divisibles sur \mathbb{N} sont exactement les lois de Poisson composées.

Démonstration: Le lecteur trouvera la preuve, qui utilise les fonctions génératrices, dans [FEL¹ 288-290] □

1.2.4 Processus de Poisson

Même si les processus aléatoires ne figurent pas au programme du tronçon commun, il est possible d'aborder dans cette leçon la notion de processus de Poisson car la construction en est élémentaire : voir par exemple [OUV² 171-177].

Une application des processus de Poisson qui peut faire l'objet d'un développement est le «paradoxe de l'autobus» [COT 98-101], appelé plus classiquement *paradoxe de l'inspection* dans d'autres ouvrages.

1.3 Approximation de la loi binomiale par la loi de Poisson

1.3.1 Deux théorèmes-limites

Le résultat fondamental faisant le lien entre loi binomiale et loi de Poisson est le *théorème de Poisson* :

Théorème 1.3.1

$$np_n \xrightarrow[n \rightarrow +\infty]{} \lambda \implies B(n, p_n) \xrightarrow[n \rightarrow +\infty]{(e)} \mathcal{P}(\lambda)$$

La convergence dont il est ici question est la *convergence étroite* qui s'exprime simplement dans le cas de lois sur \mathbb{N} sous la forme :

$$\forall k \in \mathbb{N} \quad B(n, p_n)(\{k\}) \xrightarrow[n \rightarrow +\infty]{} \mathcal{P}(\lambda)(\{k\})$$

Le théorème de Poisson admet la généralisation suivante, appelée *théorème des événements rares* [OUV² 321-322] :

Théorème 1.3.2 *Pour tout $n \in \mathbb{N}^*$, on considère une famille d'événements indépendants $(A_{n,j})_{1 \leq j \leq m_n}$ et la variable aléatoire $S_n := \sum_{j=1}^{m_n} \mathbf{1}_{A_{n,j}}$.*

En notant $p_{n,j} := P(A_{n,j})$, on fait les trois hypothèses suivantes :

$$m_n \xrightarrow[n \rightarrow +\infty]{} +\infty \quad ; \quad \max_{1 \leq j \leq m_n} p_{n,j} \xrightarrow[n \rightarrow +\infty]{} 0 \quad ; \quad \sum_{j=1}^{m_n} p_{n,j} \xrightarrow[n \rightarrow +\infty]{} \lambda > 0.$$

Alors la suite (S_n) converge en loi vers la loi de Poisson de paramètre λ , i.e.

$$P_{S_n} \xrightarrow[n \rightarrow +\infty]{(e)} \mathcal{P}(\lambda)$$

On démontre ce théorème en utilisant les fonctions génératrices.

Le théorème des événements rares explique pourquoi la loi de Poisson apparaît dans de nombreux phénomènes naturels : [FEL¹ 159-164] donne beaucoup d'exemples. [FOA 73] donne un exemple et en tire une application statistique à la médecine.

1.3.2 Vitesse de convergence dans le thm de Poisson

Considérons une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ telle que, pour tout $n \in \mathbb{N}^*$, $X_n \sim B(n, \lambda/n)$. D'après le théorème de Poisson, (X_n) converge en loi vers $\mathcal{P}(\lambda)$, ce qui s'écrit :

$$\forall k \in \mathbb{N} \quad P(X_n = k) \xrightarrow{n \rightarrow +\infty} \frac{\lambda^k}{k!} \exp(-\lambda)$$

Le résultat suivant [SHI] précise dans un sens très fort la vitesse à laquelle a lieu cette convergence.

Proposition 1.3.3

$$\sum_{k \geq 0} \left| P(X_n = k) - \frac{\lambda^k}{k!} \exp(-\lambda) \right| \leq \frac{2\lambda^2}{n}$$

Il est remarquable que l'on soit capable d'estimer la somme des valeurs absolues des erreurs commises en approchant chacun des poids. En particulier, il est facile de déduire de la Proposition que pour toute fonction bornée f de \mathbb{R} dans lui-même :

$$\left| E[f(X_n)] - \sum_{k \geq 0} f(k) \frac{\lambda^k}{k!} \exp(-\lambda) \right| \leq \frac{2\lambda^2}{n} \times \|f\|_\infty$$

Règle empirique : On peut approcher de façon satisfaisante la loi $B(n, p)$ par la loi $\mathcal{P}(np)$ dès que $n \geq 50$ et $p \leq 0, 1$.

Notons que l'on peut aussi utiliser cette approximation dans le cas $n \geq 50$ et $p \geq 0, 9$ en intervertissant pile et face (ou de façon plus générale, en comptant les échecs à la place des succès).

Si ces conditions ne sont pas satisfaites, on envisagera l'approximation de la loi $B(n, p)$ par la loi gaussienne $\mathcal{N}(np, np(1-p))$: voir section 2.3.1 page 14.

Chapitre 2

Le jeu de pile ou face (suites de v.a. de Bernoulli indépendantes)

2.1 Suite finie

Construction d'un jeu de pile ou face à n jets [OUV¹ 66].
Loi binomiale $B(n, p)$: voir sections 1.1.1 à 1.1.3.

2.2 Suite infinie

2.2.1 Construction

Elle est nettement plus subtile que celle d'une suite finie.
Dans le cas d'un pile ou face équitale ($p = 1/2$), il existe une construction astucieuse sur l'espace $(\Omega = [0, 1[, \mathcal{B}([0, 1[))$ muni de la mesure de Lebesgue : elle fait appel au développement en base 2 d'un nombre réel, cf. [REV 47] ou [OUV² 55].

Dans le cas $p \neq 1/2$, on utilise la construction générale d'une suite de variables aléatoires indépendantes de lois données : voir section 3.1 page 17.

2.2.2 Loi binomiale négative

Dans un jeu de pile ou face (pile avec proba p , face avec proba $q = 1 - p$), soit T_r le nombre de résultats «face» obtenus lorsque «pile» apparaît pour la r -ième fois ($r \in \mathbb{N}^*$). Alors, on a [FOA 74-75] :

$$\forall k \in \mathbb{N} \quad P(T_r = k) = \binom{r+k-1}{r-1} p^r q^k = \binom{-r}{k} p^r (-q)^k$$

Ceci définit la loi binomiale négative¹ de paramètres $r \in \mathbb{N}^*$ et $p \in]0, 1[$, que nous noterons $B^-(r, p)$. Son espérance vaut [JP 33] : $E[T_r] = r(1-p)/p$.

2.2.3 Ruine du joueur

Un joueur gagne (resp. perd) 1 dinar quand la pièce tombe sur pile (resp. face). Il décide de continuer à jouer jusqu'à ce qu'il ait gagné a dinars ou perdu b dinars ($a \in \mathbb{N}^*, b \in \mathbb{N}^*$). Autrement dit, si nous notons S_n son gain algébrique au bout du n -ième coup, le jeu s'arrête à l'instant (aléatoire) :

$$T_{a,b} = \min\{n \in \mathbb{N}^*, S_n = a \text{ ou } S_n = -b\}$$

C'est un cas particulier de *marche aléatoire avec double barrière*.

On montre que $T_{a,b} < +\infty$ presque sûrement et même que $T_{a,b}$ admet des moments de tous les ordres [DAC¹ cours 53-54].

2.3 Résultats asymptotiques

2.3.1 Théorème de De Moivre-Laplace

Théorème 2.3.1 Soit $p \in]0, 1[$ fixé et $(S_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires telle que $S_n \sim B(n, p)$ pour tout $n \in \mathbb{N}^*$.

Alors, nous avons la convergence en loi suivante :

$$\frac{S_n - np}{\sqrt{np(1-p)}} \longrightarrow \mathcal{N}(0, 1)$$

Le résultat suivant, qui est un cas particulier du théorème de Berry-Esseen, précise la vitesse à laquelle cette convergence en loi a lieu.

Théorème 2.3.2 Si F_n est la fonction de répartition de $\frac{S_n - np}{\sqrt{np(1-p)}}$ et si F est celle de la loi $\mathcal{N}(0, 1)$, on a :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{C[p^2 + (1-p)^2]}{\sqrt{np(1-p)}},$$

où $C \simeq 0,7975$.

1. encore appelée *loi de Pascal*

Un simple changement de variable affine nous donne le même majorant uniforme si l'on compare la fonction de répartition de S_n , donc de la loi $B(n, p)$, et la fonction de répartition de la loi $\mathcal{N}(np, np(1-p))$. Ceci nous fournit une approximation gaussienne de la loi $B(n, p)$.

Néanmoins, si $np(1-p)$ est trop petit, le théorème de Berry-Esseen ne nous donne que peu de garantie sur la qualité de cette approximation. Effectivement, la loi $B(n, p)$ est dans ce cas trop asymétrique pour que l'approximation gaussienne soit valable. C'est pour éviter cet écueil que l'on utilise la règle empirique suivante.

Règle empirique : On peut approcher la loi $B(n, p)$ par la loi gaussienne $\mathcal{N}(np, np(1-p))$ dès que $np(1-p) \geq 10$.

Lorsqu'on applique cette règle empirique, on introduit une «correction de continuité» pour tenir compte du fait que l'on approche une loi discrète par une loi continue. Plus précisément, si $X \sim B(n, p)$ et $Y \sim \mathcal{N}(np, np(1-p))$, nous écrivons :

$$\forall k \in \mathbb{N} \quad P(X = k) \approx P\left(k - \frac{1}{2} \leq Y \leq k + \frac{1}{2}\right)$$

2.3.2 Application statistique : intervalle de confiance

Le théorème de De Moivre-Laplace nous donne la construction d'un intervalle de confiance de niveau asymptotique 95% (par exemple) pour estimer le paramètre $\theta \in]0, 1[$ lorsqu'on dispose d'un n -échantillon de loi $B(\theta)$, avec n «grand» :

$$\forall \theta \in]0, 1[\quad P_\theta\left(\theta \in \left[\bar{X}_n - \frac{1,96}{2\sqrt{n}}, \bar{X}_n + \frac{1,96}{2\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow +\infty} 0,95$$

où $\bar{X}_n = S_n/n$ est la *moyenne empirique* de notre échantillon.

La valeur 1,96 qui apparaît vient de la valeur approchée bien connue des statisticiens :

$$\frac{1}{\sqrt{2\pi}} \int_{-1,96}^{1,96} e^{-x^2/2} dx \simeq 0,95$$

Cet intervalle de confiance est bien plus précis que celui obtenu par l'inégalité de Bienaymé-Tchebychev (cf. [OUV¹ 237]) mais il n'est valide que si n est suffisamment grand (règle empirique : $n \geq 30$).

Si n est petit, on peut utiliser Bienaymé-Tchebychev ou l'inégalité de grandes déviations énoncée dans la section 1.1.3 page 8.

Dans le cas « n grand», on peut encore améliorer un peu notre intervalle de confiance asymptotique en faisant une estimation de la variance inconnue $\theta(1 - \theta)$ par $\bar{X}_n(1 - \bar{X}_n)$ [BAR 146] : lorsque $n \rightarrow +\infty$,

$$P_\theta\left(\theta \in \left[\bar{X}_n - \frac{a}{\sqrt{n}}\sqrt{\bar{X}_n(1 - \bar{X}_n)}, \bar{X}_n + \frac{b}{\sqrt{n}}\sqrt{\bar{X}_n(1 - \bar{X}_n)}\right]\right) \rightarrow \int_{-a}^b \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

2.3.3 Loi du logarithme itéré

Ce résultat [FOA 251-256], qui admet des généralisations aux *marches aléatoires*, a d'abord été démontré par Khintchine en 1924 dans le cas du pile ou face. Nous considérons ici une suite indépendante de variables $(X_i)_{i \in \mathbb{N}^*}$ suivant toutes la loi de Bernoulli symétrique, c'est-à-dire :

$$P(X_i = -1) = P(X_i = 1) = \frac{1}{2}$$

Pour tout $n \in \mathbb{N}^*$, nous posons $S_n = \sum_{i=1}^n X_i$. Notons que S_n représente le gain algébrique du joueur décrit dans la section 2.2.3 au bout du n -ième coup, en supposant qu'il joue avec une pièce équilibrée ($p = 1/2$).

On déduit facilement du théorème de De Moivre-Laplace la convergence en loi suivante : $S_n/\sqrt{n} \rightarrow \mathcal{N}(0, 1)$. Nous pouvons donc dire *grosso modo* que les fluctuations du gain S_n sont d'ordre de grandeur \sqrt{n} .

La loi du logarithme itéré va nettement préciser cette idée :

Théorème 2.3.3 *Presque sûrement,*

$$\limsup \frac{S_n}{\sqrt{2n \log \log n}} = 1 \quad ; \quad \liminf \frac{S_n}{\sqrt{2n \log \log n}} = -1$$

Chapitre 3

Indépendance d'évènements et de variables aléatoires. Exemples.

3.1 Constructions

La construction d'un nombre fini de variables discrètes indépendantes est relativement facile [OUV¹ 66] : nous avons vu le cas particulier d'un jeu de pile ou face dans le chapitre précédent.

En revanche, nous avons constaté que la construction d'un jeu de pile ou face équilibré infini était nettement plus subtile.

Une fois cette difficulté surmontée, nous pouvons de façon générale construire une suite de variables aléatoires réelles (Y_j) indépendantes et de lois respectives (μ_j) données [OUV² 58].

3.2 Quelques exemples

Indicatrice d'Euler [COT 8-9].

Statistiques d'ordre [REV 50] ou [COT 53-55].

3.3 Vecteurs aléatoires à densités, fonctions caractéristiques

Simulation d'une variable gaussienne par la méthode de Box-Muller [BC 57]. Soient X, Y i.i.d. de densité $f > 0$ sur \mathbb{R}_+ et nulle sur \mathbb{R}_-^* ; on définit les variables $U := X \wedge Y$, $V := |X - Y|$. Alors U, V indépendants **ssi** f densité exponentielle [OUV² 75-79].

Indépendance et lois marginales de $(X + Y, \frac{X}{X+Y})$ lorsque X et Y sont indépendantes de lois respectives $\gamma(\lambda, a)$ et $\gamma(\lambda, b)$ et application au calcul de la densité du $\chi^2(d)$ [OUV² 74-75].

Produit de convolution, f.c. : $\mathcal{N}(m_1, \sigma_1^2) \star \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$,
 $C(a) \star C(b) = C(a + b)$

3.4 Variance, covariance, cas gaussien

Des variables indépendantes sont décorréélées (*i.e.* de covariance nulle). La réciproque est fautive.

La variance d'une somme de v.a. indépendantes est égale à la somme des variances.

La caractérisation de l'indépendance dans le cas gaussien par la nullité des «covariances croisées» et son application à un cas particulier du théorème de Cochran ([REV 90] ou [COT 211]) sont traitées dans la section 4.4.2 page 27.

3.5 Sommes de v.a. indépendantes, thm s asymptotiques

Produit de convolution. Fonction caractéristique $\phi_{X+Y}(u) = \phi_X(u) \phi_Y(u)$.

Le lemme de Borel-Cantelli et sa réciproque. Trois applications :

- Le singe dactylographe de Borel [FOA 232]
- Le *théorème de Cantelli*, *i.e.* la loi forte des grands nombres dans le cas L^4 [BAR 133] qui suffit pour prouver l'application suivante.
- Maths pures : Presque tous les réels de $[0,1]$ sont *normaux*, *i.e.* tels que leur développement décimal fait apparaître tous les entiers de 0 à 9 avec la même fréquence asymptotique 1/10 ([REV 129] ou [QUEFFÉLEC-ZUILY 550-551])

Le TLC et plusieurs applications seront traités dans le chapitre suivant.

La ruine du joueur a été abordée dans la section 2.2.3.

Chapitre 4

Lois gaussiennes et applications

4.1 La loi normale centrée réduite

Proposition 4.1.1 (Intégrale de Gauss)

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

[GOURDON 163,335] démontre cette égalité par 3 méthodes différentes :

- dérivation d’une intégrale dépendant d’un paramètre
- inégalité de convexité, changement de variables et formule de Wallis
- passage en coordonnées polaires

□

Nous suivons maintenant la présentation de Foata et Fuchs dans leur ouvrage “Calcul des probabilités” chez Dunod, 2ème édition, pages 178 à 181.

Proposition et définition 4.1.2 *L’application $g : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par :*

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

est une densité de probabilité. La loi de densité g est appelée loi de Gauss ou loi normale (centrée réduite) et notée $\mathcal{N}(0, 1)$.

Gauss a introduit cette loi en 1809 à propos d’un problème statistique d’estimation de paramètre.

Le graphe de la densité de Gauss g est une «courbe en cloche» aplatie. On le trace facilement en notant que f est paire, qu’elle admet un maximum global en $x = 0$ égal à $\frac{1}{\sqrt{2\pi}} \simeq 0,399$ et que la courbe admet deux points

d'inflexion en $x = -1$ et $x = 1$.

Si $X \sim \mathcal{N}(0, 1)$, on calcule immédiatement $E[X] = 0$ par le théorème de transfert en utilisant la parité de g et

$$\text{Var}X = E[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

par intégration par parties. C'est pourquoi on parle de loi normale *centrée réduite*.

La fonction de répartition de la loi de Gauss est donnée par

$$\forall x \in \mathbb{R} \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

On vérifie facilement que Φ est de classe C^∞ , strictement croissante et donc bijective de \mathbb{R} sur son image $]0, 1[$, telle que $\Phi(-x) = 1 - \Phi(x)$ pour tout $x \in \mathbb{R}$.

Son graphe est une «courbe en S» assez étalée, symétrique par rapport à $(0, 1/2)$, où elle admet un point d'inflexion et où la pente de sa tangente vaut $\frac{1}{\sqrt{2\pi}} \simeq 0,399$.

On n'a pas de formule plus explicite pour Φ mais cette fonction et sa bijection réciproque Φ^{-1} sont tabulées, notamment pour des applications statistiques. Une valeur que nous utiliserons dans la suite est $\Phi(1,96) \approx 0,975$.

Si $X \sim \mathcal{N}(0, 1)$, alors $P(|X| \leq x) = 2\Phi(x) - 1$ pour tout $x \geq 0$, de sorte que $P(|X| \leq 1,96) \approx 0,95$.

On peut démontrer l'équivalence suivante lorsque $x \rightarrow +\infty$:

$$1 - \Phi(x) \sim \frac{1}{x} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

Comme cela tend « vite » vers 0, on dit que la loi de Gauss a une queue de distribution peu épaisse.

Proposition 4.1.3 *La loi de Gauss admet des moments de tous les ordres. Si $X \sim \mathcal{N}(0, 1)$, nous avons, pour tout $n \in \mathbb{N}$:*

$$E[X^{2n+1}] = 0 \quad ; \quad E[X^{2n}] = \frac{(2n)!}{2^n n!}$$

Démonstration: Pour tout $k \in \mathbb{N}^*$, nous avons $|x|^k g(x) = o(1/x^2)$ quand $x \rightarrow \pm\infty$ donc X admet un moment d'ordre k donné par :

$$E[X^k] = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dt$$

Lorsque $k = 2n + 1$, cette intégrale est nulle pour des raisons de parité. On démontre par intégration par parties que $E[X^{2n}] = (2n - 1)E[X^{2n-2}]$ pour tout $n \in \mathbb{N}^*$, ce qui nous donne par récurrence la formule annoncée puisque $E[X^0] = 1$ par convention. □

Pour simuler numériquement la loi de Gauss, on utilise la proposition suivante, qui se démontre grâce à un changement de variable dans une intégrale double [OUVRARD tome 2, 67-68]

Proposition 4.1.4 (Méthode de Box-Muller) *Soient U_1 et U_2 deux v.a.r. indépendantes, de même loi uniforme sur $[0, 1]$. Nous posons :*

$$X := \sqrt{-2 \log U_1} \cos(2\pi U_2) \quad , \quad Y := \sqrt{-2 \log U_1} \sin(2\pi U_2).$$

Alors les variables aléatoires X et Y sont indépendantes et de même loi $\mathcal{N}(0, 1)$.

4.2 La loi normale générale

Proposition et définition 4.2.1 *Soient $\mu \in \mathbb{R}$ et $\sigma > 0$ deux paramètres fixés. L'application $g_{\mu, \sigma^2} : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par :*

$$\forall x \in \mathbb{R} \quad g_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

est une densité de probabilité. La loi de densité g_{μ, σ^2} est appelée loi normale de paramètres (μ, σ^2) et notée $\mathcal{N}(\mu, \sigma^2)$.

On vérifie que g_{μ,σ^2} est bien une densité en effectuant un changement de variable affine dans une intégrale simple, puis en utilisant la proposition 4.1.2. Notons que pour $\mu = 0$ et $\sigma^2 = 1$, on retrouve bien la densité de la loi $\mathcal{N}(0, 1)$ telle qu'elle a été définie dans la sous-section précédente. Nous allons maintenant préciser les rapports entre la loi normale générale et la loi normale centrée réduite.

Proposition 4.2.2 *Soient $\mu \in \mathbb{R}$ et $\sigma > 0$ deux paramètres fixés. On considère deux variables aléatoires réelles X et Y telles que $Y = \mu + \sigma X$. Alors on a l'équivalence suivante :*

$$X \sim \mathcal{N}(0, 1) \iff Y \sim \mathcal{N}(\mu, \sigma^2)$$

Démonstration: On vérifie facilement que cette équivalence résulte de la proposition ??.

□

Pour simuler numériquement la loi $\mathcal{N}(\mu, \sigma^2)$, on applique d'abord la méthode de Box-Muller pour simuler $X \sim \mathcal{N}(0, 1)$ puis l'on calcule $Y = \mu + \sigma X$.

Corollaire 4.2.3 *Soit $Y \sim \mathcal{N}(\mu, \sigma^2)$. Alors $E[Y] = \mu$ et $\text{Var}Y = \sigma^2$*

Ce résultat justifie que le premier paramètre d'une loi normale soit traditionnellement noté μ ou m (comme «moyenne» probabiliste, c'est-à-dire espérance) et le second σ^2 puisqu'il est égal à la variance de la loi.

