



Cours, Exercices et Travaux pratiques de UE Statistiques Mathématiques 2

L3 MIASHS

(Mathématiques et Informatique Appliquées aux Sciences de l'Homme et de la Société)

Intitulé du cours : L3 Mat F5

1ère partie/2

Chargé du cours

Professeur Mustapha RACHDI

Bureau : C08 du Bât. Michel Dubois

sur RDV

Unité de Formation et de Recherche
Sciences de l'Homme et de la Société

Université Grenoble Alpes
UFR SHS
BP. 47
38040 Grenoble Cedex 09

Année universitaire
2021-2022

Table des matières

Table des matières	3
1 Approximations dans le cas d'échantillons de grandes tailles	5
1.1 Convergence, grands échantillons, approximations gaussiennes	5
1.1.1 Inégalités de Markov et de Bienaymé-Tchebychev	5
1.1.2 Différent types de convergence	6
1.2 Loi des grands nombres (faible et forte)	8
1.3 Théorème central-limite ou de la limite centrée	9
1.4 Exercices	11
1.4.1 Travaux dirigés	11
1.4.2 Travaux personnels	12
1.5 Travaux pratiques	14
1.5.1 Notion de convergence	14
1.5.2 Loi des grands nombres	16
1.5.3 Théorème central-limite : TCL	18
2 Estimation paramétrique	19
2.1 Notion de statistique	20
2.2 Définition et propriétés d'un estimateur	21
2.2.1 Biais d'un estimateur	21
2.2.2 Convergence d'un estimateur	22
2.3 Estimateur optimal	22
2.3.1 Qualité d'un estimateur	22
2.3.2 Inégalité de Fréchet-Darmois-Cramer-Rao	23
2.3.3 Estimateur efficace	24
2.4 Méthodes de construction d'un estimateur	24
2.4.1 Méthode des moments	24
2.4.2 Méthode du maximum de vraisemblance	26
2.5 Estimation par intervalle	27
2.5.1 Intervalle de confiance pour une moyenne quand la variance est connue	27
2.5.2 Lecture des tables statistiques	28
2.5.3 Vers une réalité absolument pas virtuelle	29
2.6 Statistique pivotale	30
2.7 Exercices	32
2.8 Travaux pratiques	35
2.8.1 Biais et Variance	35

2.8.2	Comparaison des estimateurs du paramètre de position	37
2.8.3	Intervalles de confiance (IC)	39
2.8.4	Calcul de l'EMV à l'aide de la commande <code>mle</code>	41
2.8.5	Calcul de l'EMV à l'aide de la commande <code>mle</code> pour les deux paramètres de la loi gamma	45
3	Annexes : Lois utiles en statistique	49
3.1	Loi normale	50
3.2	Loi de chi-deux : χ^2	51
3.3	Loi de Student	53
3.4	Loi de Fisher	56
3.5	La loi gamma	58
	Bibliography	59

Chapitre 1

Approximations dans le cas d'échantillons de grandes tailles

Objectifs : au cours des chapitres précédents, dans lesquels on a introduit la notion de distribution de probabilité, on est sans doute été frappé par la parenté existant entre concepts statistiques et concepts probabilistes. A la notion de fréquence pour une distribution statistique observée correspond la notion de probabilité pour une loi de probabilité; à la notion de moyenne arithmétique d'un caractère statistique correspond la notion d'espérance mathématique d'une variable aléatoire, ...

Notons que la définition d'un système d'événements équiprobables dont les probabilités peuvent être calculées a priori (comme dans le cas des jeux de hasard) est fréquemment impossible dans les études pratiques. Ceci a conduit au développement de la théorie axiomatique du calcul des probabilités (la probabilité d'un événement est un nombre positif satisfaisant quelques conditions et axiomes). Cette théorie, comme son nom l'indique, doit être appliquée dans les études pratiques. D'où la nécessité de déterminer les valeurs numériques des probabilités des événements à partir des données observées. Il faut donc établir un lien entre les données échantillonales et les concepts abstraits de la théorie des probabilités. Ce lien est la fameuse loi des grands nombres, introduite dès le début du XVIII^e siècle par Jacques Bernoulli.

1.1 Convergence, grands échantillons, approximations gaussiennes

1.1.1 Inégalités de Markov et de Bienaymé-Tchebychev

Dans ce paragraphe nous rappelons les deux fameuses inégalités que nous utiliserions le plus souvent, à savoir, l'inégalité de Markov et l'inégalité de Bienaymé-Tchebychev.

Inégalité de Markov : soit X une variable aléatoire **positive** dont l'espérance mathématique qui existe. L'inégalité de Markov est définie pour tout $\varepsilon > 0$ par :

$$\mathbb{P}(X \geq \varepsilon \mathbb{E}(X)) \leq \frac{1}{\varepsilon} \quad \text{ou} \quad \mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

Exemple 1.1.1. Soit X une variable aléatoire suivant la loi Exponentielle de paramètre 1. Calculer le majorant de Markov puis la valeur exacte de $\mathbb{P}(X \geq 2)$.

Inégalité de Bienaymé-Tchebychev : soit X une variable aléatoire quelconque d'espérance mathématique μ_X et d'écart-type σ_X . Nous nous proposons d'étudier la probabilité P , pour que X appartienne à l'intervalle symétrique par rapport à la moyenne $[\mu_X - t\sigma_X, \mu_X + t\sigma_X]$, suivante :

$$P = \mathbb{P}(X \in [\mu_X - t\sigma_X, \mu_X + t\sigma_X]) = \mathbb{P}(|X - \mu_X| \leq t\sigma_X)$$

où t est un nombre réel qui détermine la longueur de l'intervalle.

L'inégalité de Bienaymé-Tchebychev est donnée par la formule suivante :

$$\text{pour tout } t > 0, \quad \mathbb{P}(|X - \mu_X| \geq t) \leq \frac{\sigma_X^2}{t^2}$$

Une déduction directe de cette inégalité est :

$$\mathbb{P}(|X - \mu_X| \leq t\sigma_X) \geq 1 - \frac{1}{t^2}$$

Cette inégalité s'interprète de la façon suivante : *quelque soit la loi de la variable aléatoire X , si l'on connaît la valeur de son écart-type σ_X , on peut toujours choisir t assez grand de telle sorte que la probabilité P relative à l'intervalle $\mu_X \pm t\sigma_X$ soit aussi proche de 1 que l'on désire. Autrement dit, on est quasiment certain que cette variable aléatoire X appartienne à l'intervalle ainsi déterminé.*

Exemple 1.1.2. Soit X une variable aléatoire suivant la loi normale d'espérance μ et de variance σ^2 . Calculer le majorant de Bienaymé-Tchebychev puis la valeur exacte de $\mathbb{P}(|X - \mu| \geq 1.5\sigma)$. Conclure.

L'inégalité de Bienaymé-Tchebychev permet de prouver la loi des grands nombres (cf. Section 1.2).

1.1.2 Différent types de convergence

Dans toute la suite de ce paragraphe, nous considérons une suite infinie de variables aléatoires $X_1, X_2, \dots, X_n, \dots$ qu'on notera simplement (X_n) . Avant de sombrer dans les résultats d'approximations de lois, nous définissons ce que c'est qu'une convergence stochastique. Il faut dire qu'il y a plusieurs types de convergences (nous interpréterons chacune d'elles pendant la séance du cours).

Définition 1.1.1. Soit (X_n) une suite de variables aléatoires. On dit que (X_n) converge :

- en *loi* vers X :

$$\begin{aligned} &\Leftrightarrow \forall x \in \mathbb{R}, \lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x) \\ &\Leftrightarrow \lim_{n \rightarrow +\infty} \mathcal{G}_{X_n}(x) = \mathcal{G}_X(x) \\ &\Leftrightarrow \lim_{n \rightarrow +\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \end{aligned}$$

où $\mathcal{G}_X(t) = \mathbb{E}(e^{tX})$ désigne la fonction génératrice des moments de X et f une fonction quelconque dans \mathcal{C}_b .

- en *probabilité* vers X :

$$\begin{aligned} &\Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1 \\ &\Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0. \end{aligned}$$

Exemple 1.1.3. Soit (U_n) une suite de variables aléatoires indépendantes et identiquement distribuées à $\mathcal{U}_{[0,1]}$. Soit (X_n) la suite de variables aléatoires définie par $X_n = \min(X_1, \dots, X_n)$ pour tout $n \in \mathbb{N}^*$. Montrer que (X_n) converge vers 0 en probabilité.

- en *moyenne d'ordre p* ($1 \leq p < \infty$) vers $X \Leftrightarrow$ le moment d'ordre p de X_n existe pour tout $n \in \mathbb{N}^*$ et $\lim_{n \rightarrow +\infty} \mathbb{E}|X_n - X|^p = 0$

Exemple 1.1.4. Soit (X_n) une suite de variables aléatoires indépendantes et identiquement distribuées d'espérance μ et de variance σ^2 . Montrer que la moyenne empirique converge vers μ en probabilité et en moyenne quadratique.

- *presque-sûrement (p.s.)* vers X :

$$\begin{aligned} &\Leftrightarrow \mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1 \\ &\Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}\left(\sup_{m \geq n} |X_m - X| < \varepsilon\right) = 1 \\ &\Leftrightarrow \forall \varepsilon > 0, \sum_{n=1}^{+\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty \end{aligned}$$

Remarque 1.1.1. Les remarques suivantes sont en fait des propriétés relatives à la définition de la convergence stochastique.

- Théorème de Slutsky : si $X_n \rightarrow X$ presque sûrement et g est une fonction continue, alors $g(X_n) \rightarrow g(X)$ presque sûrement.
- Si $X_n \rightarrow X$ presque sûrement et $Y_n \rightarrow Y$ presque sûrement et g est une fonction continue sur \mathbb{R}^2 alors $g(X_n, Y_n) \rightarrow g(X, Y)$ presque sûrement.

- On a les relations suivantes :

$$\begin{array}{c} \text{Convergence presque-sûre} \Rightarrow \text{Convergence en probabilité} \Leftarrow \text{Convergence en moyenne d'ordre } p \\ \Downarrow \\ \text{Convergence en loi} \end{array}$$

En outre, la convergence en probabilité est équivalente à la convergence en loi dans le cas de la convergence vers une constante. Notons aussi que, dans le cas général, il n'y a pas, entre convergence en m.q. et convergence p.s., de domination de l'une sur l'autre.

Nous énonçons, dans le paragraphe suivant, les principaux résultats de ce chapitre, à savoir, la loi des grands nombres et le théorème central-limite ou de la limite centrée.

1.2 Loi des grands nombres (faible et forte)

Soit (X_n) une suite de variables aléatoires, indépendantes deux à deux, et qui admettent la même espérance mathématique et le même écart-type (c'est-à-dire les mêmes moments d'ordres 1 et 2). Alors :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mathbb{E}(X), \text{ en probabilité, quand } n \rightarrow +\infty$$

Remarque 1.2.1. (*Loi forte des grands nombres*)

Si (X_n) est une suite de variables intégrables mutuellement indépendantes et identiquement distribuées, alors la suite \bar{X}_n tend vers $\mathbb{E}(X)$ presque sûrement.

Remarque 1.2.2. La loi des grands nombres n'a pas d'intérêt pratique pour le calcul statistique, contrairement au théorème central limite (cf. Pragraphe suivant) qui vient préciser la façon dont la moyenne empirique \bar{X}_n converge vers $\mathbb{E}(X)$. Ce théorème est à la base de nombreuses propriétés essentielles des échantillons en statistique.

Exemple 1.2.1. *Convergence de la fréquence observée d'un événement vers sa probabilité :*

On considère le tirage d'un échantillon d'effectif n dans une population comprenant des individus A en proportion p et des individus B en proportion $q = 1 - p$. Si le tirage de l'échantillon est effectué avec remise¹, la fréquence $f_n = X/n$ des individus A observés sur l'échantillon a pour espérance mathématique p et pour écart-type $\sigma = \sqrt{pq/n}$

Une population contient une proportion $p = 0.4$ d'éléments A. On désire que la fréquence f_n des éléments A observés sur l'échantillon, se trouve, avec la probabilité d'au moins 0.99, dans l'intervalle $p \pm 0.01$.

¹. cela peut se faire dans le cas d'un échantillon tiré sans remise : l'espérance sera p et l'écart-type $\sigma = \sqrt{pq/n}\sqrt{(N-n)(N-1)}$

C'est ce qu'on appelle la *loi des grands nombres* : il suffit de tirer un échantillon d'un effectif suffisant dans une population (comportant une proportion p d'individus A) pour que la fréquence observée f_n des individus A soit "presque sûrement" très voisine de la probabilité p . L'inconvénient est qu'il n'y a pas de certitude absolue pour que f_n se trouve dans un intervalle autour de p : la probabilité qu'il n'en soit pas ainsi est d'au plus $1/t^2$.

On dit que la fréquence observée d'un événement converge en probabilité vers la probabilité de cet événement, lorsque n augmente indéfiniment :

$$f_n \rightarrow p, \text{ en probabilité.}$$

1.3 Théorème central-limite ou de la limite centrée

Soit (X_n) une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.). Supposons que pour tout $n \in \mathbb{N}^*$, X_n admet une espérance μ et un écart-type σ . Alors la suite de variables aléatoires $(\sqrt{n}(\bar{X}_n - \mu)/\sigma)$ vérifie :

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow \mathcal{N}(0, 1), \text{ en loi.}$$

Ceci nous amène à une des applications fondamentales dans les calculs statistiques.

Exemple 1.3.1. Dans l'exemple précédent, un échantillon de 240000 individus coûte assez cher, en pratique, pour obtenir avec une probabilité de 99%, une **estimation** de p au 1/100 près :

$$\mathbb{P}(|f_n - p| \leq 0.01) \geq 0.99$$

- On peut déterminer la valeur de la variable aléatoire normale centrée-réduite telle qu'il y ait 99 chances sur 100 pour que f_n se trouve dans l'intervalle $p \pm t\sqrt{pq/n}$.
- Etant trouvé t , on peut déterminer la taille de l'échantillon n qui permet d'avoir la précision désirée.
- Conclusion

Exemple 1.3.2. Application fondamentale

Soit \bar{X}_n la moyenne empirique d'un n -échantillon X_1, \dots, X_n de loi mère quelconque, de moyenne μ et de variance σ^2 . Alors, si n est assez grand \bar{X}_n suit approximativement une loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ ce que l'on note :

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Dans tous les cas (ou presque) $n \geq 30$ suffit pour obtenir des approximations de probabilités 10^{-2} près. Pour une loi continue un seul mode sans queues de distribution trop allongées un n plus petit² pourrait même suffire. Si la loi mère est gaussienne \bar{X}_n est exactement gaussienne pour tout n .

2. $n \geq 5 !!!$

Un cas particulier du Théorème central-limite est le théorème dit de Moivre-Laplace.

Théorème 1.3.1. *Si (X_n) est une suite de variables aléatoires indépendantes telles que pour tout $n \in \mathbb{N}$, $X_n \sim \mathcal{B}(n, p)$ où $p \in]0, 1[$ fixé, alors la variable aléatoire centrée-réduite :*

$$Z_n = \frac{X_n - np}{\sqrt{npq}} \approx \mathcal{N}(0, 1)$$

Dans la pratique, cette approximation semble bonne dès que $np(1 - p) > 9!!$

Exemple 1.3.3. *Approximation de la loi binomiale par la loi normale :*

Une épreuve consiste à lancer une pièce de monnaie (supposée bien équilibrée) 800 fois de suite et à noter le nombre de fois où face est apparue. On note N la variable aléatoire égale au nombre de faces observées. Quelle est la probabilité pour que ce nombre soit entre 390 et 420 ?

Remarque 1.3.1. Comme pour la loi des grands nombres il existe différentes versions du théorème central limite partant de conditions plus ou moins restrictives. En particulier il n'est pas nécessaire que les v. a. soient de même loi ni même qu'elles soient indépendantes dans la mesure où leur degré de dépendance reste faible. Ceci explique que certains phénomènes naturels répondent bien à un modèle gaussien du fait que la variable étudiée résulte de l'addition d'effets aléatoires multiples.

Ainsi on peut établir un comportement asymptotique gaussien pour d'autres types de statistiques dans la mesure où elles sont des moyennes de v.a. qui, sans être nécessairement indépendantes pour n fini, tendent à être i.i.d. quand $n \rightarrow \infty$. En particulier ceci est vrai pour la variance de l'échantillon S_n^2 pour laquelle les éléments $(X_i - \bar{X}_n)$ (et donc leurs carrés) tendent à devenir indépendants du fait que \bar{X}_n converge vers μ . Il est toutefois nécessaire que la variance de S_n^2 existe et il suffit pour cela que μ_4 existe pour la loi mère (cf. le calcul de la variance de la distribution d'échantillonnage de S_n^2).

Toutes les lois qui peuvent être définies comme résultant d'une somme de variables aléatoires i.i.d. tendent à être gaussiennes quand le nombre de termes augmente. C'est évidemment le cas de la loi binomiale $\mathcal{B}(n, p)$ quand $n \rightarrow \infty$.

1.4 Exercices

1.4.1 Travaux dirigés

Exercice 1. *Questions du cours.*

Répondre, en justifiant la réponse donnée, aux questions suivantes.

1. Les majorants des probabilités dans les inégalités de Markov et de Bienaymé-Tchebychev ont des valeurs toujours inférieurs à 1.
2. Si (X_n) est une suite de variables aléatoires telle que $\mathbb{E}(X_n) = \mu$ et $\lim_{n \rightarrow +\infty} \text{var}(X_n) = 0$, alors $X_n \rightarrow \mu$ en probabilité.
3. Si la suite (X_n) converge en probabilité vers la variable aléatoire X alors, la suite $\ln(X_n)$ (resp. $\exp(X_n)$) converge en probabilité vers $\ln(X)$ (resp. $\exp(X)$).
4. Si $X_n \rightarrow a$ en loi alors $X_n \rightarrow a$ en probabilité.
5. La convergence en moyenne quadratique implique la convergence en loi.

Exercice 2. Soit X_1, \dots, X_n, \dots des variables aléatoires indépendantes et identiquement distribuées à X de densité de probabilité :

$$f(x) = \begin{cases} \exp(-(x - \theta)) & \text{si } x \geq \theta \\ 0 & \text{sinon} \end{cases}$$

où θ est un nombre réel positif donné. On définit la variable aléatoire m_n par :

$$m_n = \min\{X_1, \dots, X_n\}$$

1. Déterminer la loi de probabilité de m_n .
2. Calculer $\mathbb{E}(m_n)$.
3. Montrer, par l'application de l'inégalité de Markov, que $m_n \rightarrow \theta$, en probabilité.

Exercice 3. Soit X une variable aléatoire de densité de probabilité :

$$f(x) = \begin{cases} \frac{1}{4}(1 + 3x^2) & \text{si } -1 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer un intervalle de confiance de la forme $[-a, a]$ qui contienne X avec une probabilité supérieure à 0.75.
2. Calculer la probabilité exacte de cet intervalle.

Exercice 4. Soit X une variable aléatoire admettant une densité de probabilité f . On suppose que X ne prend que des valeurs positives ou nulles, et que X admet une espérance mathématique $\mu \neq 0$.

1. Démontrer que pour tout $\varepsilon > 0$:

$$\mathbb{P}(X > \varepsilon\mu) \leq \frac{1}{\varepsilon}$$

2. On note q le 3ème quartile de X , c'est-à-dire le nombre q tel que $F(q) = 3/4$, où F est la fonction de répartition de X . Démontrer que $q \leq 4\mu$.

Exercice 5. Soit (X_n) une suite de variables aléatoires indépendantes et dont la loi de probabilité est définie, pour tout $n \in \mathbb{N}^*$, par :

$$\mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{2n^2} \quad \text{et} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2}$$

1. Etudier la convergence en probabilité de (X_n) .
2. Etudier la convergence en moyenne d'ordre 1 et en moyenne quadratique de (X_n) .
3. Peut-on appliquer le théorème central-limite ?
4. Peut-on appliquer la loi des grands nombres ?

Exercice 6. Un appareil électronique contient 3 accumulateurs. Pour que l'appareil fonctionne il faut que les 3 accumulateurs fonctionnent. On admet que la durée de vie d'un accumulateur suit une loi exponentielle de moyenne 2 ans et que les durées des 3 éléments sont indépendantes.

1. Quelle est la loi de la durée de fonctionnement de l'appareil ?
2. Quelle est sa moyenne ?
3. Quelle est la probabilité qu'elle soit supérieure à 1 an ?