Corollaire 4.2.4 *La fonction de répartition Φ_{μ,σ^2} de la loi $\mathcal{N}(\mu, \sigma^2)$ est donnée par :*

$$\forall x \in \mathbb{R} \quad \Phi_{\mu,\sigma^2}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Démonstration: En reprenant les notations de la proposition 4.2.2, nous avons, pour tout $x \in \mathbb{R}$:

$$\Phi_{\mu,\sigma^2}(x) = P(Y \leq x) = P(\mu + \sigma X \leq x) = P\left(X \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

□

Avant d'établir une importante propriété de stabilité de la loi normale générale, nous introduisons la notion de *produit de convolution* de deux densités de probabilité.

Proposition 4.2.5 *Soient X et Y deux v.a.r. indépendantes admettant des densités respectives f et g . Alors la v.a.r. $X+Y$ admet pour densité le produit de convolution de f et g , noté $f \star g$ et défini par :*

$$\forall x \in \mathbb{R} \quad f \star g(x) = \int_{-\infty}^{+\infty} f(x-y)g(y) dy.$$

Démonstration: cf. [OUVRARD tome 1, 203-204]. □

La proposition suivante énonce la stabilité de la loi normale par produit de convolution.

Proposition 4.2.6 *Soient X et Y deux v.a.r. indépendantes de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$, avec $(m_1, m_2) \in \mathbb{R}^2$ et $(\sigma_1, \sigma_2) \in (\mathbb{R}_+^*)^2$. Alors la v.a.r. $X + Y$ suit la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Le résultat essentiel établi par cette proposition est que la somme de deux variables gaussiennes indépendantes est encore gaussienne. Il est alors facile de déterminer ses paramètres en calculant l'espérance et la variance.

Démonstration: Commençons par réduire le problème au cas où X est centrée réduite et Y centrée en supposant démontrée la proposition suivante : si $X' \sim \mathcal{N}(0, 1)$ et $Y' \sim \mathcal{N}(0, s^2)$, avec $s > 0$, sont indépendantes, alors $X' + Y' \sim \mathcal{N}(0, 1 + s^2)$. Nous pouvons alors en déduire la proposition dans le cas général en définissant les variables aléatoires réelles :

$$X' = \frac{X - m_1}{\sigma_1} \quad , \quad Y' = \frac{Y - m_2}{\sigma_1} \quad ,$$

si bien que les hypothèses précédentes sont satisfaites avec $s = \sigma_2/\sigma_1$ et :

$$X + Y = m_1 + \sigma_1 X' + m_2 + \sigma_1 Y' = m_1 + m_2 + \sigma_1 (X' + Y').$$

Par transformation affine de la variable $X' + Y' \sim \mathcal{N}(0, 1 + (\sigma_2^2/\sigma_1^2))$, nous obtenons alors la conclusion dans le cas général.

Nous supposons donc désormais $m_1 = m_2 = 0$, $\sigma_1 = 1$ et $\sigma_2 = s > 0$. D'après la proposition 4.2.5, la variable aléatoire réelle $X + Y$ admet une densité proportionnelle à :

$$\int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2}{2}} e^{-\frac{y^2}{2s^2}} dy = \int_{-\infty}^{+\infty} \exp\left(-\frac{s^2 x^2 - 2s^2 xy + (1 + s^2)y^2}{2s^2}\right) dy.$$

Cette dernière expression s'écrit encore :

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(1+s^2)\left(y - \frac{s^2}{1+s^2}x\right)^2}{2s^2}\right) \exp\left(\left(\frac{s^2}{1+s^2} - 1\right)\frac{x^2}{2}\right) dy,$$

ou encore, puisque l'intégration porte sur la variable y ,

$$\exp\left(-\frac{x^2}{2(1+s^2)}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{(1+s^2)\left(y - \frac{s^2}{1+s^2}x\right)^2}{2s^2}\right) dy.$$

Nous constatons que cette dernière intégrale est une constante en effectuant, à x fixé, le simple changement de variable $u = y - \frac{s^2}{1+s^2}x$.

Nous avons donc prouvé que la variable aléatoire réelle $X + Y$ admet une densité proportionnelle à $\exp(-x^2/(2(1+s^2)))$. Nous en déduisons que la seule constante de proportionnalité possible pour que ce soit effectivement une densité de probabilité vaut $1/\sqrt{2\pi(1+s^2)}$ et que $X + Y \sim \mathcal{N}(0, 1+s^2)$. \square

Pour une présentation de la loi normale dans \mathbb{R}^d , encore appelée loi de Laplace-Gauss en dimension d , le lecteur pourra lire [ESC 147ss] qui en fait une présentation cadrant bien avec le programme officiel.

4.3 Approximation normale de la loi binomiale

Historiquement, Abraham de Moivre, mathématicien anglais d'origine française, avait mis en évidence dès 1733 une approximation normale de la loi binomiale en étudiant un modèle de pile ou face avec une pièce équilibrée. Sa preuve reposait sur des calculs laborieux d'estimation des coefficients binomiaux.

Pierre-Simon de Laplace avait généralisé ce résultat en 1820 au cas d'une pièce éventuellement biaisée. Il avait également mis en évidence le lien entre la loi normale et la théorie des erreurs d'observation, dont nous reparlerons dans le chapitre suivant.

Nous suivons ici la présentation d'[OUVRARD tome 1, 228-229].

Théorème 4.3.1 (de Moivre, Laplace) *Considérons $p \in]0, 1[$ et, pour tout $n \in \mathbb{N}^*$, une variable aléatoire réelle S_n qui suit la loi binomiale $B(n, p)$. Nous noterons \tilde{S}_n la variable centrée réduite associée à S_n , i.e.*

$$\tilde{S}_n = \frac{S_n - E[S_n]}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Alors nous avons la convergence suivante, uniforme sur tous les intervalles réels I :

$$\sup_I \left| P(\tilde{S}_n \in I) - \int_I \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Autrement dit, quand n est «grand», \tilde{S}_n suit une loi approximativement égale à la loi normale centrée réduite, ce que nous notons $\tilde{S}_n \approx \mathcal{N}(0, 1)$.

En utilisant un changement de variable affine dans une intégrale simple, nous pouvons dire de façon équivalente que $S_n \approx \mathcal{N}(np, np(1-p))$, toujours pour n «grand». Plus précisément, nous avons le résultat suivant :

Corollaire 4.3.2 *Considérons $p \in]0, 1[$ et, pour tout $n \in \mathbb{N}^*$, une variable aléatoire réelle S_n qui suit la loi binomiale $B(n, p)$. Alors nous avons la convergence suivante, uniforme sur tous les intervalles réels I :*

$$\sup_I \left| P(S_n \in I) - \int_I \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(t-np)^2}{2np(1-p)}\right) dt \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Pour pouvoir utiliser cette approximation gaussienne de la loi binomiale, il nous reste à préciser ce que signifie n «grand» ! Le théorème de Berry¹-Esseen², dont nous donnerons l'énoncé général dans le chapitre suivant, met en évidence le rôle joué par la quantité $np(1-p)$ dans la qualité de l'approximation.

Proposition 4.3.3 (Berry-Esseen, cas binomial) *Nous reprenons les hypothèses et les notations du théorème 4.3.1 et nous notons en outre $F_{\tilde{S}_n}$ la fonction de répartition de la variable \tilde{S}_n et Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Nous avons alors, pour tout $n \in \mathbb{N}^*$:*

$$\sup_{x \in \mathbb{R}} |F_{\tilde{S}_n}(x) - \Phi(x)| \leq \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}.$$

Si $np(1-p)$ est suffisamment grand, le théorème de Berry-Esseen garantit que l'approximation de la loi binomiale par la loi normale sera bonne. Dans la pratique, certaines règles empiriques existent, variables d'ailleurs d'un ouvrage à l'autre ! Une d'entre elles dit que l'approximation de la loi $B(n, p)$ par la loi $\mathcal{N}(np, np(1-p))$ est considérée comme satisfaisante quand $np(1-p) > 18$ (référence : Dacunha-Castelle, Duflo, tome 1).

1. Andrew C. Berry

2. Carl-Gustav Esseen (1918-2001), mathématicien suédois

Si $np(1-p)$ est trop petit pour que l'on puisse utiliser l'approximation normale, on pourra alors utiliser avantageusement l'approximation de la loi binomiale par la loi de Poisson. OUVRARD écrit que cette dernière est justifiée lorsque $n \geq 30$ et $p \leq 0,1$. Notons que si $p \geq 0,9$, l'approximation poissonnienne est aussi utilisable en comptant les échecs à la place des succès (ou les faces au lieu des piles!).

Le lecteur pourra trouver des applications pratiques de l'approximation gaussienne d'une loi binomiale dans l'exercice 7.4 du livre d'Ouvrard déjà cité ou encore dans [REV 164-165] (exercice sur un serveur informatique).

4.4 Le cas de la dimension d : loi normale multivariée

4.4.1 Définition. Loi gaussienne dans \mathbb{R}^d

- Un vecteur aléatoire (X_1, \dots, X_d) est dit gaussien si toute combinaison linéaire $\sum a_i X_i$ est une v.a.r. gaussienne.
En particulier, les composantes X_i sont gaussiennes mais la réciproque est fautive : voir l'exercice 4.5.1.
Néanmoins, il y a une réciproque partielle : si les X_i sont mutuellement indépendantes et gaussiennes alors le vecteur aléatoire (X_1, \dots, X_d) est gaussien.
- Si X est un vecteur gaussien de dimension d et A une matrice de taille $d' \times d$ alors $Y = AX$ est un vecteur gaussien de dimension d' .
- Pour $X = (X_1, \dots, X_d)$ vecteur aléatoire *quelconque*, on définit son vecteur moyenne :

$$E[X] = (E[X_1], \dots, E[X_d])^*$$

et sa matrice de covariances :

$$K_X = \left[Cov(X_i, X_j) \right]_{1 \leq i, j \leq d}.$$

- Avec les définitions précédentes, la fonction caractéristique d'un vecteur aléatoire gaussien de dimension d vaut :

$$\forall u \in \mathbb{R}^d, \quad \varphi_X(u) = \exp(iu^* E[X] - \frac{1}{2} u^* K_X u).$$

En particulier, une loi gaussienne vectorielle est déterminée par sa moyenne m et sa matrice de covariances K .

4.4. LE CAS DE LA DIMENSION D : LOI NORMALE MULTIVARIÉE 27

Réciproquement, pour tout $m \in \mathbb{R}^d$ et pour toute matrice K de taille $d \times d$ symétrique semi-définie positive, il existe un vecteur gaussien de moyenne m et de matrice de covariances K .

On peut donc parler de la loi $\mathcal{N}_d(m, K)$.

- Si K est inversible, alors la loi $\mathcal{N}_d(m, K)$ a pour densité :

$$f(x) = (2\pi)^{-d/2} |K|^{-1/2} \exp(-(x - m)^* K^{-1} (x - m) / 2).$$

Sinon, la loi est portée par un sous-espace affine propre de \mathbb{R}^d et ne saurait donc avoir de densité : on parle alors de loi gaussienne dégénérée.

4.4.2 Critère d'indépendance

Théorème 4.4.1 (Indépendance dans le cas gaussien) *Nous supposons que $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ est un vecteur gaussien. Alors il y a équivalence entre les trois propriétés suivantes :*

1. les vecteurs $X = (X_1, \dots, X_m)$ et $Y = (Y_1, \dots, Y_n)$ sont indépendants
2. Pour tous $1 \leq i \leq m, 1 \leq j \leq n$, on a : $\text{Cov}(X_i, Y_j) = 0$
3. La matrice de covariances de (X, Y) est diagonale par blocs :

$$K_{(X,Y)} = \left(\begin{array}{c|c} K_X & 0 \\ \hline 0 & K_Y \end{array} \right).$$

††† C'est le "grand vecteur" (X, Y) qui doit être gaussien et non pas seulement chacun des sous-vecteurs X et Y : voir l'exercice 4.5.1 pour un contre-exemple.

- Tout ce qui vient d'être dit se généralise de 2 blocs à k blocs.
- En particulier, si $X = (X_1, \dots, X_d)$ est un vecteur gaussien, alors on a l'équivalence suivante :
Les v.a.r. X_1, \dots, X_d sont indépendantes **ssi** la matrice de covariances K_X est diagonale.
- Insistons sur un cas encore plus particulier :
Les v.a.r. X_1, \dots, X_d sont i.i.d. de loi commune $\mathcal{N}(0, 1)$ **ssi** $X = (X_1, \dots, X_d)$ est un vecteur gaussien centré et de matrice de covariances I_d .

à titre d'application de ce théorème, nous présentons un résultat qui est un cas particulier du théorème de Cochran [BC 242-243] :

Proposition 4.4.2 *Soit (X_1, \dots, X_n) un vecteur aléatoire de loi $\mathcal{N}(0, Id_n)$. On appelle moyenne empirique, resp. variance empirique corrigée associée à*

(X_1, \dots, X_n) la variable aléatoire réelle :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{resp.} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Alors, $\bar{X}_n \sim \mathcal{N}(0, \frac{1}{n})$, $R_n := (n-1)S_n^2 \sim \chi^2(n-1)$ et ces deux variables sont indépendantes.

Démonstration: Nous utiliserons le résultat suivant : La transformée de Laplace de la loi $\chi^2(n)$ est définie pour $t > -\frac{1}{2}$ et vaut $(\frac{1}{1+2t})^{n/2}$.

Le vecteur aléatoire $(\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ est gaussien en tant que transformé linéaire d'un vecteur gaussien. Or, pour tout $1 \leq k \leq n$, on a :

$$\text{Cov}(\bar{X}_n, X_k - \bar{X}_n) = \text{Cov}(\bar{X}_n, X_k) - \text{Var}\bar{X}_n = \frac{1}{n} - \frac{1}{n} = 0,$$

d'où l'indépendance entre \bar{X}_n et $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$. On en déduit immédiatement l'indépendance demandée.

Notant maintenant que $R_n + n\bar{X}_n^2 = \sum_{i=1}^n X_i^2$, nous obtenons la relation suivante entre transformées de Laplace :

$$\mathcal{L}_{R_n}(t) \times \mathcal{L}_{n\bar{X}_n^2}(t) = \mathcal{L}_{\sum_{i=1}^n X_i^2}(t), \quad t > -\frac{1}{2}$$

d'où

$$\mathcal{L}_{R_n}(t) = \frac{(\frac{1}{1+2t})^{\frac{n}{2}}}{(\frac{1}{1+2t})^{\frac{1}{2}}} = (\frac{1}{1+2t})^{\frac{n-1}{2}}$$

et donc $R_n \sim \chi^2(n-1)$. □

Remarques :

1. On définit la *loi de Student* à $k \in \mathbb{N}^*$ *degrés de liberté* et l'on note $T(k)$ la loi de la v.a.r. $\frac{X}{\sqrt{\frac{Y}{k}}}$, avec $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(k)$ indépendantes.

D'après ce qui précède, la loi de $\frac{\sqrt{n}\bar{X}_n}{S_n}$ vaut $T(n-1)$.

Si l'on remplace notre n -échantillon de loi $\mathcal{N}(0, 1)$ par un n -échantillon de loi $\mathcal{N}(m, \sigma^2)$, on constate facilement que la variable $\frac{\sqrt{n}(\bar{X}_n - m)}{S_n}$ suit encore la loi $T(n-1)$. En statistique, cela permet en particulier de construire un intervalle de confiance pour m lorsque σ est inconnue [DAC¹ cours 122] ou [LEJ 144-145].

2. Réciproquement, si un n -échantillon (X_1, \dots, X_n) est tel que sa moyenne empirique \bar{X}_n et sa variance empirique corrigée S_n^2 sont indépendantes, alors la loi commune aux X_i est une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ [COT 211-214]

4.5 Exercices corrigés.

4.5.1 Un contre-exemple.

Soit X une v.a.r. de loi $\mathcal{N}(0, 1)$ et soit $a > 0$. On pose :

$$Y^a = X\mathbf{1}_{\{|X| < a\}} - X\mathbf{1}_{\{|X| \geq a\}}.$$

1. La v.a.r. Y^a est-elle gaussienne ? Le couple (X, Y^a) est-il gaussien ?
2. Montrer qu'il existe $b > 0$ tel que $\frac{1}{\sqrt{2\pi}} \int_0^b t^2 e^{-\frac{t^2}{2}} dt = \frac{1}{4}$.
Calculer $\text{Cov}(X, Y^b)$. Les variables X et Y^b sont-elles indépendantes ?

Corrigé :

1. Pour toute fonction f borélienne bornée, on a :

$$\begin{aligned} E[f(Y^a)] &= E[f(X)\mathbf{1}_{\{|X| < a\}} + f(-X)\mathbf{1}_{\{|X| \geq a\}}] \\ &= \int_{\{|x| < a\}} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx + \int_{\{|x| \geq a\}} f(-x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \end{aligned}$$

Le changement de variable $x' = -x$ dans la seconde intégrale nous permet alors d'obtenir :

$$E[f(Y^a)] = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

ce qui prouve que $Y^a \sim \mathcal{N}(0, 1)$.

Puisque $X + Y^a = 2X\mathbf{1}_{\{|X| < a\}}$ et $X \neq 0$ p.s., on a :

$$P(X + Y^a = 0) = P(|X| \geq a) \in]0, 1[.$$

On en conclut que $X + Y^a$ ne peut être une v.a.r. gaussienne et donc que (X, Y^a) n'est pas un couple gaussien.

2. L'application définie sur $[0, +\infty[$ par $G(u) = \frac{1}{\sqrt{2\pi}} \int_0^u t^2 e^{-\frac{t^2}{2}} dt$ est continue et strictement croissante, donc bijective sur son image $[0, \frac{1}{2}[$ (en effet, la limite à l'infini vaut $\frac{1}{2}E[X^2] = \frac{1}{2}$); on prend $b = G^{-1}(\frac{1}{4})$. Notons qu'on a alors :

$$\frac{1}{\sqrt{2\pi}} \int_0^b t^2 e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_b^{+\infty} t^2 e^{-\frac{t^2}{2}} dt.$$

On calcule d'autre part :

$$\text{Cov}(X, Y^b) = E[XY^b] = E[X^2\mathbf{1}_{\{|X| < b\}} - X^2\mathbf{1}_{\{|X| \geq b\}}]$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^b t^2 e^{-\frac{t^2}{2}} dt - \frac{2}{\sqrt{2\pi}} \int_b^{+\infty} t^2 e^{-\frac{t^2}{2}} dt.$$

D'après la remarque précédente, cette covariance est donc nulle ; néanmoins X et Y^b ne sont pas indépendantes puisque $|X| = |Y^b|$ *p.s.*

4.5.2 Indépendance gaussienne : un exemple très simple.

Soit (X, Y) un vecteur gaussien de matrice de covariance :

$$K = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \rho \in [0, 1].$$

Montrer que $X + Y$ et $X - Y$ sont deux v.a. gaussiennes indépendantes.

Corrigé succinct : Le couple $(X + Y, X - Y)$ est gaussien en tant que transformé linéaire du couple gaussien (X, Y) .

Or, on a : $\text{Cov}(X + Y, X - Y) = \text{Var}X - \text{Var}Y = 0$, ce qui entraîne l'indépendance.

4.5.3 Limite en loi d'une suite gaussienne.

1. Montrer que toute limite en loi d'une suite de variables aléatoires gaussiennes est une variable aléatoire gaussienne de moyenne (resp. variance) la limite des moyennes (resp. variances).
2. En déduire que si (X_n) est un processus gaussien indexé par \mathbb{N}^* et tel que $X_n \xrightarrow{(P)} X$, alors on a : $X_n \xrightarrow{(L^2)} X$.

Corrigé :

1. Considérons une suite $(X_n)_{n \in \mathbb{N}^*}$ qui converge en loi vers une v.a.r. X et supposons que pour tout $n \in \mathbb{N}^*$, $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$.

On a alors, pour tout $t \in \mathbb{R}$:

$$\varphi_{X_n}(t) = \exp(itm_n - \frac{t^2\sigma_n^2}{2}) \longrightarrow \varphi_X(t). \quad (4.1)$$

En particulier, on a $|\varphi_X(t)| = \lim \exp(-\frac{t^2\sigma_n^2}{2})$. Or φ_X est continue et telle que $\varphi_X(0) = 1$, ce qui nous permet de choisir $t_0 \neq 0$ tel que $\varphi_X(t_0) \neq 0$; passant au logarithme dans l'égalité précédente prise en $t = t_0$, nous en déduisons que la suite (σ_n^2) est convergente. Notons σ^2 sa limite.

D'après ce que l'on vient de montrer, la suite $\exp(\frac{t^2\sigma_n^2}{2})\varphi_{X_n}(t) = \exp(itm_n)$ est convergente pour tout $t \in \mathbb{R}$; nous allons en déduire que la suite (m_n) est convergente.

Pour cela, posons $\underline{m} = \underline{\lim} m_n$ et $\overline{m} = \overline{\lim} m_n$; dans un premier temps, nous allons montrer par l'absurde que $\overline{m} < \infty$. Si ce n'était pas le cas, on pourrait extraire une sous-suite (m_{n_k}) qui diverge vers $+\infty$.

En tout point $a \in \mathbb{R}$ tel que $P(X = a) = 0$, on aurait alors :

$$P(X_{n_k} \leq a) \longrightarrow P(X \leq a).$$

Le membre de gauche étant égal à $P(m_{n_k} + \sigma_{n_k}Y \leq a)$, avec $Y \sim \mathcal{N}(0, 1)$, on en déduit que $P(X \leq a) = 0$.

Comme a peut être choisi arbitrairement grand, cela contredit le fait que la fonction de répartition de X a pour limite 1 en $+\infty$.

On a donc $\overline{m} < \infty$ et, par un raisonnement symétrique, $\underline{m} > -\infty$.

La convergence de la suite $(\exp(itm_n))$ implique alors l'égalité : $\exp(it\underline{m}) = \exp(it\overline{m})$.

On en déduit que, pour tout $t \in \mathbb{R}$,

$$t(\overline{m} - \underline{m}) \equiv 0 \pmod{2\pi},$$

ce qui n'est bien sûr possible que si $\overline{m} = \underline{m}$.

Ainsi, la suite (m_n) est convergente et nous noterons sa limite m .

La convergence (4.1) nous donne alors immédiatement, pour tout $t \in \mathbb{R}$, $\varphi_X(t) = \exp(itm - \frac{t^2\sigma^2}{2})$, ce qui prouve que $X \sim \mathcal{N}(m, \sigma^2)$.

2. On a les implications :

$$X_n \xrightarrow[n \rightarrow +\infty]{(P)} X \implies X_n - X \xrightarrow[n \rightarrow +\infty]{(P)} 0 \implies X_n - X \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0.$$

Pour $n \in \mathbb{N}^*$ fixé, la v.a.r. $X_n - X$ est limite en probabilité (et donc en loi) de la suite $(X_n - X_k)_{k \in \mathbb{N}^*}$; or cette suite est gaussienne puisque (X_n) est un processus gaussien indexé par \mathbb{N} . D'après la question précédente, on en déduit que la v.a.r. $X_n - X$ est gaussienne.

Comme $n \in \mathbb{N}^*$ était arbitraire, on peut de nouveau utiliser la question précédente pour dire que $(X_n - X)_{n \in \mathbb{N}^*}$ est une suite de gaussiennes tendant vers 0, donc que :

$$E[X_n - X] \longrightarrow 0 \text{ et } \text{Var}(X_n - X) \longrightarrow 0.$$

Il suffit alors décrire l'égalité :

$$E[(X_n - X)^2] = (E[X_n - X])^2 + \text{Var}(X_n - X)$$

pour en déduire que $X_n \xrightarrow{(L^2)} X$.

Chapitre 5

Notions de convergence d'une suite de variables aléatoires réelles

5.1 Diverses notions de convergence

Définitions, tableau des implications, contre-exemples.

5.2 Lois des grands nombres

5.2.1 Loi faible

L'inégalité de Bienaymé-Tchebychev permet de prouver la loi faible dans L^2 , ainsi que le théorème de Weierstrass par les polynômes de Bernstein (cf. section 1.1.2 page 7).

5.2.2 Loi forte

Elle se démontre élémentairement dans le cas L^4 : c'est le théorème de Cantelli [BAR 133].

Quelques applications de la loi forte :

- (\bar{X}_n) , resp. (\hat{S}_n^2) sont des estimateurs fortement convergents de la moyenne, resp. de la variance [RS 8-9]
- Maths pures : Presque tous les réels de $[0,1]$ sont normaux, i.e. tels que leur développement décimal fait apparaître tous les entiers de 0 à 9 avec la même fréquence asymptotique $1/10$ [REV 129] ou [QUEFFÉLEC-ZUILY 550-551]

- Processus de renouvellement [DAC² cours 180] : lorsque $t \rightarrow \infty$,

$$\frac{N_t}{t} \xrightarrow{\text{p.s.}} \frac{1}{m}$$

5.3 Théorème-limite central et raffinements

5.3.1 En dimension 1

- On peut lever l'hypothèse d'équidistribution si l'on considère une suite indépendante et bornée dans $L^{2+\epsilon}$. En supposant les variables centrées (ce qui ne fait pas perdre de généralité), le TLC s'écrit alors [JP 189] :

$$\frac{S_n}{\sigma(S_n)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Autrement dit, dès que les fluctuations d'un phénomène aléatoire résultent de la superposition de nombreuses micro-fluctuations indépendantes, on peut modéliser (au moins approximativement) ce phénomène par une variable aléatoire réelle gaussienne

- La condition (plus technique) de *Lindeberg* :

$$\forall \epsilon > 0 \quad \frac{1}{\sigma^2(S_n)} \sum_1^n \int_{\{|X_k| > \epsilon \sigma(S_n)\}} X_k^2 dP \xrightarrow[n \rightarrow +\infty]{} 0$$

permet aussi de lever l'hypothèse d'équidistribution, les variables étant simplement supposées centrées, indépendantes et dans L^2 . [REV 163] ou [FOA 244-245]

- Le *théorème de Berry-Esseen* précise à quelle vitesse a lieu la convergence énoncée par le TLC [JP 191] :

Soit (X_n) une suite de v.a.r. i.i.d. admettant un moment d'ordre 3. On suppose ces variables centrées, de variance $E[X_1^2] = \sigma^2 > 0$. Notons $\rho = E[|X_1|^3]$, F_n f.r. de $\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}$, et F f.r. de la loi $\mathcal{N}(0, 1)$. Alors,

$$\forall n \in \mathbb{N}^* \quad \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{C\rho}{\sigma^3\sqrt{n}} \quad ,$$

où la constante C a pour valeur approximative 0,7975.

5.3.2 TLC multidimensionnel

énoncé dans [JP 190]. Il se déduit facilement du TLC unidimensionnel en utilisant les f.c. et le théorème de convergence de Lévy.

Le *théorème de la limite locale* : Soit (X_n) une suite de vecteurs aléatoires indépendants et de même loi admettant un moment d'ordre 2 et une densité, ainsi qu'une fonction caractéristique φ_{X_1} intégrable. Alors les vecteurs aléatoires

$$\frac{S_n - E[S_n]}{\sqrt{n}}$$

admettent des densités qui convergent uniformément vers la densité de la loi $\mathcal{N}(0, K_{X_1})$. [REV 165]

5.4 Applications du TLC

- Approximation gaussienne d'une loi binomiale : voir section 2.3.1 page 14 et section 4.3 page 24. à propos de l'approximation poissonnienne, noter que $\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda)$ dès que $\lambda \geq 20$. [LEJ 85]
- Le TLC pour une suite i.i.d. de variables exponentielles est équivalent à la formule de Stirling $n! \sim \sqrt{2\pi n} n^n e^{-n}$ [FOA 243].
- la formule de Bernstein [FOA 242] :

$$e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \xrightarrow{n \rightarrow +\infty} \frac{1}{2}$$

et son application à la première «réapparition» d'une boule lors d'un tirage dans une urne contenant n boules numérotées [FOA 273-274]

- Une application du TLC multidimensionnel en statistique est la construction du test du χ^2 : voir chapitre 12.
- Fonction de répartition empirique, théorème de Glivenko-Cantelli et test de Kolmogorov-Smirnov [OUV² 116-120]
- Processus de renouvellement : lorsque $t \rightarrow \infty$, on a

$$\sqrt{t} \left(\frac{N_t}{t} - \frac{1}{m} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2}{m^3} \right)$$

[DAC¹ exo 4.4.11]

- Construction d'un intervalle de confiance (voir section 2.3.2 page 15 et section suivante) : [RS 25-26] ou [LEJ 144-145].
- Méthodes de Monte-Carlo [REV 128 et 179] [RS 105-107] : voir dernière section de ce chapitre.

5.5 Intervalles de confiance

5.5.1 Quelques notions de statistique

Pour montrer la différence entre l'objet du calcul des probabilités et celui de la statistique, prenons l'exemple d'un jeu de pile ou face dans lequel on jette la pièce (pas nécessairement équilibrée) un "grand nombre" de fois. Nous introduisons donc le *modèle canonique* :

$$\Omega = \{0, 1\}^N \quad ; \quad \mathcal{A} = \mathcal{P}(\Omega) \quad ; \quad P = B(p)^{\otimes N}, \quad p \in]0, 1[$$

Le probabiliste connaît p et cherche à en déduire des propriétés sur le résultat de l'expérience. Par exemple, si nous appelons X_i la i^e projection canonique (définie par : $\forall \omega = (\omega_1, \dots, \omega_N) \in \Omega, \quad X_i(\omega) = \omega_i$), la loi forte des grands nombres nous donne :

$$P(d\omega)\text{-p.s.}, \quad \frac{X_1(\omega) + \dots + X_N(\omega)}{N} \approx p$$

Notons que ceci est une propriété du résultat ω de l'expérience puisqu'elle s'écrit encore :

$$\frac{\omega_1 + \dots + \omega_N}{N} \approx p$$

La théorie des probabilités nous permet de dire que si l'expérience nous donne un résultat ω , il a "de fortes chances" de vérifier cette propriété.

Pour le probabiliste, la probabilité P est donc connue et il s'agit d'en déduire des propriétés du résultat ω de l'expérience. Le statisticien fait la démarche en sens inverse : il connaît le résultat ω de l'expérience et il en déduit des propriétés de la probabilité P qui est inconnue.

Ainsi, dans l'exemple précédent, supposons que le paramètre p qui intervient dans la définition de la probabilité P est inconnu mais que nous disposons du résultat de N jets de pièce successifs, c'est-à-dire que nous ayons observé un ω pour l'expérience considérée. Se basant sur la loi forte des grands nombres, le statisticien pourra proposer l'estimation suivante du paramètre p :

$$\hat{p}(\omega) = \frac{X_1(\omega) + \dots + X_N(\omega)}{N}$$

De façon plus générale, le statisticien part d'une famille de probabilités $(P_\theta)_{\theta \in \Theta}$ parmi laquelle il cherche à déterminer celle qui modélisera le mieux le phénomène observé. En *statistique paramétrique*, on suppose que l'ensemble Θ est inclus dans \mathbb{R}^d pour un certain $d \in \mathbb{N}^*$. Au vu du résultat ω de l'expérience, le statisticien propose une valeur $\hat{\theta}(\omega)$ pour estimer le paramètre θ

(ou encore pour estimer une fonction $f(\theta)$ de ce paramètre). Dans l'exemple précédent, nous avons $\Theta =]0, 1[$ et $P_\theta = B(\theta)^{\otimes N}$.

Nous demanderons à cette fonction $\hat{\theta} : \Omega \rightarrow \Theta$ d'être mesurable et nous l'appellerons un *estimateur* de θ . Dans notre exemple, nous avons donc $\hat{\theta} = \bar{X}_n$.

Notons que faire de la statistique nous demande de travailler avec toute une famille de probabilités à la fois et que nous serons donc amenés à écrire des expressions telle que :

$$\forall \theta \in \Theta \quad P_\theta(|\hat{\theta} - \theta| \geq \epsilon) < \alpha$$

Par exemple, l'estimateur $\hat{\theta}$ est dit *sans biais* s'il vérifie $E_\theta[\hat{\theta}] = \theta$ pour tout $\theta \in \Theta$, où E_θ désigne l'espérance calculée sous la probabilité P_θ .

5.5.2 L'estimateur moyenne empirique

Dans ce paragraphe, nous allons nous intéresser plus particulièrement à l'estimateur \bar{X}_N , appelé *moyenne empirique*. Nous allons constater en effet qu'il possède des propriétés intéressantes lorsque nous souhaitons estimer la moyenne d'une loi.

Supposons donc maintenant que $P_\theta = \mu_\theta^{\otimes N}$, avec μ_θ loi sur \mathbb{R} admettant une espérance $m(\theta)$. Nous souhaitons estimer la valeur de $m(\theta)$.

Dans l'exemple précédent, nous avons $\mu_\theta = B(\theta)$ donc $m(\theta) = \theta$.

Citons un autre exemple : $\Omega = \mathbb{R}^N$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, $P_\theta = \text{Exp}(\theta)^{\otimes N}$, $\Theta = \mathbb{R}_+^*$; dans ce cas, $m(\theta) = 1/\theta$.

Une propriété intéressante de l'estimateur \bar{X}_N de $m(\theta)$ est qu'il est *sans biais*, c'est-à-dire qu'il donne la "bonne" valeur en moyenne en ce sens que :

$$\forall \theta \in \Theta \quad E_\theta[\bar{X}_N] = \frac{1}{N} \sum_{i=1}^N E_\theta[X_i] = m(\theta)$$

Supposons maintenant que nous pouvons répéter indéfiniment la même procédure dans des conditions identiques (par exemple jouer à pile ou face indéfiniment). Nous prenons alors plutôt pour *modèle statistique* :

$$\Omega = \mathcal{E}^{\mathbb{N}} \quad , \quad \mathcal{A} = \sigma(X_n, n \in \mathbb{N}) \quad , \quad P_\theta = \mu_\theta^{\otimes \mathbb{N}}$$

où $\mathcal{E} = \{0, 1\}$ ou \mathbb{R} ou $\mathbb{R}^d \dots$, X_n est la n^e injection canonique définie par $X_n(\omega) = \omega_n$ et la probabilité $\mu_\theta^{\otimes \mathbb{N}}$ est l'unique probabilité sous laquelle la suite $(X_n)_{n \in \mathbb{N}}$ est indépendante identiquement distribuée de loi commune μ_θ (nous admettons l'existence et l'unicité de cette probabilité, qui résultent du

théorème des classes monotones ainsi que d'un théorème d'extension énoncé par Kolmogorov).

Dans un tel modèle statistique, nous disposons d'une suite d'estimateurs de $m(\theta)$: la suite des moyennes empiriques $(\bar{X}_n)_{n \in \mathbb{N}^*}$.

Une propriété intéressante de cette suite d'estimateurs est qu'elle est *consistante* (ou *convergente*, ce qui est synonyme) c'est-à-dire qu'elle donne la "bonne" valeur asymptotiquement en ce sens que, d'après la loi forte des grands nombres :

$$\forall \theta \in \Theta \quad P_\theta(d\omega)\text{-p.s.}, \quad \bar{X}_n(\omega) \longrightarrow m(\theta)$$

C'est à ce stade qu'intervient le théorème-limite central car dans la pratique, il est totalement insuffisant d'avoir une suite d'estimateur convergente si nous n'avons aucune idée de la vitesse à laquelle elle converge. C'est toute la question du degré de précision de notre estimateur qui est en jeu et le théorème-limite central va nous permettre d'y apporter une réponse grâce à la notion d'*intervalle de confiance*.

5.5.3 Intervalles de confiance

Afin de pouvoir appliquer le théorème-limite central, nous supposons désormais que pour tout $\theta \in \Theta$, la loi μ_θ admet un moment d'ordre 2 et nous notons alors $\sigma^2(\theta)$ sa variance. Nous avons alors :

$$\text{Sous } P_\theta, \quad \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

La loi $\mathcal{N}(0, 1)$ n'ayant pas de masse ponctuelle, nous en déduisons que pour tous réels $a \leq b$, nous avons la convergence suivante lorsque n tend vers l'infini :

$$P_\theta \left[a \leq \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \leq b \right] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Il est facile de vérifier que pour tout $0 < \alpha \leq 1$, il existe un unique $\phi_\alpha \in \mathbb{R}_+$ tel que :

$$\frac{2}{\sqrt{2\pi}} \int_{\phi_\alpha}^{+\infty} e^{-\frac{x^2}{2}} dx = \alpha$$

Nous en déduisons que, pour tout $\theta \in \Theta$, la convergence suivante a lieu lorsque n tend vers l'infini :

$$P_\theta \left[-\phi_\alpha \leq \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \leq \phi_\alpha \right] \longrightarrow 1 - \alpha$$

Cette convergence s'écrit encore :

$$\forall \theta \in \Theta, \quad P_\theta \left[m(\theta) \in \left[\bar{X}_n - \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}} \right] \right] \rightarrow 1 - \alpha$$

L'intervalle (aléatoire) $\left[\bar{X}_n - \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}} \right]$ est appelé intervalle de confiance pour $m(\theta)$ de niveau de confiance asymptotique $1 - \alpha$ (ou encore de niveau d'erreur asymptotique α). Cela signifie que, si nous affirmons que $m(\theta)$ est dans cet intervalle, la probabilité que nous fassions erreur est proche de α lorsque n est "grand".

Par exemple, si $\alpha = 0,05$, les tables de la loi normale (classiques en statistique, voir par exemple les dernières pages de [DAC¹ exo]) nous fournissent la valeur $\phi_\alpha \sim 1,96$; pour $\alpha = 0,002$, nous obtenons $\phi_\alpha \sim 3,09$.

Il est intéressant de faire la comparaison entre l'information qui nous est donnée par le théorème-limite central et celle que nous fournit l'inégalité de Bienaymé-Tchebichev, à savoir :

$$P_\theta \left[-\phi_\alpha \leq \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \leq \phi_\alpha \right] \geq 1 - \frac{1}{\phi_\alpha^2}$$

Nous constatons que le théorème-limite central nous donne un renseignement beaucoup plus précis. Notons cependant que l'inégalité de Bienaymé-Tchebichev est vraie même si n est petit.

Un problème subsiste concernant $\left[\bar{X}_n - \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}} \right]$, intervalle de confiance que nous venons de déterminer : le paramètre θ y intervient au travers du facteur $\sigma(\theta)$, or précisément θ est un paramètre inconnu que nous souhaitons évaluer (en estimant $m(\theta)$) donc nous risquons de tourner en rond !

Dans certaines situations, il est possible de majorer $\sigma(\theta)$ indépendamment de θ , ce qui nous permet de construire un intervalle de confiance dans lequel le paramètre θ n'intervient plus.

Par exemple, si $\mu_\theta = B(\theta)$, nous avons $\sigma^2(\theta) = \theta(1 - \theta) \leq 1/4$ pour tout $\theta \in \Theta = [0, 1]$. Nous en déduisons que $\left[\bar{X}_n - \frac{1}{\sqrt{n}}, \bar{X}_n + \frac{1}{\sqrt{n}} \right]$ est un intervalle de confiance pour $m(\theta) = \theta$ de niveau de confiance asymptotique 95%.

Dans le cas général, nous introduisons la *variance empirique* définie par :

$$\Sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Une certaine propriété de stabilité de la convergence en loi, dite lemme de Slutsky, permet de démontrer que l'intervalle aléatoire suivant (dans lequel nous avons remplacé l'écart-type théorique $\sigma(\theta)$ par l'écart-type empirique

Σ_n , ce qui supprime toute dépendance en le paramètre θ) :

$$\left[\bar{X}_n - \frac{\Sigma_n \phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\Sigma_n \phi_\alpha}{\sqrt{n}} \right]$$

est encore un intervalle de confiance pour $m(\theta)$ de niveau de confiance asymptotique $1 - \alpha$.

5.6 Méthodes de Monte-Carlo

On appelle méthode de Monte-Carlo¹ toute méthode visant à calculer une valeur numérique en utilisant des techniques probabilistes.

Le problème de départ peut être *a priori* de nature complètement déterministe, comme par exemple le calcul approché de l'intégrale :

$$I = \int_{[0,1]^d} f(x) dx, \text{ où } f \text{ est une application borélienne de } [0, 1]^d \text{ dans } \mathbb{R},$$

mais nous allons y introduire artificiellement un aspect aléatoire en remarquant que cette intégrale s'écrit encore, d'après le théorème de transfert,

$$I = E[f(X)], \text{ où } X \text{ est un vecteur aléatoire uniforme sur } [0, 1]^d.$$

La loi forte des grands nombres implique alors que, si nous disposons d'une suite $(X_n)_{n \in \mathbb{N}^*}$ indépendante et de même loi uniforme sur $[0, 1]^d$:

$$\frac{f(X_1) + \dots + f(X_n)}{n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} I$$

Notre méthode de Monte-Carlo consistera donc à simuler un nombre "suffisamment grand" N de vecteurs aléatoires uniformes sur $[0, 1]^d$ indépendants (ce qui revient à simuler Nd variables aléatoires uniformes sur $[0, 1]$ indépendantes) et à prendre pour approximation :

$$I \approx \frac{f(X_1) + \dots + f(X_N)}{N}$$

Dans la pratique, toute la question est de savoir ce que l'on entend par N "suffisamment grand" !

Si les variables aléatoires $f(X_i)$ sont dans $L^2(\Omega, \mathcal{A}, P)$, ou de façon équivalente si $f \in L^2([0, 1]^d, \mathcal{B}([0, 1]^d), \lambda_d)$, le théorème-limite central nous donne la convergence suivante lorsque N tend vers l'infini :

$$\frac{\sqrt{N}}{\sigma} \left(\frac{f(X_1) + \dots + f(X_N)}{N} - I \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad ,$$

1. quartier de la principauté de Monaco connu pour ses jeux de hasard

où nous avons noté :

$$\sigma^2 = \text{Var}(f(X_i)) = \int_{[0,1]^d} f^2(x) dx - \left(\int_{[0,1]^d} f(x) dx \right)^2 = \int_{[0,1]^d} f^2 d\lambda_d - I^2$$

Nous constatons donc que l'erreur d'approximation commise est d'ordre de grandeur σ/\sqrt{N} .

En dimension $d = 1$, faisons la comparaison avec une méthode numérique classique d'approximation des intégrales, par exemple la *méthode des trapèzes*, qui nous donne l'approximation :

$$\int_{[0,1]} f(x) dx \approx T_N(f) := \frac{f(0) + f(1)}{N} + \frac{\sum_{i=1}^{N-1} f\left(\frac{i}{N}\right)}{N}$$

Sous l'hypothèse $f \in C^2([0,1], \mathbb{R})$, nous avons la majoration suivante de l'erreur d'approximation :

$$\left| \int_{[0,1]} f(x) dx - T_N(f) \right| \leq \frac{1}{12N^2} \max_{[0,1]} |f''|.$$

Nous constatons donc que cette méthode numérique simple est beaucoup plus efficace que celle de Monte-Carlo lorsque f est régulière, plus précisément de classe C^2 .

En revanche, si f est peu régulière, la méthode des trapèzes risque de fournir une approximation de mauvaise qualité (faire un petit dessin pour voir sur quelle idée intuitive repose cette méthode). Il en va différemment pour la méthode de Monte-Carlo puisque le théorème-limite central n'impose qu'une hypothèse de régularité très faible sur f , qui est supposée borélienne.