1.4.2 Travaux personnels

Exercice 7. Une machine en fonctionnement normal produit 9% de pièces défectueuses. Un contrôle de qualité consiste à prélever 120 pièces au hasard.

1. Quelle est la loi du nombre de pièces défectueuses ?
 2. Expérience faite, 22 pièces s'avèrent être défectueuses. A quel quantile correspond cette valeur sur la loi précédente ?
- Indication : on recourra à l'approximation gaussienne avec correction de continuité.
3. Qu'en conclure quand au fonctionnement de la machine ?

Exercice 8. Soit (U_n) une suite de variables aléatoires indépendantes et de même loi définie par :

$$\mathbb{P}(U_n = 1) = p \quad \text{et} \quad \mathbb{P}(U_n = -1) = q = 1 - p \quad \text{avec} \quad 0 < p < 1$$

Soit la suite de v.a. (V_n) définie par :

$$V_n = \prod_{i=1}^n U_i$$

1. Déterminer la loi de probabilité exacte de V_n .
 2. Déterminer la loi de probabilité limite V_n .
- Indication : calculer $\mathbb{E}(V_n)$ de deux manières.

Exercice 9. Soit (X_n) une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ que X , où σ est un nombre positif donné.

Montrer que la suite (T_n) définie par :

$$T_n = \frac{1}{n} \sum_{i=1}^n |X_i|$$

converge en moyenne quadratique vers une limite que l'on déterminera.

1.5 Travaux pratiques

La séance de TP se fait sous l'environnement Windows, sauf si l'on a une nette préférence pour Linux. Pour commencer la séance :

1. Créer un répertoire **TP_L3MIASHS_Stat2** sur le bureau.
2. Lancer ensuite R et modifier le répertoire de travail en allant dans **Fichier -> Changer le Répertoire Courant** et en choisissant le répertoire **Bureau/TP_L3MIASHS_Stat2** qui est déjà créé.
3. Ouvrir une fenêtre d'éditeur **Fichier -> Nouveau Script**.
4. Sauver le fichier dans le répertoire courant, par exemple, sous le nom **TP1.R** : **Fichier -> Sauver sous**
5. Pour les différentes questions, vous pouvez utiliser un “copier-coller” à partir de ce document. *Il est fortement recommandé de saisir toutes les commandes dans la fenêtre de l'éditeur que l'on a ouverte.* Pour exécuter les commandes saisies, il suffit de les sélectionner avec la souris et d'appuyer simultanément sur les touches **Ctrl et R**.
6. Pour inclure des commentaires dans le programme, ce qui est fortement recommandé, on doit utiliser le caractère **#**. Tout ce qui suit le caractère **#** sera négligé lors de l'exécution.
7. Penser à sauvegarder régulièrement le contenu du fichier **TP1.R** en appuyant sur les touches **Ctrl et S**.

1.5.1 Notion de convergence

Avant de considérer des propriétés de convergence liées à certaines suites de v. a., nous évoquons des inégalités classiques concernant les variables aléatoires en général. Nous reviendrons dans l'exercice qui suit sur l'inégalité de Tchebychev qui, moyennant des hypothèses légères, majore la probabilité qu'une variable aléatoire s'écarte de son espérance.

Théorème 1.5.1. *Soit une variable aléatoire X d'espérance μ et de variance $\text{var}(X) = \sigma^2$ finie. Alors, pour tout $t > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| > t) \leq \frac{\text{var}(X)}{t^2}. \quad (1.5.1)$$

Exercice 10. Soit $X \sim \mathcal{U}_{[0,1]}$.

1. Rappeler $\mathbb{E}(X)$ et $\text{var}(X)$. Que donne l'inégalité de Tchebychev dans ce cas ?
2. En simulant $N = 1000$ réalisations de X , approcher $\mathbb{P}(|X - 1/2| > t)$ pour $t = 0.3$.
3. Dans un graphe, comparer les membres (de gauche et de droite) mis en jeu dans l'inégalité de Tchebychev (1.5.1) en donnant à t des valeurs comprises entre 0.01 et 1 par exemple.

Les inégalités du type (1.5.1) sont d'une grande utilité théorique, en particulier pour étudier les convergences d'estimateurs (cf. Chapitre suivant). Les propriétés de convergence sont les premières qualités requises pour un estimateur. Elles spécifient son comportement lorsque la taille n de l'échantillon tend vers l'infini. Comme on considère des variables aléatoires, la notion de convergence s'énonce en termes probabilistes. Il existe différentes formes de convergence parmi lesquelles

1. *la convergence en loi* : la suite de variables aléatoires (X_n) converge en loi vers la variable aléatoire X si, pour tout réel x où la fonction de répartition F de X est continue, on a

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

où $F_n(\cdot)$ désigne la fonction de répartition de X_n . On note alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$.

2. *la convergence en probabilité* : la suite de variables aléatoires (X_n) converge en probabilité vers la variable aléatoire X si, pour tout $\varepsilon > 0$,

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

3. *la convergence presque-sûre* : la suite de variables aléatoires (X_n) converge presque sûrement vers la variable aléatoire X si

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1.$$

Remarque 1.5.1. Parmi les trois types de convergence sus-mentionnées, la convergence presque-sûre est la forme la plus forte de convergence dans la mesure où elle implique la convergence en probabilité, qui elle-même implique la convergence en loi³.

Avant de nous intéresser aux propriétés de convergence d'estimateurs qui seraient abordées dans le chapitre suivant, nous revenons, dans ce qui suit, sur la notion de convergence telle qu'elle a été abordée en Analyse Mathématique au travers de l'exemple d'une série numérique. Ceci doit permettre de faire une comparaison avec les notions de convergence telles qu'on les concçoit en probabilité et en statistique.

Exercice 11. (Convergence d'une série numérique)

On rappelle que la suite $S_n = \sum_{k=1}^n \frac{1}{k^2}$ converge vers $\pi^2/6$. Illustrer cette convergence c'est-à-dire le fait que

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ tel que, } \forall n \geq n_0 : \sum_{k=1}^n \left| S_n - \frac{\pi^2}{6} \right| < \varepsilon.$$

Nous en venons aux notions de convergence abordées en probabilité. En guise d'exemple d'introduction, nous nous intéressons tout d'abord aux propriétés de convergence de la borne inférieure d'une suite de v.a. suivant la loi uniforme :

Soit U_1, \dots, U_n, \dots des variables aléatoires indépendantes et de même loi uniforme sur $[0, 1]$. On pose $T_n = \inf(U_1, \dots, U_n)$. On peut montrer facilement que T_n converge en probabilité vers 0. En effet, pour $U \sim \mathcal{U}_{[0,1]}$, on a, pour tout $t \in [0, 1]$

$$\mathbb{P}(U > t) = 1 - F_U(t) = 1 - t.$$

3. Le terme de convergence est souvent associé à la convergence en probabilité alors que l'on parle de convergence forte pour la convergence presque-sûre.

Ainsi, pour tout $\varepsilon > 0$,

$$\mathbb{P}(T_n \leq \varepsilon) = 1 - \mathbb{P}(T_n > \varepsilon) = 1 - \mathbb{P}(U_1 > \varepsilon, \dots, U_n > \varepsilon).$$

L'indépendance donne alors $\mathbb{P}(T_n \leq \varepsilon) = 1 - (1 - \varepsilon)^n$ duquel on déduit facilement $\mathbb{P}(T_n > \varepsilon) = (1 - \varepsilon)^n$, dont la limite est 0.

Exercice 12. (Minimum de lois uniformes)

Soit U_1, \dots, U_n, \dots des variables aléatoires indépendantes et de même loi uniforme sur $[0, 1]$. On pose $T_n = \inf(U_1, \dots, U_n)$. Vérifier par des simulations que T_n converge vers la borne inférieure de la loi mère à savoir 0. Pour cela,

1. Générer $N = 100$ nombres aléatoires suivant une loi uniforme sur $[0, 1]$.
2. Représenter ensuite les points (n, T_n) pour $n = 1, \dots, N$ et vérifier ainsi la convergence vers 0, borne inférieure de l'intervalle $[0, 1]$.

On observe que lorsque n augmente, la suite T_n se rapproche de plus en plus de 0. Ceci tend à montrer la convergence de T_n vers 0. Par ailleurs, est-on certain (comme pour l'exercice précédent) que pour un “couloir/bande” de largeur $\varepsilon > 0$ aussi petite soit-elle, on finira, en augmentant n , par avoir les observations de T_n appartenant à ce “couloir/bande” ?

La réponse est “NON”. En effet, il existe une multitude de tirages possibles pour lesquels T_n ne tend pas vers 0 : un cas simple à envisager est celui où les U_i tirés ne sont jamais inférieurs à une valeur particulière. On ne peut donc pas garantir l'existence d'un indice n_0 au-delà duquel T_n sera aussi proche que souhaité de la valeur cible 0. En revanche, le fait que T_n converge en probabilité vers 0 assure que la probabilité d'observer un échantillon qui sort de ce “couloir/bande” a tendance à baisser avec n , et ainsi tomber à 0 pour n infiniment grand.

1.5.2 Loi des grands nombres

La loi des grands nombres assure la convergence en probabilité de la moyenne empirique d'une suite de variables aléatoires i.i.d. vers l'espérance de la loi commune. La loi forte en assure la convergence presque-sûre moyennant une hypothèse supplémentaire (indépendance mutuelle).

Théorème 1.5.2. (Loi forte/faible des grands nombres)

- Loi faible : Soit (X_n) une suite de variables aléatoires indépendantes et de même espérance $\mathbb{E}(X)$. On a la convergence en probabilité de (\bar{X}_n) vers $\mathbb{E}(X)$:

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|\bar{X}_n - \mathbb{E}(X)| > \varepsilon) = 0$$

où $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est la moyenne empirique.

- Loi forte : Soit (X_n) une suite de variables aléatoires indépendantes, intégrables et de même loi. On a la convergence presque-sûre de \bar{X}_n vers $\mathbb{E}(X)$:

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} \bar{X}_n = \mathbb{E}(X)\right) = 0.$$

Exercice 13. (Loi(s) des grands nombres)

1. Générer une suite $(X_k)_{k \in \mathbb{N}}$ de variables aléatoires i.i.d. selon une même loi (Bernoulli, Normale, ...) admettant une espérance et une variance.
2. Définir la moyenne $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ pour un échantillon de taille n i.e., $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$.
3. Représenter la suite des points (n, \bar{x}_n) en fonction de n .