Ainsi, dans le cas d'une fonction f très irrégulière, la méthode de Monte-Carlo peut donner une meilleure approximation de l'intégrale qu'une méthode numérique classique.

Plaçons-nous maintenant en dimension $d \geq 2$. Quelle est l'influence de la dimension sur l'efficacité des méthodes numériques et de la méthode de Monte-Carlo ?

On peut dire *grosso modo* qu'une méthode numérique classique d'ordre N calculera les valeurs de la fonction f en N^d points "bien choisis" dans $[0,1]^d$ puis en déduira une approximation de l'intégrale. Une méthode de Monte-Carlo d'ordre N reposera sur la simulation de Nd variables aléatoires uniformes sur $[0,1]$ indépendantes, pour obtenir N vecteurs aléatoires uniformes sur $[0,1]^d$ indépendants. En les valeurs prises par ces vecteurs aléatoires, c'est-à-dire en N points de $[0,1]^d$, l'algorithme calculera alors les valeurs prises par f puis en fera la moyenne arithmétique.

Nous constatons donc la sensibilité d'une méthode de Monte-Carlo à la dimension d est beaucoup moindre que celle d'une méthode numérique classique. Dans la pratique, dès la dimension 3 ou 4, une méthode de Monte-Carlo peut s'avérer plus efficace qu'une méthode numérique classique.

Terminons cette introduction aux méthodes de Monte-Carlo en notant que l'erreur d'approximation, d'ordre de grandeur σ/\sqrt{N} , sera d'autant plus satisfaisante que la variance σ^2 sera petite.

Il existe donc des techniques, dites de *réduction de variance*, qui consistent à modifier un peu la méthode de Monte-Carlo présentée dans ce paragraphe pour faire chuter la variance des variables aléatoires intervenant dans l'approximation de l'intégrale et donc accélérer la convergence.

Chapitre 6

Utilisation des transformées de Laplace et de Fourier en calcul des probabilités

Le produit de convolution apparaît naturellement en probabilités lorsqu'on additionne des variables aléatoires indépendantes. Il va intervenir dans chacune des sections suivantes.

6.1 Cas discret : fonctions génératrices

La fonction génératrice d'une variable aléatoire X à valeurs dans \mathbb{N} est un cas simple de transformation de Laplace, à un changement de variable près, puisque :

$$\forall u \in]0, 1[\quad g_X(u) = E[u^X] = E[e^{(\log u)X}]$$

Voici cinq applications des fonctions génératrices (de la plus élémentaire à la plus élaborée) :

- Si $X \sim \mathcal{P}(\lambda)$, sa f.g. vaut $E[u^X] = \exp(\lambda(u - 1))$. Ceci permet de retrouver $\mathcal{P}(\lambda) \star \mathcal{P}(\mu) = \mathcal{P}(\lambda + \mu)$
- Le théorème des événements rares (cf. théorème 1.3.2 page 10).
- Les lois infiniment divisibles sur \mathbb{N} sont exactement les lois de Poisson composées [*the compound Poisson distribution dans* FELLER¹ 288-290]
- Théorème de Raïkov : voir section 1.2.2 page 8.
- Processus de Galton-Watson : étude de la probabilité d'extinction en fonction du nombre moyen m d'enfants d'un individu donné. Voir chapitre 16.

6.2 Transformation de Laplace

6.2.1 Injectivité, lien avec la convolution

En utilisant l'injectivité de la transformation de Laplace sur les lois de probabilités portées par \mathbb{R}_+ , on retrouve facilement le résultat suivant :

$$\gamma(\lambda, a) \star \gamma(\lambda, b) = \gamma(\lambda, a + b)$$

6.2.2 Inégalité de Hoeffding

Cette inégalité [OUV² 132-135] n'est pas asymptotique mais vraie pour toute valeur de n .

Proposition 6.2.1 *Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite i.i.d. telle qu'il existe une constante $K > 0$ avec $P(|X_1| \leq K) = 1$. On note $m = E[X_1]$ et $\bar{X}_n = (X_1 + \dots + X_n)/n$ pour tout $n \in \mathbb{N}^*$. Alors, pour tout $\epsilon > 0$ et tout $n \in \mathbb{N}^*$, on a l'inégalité :*

$$P(|\bar{X}_n - m| \geq \epsilon) \leq 2 \exp\left(-\frac{n\epsilon^2}{2K^2}\right)$$

Démonstration: On peut supposer les variables X_i centrées sans perte de généralité. Pour tout $\lambda \geq 0$, on définit la «fonction des cumulants» de X_i par $\Lambda_i(\lambda) = \log E[\exp(\lambda X_i)]$; notons qu'elle ne dépend pas de l'indice i puisque les variables sont identiquement distribuées et donc nous la noterons simplement $\Lambda(\lambda)$. L'inégalité de Jensen nous donne immédiatement que Λ est à valeurs dans \mathbb{R}_+ .

En utilisant $|X_i| \leq K$ p.s., on justifie facilement les dérivations sous le signe espérance qui nous donnent, pour tout $\lambda \geq 0$:

$$\Lambda'(\lambda) = \frac{E[X_i \exp(\lambda X_i)]}{E[\exp(\lambda X_i)]},$$

$$\Lambda''(\lambda) = \frac{E[X_i^2 \exp(\lambda X_i)] E[\exp(\lambda X_i)] - \{E[X_i \exp(\lambda X_i)]\}^2}{\{E[\exp(\lambda X_i)]\}^2}.$$

En particulier, on a $\Lambda(0) = 0$, $\Lambda'(0) = 0$ (car les variables sont centrées) et pour tout $\lambda \geq 0$,

$$\Lambda''(\lambda) \leq \frac{E[X_i^2 \exp(\lambda X_i)]}{E[\exp(\lambda X_i)]} \leq K^2.$$

La formule de Taylor nous permet d'en déduire que, pour tout $\lambda \geq 0$,

$$0 \leq \Lambda(\lambda) \leq K^2 \lambda^2 / 2.$$

Or l'inégalité de Markov nous donne :

$$P\left(\frac{S_n}{n} \geq \epsilon\right) = P(e^{\lambda S_n} \geq e^{\epsilon \lambda n}) \leq e^{-\epsilon \lambda n} (E[e^{\lambda X_1}])^n.$$

On a donc : $P(\bar{X}_n \geq \epsilon) \leq \exp(-\epsilon \lambda n + nK^2 \lambda^2 / 2)$.

On optimise cette inégalité en choisissant $\lambda = \frac{\epsilon}{K^2}$ d'où le majorant $\exp(-\frac{\epsilon^2 n}{2K^2})$.

En écrivant la même inégalité pour les variables $(-X_i)$, on aboutit à la conclusion.

□

Ce résultat permet de construire des intervalles de confiance en statistique. [RS 23-24]

6.2.3 Grandes déviations

Une application nettement plus élaborée de la transformation de Laplace se rencontre dans les théorèmes de *grandes déviations*, qui sont au programme de l'option.

[REV 172-178]

[TOU chap 3]

[BC 88-91] compare TLC et principe des grandes déviations dans le cas d'une suite de Bernoulli de paramètre p .

6.3 Transformation de Fourier

L'injectivité de la transformation de Fourier sur les lois permet de démontrer facilement :

$$\mathcal{N}(m_1, \sigma_1^2) \star \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$$

Il faut bien sûr parler du *théorème-limite central* qui est une application essentielle de la caractérisation de la convergence en loi par la convergence simple des fonctions caractéristiques.

Le même genre de raisonnement nous permet de caractériser la loi gaussienne centrée comme suit : Si X et Y sont indépendantes de même loi μ centrée, admettant un moment d'ordre deux σ^2 et telle que $(X + Y)/\sqrt{2}$ a pour loi μ , alors $\mu = \mathcal{N}(0, \sigma^2)$ [OUVRARD tome 2, p.278-280].

Une autre caractérisation de la loi gaussienne, qui se démontre à l'aide des fonctions caractéristiques, est le *théorème de Bernstein* [OUVRARD tome 2

p.280-283] : Si X et Y sont des v.a.r. indépendantes telles que les v.a.r. $X + Y$ et $X - Y$ soient indépendantes, alors X et Y sont des variables gaussiennes.

Citons encore le *théorème de Cramer-Lévy* (à comparer au théorème de Raïkov) : Si X et Y sont des v.a.r. indépendantes telles que $X + Y$ est gaussienne, alors X et Y sont des variables gaussiennes. [BJ 200]

C'est en calculant la fonction caractéristique d'un vecteur gaussien quelconque que l'on démontre l'importante propriété suivante :

Soient X et Y des vecteurs aléatoires de dimension respectives d et d' tels que (X, Y) est un vecteur gaussien. Alors X et Y sont indépendants si et seulement si toutes les covariances croisées sont nulles, i.e.

$$\forall i = 1, \dots, d \quad \forall j = 1, \dots, d' \quad \text{Cov}(X_i, Y_j) = 0$$

Une application de la formule d'inversion de Fourier est le calcul de la fonction caractéristique d'une variable de Cauchy à partir de la transformée de Fourier de la densité de Laplace de paramètre $a > 0$: $\frac{a}{2} \exp(-a|x|)$. On en déduit alors facilement : $C(a) \star C(b) = C(a + b)$.

6.4 Exercices sur les transformées de Laplace et Cramer

6.4.1 Domaine de définition de \mathcal{L}_X

- Vérifier la valeur du domaine de définition I_X de la transformée de Laplace \mathcal{L}_X dans les cas suivants :
 - Si $X \sim \mathcal{N}(m, \sigma^2)$, $I_X = \mathbb{R}$.
 - Si $X \sim \text{Exp}(\lambda)$, $I_X =] - \infty, \lambda[$.
 - Si $X \sim \text{Laplace}(\lambda)$ (densité $\frac{\lambda}{2} \exp(-\lambda|x|)$, $I_X =] - \lambda, \lambda[$.
 - Si $X \sim \text{Pareto}(a)$ (densité $\frac{a}{x^{a+1}} \mathbf{1}_{x \geq 1}$), $I_X =] - \infty, 0]$.
 - Si $X \sim C(a)$, $I_X = \{0\}$.
- Trouver un exemple de densité pour la variable X telle que I_X soit un intervalle semi-ouvert.

6.4.2 Calculs explicites de transformées de Laplace et Cramer

Vérifier les affirmations suivantes :

6.4. EXERCICES SUR LES TRANSFORMÉES DE LAPLACE ET CRAMER 47

1. Si $X \sim B(p)$ avec $p \in]0, 1[$, alors $I_X = \mathbb{R}$, $L_X(u) = pe^u + 1 - p$,
 $\Lambda_X^*(u) = u \log(\frac{u}{p}) + (1 - u) \log(\frac{1-u}{1-p})$ si $u \in [0, 1]$, $\Lambda_X^*(u) = +\infty$ sinon.
 Enfin, on a $] \alpha_X, \beta_X [=]0, 1[$.
2. Si $X \sim P(\lambda)$, alors $I_X = \mathbb{R}$, $L_X(u) = \exp(-\lambda(1 - e^u))$,
 $\Lambda_X^*(u) = u \log(\frac{u}{\lambda}) + \lambda(1 - \frac{u}{\lambda})$ si $u \geq 0$, $\Lambda_X^*(u) = +\infty$ si $u < 0$
 et $] \alpha_X, \beta_X [=]0, +\infty[$.
3. Si $X \sim \mathcal{N}(m, \sigma^2)$, alors $L_X(u) = \exp(mu + \frac{\sigma^2 u^2}{2})$,
 $\Lambda_X^*(u) = \frac{(u-m)^2}{2\sigma^2}$ et $] \alpha_X, \beta_X [=] - \infty, +\infty[$.
4. Si $X \sim Exp(\lambda)$, alors $L_X(u) = \frac{\lambda}{\lambda - u}$,
 $\Lambda_X^*(u) = \lambda u - 1 - \log(\lambda u)$ si $u > 0$, $\Lambda_X^*(u) = +\infty$ si $u \leq 0$
 et $] \alpha_X, \beta_X [=]0, +\infty[$.
5. Si $X \sim C(a)$, alors $\Lambda_X^*(u) = +\infty$ si $u \neq 0$ et $\Lambda_X^*(0) = 0$.
 Noter que dans ce cas, la notation $] \alpha_X, \beta_X [$ n'a pas de sens.

6.4.3 Comportement asymptotique de la transformée de Cramer

1. On suppose que I_X est un voisinage de 0. Montrer que dans ce cas,

$$\lim_{|a| \rightarrow \infty} \Lambda_X^*(a) = +\infty$$

Indication : Choisir $u > 0$ tel que $\Lambda_X(u) < \infty$ et $\Lambda_X(-u) < \infty$ puis utiliser la minoration :

$$\forall a \in \mathbb{R} \quad \Lambda_X^*(a) \geq (au - \Lambda_X(u)) \vee (-au - \Lambda_X(-u))$$

2. Montrer que si $I_X = \mathbb{R}$, on a même :

$$\lim_{|a| \rightarrow \infty} \frac{\Lambda_X^*(a)}{|a|} = +\infty$$

6.4.4 Transformée de Cramer identiquement nulle

Montrer qu'on a l'équivalence suivante :

$$\Lambda_X^* \equiv 0 \Leftrightarrow I_X = \{0\}$$

Indication : Dans le sens direct, montrer que l'hypothèse implique :
 $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}, \Lambda_X(u) \geq au$ et conclure.

Chapitre 7

Exemples de lois et de leur utilisation en probabilités

Chapitre 8

Les lois usuelles, leurs applications pratiques et les techniques de simulation de variables aléatoires

Notons d'emblée que dans toutes les leçons, nous faisons comme si l'instruction `rand` de Matlab nous fournissait un générateur aléatoire parfait de variables uniformes sur $[0, 1]$; faire preuve d'esprit critique à ce sujet nous mènerait en effet dans un thème très vaste...

Nous souhaitons ici simuler des variables gaussiennes centrées réduites à partir de ces variables uniformes sur $[0, 1]$. Une première méthode qui découle immédiatement du TLC consiste à poser :

$$X = \sqrt{\frac{12}{p}} \left(\sum_{i=1}^p U_i - \frac{p}{2} \right),$$

qui nous donne approximativement une gaussienne pour p «suffisamment grand».

Une autre méthode, dite de Box-Muller, nous dit que, par changement de variables en coordonnées polaires, les v.a.r.

$$Y_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2) \quad , \quad Y_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

sont exactement des variables gaussiennes centrées réduites indépendantes.

Pour valider nos simulations, nous allons faire appel au test de Kolmogorov-Smirnov dont nous allons présenter maintenant le principe.

8.1 Introduction au test de Kolmogorov-Smirnov.

Si les $(X_i)_{1 \leq i \leq n}$ sont des v.a.r. i.i.d. de loi μ , on définit leur fonction de répartition empirique par :

$$\forall t \in \mathbb{R}, F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}}$$

Si F est la fonction de répartition de la loi μ et si l'on dispose maintenant d'une suite $(X_i)_{i \in \mathbb{N}^*}$ i.i.d. de loi μ , la LGN nous donne immédiatement :

$$\forall t \in \mathbb{R}, p.s., F_n(t) \rightarrow F(t)$$

(on peut en fait inverser le $\forall t$ et le $p.s.$).

Le théorème de Kolmogorov-Smirnov nous permet de contrôler la vitesse de convergence d'une façon précise sous la seule hypothèse que μ soit diffuse (i.e. F continue). On a alors :

$$\forall u \in \mathbb{R}, \lim_{n \rightarrow \infty} P\left(\sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - F(t)| \leq u\right) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 u^2}.$$

La somme $S(u)$ de la série de droite est tabulée : on a par exemple $S(1,22) \sim 90\%$,

$S(1,358) \sim 95\%$, $S(1,628) \sim 99\%$. Nous admettrons qu'il est raisonnable d'estimer la limite du membre de gauche par la valeur obtenue pour $n = 100$.

Si l'on dispose d'un échantillon de taille 100 d'une loi ν inconnue et que l'on souhaite tester l'hypothèse $H_0 : \nu = \mu$ avec un niveau d'erreur asymptotique 10% par exemple, le théorème précédent nous permet de procéder comme suit :

On calcule la f.r. empirique G_{100} associée à notre échantillon et on la compare à la f.r. F de notre loi théorique μ en estimant : $\sup_{t \in \mathbb{R}} \sqrt{100} |G_{100}(t) - F(t)|$. Si la quantité obtenue est supérieure à 1,22, on rejette l'hypothèse ; sinon, on l'accepte.

8.2 Application sur machine.

1. Simuler 100 variables i.i.d. $\mathcal{N}(0,1)$ par la méthode des p uniformes.
2. Tracer un graphe rendant compte de la f.r. empirique correspondante : on utilisera le fait que cette fonction croît uniquement par sauts de $1/100$ aux points dont les abscisses sont données par l'échantillon réordonné (se servir de l'instruction `sort`).

3. Représenter sur le même graphe la f.r. de la loi $\mathcal{N}(0, 1)$.
4. Estimer la statistique de Kolmogorov-Smirnov ; afficher sa valeur et la conclusion du test pour un niveau d'erreur asymptotique de 10% (faire `help if`, `help disp`). On pourra faire des essais pour $p = 12$ puis pour des valeurs plus petites.
5. Reprendre ces opérations pour la méthode de Box-Muller. On pourra représenter sur une même figure les graphes correspondant aux deux méthodes différentes (voir l'instruction `subplot`).

Autre activité proposée : simuler une loi de Poisson et faire un test du χ^2 pour valider ou non la simulation.

Chapitre 9

Lois des grands nombres ; loi forte
des grands nombres, application à
l'estimation

Chapitre 10

Convergence en loi des variables aléatoires réelles, le TLC et ses applications

Le jury suggère aussi des théorèmes où apparaissent d'autres lois limites que les gaussiennes. Par exemple, l'approximation de la loi binomiale par la loi de Poisson.

Ou encore les lois limites des extrêmes : si les (X_i) sont i.i.d. de f.r. F , $M_n = \sup(X_1, \dots, X_n) \xrightarrow{\text{p.s.}} \text{ess-sup}$ avec $\text{ess-sup} = \inf\{x, F(x) = 1\} \leq \infty$.

Si une suite de la forme $\frac{M_n - b_n}{a_n}$ converge vers une loi qui n'est pas une Dirac, cette loi limite ne peut prendre que trois formes (déterminées par leur f.r.) [COT 140-143] qui renvoie à Billingsley, Probability and measure, p.195

Chapitre 11

Exemples d'utilisation de la fonction de répartition empirique dans les problèmes de modélisation

11.1 Éléments de comparaison entre le test du χ^2 et le test de Kolmogorov-Smirnov

Le test de K-S est toujours consistant (ce qui signifie que la fonction puissance tend vers 1 lorsque $n \rightarrow \infty$ en tout point $\theta \in \Theta_1$); le test du χ^2 ne l'est que dans le cas d'un espace d'états fini : dès qu'il y a choix de classes (E infini), on perd cette propriété.

Le test de K-S ne peut s'appliquer qu'à une loi sur \mathbb{R} puisqu'il fait appel à la f.r. ; par contre, le test du χ^2 peut s'appliquer à des situations beaucoup plus générales (\mathbb{R}^d ou autres) : il suffit en effet partitionner l'espace en un nombre fini de classes.

[à revoir suite à discussion avec François] Dans le cas où la "vraie" loi est différente de la loi théorique, la statistique de K-S diverge vers l'infini plus lentement que celle du χ^2 : la première à vitesse proportionnelle à \sqrt{n} , la seconde à n . En général, il faudra donc tester des échantillons plus grands par K-S que par le χ^2 pour obtenir "raisonnablement souvent" la bonne réponse, à savoir que l'on rejette l'hypothèse H_0 .

Supposons par exemple que nous soyons dans la situation suivante : nous

disposons d'un n -échantillon de loi $U([- \sqrt{3}, \sqrt{3}])$ (extrémités choisies pour obtenir une loi centrée réduite) et nous voulons tester l'hypothèse H_0 : «La loi μ (dont on observe un n -échantillon) vaut $\mathcal{N}(0, 1)$ »

Notons F , resp. G la f.r. de la loi $U([- \sqrt{3}, \sqrt{3}])$, resp. $\mathcal{N}(0, 1)$ et désignons par F_n la f.r. empirique de notre échantillon.

Rappelons ce qui a déjà été vu en introduction au test de Kolmogorov-Smirnov dans le TP "Simulation de v.a. ; applications." :

$$\forall u \in \mathbb{R}, \lim_{n \rightarrow \infty} P\left(\sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - F(t)| \leq u\right) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 u^2}.$$

La somme $S(u)$ de la série de droite est tabulée ; on a par exemple : $S(0,83) \sim 50\%$, $S(1,22) \sim 90\%$, $S(1,36) \sim 95\%$, $S(1,63) \sim 99\%$.

Nous admettrons qu'il est raisonnable d'estimer la limite du membre de gauche par une valeur obtenue pour $n \geq 100$.

Dans notre cas particulier, pour $n \geq 100$, on a :

$$P\left(\sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - F(t)| \leq 1,63\right) \sim 0,99.$$

Pour tester notre hypothèse H_0 avec un niveau de confiance asymptotique de 99%, nous allons calculer $\sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - G(t)|$ et nous rejetterons l'hypothèse si cette quantité est supérieure à 1,63.

Nous remarquons maintenant que nous avons l'inégalité suivante :

$$\sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - G(t)| \leq \sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - F(t)| + \sup_{t \in \mathbb{R}} \sqrt{n} |F(t) - G(t)|$$

D'après ce qui précède, le premier terme du membre de droite sera majoré par 0,83 avec une chance sur deux ; si l'on est dans un tel cas, il est indispensable pour rejeter l'hypothèse que le second terme dépasse $1,63 - 0,83 = 0,8$.