Plutôt qu'une illustration de la convergence de la moyenne empirique vers l'espérance de la loi mère, le graphe obtenu à la Figure 1.1 illustre la stabilisation de cette moyenne empirique vers cette espérance en fonction de l'augmentation de n .

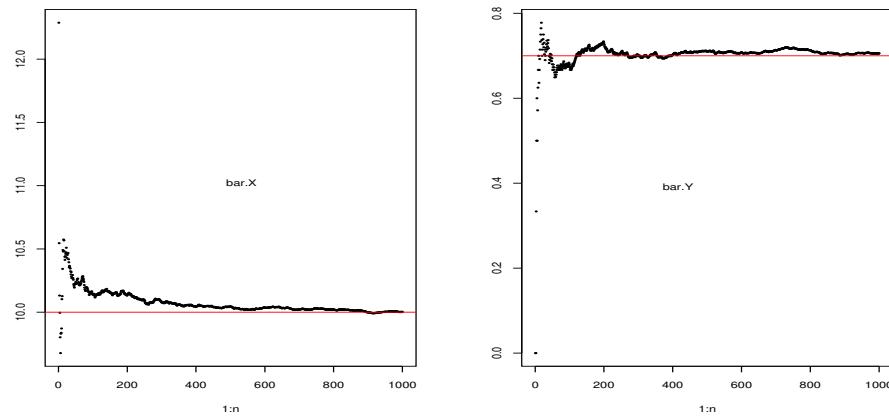


FIGURE 1.1 – Illustration de la loi des grands nombres

Pour compléter l'exercice précédent, précisons qu'il est plus difficile d'illustrer rigoureusement les notions de convergence qu'elles soient en probabilité ou presque-sûre. Les graphes réalisés plus haut représentent une réalisation de moyennes d'échantillon de taille n avec $n \leq 1000$. Dans les cas représentés dans la Figure 1.1, on observe en effet que lorsque n est assez grand, la moyenne devienne proche de μ . Contrairement à ce qui se passe dans l'étude des séries numériques (cf. Exercice 11), les sommes sont ici aléatoires. On ne peut donc pas garantir l'existence d'un indice à partir duquel les moyennes seront à coup sûr aussi proches que souhaité de la limite μ . La loi des grands nombres exprime en revanche que, lorsque la taille de l'échantillon n augmente, la probabilité d'observer un échantillon qui sort du "couloir/bande" diminue, jusqu'à tomber à 0 pour n infiniment grand.

Exercice 14. (Suites convergeants vers σ^2)

Comme dans l'Exercice 13, illustrer graphiquement la convergence, vers σ^2 , des statistiques

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

1.5.3 Théorème central-limite : TCL

Le théorème central-limite précise le comportement asymptotique de la moyenne empirique puisqu'il donne la loi limite de cette variable aléatoire. Il s'énonce de la manière suivante.

Théorème 1.5.3. *Soit (X_n) une suite de variables aléatoires i.i.d. selon une loi commune à celle de X d'espérance $\mathbb{E}(X)$ et de variance σ^2 . Alors :*

$$\frac{1}{n} \left(\frac{X_1 + \dots + X_n - n\mu}{\sigma} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \quad (1.5.2)$$

La statistique/le membre de gauche de l'équation (1.5.2) s'écrit $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$. Ainsi, sous une forme plus intuitive (mais moins rigoureuse), le sens concrét de ce théorème est que la moyenne de l'échantillon aléatoire suit approximativement une loi normale centrée en μ et de variance σ^2/n . Cette approximation s'améliore avec la taille de l'échantillon (on recommande souvent $n \geq 30$ pour que l'approximation soit considérée comme satisfaisante). L'aspect remarquable de ce résultat est que cela est vrai quelque soit la loi de la v.a. X (pour peu qu'elle admette une variance).

Exercice 15. (Théorème central-limite : TCL)

Dans le but d'illustrer le TCL :

1. Considérer une v.a. X d'espérance μ et de variance σ^2 .
2. Générer K échantillons de taille n de la v.a. X .
3. Représenter la distribution des moyennes des échantillons.

On pourra prendre $X \sim \mathcal{N}(\mu, \sigma^2)$, $X \sim \mathcal{U}_{[0,1]}$, $X \sim \mathcal{E}(\lambda)$ et $X \sim \mathcal{B}(p)$, par exemple.

Exercice 16. (Approximation de la loi binomiale par la loi normale)

Soit (X_n) une suite de v.a. avec $X_n \sim \mathcal{B}(n, p)$. Illustrer le fait que

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

(Théorème de Moivre-Laplace).

Chapitre 2

Estimation paramétrique

Objectifs : dans tout problème statistique, on dispose d'une observation x d'un élément (variable ou vecteur) aléatoire X . La statistique inférentielle associe à cette observation un *modèle statistique* ou une *structure statistique* $(\Omega, \mathcal{A}, \mathcal{P})$, où :

- Ω est l'espace des observations : ensemble des valeurs possibles de X
- \mathcal{A} est la tribu des événements observables associés à Ω
- \mathcal{P} est une famille de lois de probabilité possibles pour X , définie sur \mathcal{A} , qu'on identifiera à l'ensemble \mathcal{F} des fonctions de répartition possibles pour X

Dans le cadre de la *statistique paramétrique*, on suppose que \mathcal{F} est en bijection avec un ensemble de paramètres appartenant à un espace de dimension finie :

$$\mathcal{F} = \{F(\cdot, \theta) \text{ où } \theta \in \Theta \subset \mathbb{R}^k\}$$

Par exemple, si l'observation est celle de n variables aléatoires X_1, \dots, X_n indépendantes et de même loi normale $\mathcal{N}(\mu, \sigma^2)$, on a :

$$\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+ \subset \mathbb{R}^2 \text{ et } F(x, \theta) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du$$

Le but de la *statistique inférentielle paramétrique* est alors de déduire des observations le maximum de renseignements sur θ : estimation ponctuelle, intervalle de confiance, tests d'hypothèses, ...

Si l'on connaît la valeur de θ , on peut maîtriser la loi de probabilité $F(\cdot, \theta)$. Mais, en pratique la valeur de θ est inconnue. Pour déterminer la valeur de θ , que l'on utilisera en pratique, on procède de la manière suivante : on utilise un échantillon (des variables aléatoires indépendantes et identiquement distribuées i.e., i.i.d.) de cette loi. A partir des valeurs observées x_1, \dots, x_n de cet échantillon, on calcule une certaine valeur numérique que l'on considérera comme une valeur approchée de θ et qu'on appelle une *estimation de θ* .

Dans ce chapitre, nous nous penchons donc sur cette question, et plus précisément sur la définition de deux types d'estimations : l'estimation ponctuelle et l'estimation par intervalle de confiance.

Exemple 2.0.1. Un problème d'inférence concernant un paramètre θ comme, par exemple, une proportion p (pourcentage de fumeurs, protestants, chômeurs, etc.) ou un taux moyen λ (de naissances par jour, de collisions par semaine sur une route, d'appels par heure pour un standard téléphonique, ...) ou la valeur moyenne μ d'une caractéristique (poids, volume, salaire, temps d'attente entre deux accidents ou deux appels, etc.), est approché par un modèle :

$$X \sim F(x, \theta)$$

où F est la fonction de répartition d'une variable aléatoire X dont la distribution dépend de θ .

2.1 Notion de statistique

On se donne n variables aléatoires indépendantes X_1, \dots, X_n suivant une loi identique à celle de X . Les variables X_1, \dots, X_n constituent alors un *échantillon aléatoire* d'une population de répartition $F(x, \theta)$.

Définition 2.1.1. Soit X_1, \dots, X_n un échantillon aléatoire provenant d'une population $X \sim F(x, \theta)$. Toute variable aléatoire qui s'exprime en fonction de X_1, \dots, X_n est appelée une *statistique*.

Exemple 2.1.1.

$\widehat{\theta}_1(X_1, \dots, X_n) = \bar{X}_n$: Moyenne échantillonnale/empirique
$\widehat{\theta}_2(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = S'_n{}^2$: Variance échantillonnale non corrigée/empirique
$\widehat{\theta}_3(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = S_n^2$: Variance échantillonnale corrigée
$\widehat{\theta}_4(X_1, \dots, X_n) = \frac{\text{nombre de } X_i \leq x}{n} = F_n(x)$: Fonction de répartition échantillonnale/empirique
$\widehat{\theta}_5(X_1, \dots, X_n) = \max\{X_1, \dots, X_n\} = X_{(n)}$: n -ième statistique d'ordre
$\widehat{\theta}_6(X_1, \dots, X_n) = \min\{X_1, \dots, X_n\} = X_{(1)}$: 1ère statistique d'ordre
$\widehat{\theta}_7(X_1, \dots, X_n) = X_{(n)} - X_{(1)}$: Etendue échantillonnale

Chaque réalisation x_1, \dots, x_n de l'échantillon aléatoire X_1, \dots, X_n , qu'on appelle les *données statistiques* ou *observations*, produit une valeur de chacune des statistiques.

Exemple 2.1.2. Les données

$$4.3 \quad 2.2 \quad 2.7 \quad 2.8$$

produiront les valeurs $\bar{x}_4 = 3$, $s_4^2 = 0.82$ et $F_4(2.3) = 0.25$.

Le traitement théorique d'un problème d'inférence portant sur une population $X \sim F(x, \theta)$ consiste à choisir une statistique appropriée (par exemple, \bar{X}_n , X , $S'_n{}^2$, $X_{(n)}$, ...) et à associer à chaque valeur de la statistique choisie une *décision* à propos du paramètre inconnu. La *décision* peut prendre différentes formes, trois desquelles seront traitées dorénavant :

1. *Estimation ponctuelle* : on peut décider que le paramètre a telle ou telle valeur.

2. *Estimation par intervalle* : on peut décider que le paramètre se trouve vraisemblablement dans tel ou tel intervalle.
3. *Test d'hypothèses* : On peut décider que la valeur du paramètre est ou n'est pas égale à un nombre fixé d'avance.

2.2 Définition et propriétés d'un estimateur

Définition 2.2.1. L'*estimation ponctuelle* consiste à trouver un *estimateur* $\hat{\theta}_n$ d'un paramètre inconnu θ , c'est-à-dire une statistique dont les valeurs auraient tendance, en un certain sens, à s'approcher du paramètre inconnu.