Notons bien que $e = \sup_{t \in \mathbb{R}} |F(t) - G(t)|$ est parfaitement déterminé. On peut facilement en calculer une valeur approchée en Matlab et l'on obtient $e \sim 0,057$.

Ainsi, si l'on veut rejeter (à raison !) avec plus d'une chance sur deux, il est indispensable que $0,057\sqrt{n} \geq 0,8$, ce qui nous donne $n \geq 196$. On peut dire de façon équivalente que la condition $n \geq 196$ est nécessaire pour que la puissance du test dépasse 50%.

Conclusion : les f.r. des lois $U([- \sqrt{3}, \sqrt{3}])$ et $\mathcal{N}(0, 1)$ n'étant pas si différentes que cela, il faut un échantillon de taille relativement grande pour que le test de Kolmogorov-Smirnov puisse les distinguer.

11.1. ÉLÉMENTS DE COMPARAISON ENTRE LE TEST DU χ^2 ET LE TEST DE KOLMOGORO

Pour terminer, notons que l'on peut construire un test de Kolmogorov-Smirnov non asymptotique. Il est basé sur l'idée suivante :

La loi de la v.a.r. $D_n = \sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$ ne dépend en réalité pas de la loi théorique sous-jacente.

Ceci est démontré par exemple dans [OUV² 119-120]. La loi de D_n est tabulée pour différentes valeurs de n , ce qui permet de construire un test de niveau de confiance donné pour un n -échantillon, même si n est "petit".

On pourra trouver une table de la loi de D_n en dernière page de [DAC¹ exo]. Un exemple d'application est donné en bas de la page 80 du même ouvrage.

Chapitre 12

Application du test d'adéquation χ^2 pour un vecteur multinomial en modélisation

12.1 Test d'adéquation à une loi donnée sur un espace d'états fini

Nous considérons n répétitions d'une même expérience aléatoire qui produit un résultat dans un ensemble fini, par exemple $\{1, \dots, k\}$. Nous avons des raisons de supposer que la loi sur $\{1, \dots, k\}$ qui gouverne cette expérience est donnée par $p = (p^1, \dots, p^k)$, avec $p^i \geq 0$ pour tout $1 \leq i \leq k$ et $p^1 + \dots + p^k = 1$, mais nous voudrions confirmer ou infirmer cette hypothèse au regard des valeurs $(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$ observées au cours des n répétitions de l'expérience, c'est-à-dire tester l'*ajustement* de la réalité expérimentale à la loi dont nous avons fait l'hypothèse *a priori*.

Pour ce faire, une première étape consiste à introduire une sorte de distance entre les lois de probabilité sur l'ensemble $\{1, \dots, k\}$, l'idée étant de regarder ensuite si la loi empirique (définie ci-dessous et calculable à partir du résultat de l'expérience) est proche ou éloignée de la loi théorique p dont nous avons fait l'hypothèse.

Définition 12.1.1 Soient p et q deux lois de probabilité sur $\{1, \dots, k\}$. On appelle pseudo-distance du χ^2 entre p et q la quantité :

$$d_{\chi^2}(p, q) = \sum_{i=1}^k \frac{(p^i - q^i)^2}{p^i}$$

On parle de «pseudo-distance» du χ^2 car ce n'est pas du tout une distance au sens des espaces métriques : elle n'est visiblement pas symétrique et il est facile de constater qu'elle ne vérifie pas non plus l'inégalité triangulaire. En réalité, son seul rapport avec une distance est la propriété suivante :

$$d_{\chi^2}(p, q) = 0 \Leftrightarrow p = q$$

Remarquons que cette pseudo-distance a tendance à surévaluer les différences entre p et q sur les entiers i où p_i est petit : nous chercherons à limiter ce phénomène dans la suite en imposant des conditions telles que $np_i \geq 5$ pour tout $i \in \{1, \dots, k\}$.

La deuxième étape de construction du test qui va nous permettre de confirmer ou infirmer notre hypothèse consiste à comparer la loi théorique p avec la loi empirique \bar{p}_n que nous définissons comme suit :

Définition 12.1.2 Si (X_1, \dots, X_n) est la variable aléatoire modélisant les n répétitions de notre expérience, nous posons :

$$\forall i \in \{1, \dots, k\}, \quad N_n^i = \sum_{j=1}^n \mathbf{1}_{\{X_j=i\}}$$

Nous appelons alors loi empirique et nous notons \bar{p}_n la loi sur $\{1, \dots, k\}$ définie par :

$$\forall i \in \{1, \dots, k\}, \quad \bar{p}_n^i = \frac{N_n^i}{n}$$

Notons que la valeur de \bar{p}_n dépend du résultat ω de l'expérience, d'où le qualificatif empirique. En toute rigueur, c'est d'ailleurs $\bar{p}_n(\omega)$ (et non pas \bar{p}_n) qui est une loi de probabilité sur $\{1, \dots, k\}$.

Définition 12.1.3 On appelle χ^2 d'ajustement la variable aléatoire suivante :

$$nd_{\chi^2}(p, \bar{p}_n) = n \sum_{i=1}^k \frac{(p^i - \bar{p}_n^i)^2}{p^i} = \sum_{i=1}^k \frac{(np^i - N_n^i)^2}{np^i}$$

Rappelons maintenant la définition de la loi du χ^2 à d degrés de liberté. Nous noterons $\|\cdot\|$ la norme euclidienne dans \mathbb{R}^d .

Définition 12.1.4 Considérons $Z = (Z_1, \dots, Z_d)$ un vecteur aléatoire dont les composantes sont i.i.d. de loi commune $\mathcal{N}(0, 1)$.

Alors la loi de la v.a.r. $\|Z\|^2$ est appelée loi du χ^2 à d degrés de liberté et notée $\chi^2(d)$. Elle est égale à la loi $\gamma(\frac{1}{2}, \frac{d}{2})$ et admet donc pour densité :

$$g_{\frac{1}{2}, \frac{d}{2}}(x) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} e^{-\frac{x}{2}} x^{\frac{d}{2}-1} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

Le résultat essentiel qui va nous permettre de construire le test dit du χ^2 est le suivant :

Proposition 12.1.5 (Pearson) *Si pour tout $n \in \mathbb{N}^*$, le vecteur aléatoire (X_1, \dots, X_n) suit la loi $p^{\otimes n}$, alors la convergence suivante a lieu :*

$$nd_{\chi^2}(p, \bar{p}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k-1)$$

Démonstration: La démonstration de ce résultat, que le lecteur pourra consulter dans [BC 254], fait appel à la version vectorielle (ou «multivariée») du théorème-limite central.

Remarquons simplement que si (X_1, \dots, X_n) suit la loi $p^{\otimes n}$, c'est-à-dire si les variables X_i sont i.i.d. et de loi commune p , alors le théorème-limite central (en dimension 1) implique la convergence suivante :

$$\frac{N_n^i - np^i}{\sqrt{np^i}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1 - p_i)$$

Il est alors assez intuitif que la loi limite du χ^2 d'ajustement, c'est-à-dire de la variable suivante :

$$nd_{\chi^2}(p, \bar{p}_n) = \sum_{i=1}^k \left(\frac{N_n^i - np^i}{\sqrt{np^i}} \right)^2$$

soit une loi du χ^2 mais on pourrait penser que celle-ci a k degrés de liberté, alors que la vraie loi limite est $\chi^2(k-1)$. En fait, la “perte” d'un degré de liberté peut se comprendre en constatant que nos variables ne sont pas totalement “libres” puisqu'il existe entre elles la relation linéaire suivante :

$$\sum_{i=1}^k (N_n^i - np^i) = 0$$

□

La proposition de Pearson nous dit donc que, si l'expérience est bien gouvernée par la loi p supposée, alors le χ^2 d'ajustement suit une loi proche de $\chi^2(k-1)$ lorsque le nombre n de répétitions de l'expérience devient “grand”.

En revanche, si l'expérience est en réalité régie par une loi $q \neq p$, alors il existe $1 \leq i \leq k$ tel que $q^i \neq p^i$ et la loi des grands nombres implique la convergence suivante lorsque n tend vers l'infini :

$$d_{\chi^2}(p, \bar{p}_n) = \sum_{i=1}^k \frac{(p^i - \bar{p}_n^i)^2}{p^i} \xrightarrow{\text{p.s.}} d_{\chi^2}(p, q) > 0$$

d'où nous déduisons le comportement asymptotique du χ^2 d'ajustement :

$$nd_{\chi^2}(p, \bar{p}_n) \xrightarrow{\text{p.s.}} +\infty \quad (12.1)$$

C'est la différence entre ces deux comportements asymptotiques qui va nous permettre de tester l'hypothèse H_0 : «L'expérience est gouvernée par la loi p » contre l'hypothèse alternative H_1 : «L'expérience est gouvernée par une loi $q \neq p$ ». Passons à la construction effective du test :

Si nous notons F_{k-1} la fonction de répartition de la loi $\chi^2(k-1)$, nous prouvons facilement que F_{k-1} est une bijection de \mathbb{R}_+ sur $[0, 1[$ en constatant que la densité de la loi $\chi^2(k-1)$ est strictement positive sur \mathbb{R}_+^* et nulle sur \mathbb{R}_- .

Par conséquent, pour tout $\alpha \in]0, 1]$, il existe un unique $c_\alpha \in \mathbb{R}_+$ tel que $\chi^2(k-1)(]c_\alpha, +\infty[) = \alpha$ et l'on a $c_\alpha = F_{k-1}^{-1}(1-\alpha)$.

Si l'hypothèse H_0 est réalisée, donc si l'expérience est régie par la loi p , la proposition de Pearson implique (la loi $\chi^2(k-1)$ n'admettant pas de masse ponctuelle) que la convergence suivante a lieu lorsque n tend vers l'infini :

$$P_p(nd_{\chi^2}(p, \bar{p}_n) > c_\alpha) \longrightarrow \alpha \quad (12.2)$$

En revanche, si c'est l'hypothèse H_1 qui est réalisée, donc si l'expérience est gouvernée par une loi $q \neq p$, alors le comportement asymptotique (12.1) entraîne la convergence suivante lorsque n tend vers l'infini :

$$P_q(nd_{\chi^2}(p, \bar{p}_n) > c_\alpha) \longrightarrow 1 \quad (12.3)$$

Nous pratiquons donc notre test comme suit :

Nous choisissons une valeur $\alpha \in]0, 1]$ (typiquement α est petit car, comme nous allons le voir, il représente le niveau d'erreur du test) et nous en déduisons la valeur c_α . Pour le résultat ω de l'expérience que nous observons, nous calculons la valeur du χ^2 d'ajustement :

$$nd_{\chi^2}(p, \bar{p}_n(\omega)) = \sum_{i=1}^k \frac{(np^i - N_n^i(\omega))^2}{np^i}$$

Nous comparons alors cette valeur à c_α pour conclure :

- Si $nd_{\chi^2}(p, \bar{p}_n(\omega)) > c_\alpha$, alors nous rejetons l'hypothèse H_0 .
- Si $nd_{\chi^2}(p, \bar{p}_n(\omega)) \leq c_\alpha$, alors nous acceptons l'hypothèse H_0 .

De façon générale, lorsque nous pratiquons un test statistique, notre conclusion peut être erronée de deux façons différentes :

Erreur de 1ère espèce : Je rejette l'hypothèse H_0 alors qu'elle est satisfaite en réalité. Sa probabilité est appelée risque de première espèce ou «risque α ».

Erreur de 2nde espèce : J'accepte l'hypothèse H_0 alors qu'elle n'est pas satisfaite en réalité. Sa probabilité est appelée risque de seconde espèce ou «risque β ».

Dans de nombreuses situations pratiques, ces deux types d'erreurs ne sont pas symétriques et l'on choisit alors systématiquement l'hypothèse H_0 de sorte que l'erreur de première espèce soit plus grave que l'erreur de seconde espèce. Par exemple, si je teste le câble d'un ascenseur supposé pouvoir accueillir 10 personnes (750kg) et si je note M la masse critique à partir de laquelle le câble casse, je choisirai $H_0 = \ll M \leq 750 \gg$ et $H_1 = \ll M > 750 \gg$ et non l'inverse. L'erreur de 1ère espèce conduirait des usagers de l'ascenseur au grand plongeon : c'est ce risque que je veux absolument maîtriser. L'erreur de 2nde espèce conduirait à des réparations inutiles sur l'ascenseur : je veux l'éviter mais elle est moins grave que la première.

Dans le test du χ^2 que nous venons de construire, la convergence (12.2) se traduit comme suit : la probabilité de commettre une erreur de 1ère espèce est asymptotiquement égale à α . On dit qu'on a construit un test de niveau d'erreur asymptotique α (ou de niveau de confiance asymptotique $1 - \alpha$). Quant à la convergence (12.3), sa traduction est plus vague : lorsque n devient grand, la probabilité de commettre une erreur de 2nde espèce devient "petite" mais nous ne maîtrisons pas la vitesse à laquelle cette convergence se produit. On dit que la *puissance* du test, c'est-à-dire la probabilité de rejeter l'hypothèse H_0 quand elle n'est effectivement pas satisfaite dans la réalité, tend vers 1 lorsque n tend vers l'infini, ou encore que le test est *convergent*.

La proposition 12.1.5 énonçant un résultat asymptotique, quand pouvons-nous considérer qu'elle donne une bonne approximation dans la pratique ? La réponse, basée sur des considérations empiriques, consiste à exiger que $n \geq 30$ et que tous les effectifs théoriques soient supérieurs ou égaux à 5. Voici l'idée de la justification théorique de cette deuxième exigence (le choix du nombre 5 comme seuil étant empirique) :

Quand on approche une binomiale $B(n, p)$ par une gaussienne via le TLC, le théorème de Berry-Esseen nous donne pour borne de la différence entre la vraie fonction de répartition et la fonction de répartition gaussienne une quantité de la forme : $C(p^2 + (1 - p)^2) / \sqrt{np(1 - p)}$.

Ceci suggère que l'approximation pourrait ne pas être bonne lorsque np est petit. On vérifie numériquement que cette difficulté existe bel et bien et que dans ce cas, on a plutôt intérêt à faire une approximation par la loi de Poisson de paramètre $\lambda = np$.

Le même phénomène se reproduit dans le contexte du χ^2 qui est lui aussi basé sur une convergence vers une loi gaussienne donnée par le TLC vectoriel : si l'un des effectifs théoriques np_i est trop petit (< 5 empiriquement), l'attraction poissonnienne vient supplanter l'attraction gaussienne et la loi du χ^2 n'est pas une bonne approximation de la loi réelle du χ^2 d'ajustement.

Notons que la condition $np_i \geq 5$ pour tout $i = 1, \dots, k$ entraîne $n \geq 30$ dès que le cardinal k de l'espace d'états est supérieur ou égal à 6 ; il faudra donc être vigilant simplement dans le cas d'un espace d'états «petit». Retenons que nous pouvons utiliser le test du χ^2 lorsque la règle empirique suivante est satisfaite :

$$\boxed{n \geq 30 \text{ et } \forall i = 1, \dots, k, \quad np^i \geq 5} \quad (12.4)$$

Exemple : Un exemple classique d'application du test du χ^2 consiste à vérifier la validité du raisonnement mené par Mendel [REV 168].

Exercice : Pile ou face biaisé ou non ? On effectue 200 jets d'une pièce de monnaie et l'on obtient 110 piles. Tester l'hypothèse H_0 : «La pièce est équilibrée» par un χ^2 d'ajustement de niveau asymptotique 5% . Reprendre l'exercice avec 2000 jets donnant lieu à 1100 piles.

12.2 Test d'adéquation à une loi donnée sur un espace d'états infini

Nous souhaitons maintenant généraliser le test du χ^2 au cas où l'ensemble E des résultats possibles pour l'expérience est infini. Nous pouvons alors adapter notre méthode comme suit :

Nous choisissons une partition finie (E_1, \dots, E_k) de l'ensemble E . Si ν est la loi sur E supposée gouverner l'expérience, nous posons $p^i = \nu(E_i)$ et nous comptons maintenant le nombre de fois où l'on tombe dans la classe E_i au cours des n répétitions de l'expérience :

$$N_n^i(\omega) = \sum_{j=1}^n \mathbf{1}_{\{X_j(\omega) \in E_i\}}$$

Le reste du test se déroule comme précédemment.

Notons qu'avec cette méthode, nous ne testons pas réellement l'adéquation des données empiriques à la loi ν mais uniquement leur adéquation aux valeurs de ν sur les différentes classes E_i . Ce test ne peut distinguer deux lois

qui chargent les classes E_i de la même façon.

Le choix des classes E_i n'est pas du tout innocent puisque la règle empirique (12.4) doit être respectée. Ainsi, les effectifs théoriques np_i des différentes classes E_i doivent être tous supérieurs ou égaux à 5 (noter que np_i est appelé effectif théorique car c'est l'espérance de N^i dans le cas où l'expérience est vraiment gouvernée par la loi ν). L'hypothèse $n \geq 30$ n'est à vérifier que si l'on a moins de cinq classes.

Le livre «Goodness-of-fit technique» de D'Agostino et Stephens, éditions Dekker, fournit d'autres renseignements intéressants sur le choix des classes ; il est néanmoins compliqué de les justifier théoriquement. En particulier, une situation avantageuse est celle où les classes sont choisies équiprobables sous l'hypothèse nulle, i.e. telles que tous les effectifs théoriques np_i soient égaux (page 69) : le test est en effet alors sans biais et des études empiriques ont montré que l'approximation de la vraie loi de notre statistique sous H_0 par la loi du χ^2 (qui est la loi limite) était particulièrement bonne dans ce cadre. Nous n'insisterons pas plus sur ce sujet qui concerne plus les statisticiens «orfèvres» du test du χ^2 que les agrégatifs.

Exercice : Avec une loi non discrète. Après 1000 répétitions d'une expérience, on obtient la répartition suivante dans différentes classes :

<i>Classe</i> :	[0 ; 0,5]	[0,5 ; 1]	[1 ; 1,5]	[1,5 ; 2]	[2 ; 2,5]	[2,5 ; 3]	[3 ; 3,5]	[3,5 ; 4]
<i>Effectif</i> :	197	220	112	115	71	94	61	45

<i>Classe</i> :	[4 ; 4,5]	[4,5 ; 5]	[5 ; 5,5]	[5,5 ; 6]
<i>Effectif</i> :	36	24	9	16

Appliquer un test d'ajustement de niveau voisin de 1% pour la loi ν définie ainsi : soit X une v.a. uniforme sur $\{0, \dots, 6\}$ et Y une v.a. indépendante de X et uniforme sur $[0, 1]$; alors ν est la loi de XY .

12.3 Test d'adéquation à une famille de lois sur un espace d'états fini

12.3.1 Principe général

L'idée est de faire d'abord une estimation du paramètre θ par un certain estimateur $\hat{\theta}_n$ puis de tester l'ajustement à la loi obtenue pour cette valeur du paramètre en appliquant la méthode habituelle de la première section.

On dispose d'un théorème général nous disant que si $\hat{\theta}_n$ est un estimateur du maximum de vraisemblance (EMV), sous certaines hypothèses de régularité, on garde la convergence mais vers une loi du $\chi^2(r - 1 - k)$, où r est le cardinal de l'espace d'états et k le nombre de paramètres préalablement estimés (on perd donc autant de degrés de liberté que l'on estime de paramètres). La démonstration de ce théorème fait appel à des notions statistiques qui ne sont pas au programme [DAC² cours 112-114]

On peut néanmoins faire une démonstration *ad hoc* dans deux cas particuliers : le test de symétrie [TOU 134-137] et le test d'indépendance [TOU 138-141].

12.3.2 Cas particulier : Test du χ^2 d'indépendance

Exercice : Un examen est ouvert à des candidats qui ont suivi des filières différentes : économie, informatique, mathématiques. On désire savoir si le choix d'une filière par un étudiant influe sur sa réussite à cet examen. Pour cela, on dispose des résultats obtenus l'année précédente par 286 étudiants d'origines diverses ; on les a regroupés dans le tableau suivant :

	Eco	Info	Maths
Succès	41	59	54
Echec	21	36	75

Avec un niveau de confiance voisin de 5%, quelle est votre conclusion ?

Pour des compléments sur le test du χ^2 , par exemple dans le cas de l'ajustement à une famille de lois, le lecteur peut consulter le tome 1 du livre de cours "Probabilités et statistiques" par Dacunha-Castelle et Duflo, éditions Masson, chapitre 5 intitulé 'Échantillons gaussiens, régression et analyse de la variance', paragraphe intitulé 'Le test du χ^2 ' (pages 135 à 137 dans la 2ème édition).

12.4 Test d'adéquation à une famille de lois sur un espace d'états infini

Ce cas est difficile, même si l'espace d'états est simplement dénombrable : par exemple, l'ajustement à une famille de lois de Poisson $P(\theta)$ avec estimation de θ par la moyenne empirique des données nous donne une convergence du χ^2 d'ajustement vers une loi limite qui n'est plus une loi du χ^2 !

Comme dans la section 12.2, on commence par se ramener au cas espace d'états fini en partitionnant l'espace d'états en un nombre fini de classes...et

l'on a toujours la difficulté du choix. Mais à cette étape du traitement du problème apparaît une difficulté supplémentaire quand il s'agit d'estimer θ . En effet, nous étant ramenés au cas fini, nous pouvons appliquer le théorème de la section 12.3, à condition de prendre pour $\hat{\theta}_n$ un EMV correspondant à nos NOUVELLES données, i.e. celles qui nous précisent uniquement les numéros des classes auxquelles appartiennent nos données de départ. Formellement, si notre échantillon de départ s'écrivait (X_1, \dots, X_n) , nous devons maintenant remplacer les données X_i par les

$$Y_i = \sum_{j=1}^r j \mathbf{1}_{\{X_i \in C_j\}},$$

où C_1, \dots, C_r sont les classes formant notre partition finie.