Exemple 2.2.1. La moyenne arithmétique $\hat{\theta}_n = \bar{X}_n$ est généralement utilisée comme estimation de l'espérance mathématique $\theta = \mu$ des variables X_1, \dots, X_n , et la variance échantillonnale corrigée :

$$\hat{\theta}_n = S_n^2(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

est utilisée comme estimation de leur variance $\theta = \sigma^2$.

La loi de la variable aléatoire $\hat{\theta}_n(X_1, \dots, X_n)$ dépend de celle de X , et donc de θ , et chaque réalisation $\hat{\theta}_n(x_1, \dots, x_n)$ sera une estimation de θ .

2.2.1 Biais d'un estimateur

Pour pouvoir considérer $\hat{\theta}_n(x_1, \dots, x_n)$ comme une valeur approchée de θ , il faut que les valeurs prises par la variable aléatoire $\hat{\theta}_n$ ne s'écartent pas trop de la valeur, fixe, de θ . Comme $\hat{\theta}_n$ est une variable aléatoire, on peut se référer, par exemple, à sa valeur moyenne, ce qui nous amène à définir la notion de *biais* d'un estimateur comme étant l'écart entre sa moyenne et la vraie valeur du paramètre :

$$b_n(\theta) = \mathbb{E}_{\theta}(\hat{\theta}_n) - \theta$$

Définition 2.2.2. Un estimateur $\hat{\theta}_n$ de $\theta \in \Theta$ est dit *sans biais*, si pour tout entier positif n :

$$\mathbb{E}_{\theta}(\hat{\theta}_n) = \theta$$

Un estimateur $\hat{\theta}_n$ de θ est dit *asymptotiquement sans biais* si :

$$\mathbb{E}_{\theta}(\hat{\theta}_n) \rightarrow \theta \text{ quand } n \rightarrow +\infty$$

Exemple 2.2.2. Soit X_1, \dots, X_n un échantillon de X d'espérance μ et de variance σ^2 .

1. \bar{X}_n est un estimateur sans biais de μ .
2. S'_n est un estimateur asymptotiquement sans biais de σ^2 .

2.2.2 Convergence d'un estimateur

Les notions étudiées dans le chapitre précédent nous seront très utiles pour définir la notion de convergence d'un estimateur.

Définition 2.2.3. Un estimateur $\hat{\theta}_n$ est *convergent* si la suite des variables aléatoires $(\hat{\theta}_n)$ converge en probabilité vers la valeur du paramètre $\theta \in \Theta$:

$$\begin{aligned}\hat{\theta}_n \rightarrow \theta \text{ en probabilité} &\iff \forall \varepsilon > 0, \mathbb{P}(|\hat{\theta}_n - \theta| < \varepsilon) \rightarrow 1, \text{ quand } n \rightarrow +\infty \\ &\iff \forall \varepsilon > 0, \mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \rightarrow 0, \text{ quand } n \rightarrow +\infty\end{aligned}$$

Ce qui nous amène au résultat suivant qui relie la notion de convergence d'estimateur sans biais à la notion de convergence vers 0 de la suite des variances.

Théorème 2.2.1. *Tout estimateur sans biais, ou asymptotiquement sans biais, dont la variance tend vers zéro, quand n tend vers l'infini, est convergent.*

Exemple 2.2.3. $\bar{X}_n \rightarrow \mu$ en moyenne quadratique.

2.3 Estimateur optimal

On se pose la question suivante : est ce qu'il n'existe qu'un seul estimateur pour un paramètre donné ? Autrement dit, est ce qu'un estimateur est unique ? La réponse est “NON” :

Exemple 2.3.1. Soit X_1, \dots, X_n un échantillon d'un caractère X d'espérance μ et de variance σ^2 . Les statistiques X_2 , $2X_3 - X_1$, \bar{X}_n sont des estimateurs sans biais de μ .

Ceci motive la question suivante : comment choisir, donc, un estimateur ? ou comment détecter le meilleur estimateur parmi un ensemble d'estimateurs du même paramètre ?

Dans le paragraphe suivant, nous apportons des éléments de réponse à cette question.

2.3.1 Qualité d'un estimateur

La qualité d'un estimateur peut se mesurer par l'*erreur quadratique moyenne*, définie pour tout $\theta \in \Theta$ par :

$$\text{EQM}(\hat{\theta}_n) = \text{MSE}(\hat{\theta}_n) = \mathbb{E}_{\theta}(\hat{\theta}_n - \theta)^2 = b_n^2(\theta) + \text{var}_{\theta}(\hat{\theta}_n)$$

Dans le cas particulier d'un estimateur sans biais, cette erreur quadratique se confond avec la variance de l'estimateur. Si dans l'erreur totale d'estimation on privilégie l'erreur structurelle, mesurée par $b_n^2(\theta)$, on fera le choix d'un estimateur sans biais et l'erreur d'estimation se réduira à l'erreur statistique mesurée par la variance de l'estimateur. Mais, si l'on a à disposition deux estimateurs sans biais, comment privilégie-t-on l'un sur l'autre ? Pour répondre à cette question, il suffit de savoir que l'erreur quadratique moyenne n'est autre que la variance de l'estimateur. Donc, le meilleur estimateur est celui qui provoque la plus petite erreur quadratique c.-à-d. celui qui a la plus petite variance. On écrit donc :

Si $\hat{\theta}_n$ et $\hat{\theta}'_n$ sont deux estimateurs sans biais du même paramètre θ , alors $\hat{\theta}_n$ est plus efficace que $\hat{\theta}'_n$ si, et seulement si, $\text{var}_\theta(\hat{\theta}_n) \leq \text{var}_\theta(\hat{\theta}'_n)$.

Exemple 2.3.2. Revenir à l'Exemple 2.2.3.

2.3.2 Inégalité de Fréchet-Darmois-Cramer-Rao

Dans ce paragraphe, nous définissons la notion de vraisemblance qui nous permettra de définir ce que c'est qu'un estimateur *efficace* et d'introduire une des méthodes de construction d'un estimateur (cf. paragraphe suivant).

Définition 2.3.1. On appelle *vraisemblance* de l'échantillon X_1, \dots, X_n , la loi de probabilité de ce n -uplet, notée¹ $L(x_1, \dots, x_n; \theta)$, et définie par :

$$L(x_1, \dots, x_n; \theta) = \begin{cases} \prod_{i=1}^n \mathbb{P}(X_i = x_i; \theta) & \text{si } X \text{ est une variable aléatoire discrète} \\ \prod_{i=1}^n f(x_i; \theta) & \text{si } X \text{ est une variable aléatoire continue} \\ & \text{de densité de probabilité } f(x; \theta) \end{cases}$$

$\ln(L(x_1, \dots, x_n; \theta))$ est dite *fonction de score* (ou *score*) de la vraisemblance L .

La technique d'estimation de θ par le maximum de vraisemblance, qui a été introduite par Fisher, consiste à choisir la valeur maximisant la vraisemblance des observations X :

$$\mathbb{E}_\theta \left(\frac{\partial \ln(f(X; \theta))}{\partial \theta} \right) = 0 \quad (2.3.1)$$

L'information de Fisher est une notion statistique qui a été également introduite par R.A. Fisher. Celle-ci quantifie l'information relative à un paramètre contenue dans une distribution.

La quantité, dite *d'information de Fisher*, est donc définie comme étant la variance associée (au maximum obtenu) à la solution de (2.3.1) :

$$I_n(\theta) = \mathbb{E}_\theta \left(\frac{\partial \ln(L)}{\partial \theta} \right)^2 \quad (2.3.2)$$

D'où le théorème fondamental de ce paragraphe :

Théorème 2.3.1. *Sous les hypothèses de Cramer-Rao, en particulier si $X(\Omega)$ est indépendant du paramètre à estimer θ , pour tout estimateur sans biais $\hat{\theta}_n$ de θ , on a :*

$$\text{var}_\theta(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} = B_F(\theta)$$

1. *L* comme Likelihood en anglais qui veut dire vraisemblance

La quantité $B_F(\theta)$ est dite la borne inférieure de Fréchet-Darmois-Cramer-Rao.

Notons que dans les conditions d'application de ce théorème, en particulier si $X(\Omega)$ est indépendant du paramètre à estimer θ , on obtient une expression équivalente de la quantité d'information de Fisher (2.3.2), qui est généralement plus simple à calculer :

$$I_n(\theta) = \mathbb{E}_\theta \left(-\frac{\partial^2 \ln(L)}{\partial \theta^2} \right)$$

Exemple 2.3.3. Soit X une variable aléatoire de loi exponentielle de paramètre $1/\theta$ ou de loi $\Gamma(1, 1/\theta)$, avec $\theta > 0$, de densité de probabilité :

$$f(x, \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \text{ pour } x > 0$$

La vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \exp\left(\frac{1}{\theta} \sum_{i=1}^n x_i\right)$$

et la quantité d'information de Fisher est :

$$I_n = \frac{n}{\theta^2}$$

2.3.3 Estimateur efficace

L'efficacité d'un estimateur se mesure donc par la comparaison de sa variance à la borne de Fréchet-Darmois-Cramer-Rao. D'où la définition :

Définition 2.3.2. Un estimateur sans biais $\hat{\theta}_n$ est dit *efficace* si sa variance est égale à la borne inférieure de Fréchet-Darmois-Cramer-Rao :

$$\text{var}_\theta(\hat{\theta}_n) = \frac{1}{I_n(\theta)}$$

Exemple 2.3.4. Reprenons l'exemple précédent de la loi exponentielle de paramètre $1/\theta$. La statistique $\hat{\theta}_n = \bar{X}_n$ est un estimateur efficace de μ .

2.4 Méthodes de construction d'un estimateur

Il existe, en effet, plusieurs méthodes d'estimation. Nous étudions ici deux méthodes : la *méthode des moments* et la *méthode du maximum de vraisemblance*.

2.4.1 Méthode des moments

Cette méthode part, elle aussi, du principe évoqué plus haut, à savoir, qu'il est naturel d'estimer une moyenne de la population par une moyenne échantillonnale i.e., on estime la moyenne $\mu = \mathbb{E}(X_i)$ par \bar{X}_n .

Supposons que θ , le paramètre qu'on veut estimer, soit le seul paramètre inconnu et que μ soit une fonction de θ : $\mu = \varphi(\theta)$.