Or calculer un EMV pour ces données-là devient très difficile voire impossible ! Dans la pratique, on remplace donc $\hat{\theta}_n$ par un EMV $\tilde{\theta}_n$ sur les données initiales (par exemple la moyenne empirique dans le cas d'une famille de lois de Poisson) mais ce n'est pas du tout anodin !

On peut montrer que dans ce cas, le χ^2 d'ajustement converge toujours en loi mais vers une autre loi qui est celle d'une v.a.r. de la forme :

$$\sum_{i=1}^{r-k-1} Y_i^2 + \sum_{i=r-k}^{r-1} \lambda_i Y_i^2,$$

où les Y_i sont i.i.d. $\mathcal{N}(0, 1)$ et les λ_i sont des coefficients compris entre 0 et 1.

Si l'on calcule notre région de rejet $]c_\alpha, +\infty[$ pour un niveau de confiance $1 - \alpha$ sur la base d'une loi limite $\chi^2(r - k - 1)$, on va donc commettre une erreur ; un instant de réflexion nous permet de constater que le niveau de confiance réel $1 - \alpha^*$ du test ainsi construit va être inférieur au niveau $1 - \alpha$ désiré.

Néanmoins, dans certains cas, l'erreur sera peu importante. Par exemple, dans le cas d'une famille de lois de Poisson $\mathcal{P}(\theta)$, une simulation avec les classes $C_1 = \{0\}, C_2 = \{1\}, C_3 = \{2, 3, 4, \dots\}$ et le niveau d'erreur souhaité $\alpha = 0.05$ va donner lieu lorsque les données suivent en fait une loi $\mathcal{P}(1)$ à un niveau d'erreur réel $\alpha^* = 0.054$.

Par contre, dans le cas d'une famille gaussienne $\mathcal{N}(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma^2)$, une simulation avec les classes $] - \infty, -1],] - 1, 0],]0, 1],]1, +\infty[$ et $\alpha = 0.05$ donne pour un échantillon de loi $\mathcal{N}(0; 6.25)$ un niveau réel d'erreur $\alpha^* \geq 0.12$.

On s'est donc éloigné du but d'une façon plus ennuyeuse dans ce cas. On peut estimer numériquement les coefficients qui déterminent la vraie loi limite et l'on trouve $\lambda_1 \sim 0.8$ et $\lambda_2 \sim 0.2$. Grossièrement, le premier coefficient

étant proche de 1 et le second assez petit, on peut dire que la vraie loi limite ressemble plus à une $\chi^2(2)$ qu'à la loi $\chi^2(1)$ utilisée pour calculer la région de rejet ; l'erreur commise devient donc non négligeable.

Le lecteur souhaitant des détails sur cette question peut consulter l'article (daté de 1954) de Chernoff et Lehmann intitulé "The use of maximum likelihood estimates in χ^2 tests for goodness of fit" dans la revue "The Annals of Mathematical Statistics" volume 25 pages 579-586.

Une référence rigoureuse sur le test du χ^2 : "Asymptotic Statistics" par Van der Vaart, Cambridge University Press

12.5 Illustrations numériques

Appliquer le test de base du χ^2 à l'expérience de Mendel [REV 168].

Simuler la loi multinomiale $\mathcal{M}(n; p_1, \dots, p_k)$ [BC 301].

Illustrer la convergence énoncée par Pearson dans le cas favorable ($np_i \geq 5$) puis dans le cas défavorable ($\exists i, np_i < 5$) en faisant une comparaison entre histogramme et densité du χ^2 .

Tester le générateur aléatoire (d'une loi $U([0, 1])$) de Matlab.

Appliquer le test d'indépendance du χ^2 sur un exemple donné.

Chapitre 13

Modélisation d'une durée de vie et application à la fiabilité

Chapitre 14

Applications de la théorie des chaînes de Markov à espace d'états fini

Chapitre 15

À partir d'un problème issu de la modélisation, présenter et illustrer la méthode de Monte Carlo pour le calcul d'intégrales multiples

Chapitre 16

Évolution de la taille d'une population lorsque la loi de reproduction est homogène

[REV 63-65]

[*extinction probabilities in branching processes* dans FELLER]

[COT 72-74 et 305-307]

Chapitre 17

Modèle linéaire gaussien ;
utilisation en situation de
modélisation

Chapitre 18

Simulations en Matlab

18.1 Introduction à Matlab

18.1.1 Création d'une fonction.

Pour créer une fonction MATLAB, on pourra suivre la procédure suivante :

1. Ouvrir un fichier "M-file" (barre de commande supérieure : File → New → M-file).
2. Écrire dans ce fichier les différentes instructions qui vont définir la fonction désirée **Nom-Fonction**.

La 1ère ligne est nécessairement de la forme : `function y = Nom-Fonction(x)`, puis, après un retour à la ligne, une succession de commandes appelant à un moment ou à un autre l'argument **x**. Le résultat de **Nom-Fonction** sera **y**.

Il n'est pas nécessaire qu'une fonction appelle un argument. Dans ce dernier cas, la syntaxe est simplement : `function y = Nom-Fonction`. Dans le cas du tracé d'une figure, il n'est pas non plus nécessaire de donner un nom au résultat : la 1ère ligne se réduit à : `function Nom-Fonction(x)` ou `function Nom-Fonction`.

3. Sauvegarder le fichier (barre de commande supérieure : File → Save as... → taper **Nom-Fonction.m**). Il est essentiel de donner le même nom au fichier qu'à la fonction.
4. Revenir à la fenêtre de commande -c'est-à-dire la feuille de calcul- (chercher dans la barre de commande supérieure : Windows → MATLAB Command Window).

5. Exécuter le programme `Nom-Fonction` en appelant directement la fonction `Nom-Fonction` dans la feuille de calcul au même titre désormais que les autres fonctions classiques.

Ne pas hésiter à travailler simultanément sur un "M-file" et sur la feuille de calcul. Vous pouvez également utiliser les raccourcis clavier (touches "Ctrl-quelque chose") au lieu de cliquer avec la souris.

Au cours de l'écriture d'une fonction, on gagne souvent à procéder par étapes en faisant un test à chaque fois. Si le signe `%` est placé devant une ligne, Matlab la passera sans l'exécuter (pratique pour faire des tests).

Pour connaître la définition ou la syntaxe d'une fonction prédéfinie `Matlab-Fonction`, taper `help Matlab-Fonction` dans la feuille de calcul (très utile!). Pour se renseigner sur un thème, par exemple l'intégration, taper `lookfor integration`.

Lorsque Matlab exécute un programme, il affiche toutes les variables intermédiaires (ce qui peut prendre du temps!); pour éviter cela, taper un point-virgule à la fin d'une ligne d'instruction permet de masquer toutes les variables intermédiaires qu'elle contient.

18.1.2 Commandes de base.

1. **Le calcul matriciel.** Matlab utilise comme objet élémentaire les matrices.

De façon caricaturale, un nombre réel est une matrice 1×1 . De même, un vecteur est une matrice $n \times 1$ (ou $1 \times p$).

Si $a = (a_{ij})$ est une matrice $n \times p$, alors `a(i,j)` désigne l'élément a_{ij} . Les opérations usuelles sur les matrices se font avec les opérateurs habituels `+`, `-`, `*`, `/` (multiplication par l'inverse à droite) lorsque cela a un sens.

Noter que si a est une matrice et λ est un scalaire, l'opération `a + lambda` a un sens et désigne la matrice $(a_{ij} + \lambda)$. Idem pour `a-lambda`, `a*lambda` ou `a/lambda`.

Les valeurs des matrices sont saisies entre crochets, ligne par ligne, les lignes sont séparées par des points-virgules. Ainsi, taper `a=[1,2;3,4]` crée la matrice

$$a = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Pour définir une sous-matrice, on utilise un double point : si a est la matrice $a = (1, 7, 3, 6, 5, 2, 4)$, alors `b = a(3:6)` crée la matrice $b = (3, 6, 5, 2)$ (i.e. on ne prend que les éléments de la 3e à la 6e colonne).

Inversement, on peut concaténer deux matrices **a** et **b** en créant **c=[a b]**.

Voici d'autres fonctions matricielles utiles (chercher la syntaxe grâce à l'instruction **help**) : **sum**, **prod**(somme, produit des éléments d'un vecteur), **'**(transposition), **norm**(différentes normes matricielles suivant la syntaxe) ...

Certains types de matrices sont prédéfinis : **eyes(n)** (matrice identité $n \times n$), **ones(m,n)** (matrice $m \times n$ dont tous les éléments sont des 1), **zeros(m,n)** (matrice nulle $m \times n$). Pour déclarer à Matlab qu'une matrice *A* est de taille $m \times n$, on peut par exemple taper **A=zeros(m,n)** puis modifier les coefficients de la matrice dans la suite du programme. Réciproquement, la fonction **length()** renvoie le nombre d'éléments d'un vecteur et la fonction **size()** la taille d'une matrice.

2. **Fonctions statistiques.** L'instruction **help stats** fournit l'ensemble des fonctions statistiques disponibles dans Matlab.

A connaître absolument : **rand(m,n)**(respectivement **randn(m,n)**) qui simule une matrice $m \times n$ dont les éléments sont (censées être...) des réalisations de variables indépendantes, uniformes sur $[0, 1]$ (resp. gaussiennes centrées réduites).

La fonction **cumsum()** calcule la somme cumulée des éléments d'un vecteur.

Voir aussi **mean()** pour le moyenne arithmétique d'un vecteur ligne.

3. **Boucles, incrémentation.** La syntaxe pour une boucle de $i = 1$ à n est la suivante :

```
for i=1:n
instruction
...
instruction
end .
```

La commande **z=(a:pas:b)** crée le vecteur *z* dont les éléments sont les $a + ipas$,

$$0 \leq i \leq [(b - a)/pas].$$

4. **Graphisme.** Voici la commande fondamentale : si *X* et *Y* sont deux vecteurs, **plot(X,Y,'g')** reliera par un trait de couleur verte (g=green) les points $(X(i), Y(i))$.

Le vecteur *X* est souvent de la forme **z=(a:pas:b)** ou encore **linspace(a,b,n)**,

vecteur de taille n constitué de points régulièrement espacés entre a et b .

Voir aussi `line`, qui permet de tracer une ligne (utile pour la représentation de niveaux fixes pour des simulations ; faire `help line` pour connaître la syntaxe...pas forcément naturelle).

Si l'on veut représenter plusieurs graphes sur une même figure, il faut utiliser la fonction `hold`, faute de quoi Matlab créera une nouvelle figure pour chaque graphe.

Il est prudent, lorsque l'on crée une fonction qui produira une figure, de placer au début du programme l'instruction `clf` qui réinitialisera l'écran graphique.

Pour insérer du texte en mode graphique, utiliser `title`, `xlabel` ou `ylabel`(consulter `help` pour connaître leurs usages).

5. **Saisie de variables.** L'instruction `x=input('question')` va entraîner l'affichage de question sur la feuille de calcul lors de l'exécution du programme ; la réponse entrée par l'utilisateur sera alors enregistrée comme valeur de x .
6. **Fonctions numériques.** L'usage des fonctions est standard, ainsi que le syntaxe . Voir par exemple `sqrt()`, `log10()`, `abs()`. La fonction "partie entière" s'écrit `floor()`.

18.2 Illustrations numériques de la LGN

On propose trois applications. (Les programmes sont très similaires ; il n'y a qu'un seul programme à réaliser effectivement.)

18.2.1 Première illustration

On simule n variables aléatoires X_1, \dots, X_n I.I.D. de loi commune μ telle que $m = \int_{\mathbb{R}} x\mu(dx) < \infty$. (Penser à des exemples simples : loi uniforme, exponentielle, ...). On veut montrer que la suite de variables aléatoires $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converge vers m , i.e. on veut mettre en évidence une stabilisation numérique.

1. Simuler X_1, \dots, X_n .
2. Calculer \bar{X}_i , pour $i = 1, \dots, n$.
3. Représenter graphiquement la suite \bar{X}_i , ainsi qu'une droite horizontale de hauteur m . On conseille une échelle logarithmique en abscisses.

4. Étude des fluctuations de \bar{X}_n autour de la moyenne : représenter deux bandes en qui correspondent aux bornes de confiance asymptotiques fournies par le Théorème central limite, et ceci, à un niveau de confiance α donné (par exemple, $\alpha = 0,05$ ou $\alpha = 0,01$)
5. Pour une suite de variables bornées, on peut utiliser l'inégalité de Chebyshev exponentielle pour obtenir des déviations exponentielles : dans le cas où les X_i vérifient $X_i \leq C$ p.s., pour $\lambda > 0$, proposer une majoration grossière de $\exp\{-\lambda nt\} (E\{\exp(\lambda X_1)\})^n$, et, en minimisant en λ , déduire des bornes de confiance non-asymptotiques pour la quantité $|\bar{X}_n - m|$. Comparer avec le point précédent.

18.2.2 Seconde illustration

Dans le même cadre, on s'intéresse au calcul numérique de l'intégrale $I(\varphi) = \int_{\mathbb{R}} \varphi(x)\mu(dx)$ (penser simple ici pour μ , par exemple la loi uniforme $\mu(dx) = 1_{[0,1]}(x)dx$).

Cependant, la fonction φ peut être une fonction présentant une (des) singularité(s), ce qui rend compliqué un calcul approché par une méthode numérique standard (par exemple la méthode des rectangles ou un calcul explicite de la primitive de φ si μ admet une densité par rapport à la mesure de Lebesgue).

On veut montrer que la quantité \bar{X}_n nous fournit une bonne approximation de $I(\varphi)$. Un choix typique de φ peut être une fonction présentant une singularité (e.g. $x \mapsto \frac{1}{\sqrt{1-x^2}}$ ce qui permet un calcul approché de π), des discontinuités ou tout simplement une fonction régulière au sens de l'analyse, mais oscillante, ce qui rend le calcul numérique plus compliqué : par exemple $x \mapsto \sin(1000\pi x)$.

1. Simuler $X_1, \dots, X_n \sim_{I.I.D.}$ uniformes.
2. Calculer $I_n^{(1)}(\varphi) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$.
3. Calculer $I_n^{(2)}(\varphi) = \frac{1}{n} \sum_{i=1}^n \varphi(i/n)$. Comparer $I_n^{(1)}(\varphi)$ et $I_n^{(2)}(\varphi)$.

Exemple : Le volume de la boule unité dans \mathbb{R}^d vaut :

- Si $d = 2p + 1$, avec $p \in \mathbb{N}$, $V_{2p+1} = 2^{2p+1}p!\pi^p/(2p + 1)!$
- Si $d = 2p$, avec $p \in \mathbb{N}^*$, $V_{2p} = \pi^p/p!$

Le maximum est obtenu pour $d = 5$, ce qui n'est pas intuitif.

Ne pas hésiter à étudier aussi des exemples où la méthode stochastique est "plus mauvaise" qu'une méthode déterministe (cas de fonctions régulières dont les dérivées ne sont pas trop grandes). D'un point de vue théorique,

comparer les intervalles de confiance non-asymptotiques (inégalité de type Markov exponentielle) ou asymptotiques (TLC) avec les erreurs des méthodes déterministes.

18.2.3 Troisième illustration

On simule maintenant $Y_i = |X_i|^{1-\alpha}$, $1 \leq i \leq n$ avec X_1, \dots, X_n i.i.d. suivant la loi de Cauchy de paramètre 1 et $0 < \alpha < 1$. Constater que la convergence de \bar{Y}_n est d'autant plus lente que α est proche de 0 en représentant sur un même graphique la suite (\bar{Y}_i) pour différentes valeurs de α .

Remarque: On peut aussi illustrer la non-convergence de \bar{X}_n lorsque μ n'est pas intégrable (loi de Cauchy par exemple), mais étudier la non-convergence d'une méthode numérique est plus délicat.

18.3 Théorème central limite

On propose deux illustrations numériques. Cette fois-ci encore, il n'y a qu'un seul programme à réaliser effectivement ; le programme de la seconde illustration est très proche de celui de la leçon "Loi des grands nombres". Noter qu'on ne propose pas ici l'utilisation du test du χ^2 (on attendra la leçon proprement dite), mais que c'est un bon endroit pour "le caser". (On y reviendra.)

18.3.1 Les commandes Matlab utilisées

On n'introduira que quelques fonctions statistiques supplémentaires :

La fonction `hist()` permet de tracer l'histogramme des effectifs dans différentes classes des points d'un vecteur \mathbf{X} mais elle présente un inconvénient : c'est Matlab qui choisit les extrémités de ces classes et les impute en vecteur des abscisses, ce qui ne nous simplifie pas la tâche lorsqu'on veut superposer un graphe à cet histogramme.

Plus maniable est la fonction `histc()` qui nous permet de choisir ce vecteur des abscisses (extrémités des classes rangées par ordre croissant), noté `EDGES` dans l'aide Matlab, à ceci près qu'elle ne nous renvoie pas le graphe de l'histogramme mais simplement un vecteur $\mathbf{N} = \text{histc}(\mathbf{X}, \text{EDGES})$ qui nous donne les effectifs dans chacune des classes des points de \mathbf{X} .

Si l'on veut tracer l'histogramme correspondant, il faut exécuter en plus l'instruction :

`bar(EDGES,N,'histc')` (attention, il y a une coquille dans `help histc` au sujet de cette instruction!).

Il n'y a alors aucun problème pour superposer un graphe (par exemple celui d'une densité) à condition de faire `hold` puis `plot(EDGES,...)`.

La fonction `erf()` définie par

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

fournit la fonction de répartition d'une gaussienne à transformation affine près ; voir aussi `inverf()` qui inverse la fonction `erf()` et permet de calculer des intervalles de confiances ; pour mémoire et pour la suite, aller voir la fonction `chi2cdf(x,ν)` qui calcule la probabilité pour qu'une variable aléatoire distribuée suivant une loi du χ^2 à ν degrés de liberté soit inférieure à x et son inverse `chi2inv()`.

Pour le graphisme, se souvenir que `x=a:pas:b` crée un vecteur de taille $\lfloor (b-a)/\text{pas} \rfloor$ dont les éléments sont précisément les $a + ipas$; utiliser la fonction `hold()` pour conserver une figure.

Pour améliorer la présentation d'un programme, on peut utiliser la fonction `input()`, qui permet de demander à l'utilisateur d'entrer un paramètre de la fonction.

18.3.2 Première illustration

On simule n variables aléatoires approximativement gaussiennes par la méthode des N uniformes (célèbre pour $N = 12$) : X_1, \dots, X_n . On estime la densité de la somme par un histogramme ainsi que le fonction de répartition empirique qu'on compare avec le résultat théorique de la gaussienne.

1. Simuler X_1, \dots, X_n par la méthode des N uniformes pour $N = 12$.
2. Représenter graphiquement l'histogramme des fréquences pour différents choix de fenêtres -on reviendra sur ce point- et la densité gaussienne sur le même graphique.
3. Représenter graphiquement la fonction de répartition empirique $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}$ calculée à l'aide des données et la fonction de répartition gaussienne sur le même graphique.
4. Faire varier N .

18.3.3 Seconde illustration

Dans le même cadre, on s'intéresse à la précision des estimateurs fournies par le TLC dans un problème d'estimation statistique simple.

1. Simuler $X_1, \dots, X_n \sim_{I.I.D.}$ de loi $p_\theta(x)dx = e^{\theta-x}1_{[\theta, +\infty[}(x)dx$. On simule donc l'expérience suivante : on observe un n -échantillon de la loi de densité p_θ (modèle exponentiel de translation) et on cherche à estimer le paramètre inconnu $\theta \in \mathbb{R}$.
2. calculer $E(X_1)$ et en déduire un estimateur $\hat{\theta}_n$ de θ basé sur l'observation du vecteur (X_1, \dots, X_n) qui converge vers θ . Dans quel sens ?
3. On veut raffiner les propriétés de $\hat{\theta}_n$. Calculer la loi limite de $\sqrt{n}(\hat{\theta}_n - \theta)$ via le TLC.
4. Calculer $\hat{\theta}_n$ et représenter graphiquement $\hat{\theta}_n$ en fonction de n .
5. Indiquer sur le même graphique les bornes de confiance asymptotiques données par le TCL pour différents niveaux de confiance.
6. (Éventuellement) comparer ces résultats à des bornes non-asymptotiques via des inégalités exponentielles.

18.3.4 Méthode de rejet, le cas général

On reprend les notations du cas simple. Le principe consiste à "simuler l'abscisse V_n à une échelle adaptée à f ". On suppose qu'on sait facilement simuler la loi $g(x)dx$, où g est une densité qui vérifie

$$f(x) \leq ag(x), \quad \forall x \in \mathbb{R}$$

pour une constante $a \geq 1$. Alors, si

$$T = \inf\{n \geq 1, ; aU_n g(V_n) \leq f(V_n)\},$$

(la v.a. T est p.s. finie et suit une loi géométrique, même calcul que précédemment), la v.a. $V_T = V_{T(\omega)}(\omega)$ suit la loi $f(x)dx$. En effet

$$P\{aU_n g(U_n) \leq f(V_n)\} = P\left(U_n \leq \frac{f}{ag}(V_n)\right) = \frac{1}{a} \int_{\mathbb{R}} f(v)dv = \frac{1}{a},$$

d'après les calculs -analogues- du cas simple. Puis, pour φ borélienne bornée :

$$\begin{aligned} E\{\varphi(V_T)\} &= \sum_{n \geq 1} E\{\varphi(V_n) ; aU_n g(V_n) \leq f(V_n)\} (1 - 1/a)^n \\ &= \sum_{n \geq 1} \int_{\mathbb{R}} 1_{[0,1]}(u) g(u) \varphi(u) 1_{aug(v) \leq f(v)} dudv (1/a)^n \\ &= \sum_{n \geq 1} \frac{1}{a} (1 - \frac{1}{a})^n \int_{\mathbb{R}} \varphi(v) f(v) dv = \int_{\mathbb{R}} \varphi(v) f(v) dv. \end{aligned}$$

18.4 Intervalles de confiance

On propose 3 applications. Les deux premières concernent l'estimation du paramètre θ pour la loi uniforme $\mathcal{U}([0, \theta])$ dans le cas de l'observation d'un n -échantillon, dans un cadre exact et asymptotique. La troisième application est très classique : on étudie l'estimation de la moyenne et de la variance d'un n -échantillon gaussien.