Si φ est bijective, elle admettra une application inverse qui nous permettra d'écrire $\theta = \varphi^{-1}(\mu)$. L'estimateur de θ sera donc :

$$\hat{\theta} = \varphi^{-1}(\bar{X}_n)$$

Exemple 2.4.1. Soit X_1, \dots, X_n un échantillon aléatoire simple d'une population de fonction densité de probabilité :

$$f(x; \theta) = \theta e^{-\theta x} \text{ pour } x > 0 \text{ où } \theta > 0.$$

Construire un estimateur $\hat{\theta}$ de θ , par la méthode des moments.

Remarque 2.4.1. L'estimateur obtenu par la méthode des moments n'est pas nécessairement sans biais. Dans l'exemple 2.4.1, $\hat{\theta}$ n'est pas sans biais.

La méthode des moments peut être appliquée, également, au cas d'une distribution à plus d'un paramètre. En effet, s'il est naturel d'estimer $\mu = \mathbb{E}(X_i)$ par la moyenne \bar{X}_n des X_i , il est également naturel d'estimer $\mu'_2 = \mathbb{E}(X_i^2)$ par la moyenne empirique des X_i^2 . En général, le k ème moment de la population, noté μ'_k , est défini par :

$$\mu'_k = \mathbb{E}(X_i^k)$$

On est disposés à estimer le k ème moment de la population $\mathbb{E}(X_i^k)$, par le k ème moment échantillonnaux :

$$m'_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Si la famille des distributions comprend k paramètres, $\theta_1, \dots, \theta_k$, on exprime chacun des k premiers moments μ'_1, \dots, μ'_k en fonction des k paramètres :

$$\begin{aligned} \mu'_1 &= \varphi_1(\theta_1, \dots, \theta_k) \\ &\vdots \\ \mu'_k &= \varphi_k(\theta_1, \dots, \theta_k) \end{aligned}$$

On remplace, ensuite, les μ'_i (resp. $\theta_1, \dots, \theta_n$) par leurs estimateurs m'_i (resp. $\hat{\theta}_1, \dots, \hat{\theta}_n$) et on résoud les équations :

$$\begin{aligned} m'_1 &= \varphi_1(\hat{\theta}_1, \dots, \hat{\theta}_k) \\ &\vdots \\ m'_k &= \varphi_k(\hat{\theta}_1, \dots, \hat{\theta}_k) \end{aligned}$$

Exemple 2.4.2. Soit X_1, \dots, X_n un échantillon aléatoire simple d'une population de fonction densité de probabilité $\mathcal{N}(\theta_1, \theta_2)$.

- Déterminer, par la méthode des moments, des estimateurs pour θ_1 et θ_2 .
- Ces estimateurs sont-ils sans biais ?

2.4.2 Méthode du maximum de vraisemblance

L'estimateur dit du maximum de vraisemblance se définit comme suit :

Définition 2.4.1. On appelle estimateur du maximum de vraisemblance (EMV), du paramètre θ , toute fonction $\hat{\theta}_n$ de x_1, \dots, x_n qui vérifie :

$$L(x_1, \dots, x_n; \hat{\theta}_n) = \max_{\beta \in \Theta} L(x_1, \dots, x_n; \beta) \quad (2.4.1)$$

En fait, comme la formule (2.4.1) l'indique, la vraisemblance représente la probabilité d'observer le n -uplet (x_1, \dots, x_n) pour une valeur fixée de θ . Mais pour calculer cette valeur il faut bien connaître la loi de probabilité. Notons que celle-ci dépend du paramètre θ qui est inconnu. Pour ce faire, nous allons attribuer à θ la valeur la plus vraisemblable, compte tenu de l'observation dont on dispose (qui va lui attribuer la plus forte probabilité).

Comment faire ?

On se fixe un n -uplet (x_1, \dots, x_n) et on considère la vraisemblance L comme une fonction de θ , puis on attribue à θ la valeur qui maximise cette fonction.

Remarque 2.4.2. La définition ci-dessus ne nous garantit ni l'existence, ni l'unicité d'un tel estimateur. La recherche de l'EMV peut se faire directement par la recherche du maximum de L .

Une des méthodes, les plus populaires et directe, permettant de trouver l'EMV est de trouver l'annulateur de :

$$\frac{\partial L}{\partial \theta} = 0 \text{ avec } \frac{\partial^2 L}{\partial \theta^2} < 0$$

quand L est deux fois dérivable par rapport à θ .

Remarque 2.4.3. La vraisemblance L se présente comme un produit. Comme la fonction \ln est strictement croissante, il est donc préférable de considérer le problème équivalent suivant :

$$\frac{\partial \ln L}{\partial \theta} = 0 \text{ avec } \frac{\partial^2 \ln L}{\partial \theta^2} < 0$$

Exemple 2.4.3. L'EMV pour la famille de lois exponentielles de paramètre $1/\theta$ est $\hat{\theta}_n = \bar{X}_n$.

Rappelons, par ailleurs, que si $\hat{\theta}_n$ est un EMV de θ alors $g(\hat{\theta}_n)$ est l'EMV du paramètre $g(\theta)$ pour toute fonction g .

Exemple 2.4.4. Si la variance empirique corrigée $S_n^{2'}$, qui est un estimateur sans biais de $\theta = var_\theta(X)$, est un EMV pour un modèle statistique donné, alors $S'_n = \sqrt{S_n^{2'}}$ (en prenant $g(x) = \sqrt{x}$) est un EMV de :

$$g(\theta) = \sqrt{var_\theta(X)} = \sigma_X = \sqrt{\theta}.$$

Notons que S'_n ne peut pas être un estimateur sans biais, de σ , car on aurait alors :

$$var_\theta(S'_n) = I\!E_\theta(S_n^{2'}) - I\!E_\theta^2(S'_n) = \theta - \theta = 0$$

2.5 Estimation par intervalle

Un estimateur donne une valeur unique comme estimation. La valeur obtenue a peu de chances de coïncider avec celle du vrai paramètre, qui est inconnu. L'estimation par intervalle de confiance consiste à entourer, d'un intervalle $[a, b]$, la valeur de l'estimateur et affirmer plutôt que θ se trouve dans $[a, b]$. On peut alors choisir a et b de telle sorte que la probabilité que cette proposition soit vraie soit assez élevée.

2.5.1 Intervalle de confiance pour une moyenne quand la variance est connue

Nous tâchons de montrer comment déterminer un intervalle de confiance pour une moyenne μ , ce qui permettra, en même temps, d'illustrer la procédure générale.

Pour déterminer un intervalle de confiance pour $\mu = \mathbb{E}(X)$, nous devons connaître la distribution de l'estimateur \bar{X}_n . Pour cela nous devons faire quelques suppositions sur la distribution des X_i pour $i = 1, \dots, n$.

Nous ferons la supposition suivante :

$$\text{l'échantillon } X_1, \dots, X_n \text{ est issu de } X \sim \mathcal{N}(\mu, \sigma^2).$$

Alors :

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ et } Z = \frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1).$$

Puisque nous connaissons parfaitement la distribution de Z , nous pouvons calculer la probabilité que la variable Z prenne une valeur dans tel ou tel intervalle. Nous pouvons affirmer (cf. Figure 2.1), par exemple, que :

$$\mathbb{P}\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$$

La valeur 1.96 étant choisie à l'aide de la table des fractiles de la loi $\mathcal{N}(0, 1)$, précisément pour que la probabilité à droite soit de 0.95.

On conclut que :

$$\mathbb{P}\left(\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Une façon de lire cet énoncé est la suivante :

La probabilité pour que l'intervalle $\left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right]$ recouvre la vraie moyenne μ (quelle que soit sa valeur) est de 95%.

On dit alors que l'intervalle :

$$\left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance pour μ à 95%.

2.5.2 Lecture des tables statistiques

Exemple 2.5.1. On prélève un échantillon de taille $n = 64$ de boites de conserves d'une population dont la variance des poids est $\sigma^2 = 100\text{g}^2$ et donc $\sigma/\sqrt{n} = 10/8 = 1.25$. Si l'on suppose que les poids, dans la population, sont distribués normalement, on peut affirmer que l'intervalle :

$$\left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right] = \left[\bar{X}_n - 1.96 \times 1.25, \bar{X}_n + 1.96 \times 1.25 \right] = \left[\bar{X}_n - 2.45, \bar{X}_n + 2.45 \right]$$

a une probabilité de 95% de contenir la moyenne μ (cf. Figure 2.1).

La technique que nous venons de mettre en application peut être légèrement modifiée pour obtenir des intervalles de confiance dont le niveau de confiance serait différent de 95%. Par exemple, nous pourrions vouloir déterminer un intervalle de confiance dont le niveau de confiance serait de 90% ou de 99%. Le seul ajustement à faire est de remplacer la constante 1,96 par une valeur appropriée. Pour trouver cette valeur, il faut se rappeler comment nous avions obtenu la valeur 1.96. Nous avions choisi cette valeur parce que si Z est une variable aléatoire qui suit la loi $\mathcal{N}(0, 1)$, alors $\mathbb{P}[-1.96 \leq Z \leq 1.96] = 0.95$. Entre -1.96 et $+1.96$ nous avons 95% de la surface sous la courbe normale. Par conséquent, il y a 5% de la surface en dehors de cet intervalle. A cause de la symétrie de la courbe normale, il y a 2.5% à gauche de -1.96 et 2.5% à droite de $+1.96$. Si nous voulons un intervalle de confiance à 90%, il faudrait trouver une constante c telle que $\mathbb{P}[-c \leq Z \leq c] = 0.90$. A l'extérieur de l'intervalle allant de $-c$ à c , il y aura 10%. La Figure 2.2 peut nous aider dans notre recherche :

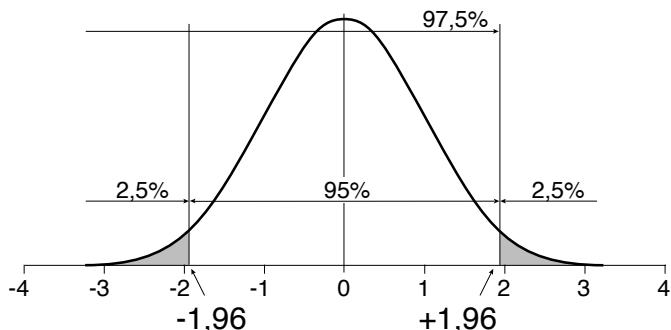


FIGURE 2.1 – Densité de probabilité de la loi $\mathcal{N}(0, 1)$ et intervalle de confiance pour $\alpha = 5\%$

En examinant la table de la loi normale standard, on constate que la surface à gauche de $c = 1.64$ est $\Phi(1.64) = 0.9495$ tandis que à gauche de $c = 1.65$ est $\Phi(1.65) = 0.9505$. Par interpolation, ou en utilisant une table plus précise, on constate que $\Phi(1.645) \approx 0.95$, et par conséquent on considère la constante c est **1.645**.

Un travail analogue montre que si nous désirons un intervalle de confiance à **99%**, il faut prendre $c = 2.576$.

D'une manière générale, nous voulons un intervalle dont le niveau de confiance est $(1 - \alpha)$ lorsque

nous acceptons une marge d'erreur α . Graphiquement, la situation est représentée comme dans la Figure 2.2.

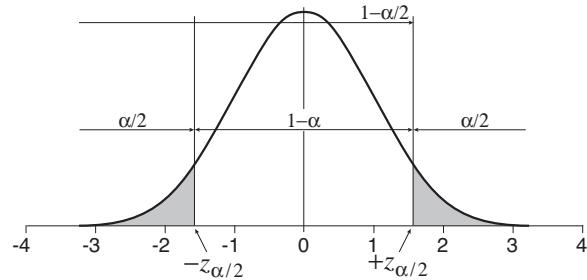


FIGURE 2.2 – Densité de probabilité de la loi $N(0, 1)$

La constante $z_{\alpha/2}$ est en fait telle que :

- (i) $\mathbb{P}\{Z \leq -z_{\alpha/2}\} = \alpha/2$
- (ii) $\mathbb{P}\{Z \geq +z_{\alpha/2}\} = \alpha/2$
- (iii) $\mathbb{P}\{-z_{\alpha/2} \leq Z \leq +z_{\alpha/2}\} = 1 - \alpha$

Cela nous conduit à la formulation générale suivante de l'intervalle de confiance. A partir de l'échantillon aléatoire X_1, \dots, X_n de X , l'intervalle de confiance de niveau $(1 - \alpha)$ pour la moyenne μ de la population mère est donné par :

$$[\bar{X}_n - z_{\alpha/2}\sigma_{\bar{X}_n}, \bar{X}_n + z_{\alpha/2}\sigma_{\bar{X}_n}] \quad \text{où } \sigma_{\bar{X}_n} = \sigma/\sqrt{n}.$$

2.5.3 Vers une réalité absolument pas virtuelle

Lorsque nous faisons une estimation ponctuelle de la moyenne μ par \bar{X}_n , nous utilisons les données que nous avons et nous calculons $\bar{X}_n = \frac{1}{n} \sum X_i$. Aucune autre information n'est nécessaire. Le désavantage de cette estimation est qu'elle ne donne aucune indication relativement à sa précision. C'est pour cette raison que nous avons introduit la notion d'intervalle de confiance. Mais, de la manière dont nous avons procédé, nous supposons connue la variance (ou d'une manière équivalente, l'écart type) de la population. Il va de soi que si cette variance n'est pas connue, il est impossible de calculer l'intervalle de confiance.

Cela nous conduit donc à un second niveau d'estimation. Afin d'indiquer la précision de notre estimation de la moyenne, nous devrons estimer, c'est le mieux que nous puissions faire, la variance de la population. Les données sont donc aussi utilisées pour estimer la variance de la population. L'estimation classique de la variance d'une population, en utilisant les données d'un échantillon de taille n , se calcule à partir de la formule suivante :

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

On dit que S_n' est la *variance échantillonnale corrigée*. Chaque échantillon étant aléatoire, S_n' est donc une variable aléatoire.

Rappelons la formule générale d'un intervalle de confiance pour la moyenne :

$$[\bar{X}_n - z_{\alpha/2} \sigma_{\bar{X}_n}, \bar{X}_n + z_{\alpha/2} \sigma_{\bar{X}_n}] \text{ où } \sigma_{\bar{X}_n} = \sigma / \sqrt{n} \text{ est l'erreur type de } \bar{X}_n.$$

La question qui se pose est la suivante :

Dans la formule, ci-dessus, donnant l'intervalle de confiance, quel est l'impact de substituer à σ^2 , la valeur de $S_n'^2$?

Nous aurions, dans ce cas, comme intervalle de confiance :

$$[\bar{X}_n - z_{\alpha/2} \hat{\sigma}_{\bar{X}_n}; \bar{X}_n + z_{\alpha/2} \hat{\sigma}_{\bar{X}_n}] \text{ où } \hat{\sigma}_{\bar{X}_n} = S_n / \sqrt{n}.$$

Il faut se rappeler que la construction de l'intervalle de confiance s'est faite sous l'hypothèse que nous échantillonons une population normale :

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} \sim \mathcal{N}(0, 1).$$

En substituant à $\sigma_{\bar{X}_n}$ une estimation $\hat{\sigma}_{\bar{X}_n}$, nous n'avons plus l'assurance que la v.a. Z_n soit distribuée selon une loi $\mathcal{N}(0, 1)$. En d'autres mots, nous ne pouvons plus utiliser la table de la loi normale pour calculer le fractile z_α ou $z_{\alpha/2}$.

Par suite, afin de résoudre ce problème, on introduirait la loi de Student (cf. Annexes)

2.6 Statistique pivotale

Une statistique $\varphi(X_1, \dots, X_n; \theta)$, qui est fonction des observations X_1, \dots, X_n et du paramètre θ , est appelée “quantité pivotale”, si sa distribution ne dépend pas du paramètre inconnu θ .

Exemple 2.6.1.

- Dans le cas d'une population $\mathcal{N}(\mu, \sigma^2)$ avec σ^2 connue, une statistique pivotale est justement :

$$\varphi = \varphi(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}.$$

- Dans le cas d'une population $\mathcal{E}(1/\theta)$, une statistique pivotale est :

$$\frac{2}{\theta} \sum_{i=1}^n X_i \sim \chi_{2n}^2. \quad (\text{Exercice})$$

C'est l'existence d'une quantité pivotale qui permet la construction d'un intervalle de confiance. En effet, le fait que la distribution de φ ne dépend plus du paramètre inconnu θ , permet de trouver deux nombres a et b , indépendants également de θ , tels que :

$$\mathbb{P}(a \leq \varphi(X_1, \dots, X_n; \theta) \leq b) = 1 - \alpha$$

où α est un nombre choisi par avance dans $[0, 1]$, normalement très petit.

Exemple 2.6.2. Si pour des valeurs fixes x_1, \dots, x_n de X , on a :

$$\psi(\theta) = \varphi(x_1, \dots, x_n; \theta)$$

est une fonction croissante de θ , alors :

$$a \leq \varphi(x_1, \dots, x_n; \theta) \leq b \Leftrightarrow a \leq \psi(\theta) \leq b \Leftrightarrow \psi^{-1}(a) \leq \theta \leq \psi^{-1}(b)$$

et si $\psi(\theta) = \varphi(x_1, \dots, x_n; \theta)$ est une fonction décroissante de θ , alors :

$$a \leq \varphi(x_1, \dots, x_n; \theta) \leq b \Leftrightarrow a \leq \psi(\theta) \leq b \Leftrightarrow \psi^{-1}(b) \leq \theta \leq \psi^{-1}(a).$$

Exemple 2.6.3. Dans l'exemple sur l'estimation d'une moyenne, ψ est la fonction

$$\psi(\mu) = \frac{(\bar{X}_n - \mu)}{\sigma_{\bar{X}_n}}$$

et ψ^{-1} est la fonction

$$\psi^{-1}(\delta) = \bar{X}_n - \delta \sigma_{\bar{X}_n}.$$

2.7 Exercices

Exercice 17. *Questions du cours.*

Répondre, en justifiant la réponse donnée, aux questions suivantes.

1. La moyenne empirique est un estimateur sans biais de la moyenne d'une loi.
2. Toute combinaison linéaire de deux estimateurs sans biais est un estimateur sans biais.
3. La variance empirique est un estimateur sans biais de la variance d'une loi.
4. Tout estimateur dont la variance tend vers zéro est convergent.
5. Si $\hat{\theta}_n$ est l'EMV pour un paramètre θ , alors $g(\hat{\theta}_n)$ est un EMV du paramètre $g(\theta)$ pour toute fonction g .

Exercice 18. Soit X_1, \dots, X_n un échantillon aléatoire tiré d'une population de moyenne μ et de variance σ^2 . Considérons l'ensemble des estimateurs de la forme :

$$\hat{\mu} = \sum_{i=1}^n a_i X_i \text{ où } a_1, \dots, a_n \text{ sont des constantes.}$$

1. Quelles conditions les constantes $a_i, i = 1, \dots, n$, devraient-elles satisfaire pour que $\hat{\mu}$ soit sans biais ?
2. Parmi tous les estimateurs sans biais de la forme $\hat{\mu} = \sum_{i=1}^n a_i X_i$, montrer que celui qui a la plus petite variance est $\bar{X}_n = \sum_{i=1}^n \frac{1}{n} X_i$.
Suggestion : Il suffit de montrer que $\sum_{i=1}^n a_i^2 \geq \frac{1}{n}$ en se servant du fait que $\sum_{i=1}^n \left(a_i - \frac{1}{n}\right)^2 \geq 0$.

Exercice 19. On doit prélever deux échantillons de deux populations indépendantes de moyennes μ_1 et μ_2 et de variances connues σ_1^2 et σ_2^2 . On a l'intention d'estimer $\mu_1 - \mu_2$ par $\bar{X}_{n_1} - \bar{Y}_{n_2}$. Le budget total est tel que les tailles n_1 et n_2 des deux échantillons doivent satisfaire $n_1 + n_2 = 100$. Comment doit-on choisir n_1 et n_2 afin que l'estimateur soit consistant ?

Exercice 20. Soit X_1, \dots, X_n un échantillon aléatoire d'une population de moyenne μ et de variance σ^2 .

1. Montrer que les estimateurs :

$$T_1(X_1, \dots, X_n) = X_1, T_2(X_1, \dots, X_n) = 2X_1 - X_2 \text{ et } T_3(X_1, \dots, X_n) = 2 \sum_{i=1}^n \frac{i}{n(n+1)} X_i$$

sont tous sans biais pour μ .