A noter que comme d'habitude, il n'y a qu'un seul programme à réaliser.

18.4.1 Première illustration

On cherche à estimer θ inconnu, $\theta \in (0, M]$, à partir de l'observation de X_1, \dots, X_n I.I.D. de loi uniforme sur $[0, \theta]$. Etant donné un niveau de confiance $1 - \alpha$ (par exemple $\alpha = 0.05$ ou 0.01), on se propose de calculer par deux méthodes des intervalles de confiance non-asymptotiques.

1. Simuler $X_1, \dots, X_n \sim_{I.I.D.}$ de loi uniforme sur $[0, \theta]$.
2. Montrer que $\hat{\theta}_n = \frac{2}{n} \sum_{i=1}^n X_i$ converge (dans quel sens ?) vers θ et majorer la quantité $\sup_{\theta \in [0, M]} E\{(\hat{\theta}_n - \theta)^2\}$. En utilisant l'inégalité de Markov, en déduire un intervalle de confiance bilatère de θ au niveau α .
3. Le représenter graphiquement en fonction de n pour différentes valeurs de α .
4. Reprendre la même majoration à l'aide d'une inégalité exponentielle. En déduire un second intervalle de confiance pour θ . Le comparer graphiquement avec celui de la question précédente pour une même valeur de α .

18.4.2 Seconde illustration

On reprend le même problème que précédemment d'un point de vue asymptotique. On introduit un second estimateur plus fin que celui basé sur la moyenne empirique.

1. Montrer que $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta^2}{12}\right)$. Montrer qu'on a aussi

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\frac{\hat{\theta}_n}{\sqrt{12}}} \rightarrow_{\mathcal{L}} \mathcal{N}(0, 1)$$

En déduire un interv. de confiance asymptotique pour θ au niveau α .

2. Le représenter graphiquement en fonction de n pour différentes valeurs de α .

3. On introduit l'estimateur $\theta_n^* = \max_{i=1, \dots, n} X_i$. Montrer que :

$$n(\theta_n^* - \theta) \rightarrow_{\mathcal{L}} \xi(\theta) \quad ,$$

où $\xi(\theta)$ a pour densité $p_\theta(x) = e^{x/\theta} \mathbf{1}_{x \leq 0}$. En déduire comme précédemment un intervalle de confiance pour θ au niveau α .

4. Comparaison graphique des deux estimateurs et commentaires.

18.4.3 Troisième illustration

On observe un n -échantillon de la loi $\mathcal{N}(m, \sigma^2)$. Les deux paramètres m et σ sont inconnus.

1. Simuler $X_1, \dots, X_n \sim_{I.I.D.}$ de loi $\mathcal{N}(m, \sigma)$.
2. Calculer des intervalles de confiance à un niveau de confiance α donné en suivant la méthode du cours.
3. Représenter les résultats graphiquement.

18.5 Test d'ajustement du χ^2 .

Pour l'illustration numérique de cette leçon, on se restreindra au test d'ajustement du χ^2 pour une hypothèse simple, et pour une distribution ne prenant qu'un nombre fini de valeurs dans $1, \dots, K$. Ceci nous permet d'éviter le choix délicat des classes pour l'approximation de la distribution théorique. Nous pouvons ainsi écarter des problèmes numériques difficiles à évaluer. Noter qu'on devra tout de même tenir compte de la 'règle' $\inf_{k=1, \dots, K} np_k > 5$, où n est le nombre d'observations et $p_k = P(X = k)$ si X suit la loi qu'on cherche à ajuster, pour pouvoir 'estimer' que les résultats asymptotiques du test du χ^2 s'appliquent.

Dans une seconde application, on raffine la procédure précédente en l'appliquant à une loi discrète quelconque (mais finie).

18.5.1 Quelques compléments de Matlab

1. **Calcul matriciel.** pour extraire la ligne l de la matrice $a = (a_{ij})$, faire `a(l, :)`. De même, faire `a(:, k)` extrait de la matrice a sa k -ième colonne. Une fonction très utile est `length()` qui renvoie le nombre d'éléments d'un vecteur (parfois plus commode que `size()`). Une autre commande essentielle (par exemple pour le calcul d'une fonction de

répartition empirique) est `find()`, qui permet de rechercher des propriétés d'un vecteur. Par exemple, `find(u==2)` renvoie les coordonnées du vecteur u dont les éléments sont égaux à 2.

2. **Graphisme.** La fonction `bar()` permet de tracer un graphique en bâtonnets (utile par exemple pour les lois discrètes empiriques). Ne pas oublier le fonction `hold` ou `hold on` pour conserver plusieurs graphiques. De même, penser à réinitialiser systématiquement l'écran graphique via la fonction `clf`. Pour insérer du texte en mode graphique, utiliser `text`, `xlabel` ou `ylabel`.
3. **Fonctions statistiques** La fonction `chi2cdf(x,n)` renvoie la valeur en x de la fonction de répartition d'une variable aléatoire distribuée selon une loi du χ^2 à n degrés. La fonction utile pour construire des tests est l'inverse `chi2inv(. , .)` (utiliser `help` pour plus d'information). Voir aussi la fonction `cumsum()` qui renvoie la somme cumulée d'un vecteur.
4. **Saisie de variables, commentaires.** La commande `x=input('commentaire')` appelle lors de l'exécution d'une fonction la variable x . On peut insérer des commentaires lors de l'exécution d'une fonction via `disp()`

18.5.2 Première illustration

On teste l'ajustement à une loi uniforme sur $\{1, \dots, K\}$. Les paramètres sont : le niveau de confiance α , le nombre de données n et K . On simule une loi uniforme sur $\{1, \dots, K_{\text{sim}}\}$. Le programme illustre quantitativement l'acceptation ou le rejet de l'hypothèse "La loi est uniforme sur $\{1, \dots, K\}$ " (selon le nombre de données, le niveau de confiance, et le choix de K_{sim}). Noter que dans cette première illustration (simple), l'alternative consiste en l'ensemble des lois uniformes sur $\{1, \dots, K'\}$ avec $K' \neq K$. (On peut très bien ignorer cet aspect des choses.) Le programme se réalise de la manière suivante :

1. Entrer la valeur théorique K et la valeur réelle K_{sim} de la loi simulée.
2. Simuler n variables $(U_i, i = 1, \dots, n)$, indépendantes, de loi uniforme sur $\{1, \dots, K_{\text{sim}}\}$.
3. Représenter graphiquement l'histogramme des fréquences (par exemple à l'aide de `bar()`) des données U_i , et tracer sur le même graphique le niveau théorique des fréquences correspondant à K .
4. Calculer la statistique du χ^2 et accepter ou rejeter l'hypothèse pour un seuil de confiance α donné (utiliser la fonction `chi2inv()`).

18.5.3 Seconde illustration

La seconde illustration n'est en fait qu'une amélioration du programme précédent. On teste l'ajustement à une loi discrète finie quelconque, qui, sans perdre de généralité, est à valeurs dans $\{1, \dots, K\}$. On la note $(p_k)_{1 \leq k \leq K}$ comme dans l'introduction. On simule une loi $(p'_k)_{1 \leq k \leq K}$. Le programme illustre alors quantitativement l'acceptation ou le rejet de l'hypothèse

$$“(p_k = p'_k, k = 1, \dots, K)”$$

(selon le nombre de données, le niveau de confiance, et le choix des p'_k) contre l'alternative : “il existe k_0 tel que $p'_{k_0} \neq p_{k_0}$.” Noter ici que pour simplifier, on se limite à des lois sur $\{1, \dots, K\}$ pour K fixé, mais là encore, ce n'est pas une restriction. Le programme se réalise de la manière suivante :

1. Entrer les valeurs théoriques p_k et les valeurs réelles p'_k . (Prendre des petites valeurs pour K , par exemple $K = 3, 4$ ou 5 .)
2. Simuler n variables $(U_i, i = 1, \dots, n)$, indépendantes, de loi $(p'_k)_{1 \leq k \leq K}$ (On pourra utiliser la méthode vue lors de la première séance).
3. Représenter graphiquement l'histogramme des fréquences (par exemple à l'aide de `bar()`) des données U_i , et tracer sur le même graphique le niveau théorique des fréquences np_K .
4. Calculer la statistique du χ^2 et accepter ou rejeter l'hypothèse pour un seuil de confiance α donné.

18.6 Méthodes de Monte-Carlo

18.6.1 Le cas multidimensionnel

On s'intéresse au calcul numérique de l'intégrale

$$I(\varphi) = \int_{\mathbb{R}^d} \varphi(x_1, \dots, x_d) \mu(dx_1 \dots, dx_d)$$

(penser simple ici pour μ , par exemple la loi uniforme $\mu(dx_1, \dots, dx_d) = 1_{[0,1]^d}(x_1, \dots, x_d) dx$ et d petit, par exemple $d = 3$). Choisir une (des) fonctions φ dont on sait calculer $I(\varphi)$ par une méthode directe (pour pouvoir comparer les méthodes !). Un choix standard est une fonction produit du type $\varphi(x_1, \dots, x_d) = \varphi_1(x_1) \dots \varphi_d(x_d)$ (Fubini !) ou une indicatrice (par exemple $1_{x^2+y^2 \leq 1}$ pour calculer une valeur approchée de π). On peut alors faire la comparaison suivante

1. Simuler $X_1, \dots, X_n \sim_{I.I.D.}$ uniformes sur $[0, 1]^d$.

2. Calculer $I_n^{(1)}(\varphi) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$.
3. Si $n = N^d$, calculer $I_n^{(2)}(\varphi) = \frac{1}{n} \sum_{(i_1, \dots, i_d)} \varphi(i_1/N, \dots, i_d/N)$. Comparer $I_n^{(1)}(\varphi)$ et $I_n^{(2)}(\varphi)$.
4. Écrire un intervalle de confiance au niveau α pour la méthode stochastique (à l'aide de l'inégalité de Chebyshev ou -mieux- du TLC si les calculs sont praticables).

Exemple : Le volume de la boule unité dans \mathbb{R}^d vaut :

- Si $d = 2p + 1$, avec $p \in \mathbb{N}$, $V_{2p+1} = 2^{2p+1} p! \pi^p / (2p + 1)!$
- Si $d = 2p$, avec $p \in \mathbb{N}^*$, $V_{2p} = \pi^p / p!$

Le maximum est obtenu pour $d = 5$, ce qui n'est pas intuitif.

18.6.2 La fonction φ présente une (des) singularité(s)

On considère le cas uni-dimensionnel pour simplifier. La singularité de φ rend compliqué un calcul approché par une méthode numérique standard (par exemple la méthode des rectangles à pas fixe où l'on perd a priori un contrôle de l'erreur).

On veut montrer que par contre la quantité \bar{X}_n nous fournit une bonne approximation de $I(\varphi)$. Un choix de φ peut être $x \mapsto \frac{1}{\sqrt{1-x^2}}$ ce qui permet un calcul approché de π , ou tout simplement une fonction régulière au sens de l'analyse, mais oscillante, ce qui rend le calcul numérique plus compliqué : par exemple $x \mapsto \sin(1000\pi x)$.

1. Simuler $X_1, \dots, X_n \sim_{I.I.D.}$ uniformes sur $[0, 1]$.
2. Calculer $I_n^{(1)}(\varphi) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$.
3. Calculer $I_n^{(2)}(\varphi) = \frac{1}{n} \sum_{i=1}^n \varphi(i/n)$. Comparer $I_n^{(1)}(\varphi)$ et $I_n^{(2)}(\varphi)$.
4. Écrire un intervalle de confiance pour la méthode déterministe.

Il faut relativiser la portée de ces résultats dans le cas où l'on peut isoler les singularités de φ (cas où φ est régulière sauf en un point.)

18.6.3 Méthode de réduction de la variance

A suivre ; se souvenir que le principe de base (en dimension 1 pour une fonction f définie sur $[0, 1]$ disons) consiste à écrire

$$\int_{[0,1]} f(x) dx = \int_{[0,1]} \frac{f(x)}{\mu(x)} \mu(x) dx$$

où $\mu(x) > 0$ est une densité sur $[0, 1]$ telle que

$$\text{Var}_{\text{uniforme}}[Z(f/\mu)] < \text{Var}_{\mu}[Z(f)],$$

avec $Z(f) = f(X) - E[f(X)]$ et X suit la loi uniforme sur $[0, 1]$ sous P_{uniforme} et la loi $\mu(x)dx$ sous P_{μ} . On peut alors simuler les X_i suivant la loi $\mu(x)dx$ par la méthode de rejet et reprendre les calculs précédents. Tout le problème est ramené au choix de μ (A suivre).

18.7 Martingales : théorème d'arrêt.

On présente une illustration numérique du théorème d'arrêt : le calcul du temps d'atteinte moyen d'un niveau a donné par une marche symétrique.

18.7.1 Compléments de Matlab

Il n'y a pas de fonction nouvelle nécessaire pour cette application. Tout au moins on peut améliorer la présentation des graphiques via des commandes comme `title` ou `subplot`. Se souvenir du connecteur logique `&` : `while (condition 1) & (condition 2) ...` exécute la boucle `while` tant que les conditions 1 et 2 sont réunies.

18.7.2 Temps d'atteinte moyen d'un niveau par une marche aléatoire

Soit $(Z_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes et uniformes sur $-1, 1$. On pose $M_0 = 0$ et $M_n = \sum_{i=1}^n Z_i$. Soit $a \geq 1$ un entier. On pose

$$\tau = \inf\{n \geq 1 ; |M_n| = a\}.$$

1. A l'aide du théorème d'arrêt, montrer que $\tau \in L^1$ et que $E(\tau) = a^2$. (Indication : le processus $M_n^2 - nE(Z_1^2)$ est une martingale.)
2. Simuler une trajectoire $(M_n)_{0 \leq n \leq \tau}$. Faire une représentation graphique (en particulier, on pourra tracer les axes $y = \pm a$).
3. Simuler N telles trajectoires indépendantes et calculer les $(\tau^j)_{j=1, \dots, N}$. Montrer que $\frac{1}{N} \sum_{j=1}^N \tau^j$ est une bonne approximation de $E(\tau)$.
4. (Compléments) Calculer le nombre minimal N de trajectoires nécessaires pour garantir avec probabilité supérieure ou égale à $1 - \alpha$ que l'approximation a une précision $\epsilon > 0$ donnée.

18.7.3 Ruine du joueur

Cette illustration est un raffinement du cas précédent ; elle ne nécessite qu'une simple modification du programme précédent. Pour plus d'information, voir Dacunha-Castelle, tome 1 Exercices, p.21.

On considère un jeu de pile ou face avec une probabilité $p > 0$ d'obtenir pile. Le joueur 1 joue "pile" et le joueur 2 "face". On note $M_0 = a$ la fortune initiale du joueur 1 et M_n sa fortune au temps n ; on note b la fortune initiale du joueur 2. Le jeu s'arrête lorsque l'un des deux joueurs est ruiné.

Si ν désigne le temps d'arrêt du jeu et ρ la probabilité de ruine du joueur 1, on montre que, si $p = 1/2$:

$$\rho = \frac{b}{a+b} \text{ et } E(\nu) = ab.$$

On dispose de formules explicites (plus compliquées) pour le cas $p \neq 1$.

1. Simuler une trajectoire $(M_n)_{0 \leq n \leq \nu}$. Faire une représentation graphique.
2. Simuler N telles trajectoires indépendantes et reprendre l'approche de l'illustration précédente pour estimer ρ et $E(\nu)$.
3. (Compléments) Construire un intervalle de confiance pour $E(\nu)$ et ρ en fonction de N et au niveau de confiance α donné.

Remarque On pourra traiter (si on a le courage) le cas $p \neq 1$ en utilisant les formules explicites (cf. Dacunha-Castelle par exemple)

$$\rho = \frac{1 - (1/p - 1)^b}{(p/(1-p))^a - (1/p - 1)^b}$$

et

$$E(\nu) = \frac{(a+b)\rho - b}{1 - 2p}.$$

18.8 Martingales : Convergence.

Aucune nouvelle commande Matlab n'est vraiment nécessaire pour cette leçon.

18.8.1 Critère de Kakutani

On se propose de vérifier un critère simple de convergence pour les martingales de la forme

$$M_n = \prod_{i=1}^n V_i$$

où les V_i sont positives, indépendantes, de moyenne 1. Alors M_n est une martingale positive, donc p.s. convergente. Si $\prod_{i=1}^{\infty} E(\sqrt{V_i}) = 0$, on vérifie facilement que la limite $M_{\infty} = 0$ p.s. On peut montrer (excellent exercice) que si $\prod_{i=1}^{\infty} E(\sqrt{V_i}) > 0$, alors $E(H_{\infty}) = 1$ (indication : on montrera que $\sqrt{M_n}$ est une suite de Cauchy dans L^2 pour montrer la convergence L^1 de M_n). L'illustration numérique peut alors être la suivante :

1. Simuler M_n où les V_i sont des variables simples à simuler, dépendant d'un paramètre tel que l'on ait facilement

$$\prod_{i=1}^{\infty} E(\sqrt{V_i}) = 0 \text{ ou } \prod_{i=1}^{\infty} E(\sqrt{V_i}) > 0$$

en fonction de ce paramètre. Penser simple ici, par exemple les V_i I.I.D., $P(V_i = \frac{1}{2}) = a$ et $P(V_i = \frac{3}{2}) = 1 - a$ où $0 < a < 1$. On vérifiera que le cas critique est atteint pour $a = \frac{\sqrt{3}-\sqrt{2}}{\sqrt{3}-1}$.

2. Représenter graphiquement M_n en fonction de n . Dans le cas où M_n converge p.s. vers 0, tracer sur le même graphique la droite $y = 0$.
3. Dans le cas où M_n ne converge pas p.s. vers 0, simuler N réalisations indépendantes $(M_n^j)_{j=1, \dots, N}$ de la martingale M_n et vérifier que $\frac{1}{N} \sum_{j=1}^N M_n^j$ est proche de 1 pour n et N grands.

18.8.2 Loi du logarithme itéré

Le but est d'illustrer numériquement la *loi du logarithme itéré* suivante : Si Y_1, \dots, Y_n sont n variables aléatoires gaussiennes indépendantes, centrées et réduites, on a :

$$\limsup \frac{\sum_{i=1}^n Y_i}{\sqrt{2n \log \log n}} = 1 \text{ p.s.}$$

1. Simuler le processus $S_n = \sum_{i=1}^n Y_i$.
2. représenter graphiquement les suites S_n et $\sqrt{2n \log \log n}$ et comparer.

18.9 Modèle linéaire gaussien.

On présente deux applications : le calcul de l'estimateur des moindres carrés dans un cas simple et un test d'appartenance à un sous-espace linéaire.

18.9.1 Quelques commandes Matlab supplémentaires

La commande `help stats` fournit la liste des fonctions statistiques essentielles (et exhaustive pour le programme de l'agreg). On regardera en particulier `regress()` qui permet de calculer une régression linéaire simple.

Pour les quantiles des lois usuelles, regarder *Critical values of Distribution Functions* dans le menu `help stats`. En particulier, `fcdf` et `finv` pour la loi de Fisher.

18.9.2 Droite des moindres carrés en régression linéaire

On observe les variables aléatoires Y_1, \dots, Y_n , avec

$$Y_i = f_{(a,b)}(i) + \sigma\epsilon_i, \quad i = 1, \dots, n$$

et où les ϵ_i sont I.I.D. normales centrées réduites. La fonction $f_{(a,b)}$ est connue au paramètre (a, b) près, et admet la représentation *linéaire* $f_{(a,b)}(x) = ax + b$. On estime le paramètre (a, b) par la méthode de maximum de vraisemblance, ou (et de manière équivalente dans le modèle linéaire gaussien) par l'estimateur des moindres carrés (\hat{a}, \hat{b}) qui minimise la fonctionnelle

$$(a, b) \mapsto \sum_{i=1}^n [Y_i - f_{(a,b)}(i)]^2.$$

1. Simuler Y_1, \dots, Y_n comme décrites ci-dessus pour un choix arbitraire de (a, b) .
2. Calculer l'estimateur des moindres carrés (\hat{a}, \hat{b}) .
3. Représenter sur un même graphique les droites $y = ax + b$ et $y = \hat{a}x + \hat{b}$.
4. En supposant σ connu, construire un intervalle de confiance pour a et b au niveau $1 - \alpha$, avec $\alpha > 0$ donné. Représentation graphique.
5. (Facultatif). On suppose σ inconnu. Construire un estimateur de σ et un intervalle de confiance pour σ au niveau $1 - \alpha$. On pourra représenter graphiquement le vecteur des résidus.

18.9.3 Test d'appartenance à un sous-espace linéaire.

On traitera ce problème sur un exemple. On considère le modèle linéaire

$$Y_i = \beta_0 + \beta_1 X_i^1 + \sigma\epsilon_i, \quad i = 1, \dots, n$$

avec les notations habituelles. Le paramètre inconnu est $\beta = (\beta_0, \beta_1)$. Vecto-riellement, le modèle s'écrit $Y = X\beta + \sigma\epsilon$, avec des notations évidentes. On veut tester l'hypothèse $\beta_1 = 0$, ce qui s'écrit

$$\beta \in V_1 = \{X\beta, C\beta = 0\}$$

où $C = (0 \ 1)$. Alors, si $V = \text{vect}\{Xt, t \in \mathbb{R}^2\}$, en définissant W_1 par la relation

$$V = V_1 \oplus W_1,$$

sous l'hypothèse, la statistique

$$T_n = \frac{\frac{\|P_{W_1}Y\|^2}{l}}{\frac{\|P_QY\|^2}{n-p}}$$

est distribuée selon une loi de Fisher à $(l, n-p)$ degrés de liberté ; $Q := I - P_V$ et P_Z désigne la projection orthogonale sur Z lorsque cela a un sens. Ici, l'indice $p - l$ désigne la dimension de V_1 qui est aussi la dimension du noyau de C et l le rang de la matrice C .

1. Simuler Y .
2. Calculer $\hat{\beta}$ par la méthode des moindres carrés. En utilisant le fait que $P_{W_1} = (P_V - P_{V_1})Y = X\hat{\beta} - P_{V_1}Y$, calculer la statistique T_n .
3. Construire un test de l'hypothèse $\beta_1 = 0$ dont l'erreur de première espèce soit contrôlée (par un niveau α arbitraire).

18.10 Fonction de répartition empirique

On pourra tester l'adéquation de lois via le test de Kolmogorov- Smirnov, vu comme une version limite fonctionnelle en loi de Glivenko-Cantelli.

18.10.1 Préliminaires Matlab

Pas de commande nouvelle vraiment nécessaire. Se souvenir de `sort`, et des comparaisons (`find` et double égalité). La fonction de répartition empirique de $\sup_{t \in [0,1]} \sqrt{n}|F_n(t) - F(t)|$ où F_n désigne la fonction de répartition empirique obtenue à partir d'un n -échantillon de la loi F est donnée analytiquement par

$$\phi(u) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 u^2}.$$

D'un point de vue numérique, cette série converge très vite pour u éloigné de l'origine (exercice : majorer le reste de la série!) et une approximation de $\phi(u)$ par $\phi_N(u) = 1 + 2 \sum_{k=1}^N (-1)^k e^{-2k^2 u^2}$, facile à calculer en Matlab pour des petites valeurs de N , sera satisfaisante.

18.10.2 Test de Kolmogorov-Smirnov

1. Simuler un n -échantillon de sa loi favorite de répartition F (par ex. uniforme, exponentielle, Cauchy, normale, et il y en a d'autres...). Calculer $F_n(t)$, la fonction de répartition empirique associée.
2. Tracer sur un même graphique les fonctions F et F_n .
3. Tracer (toujours sur le même graphe) un intervalle de confiance (approché, cf. remarque préliminaire) uniforme à un niveau de confiance α donné de la fonction (supposée inconnue ici) F que l'on estime par F_n .
4. Construire le test d'adéquation de Kolmogorov-Smirnov, dont l'erreur de première espèce est contrôlée (par un niveau de confiance α arbitraire donné). Étudier des cas où l'hypothèse est vraie (F_n obtenue à partir de la loi F) mais aussi des cas où l'hypothèse n'est pas vraie. Par exemple, simuler des variables uniformes et tester l'hypothèse $F(t) = t +$ petite perturbation. Étudier empiriquement le comportement du test en fonction de la perturbation.
5. Question subsidiaire : quelle sont les avantages respectifs des tests d'adéquation du χ^2 et de Kolmogorov-Smirnov lorsqu'on les met en compétition ?

18.11 Grandes déviations - Inégalité de Cramer-Chernov

TESTER $\theta = \theta_0$ contre $\theta = \theta_1 \neq \theta_0$ dans un modèle exponentiel. Chernov permet de contrôler l'erreur de seconde espèce. Test basé sur le max de vraisemblance.

18.12 Chaines de Markov à espace d'état fini

On considère une chaîne $(X_n)_{n \geq 0}$ à valeurs dans E fini, issue de x_0 et de matrice de transition P . Sans perdre de généralités, $E = \{1, \dots, N\}$. On propose dans une première application de simuler le comportement de X_n et

de vérifier la convergence de P^n vers sa loi stationnaire. Dans une seconde application, on étudie les fluctuations du théorème ergodique (TCL) dans le cas particulier de la chaîne d'Ehrenfest.

18.12.1 Première application

Se fixer un entier N pas trop grand ($N = 3$ ou 4 sera déjà suffisant).

1. Se donner $x_0 \in E = \{1, \dots, N\}$ et simuler $(X_i)_{0 \leq i \leq n}$ de loi initiale x_0 , pour n choisi. (On pourra utiliser l'algorithme -élémentaire- d'inversion d'une fonction de répartition dans le cas discret donné lors de la première leçon).
2. Représenter graphiquement le processus $(X_i, i = 0 \dots, n)$.
3. Choisir convenablement P pour pouvoir évaluer μ . Calculer P^n et commenter. En particulier, faire apparaître différents comportements classiques de chaînes (en pratique, il sera peut-être difficile de choisir judicieusement P et "dépasser" le cas où $N = 5$).
4. On considère un choix de n "grand". Représenter sur un même graphique la loi stationnaire μ (si elle est unique) et μ_n , où

$$\mu_n(x) = \frac{1}{n+1} \sum_{i=0}^n 1_{X_i=x}.$$

(Noter que dans les cas où l'on ne sait pas -ou l'on ne veut pas- calculer μ , l'approximation par $P^m 1_{\{x_0\}}(\cdot)$ de la loi stationnaire devrait suffire, en vue de la convergence exponentielle, et ceci, même si m n'est pas trop grand.)

18.12.2 Seconde application : chaîne d'Ehrenfest

Le modèle est le suivant : d billes sont réparties dans deux urnes A et B. On tire un nombre i uniforme entre 1 et d , et la bille i est changée d'urne. On note X_n le nombre de billes dans l'urne A après n telles opérations. Se convaincre que $(X_n)_{n \geq 0}$ est une chaîne de Markov homogène.

1. Montrer que la transition de la chaîne $(X_n)_{n \geq 0}$ vaut $p(x, y) = 0$ si $y \neq x \pm 1$, et $p(x, x+1) = \frac{d-x}{d}$, $p(x, x-1) = \frac{x}{d}$. Simuler une telle chaîne pour n assez grand (devant d).
2. Montrer que la loi stationnaire de la chaîne est une binomiale de paramètres $(d, 1/2)$.
3. Calculer $\lim_{n \rightarrow \infty} P^n$.

4. Vérifier numériquement le résultat de la question précédente par simulation, pour des petites valeurs de d ($d = 5$ par exemple), selon que l'état est pair ou impair.
5. (Question plus difficile.) On peut étudier les fluctuations du théorème ergodique lorsque l'on étudie la convergence de $\frac{1}{n} \sum_{i=1}^n f(X_i)$ à l'aide du théorème central-limite. Pour $d = 4$, la mesure stationnaire est donnée par

$$\mu = \left(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16} \right).$$

La partie pénible est le calcul de la variance asymptotique $\sigma^2(f)$ dans le théorème central limite. On a $\sigma^2(f) = \int \{g(f)^2 - P[g(f)]^2\} d\mu$, où

$$g(x) = (P - I)^{-1} [f(x) - \int f(t) \mu(dt)].$$

(L'application numérique Matlab n'est cependant pas trop difficile). On peut alors (par exemple) construire des intervalles de confiance pour $\int x \mu(dx)$ à l'aide de $\frac{1}{n} \sum_{i=1}^n X_i$ et les représenter graphiquement.

18.13 Chaines de Markov à espace d'états dénombrable

On étudie la récurrence (positive ou nulle) et la transience traite de chaînes de Markov à espace d'états au plus dénombrable. On aborde le cas des marche aléatoires qu'on avait déjà rencontré lors de l'étude des martingales, mais qui s'applique ici, et le cas des processus de vie et de mort.

18.13.1 Première application : marche aléatoire

On simule une marche aléatoire (éventuellement décentrée) sur \mathbf{Z}^d , $d \geq 1$. Dans le cas $d = 1$, on peut vérifier la transience ou la récurrence selon que la chaîne n'est pas symétrique ou est symétrique. Dans le cas $d \geq 3$, on peut vérifier numériquement la transience de la chaîne (pour n assez grand).

1. Le cas $d = 1$. Pour $n \geq 1$ et $0 < q < 1$, simuler une marche $(X_i)_{0 \leq i \leq n}$ sur \mathbf{Z} issue de x_0 et de transition $p(x, x+1) = q$ et $p(x, x-1) = 1 - q$.
2. Faire une représentation graphique et étudier le comportement de $\|X_n - x_0\|$ pour n grand.
3. Faire la même étude dans \mathbf{Z}^d pour $d \geq 2$ et dans le cas d'une marche symétrique. En particulier représenter graphiquement $(\|X_i - x_0\|)_{0 \leq i \leq n}$ et vérifier la transience de la chaîne dès que $d \geq 3$.

18.13.2 Seconde illustration : processus de vie et de mort

Soit $(p_n)_{n \geq 0}$ une suite de nombres dans $]0, 1[$. On définit une chaîne de Markov $(X_n)_{n \geq 0}$ sur \mathbf{N} de la manière suivante : sa transition est donnée par $p(0, 0) = p_0$, $p(0, 1) = 1 - p_0$ et pour $x \geq 1$: $p(x, x - 1) = p_x$, $p(x, x + 1) = 1 - p_x$.

1. Vérifier que la chaîne ainsi définie est irréductible et apériodique.
2. On peut montrer que $(X_n)_{n \geq 0}$ est récurrent positif ssi

$$C = 1 + \sum_{n \geq 1} \frac{(1 - p_0)(1 - p_1) \dots (1 - p_{n-1})}{p_1 p_2 \dots p_n} < \infty.$$

La probabilité stationnaire est alors donnée par

$$\mu(n) = C^{-1} \frac{(1 - p_0)(1 - p_1) \dots (1 - p_{n-1})}{p_1 p_2 \dots p_n}.$$

Montrer que $(X_n)_{n \geq 0}$ est transient ssi

$$\sum_{n \geq 1} \frac{p_1 p_2 \dots p_n}{(1 - p_1)(1 - p_2) \dots (1 - p_n)} < \infty.$$

Illustration numérique : promenade sur \mathbf{N} avec réflexion. On prend $p_n = p \in]0, 1[$ si $n \geq 1$.

1. Simuler $(X_i)_{0 \leq i \leq n}$ pour n suffisamment grand et faire une représentation graphique.
2. Vérifier que la chaîne est récurrente positive si $p > 1/2$, récurrente nulle si $p = 1/2$ et transiente sinon. Calculer la probabilité stationnaire dans le cas où $p < 1/2$.
3. Vérifier la convergence de X_n en loi vers μ de la même manière que 4. du paragraphe 11.1.

18.13.3 Troisième illustration : chaîne de Galton-Watson

(D'après Baldi, Mazliak, Priouret). On se donne une probabilité $\nu = (p_k)_{k \geq 0}$ sur \mathbf{N} telle que $p_0 + p_1 < 1$. On modélise le processus suivant : au temps 0 une particule donne lieu à k naissances avec probabilité p_k (éventuellement 0 particule avec probabilité p_0). Au temps 1, chaque descendant se reproduit et crée un nombre aléatoire de descendants suivant la loi de reproduction $(p_k)_{k \geq 0}$. On note X_n le nombre d'individus engendrés par cette

dynamique à la n -ième génération. Le but est d'étudier la probabilité d'extinction du système, c'est-à-dire $P(\tau_0 < \infty)$, où τ_0 est le temps d'entrée en 0.

On note $G(s) = \sum_{k \geq 0} p_k s^k$ la fonction génératrice de la loi $(p_k)_{k \geq 0}$ et $m = \sum_{k \geq 0} k p_k \leq \infty$.

1. Montrer que la transition de la chaîne est donnée par $p(i, j) = \nu^{*i}(j)$ pour $i \geq 1$ et $p(0, 0) = 1$. Pour un choix de ν et n suffisamment grand, simuler $(X_i)_{0 \leq i \leq n}$. Faire une représentation graphique.
2. Montrer que pour $0 \leq s \leq 1$ et $i \geq 0$:

$$\sum_{j=0}^{\infty} p(i, j) s^j = G(s)^i.$$

Classifier les états de la chaîne.

3. Montrer que

$$\sum_{j=0}^{\infty} p^{(n)}(i, j) s^j = G_n(s)^i,$$

où $G_1 = G$ et $G_{n+1} = G \circ G_n$.

4. Montrer que p.s. $\{Z_n = 0\} = \{\tau_0 \leq n\}$ et en déduire que

$$P(\tau_0 < \infty) = \lim_{n \rightarrow \infty} P(Z_n = 0) = \lim_{n \rightarrow \infty} G_n(0).$$

5. Montrer que G est croissante et strictement convexe dans $[0, 1]$; si $m \leq 1$, $G(t) > t$ sur $[0, 1[$ et si $m > 1$, l'équation $G(t) = t$, $0 \leq t < 1$ admet une racine unique q .
6. Soit q la plus petite racine de $G(t) = t$ dans $[0, 1]$. On a donc $q = 1$ si $m \leq 1$ et $q < 1$ et si $m > 1$. Montrer que pour $0 \leq t < q$, $G_n(t) \rightarrow q$ lorsque $n \rightarrow \infty$ en croissant alors que si $q < t < 1$, $G_n(t) \rightarrow q$ en décroissant. En déduire $P(\tau_0 < \infty)$.

18.14 Processus de renouvellement

18.14.1 Une forme faible du théorème de renouvellement

Rappelons que si $(T_n)_{n \geq 0}$ est un processus de renouvellement (avec $T_0 = 0$) de loi μ (i.e. les variables $T_i - T_{i-1}$ sont indépendantes (et de même loi μ pour $i \geq 1$)). On note $N_t = \sum_{n \geq 1} 1_{T_n \leq t}$. Alors, si $m = \int x \mu(dx)$, on a

$$\frac{1}{t} N_t \rightarrow \frac{1}{m} \text{ lorsque } t \rightarrow \infty$$

presque-surement et dans L^1 .

1. Simuler N_t pour $t \in [0, T]$, où μ est une loi d'inter arrivées que l'on choisira (par exemple exponentielles, mais on peut simuler différentes lois d'inter arrivées).
2. Représenter graphiquement N_t/t et observer une stabilisation vers $1/m$.
3. A l'aide d'une méthode de Monte-Carlo, vérifier que le résultat reste vrai en moyenne sur un grand nombre de réalisations.

18.14.2 Processus de renouvellement stationnaires

On peut tester la stationnarité du processus en faisant varier T_1 .

1. Simuler deux processus de renouvellement $(T_n^{(1)})_{n \geq 0}$ et $(T_n^{(2)})_{n \geq 0}$ indépendants de loi μ (par exemple exponentielle), avec $T_1^{(1)}$ ayant pour densité $\frac{1}{m}\mu([\cdot, \infty[)$ et $T_1^{(2)}$ suivant une loi uniforme.
2. Faire une représentation numérique simultanée de deux telles trajectoires $N_t^{(i)}$ correspondantes pour $t \leq T$ et T "grand".
3. Répéter le point 1 un "grand nombre" de fois pour estimer (Monte-Carlo) les fonctions $t \rightarrow E(N_t^{(i)})$ pour $i = 1, 2$. Faire une représentation graphique (ici, t n'a pas besoin d'être grand).
4. Conclure (on pourra en particulier représenter la droite théorique $E(N_t) = t/m$).

Table des matières

1	Loi binomiale, loi de Poisson. Applications.	7
1.1	Lois binomiales	7
1.1.1	Généralités	7
1.1.2	Application au théorème de Weierstrass	7
1.1.3	Inégalité de grandes déviations et application statistique	8
1.1.4	Loi binomiale négative	8
1.2	Loi de Poisson	8
1.2.1	Généralités	8
1.2.2	Théorème de Raïkov	8
1.2.3	Lois infiniment divisibles sur \mathbb{N}	9
1.2.4	Processus de Poisson	10
1.3	Approximation de la loi binomiale par la loi de Poisson	10
1.3.1	Deux théorèmes-limites	10
1.3.2	Vitesse de convergence dans le thm de Poisson	11
2	Le jeu de pile ou face	13
2.1	Suite finie	13
2.2	Suite infinie	13
2.2.1	Construction	13
2.2.2	Loi binomiale négative	13
2.2.3	Ruine du joueur	14
2.3	Résultats asymptotiques	14
2.3.1	Théorème de De Moivre-Laplace	14
2.3.2	Application statistique : intervalle de confiance	15
2.3.3	Loi du logarithme itéré	16
3	Indépendance d'évènements et de v.a. Exs.	17
3.1	Constructions	17
3.2	Quelques exemples	17
3.3	Vecteurs aléatoires à densités, fonctions caractéristiques	17
3.4	Variance, covariance, cas gaussien	18

3.5	Sommes de v.a. indépendantes, thm s asymptotiques	18
4	Lois gaussiennes et applications	19
4.1	La loi normale centrée réduite	19
4.2	La loi normale générale	21
4.3	Approximation normale de la loi binomiale	24
4.4	Le cas de la dimension d : loi normale multivariée	26
4.4.1	Définition. Loi gaussienne dans \mathbb{R}^d	26
4.4.2	Critère d'indépendance	27
4.5	Exercices corrigés.	29
4.5.1	Un contre-exemple.	29
4.5.2	Indépendance gaussienne : un exemple très simple. . .	30
4.5.3	Limite en loi d'une suite gaussienne.	30
5	Convergence d'une suite de v.a.r.	33
5.1	Diverses notions de convergence	33
5.2	Lois des grands nombres	33
5.2.1	Loi faible	33
5.2.2	Loi forte	33
5.3	Théorème-limite central et raffinements	34
5.3.1	En dimension 1	34
5.3.2	TLC multidimensionnel	34
5.4	Applications du TLC	35
5.5	Intervalles de confiance	36
5.5.1	Quelques notions de statistique	36
5.5.2	L'estimateur moyenne empirique	37
5.5.3	Intervalles de confiance	38
5.6	Méthodes de Monte-Carlo	40
6	Transformées de Laplace et Fourier	43
6.1	Cas discret : fonctions génératrices	43
6.2	Transformation de Laplace	44
6.2.1	Injectivité, lien avec la convolution	44
6.2.2	Inégalité de Hoeffding	44
6.2.3	Grandes déviations	45
6.3	Transformation de Fourier	45
6.4	Exercices sur les transformées de Laplace et Cramer	46
6.4.1	Domaine de définition de \mathcal{L}_X	46
6.4.2	Calculs explicites de transformées de Laplace et Cramer	46
6.4.3	Comportement asymptotique de la transformée de Cra- mer	47

6.4.4	Transformée de Cramer identiquement nulle	47
7	Exs de lois et de leur utilisation en probabilités	49
8	Lois usuelles	51
8.1	Introduction au test de Kolmogorov-Smirnov.	52
8.2	Application sur machine.	52
9	LGN, application à l'estimation	55
10	Cvg en loi, TLC et applications	57
11	Fonction de répartition empirique	59
11.1	Éléments de comparaison entre le test du χ^2 et le test de Kolmogorov-Smirnov	59
12	Test du χ^2	63
12.1	Une loi donnée sur un espace d'états fini	63
12.2	Une loi donnée sur un espace d'états infini	68
12.3	Famille de lois sur un espace d'états fini	69
12.3.1	Principe général	69
12.3.2	Cas particulier : Test du χ^2 d'indépendance	70
12.4	Famille de lois sur un espace d'états infini	70
12.5	Illustrations numériques	72
13	Durée de vie, fiabilité	73
14	Chaînes de Markov sur espace fini	75
15	Méthode de Monte Carlo	77
16	Évolution de la taille d'une population	79
17	Modèle linéaire gaussien	81
18	Simulations en Matlab	83
18.1	Introduction à Matlab	83
18.1.1	Création d'une fonction.	83
18.1.2	Commandes de base.	84
18.2	Illustrations numériques de la LGN	86
18.2.1	Première illustration	86
18.2.2	Seconde illustration	87
18.2.3	Troisième illustration	88

18.3	Théorème central limite	88
18.3.1	Les commandes Matlab utilisées	88
18.3.2	Première illustration	89
18.3.3	Seconde illustration	89
18.3.4	Méthode de rejet, le cas général	90
18.4	Intervalles de confiance	91
18.4.1	Première illustration	91
18.4.2	Seconde illustration	91
18.4.3	Troisième illustration	92
18.5	Test d'ajustement du χ^2	92
18.5.1	Quelques compléments de Matlab	92
18.5.2	Première illustration	93
18.5.3	Seconde illustration	94
18.6	Méthodes de Monte-Carlo	94
18.6.1	Le cas multidimensionnel	94
18.6.2	La fonction φ présente une (des) singularité(s)	95
18.6.3	Méthode de réduction de la variance	95
18.7	Martingales : théorème d'arrêt.	96
18.7.1	Compléments de Matlab	96
18.7.2	Temps d'atteinte moyen d'un niveau par une marche aléatoire	96
18.7.3	Ruine du joueur	97
18.8	Martingales : Convergence.	97
18.8.1	Critère de Kakutani	97
18.8.2	Loi du logarithme itéré	98
18.9	Modèle linéaire gaussien.	98
18.9.1	Quelques commandes Matlab supplémentaires	99
18.9.2	Droite des moindres carrés en régression linéaire	99
18.9.3	Test d'appartenance à un sous-espace linéaire.	99
18.10	Fonction de répartition empirique	100
18.10.1	Préliminaires Matlab	100
18.10.2	Test de Kolmogorov-Smirnov	101
18.11	Grandes déviations - Inégalité de Cramer-Chernov	101
18.12	Chaines de Markov à espace d'état fini	101
18.12.1	Première application	102
18.12.2	Seconde application : chaîne d'Ehrenfest	102
18.13	Chaines de Markov à espace d'états dénombrable	103
18.13.1	Première application : marche aléatoire	103
18.13.2	Seconde illustration : processus de vie et de mort	104
18.13.3	Troisième illustration : chaîne de Galton-Watson	104
18.14	Processus de renouvellement	105

18.14.1 Une forme faible du théorème de renouvellement 105
18.14.2 Processus de renouvellement stationnaires 106