2. Calculer leurs variances.
3. Dire pourquoi \bar{X}_n est préférable aux estimateurs T_1 , T_2 et T_3 .

Exercice 21. Soit X_1, \dots, X_n un échantillon aléatoire tiré d'une population de densité

$$f(x) = \frac{x}{\theta^2} e^{-\frac{x}{\theta}} \text{ pour } x > 0, \text{ où } \theta > 0.$$

1. Montrer que l'estimateur de θ , obtenu par la méthode des moments, est $\hat{\theta}_n = \bar{X}_n/2$.
2. L'estimateur $\hat{\theta}_n$ est-il sans biais pour θ ?
3. Déterminer la variance de $\hat{\theta}_n$.
4. L'estimateur $\hat{\theta}_n$ converge-t-il, en m.q., vers θ ?
5. Calculer la borne de Frechet de $\hat{\theta}_n$.
6. L'estimateur $\hat{\theta}_n$ est-il efficace ?

Exercice 22. Soit X_1, \dots, X_n un échantillon d'une variable aléatoire X à valeurs entières, de loi définie pour tout $k \in \mathbb{N}$ par :

$$\mathbb{P}(X = k) = \frac{\theta^k}{(1 + \theta)^{k+1}} \text{ où } \theta \text{ est un paramètre positif}$$

1. Déterminer l'EMV $\hat{\theta}_n$ de θ .
2. Etude des propriétés de $\hat{\theta}_n$:
 - (a) Montrer que $\mathbb{E}(X) = \theta$ et $\text{var}(X) = \theta(\theta + 1)$.
Indication : on peut procéder par le calcul de $\mathbb{E}(X(X - 1))$.
 - (b) Calculer $\text{var}(\hat{\theta}_n)$.
 - (c) Calculer la borne de Frechet de $\hat{\theta}_n$.
 - (d) L'estimateur $\hat{\theta}_n$ est-il efficace ?

Rappels : pour $0 < x < 1$ on a :

$$\sum x^k = 1/(1-x), \sum kx^{k-1} = 1/(1-x)^2, \sum k(k-1)x^{k-2} = 2/(1-x)^3.$$

Exercice 23. Soit X une variable aléatoire qui suit la loi de Pareto de densité de probabilité :

$$f(x) = \begin{cases} \frac{\alpha-1}{x^\alpha} & \text{si } x \geq 1 \\ 0 & \text{sinon} \end{cases}$$

où $\alpha > 1$ est un paramètre inconnu.

1. Déterminer l'EMV de α construit à partir d'un échantillon X_1, \dots, X_n de cette loi.
2. Montrer qu'il est convergent.

Exercice 24. Afin d'organiser au mieux l'accueil des groupes de visiteurs d'un parc d'attractions, on note la durée X séparant l'arrivée de deux groupes successifs. A partir des observations X_1, \dots, X_n recueillies dans une journée, on retient comme modèle la loi uniforme sur $[0, \theta]$.

1. Déterminer par la méthode du maximum de vraisemblance un estimateur, $\hat{\theta}_n$, du paramètre θ .
2. On admet que $\hat{\theta}_n = M_n = \max\{X_1, \dots, X_n\}$. Calculer $\mathbb{E}(\hat{\theta}_n)$.
3. Déduire un estimateur, $\hat{\theta}'_n$, sans biais, de θ .
 - (a) Calculer $\mathbb{E}(\hat{\theta}'_n)$ et $\text{var}(\hat{\theta}'_n)$.

(b) L'estimateur $\hat{\theta}_n'$ est-il efficace ?

Exercice 25. Soit X_1, \dots, X_n un échantillon aléatoire tiré d'une population de densité de probabilité :

$$f(x) = \frac{1}{\theta} x^{\frac{1}{\theta}-1} \text{ pour tout } 0 < x < 1 \text{ où } \theta > 0 \text{ est un paramètre inconnu.}$$

1. Trouver, par la méthode des moments, un estimateur de θ .
2. Considérer les variables aléatoires $Y_i = -\ln(X_i)$. Montrer que Y_i suit une loi exponentielle et préciser son paramètre.
3. Utiliser la question 2. pour obtenir, par la méthode des moments, un nouvel estimateur $\hat{\theta}_n$ de θ . Est-il sans biais ?
4. Déterminer un intervalle de confiance pour θ .

Exercice 26. Soit n observations d'une population de densité de probabilité :

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \text{ pour tout } x > 0 \text{ où } \theta > 0 \text{ est un paramètre inconnu.}$$

1. Montrer que \bar{X}_n est un estimateur sans biais de θ et déterminer sa variance.
2. Montrer que :

$$T_n = \frac{1}{2n} \sum_{i=1}^n X_i^2 \quad \text{et} \quad S_n^{2'} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

sont tous les deux des estimateurs sans biais de θ^2 .

3. Déterminer la variance de T_n .
4. En partant du fait que :

$$Y_n = \frac{2}{\theta} \sum_{i=1}^n X_i \sim \chi_{2n}^2$$

déterminer un intervalle de confiance pour θ .

5. Montrer que $\frac{n-1}{n\bar{X}_n}$ est un estimateur sans biais de $1/\theta$.

Exercice 27. Soit X_1, \dots, X_n un échantillon de $X \sim \mathcal{N}(0, \sigma^2)$. Considérons la statistique "écart absolu moyen" :

$$D_n = \frac{1}{n} \sum_{i=1}^n |X_i|$$

1. Calculer $\mathbb{E}(D_n)$, et montrer que $T_n = \sqrt{\frac{\pi}{2}} D_n$ est un estimateur sans biais de l'écart-type σ .
 2. Comparer les estimateurs T_n^2 et $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$.
 3. Montrer que :
- $$\sqrt{n} \frac{D_n - \mathbb{E}(|X|)}{\sigma(|X|)} \rightarrow \mathcal{N}(0, 1), \text{ en loi.}$$
4. Construire un intervalle de confiance pour σ .
 5. Si l'on observe $d_{100} = 1.084$, donner un intervalle de confiance pour σ de niveau voisin de 0.95, ainsi que sa longueur.

2.8 Travaux pratiques

1. Dans le répertoire TP_L3MIASHS_Stat2 qui a été créé sur le bureau.
2. Lancer R et modifier le répertoire de travail en allant dans **Fichier -> Changer le Répertoire Courant** et en choisissant le répertoire Bureau/TP_L3MIASHS_Stat2 qui a été créé.
3. Ouvrir une fenêtre d'éditeur **Fichier -> Nouveau Script**.
4. Sauver le fichier dans le répertoire courant sous le nom TP2.R : **Fichier -> Sauver sous**
5. Pour les différentes questions, on peut utiliser un “copier-coller” à partir de ce document. *Il est fortement recommandé de saisir toutes les commandes dans la fenêtre ouverte de l'éditeur.*
Pour exécuter les commandes saisies, il suffit de les sélectionner avec la souris et d'appuyer simultanément sur les touches **Ctrl et R**.
6. Pour inclure des commentaires dans le programme, ce qui est fortement recommandé, utiliser le caractère **#**. Tout ce qui suit le caractère **#** sera négligé lors de l'exécution.
7. Penser à sauvegarder régulièrement le contenu du fichier TP2.R en appuyant sur les touches **Ctrl et S**.

2.8.1 Biais et Variance

La convergence d'un estimateur précise le comportement de celui-ci lorsque le nombre de données n tend vers l'infini. En estimation paramétrique, on dispose en général d'un échantillon de taille n fixée. Il convient de mettre en évidence certaines propriétés des statistiques construites sur un tel échantillon aléatoire. En fait, le point de vue est différent : “faire parler” le seul n -échantillon observé dont on dispose suppose de le considérer comme une réalisation particulière, aléatoire, d'un ensemble beaucoup plus vaste de n -échantillons possibles. Ainsi une statistique observée comme la moyenne d'échantillons \bar{x} doit être vue comme la réalisation sur l'échantillon observé de la variable \bar{X} qui aurait pris des valeurs différentes sur d'autres échantillons tirés au hasard.

Parmi les propriétés attendues d'un estimateur, il y a d'une part l'absence du biais de celui-ci, qui est une qualité forte, puis sa variance minimale.

Exercice 28. (Calcul de la variance empirique avec R)

La variance empirique d'une série d'observations x_1, \dots, x_n est donnée par

$$s_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

1. La fonction **var()** de R donne-t-elle la variance empirique ? Si ce n'est pas le cas, à quelle définition est associée la fonction **var()** ?
2. Programmer une fonction qui donne la variance empirique.

La variance corrigée S_n^2 est associée (contrairement à la variance empirique) à un estimateur sans biais de la variance de population. Des illustrations très pédagogiques existent dans Wonnacot & Wonnacot (1995). Nous en reprenons la Figure 2.3 pour illustrer les définitions :

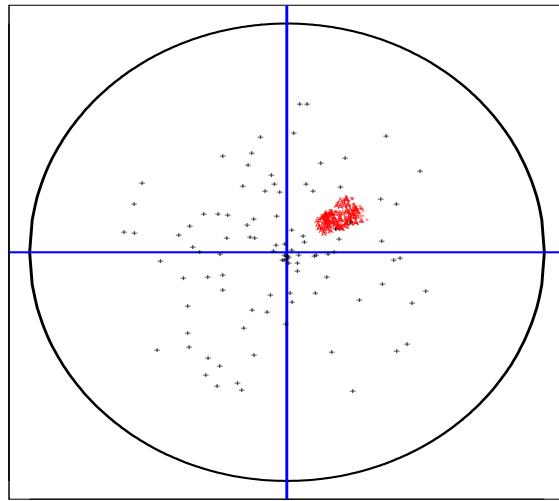


FIGURE 2.3 – Estimateur sans biais en noir, biaisé mais de faible variance en rouge

Imaginons que la valeur cible corresponde au centre du cercle. Les points noirs (resp. rouges) correspondent à 100 réalisations d'un estimateur (resp. d'un autre estimateur) de cette valeur cible. En moyenne, l'estimateur noir atteint la cible : il est sans biais. L'estimateur rouge, lui, en moyenne, se trompe : il est biaisé. En revanche sa variance est plus faible dans la mesure où les réalisations sont globalement plus proches les unes des autres.

Exercice 29. (Minimum de lois uniformes)

Soit U_1, \dots, U_n des variables aléatoires indépendantes de même loi uniforme sur $[0, 1]$. On pose $\hat{\theta}_n = \inf(U_1, \dots, U_n)$. $\hat{\theta}_n$ est-il un estimateur sans biais de la borne inférieure de l'intervalle ?

Exercice 30. (Moyenne d'un échantillon)

Montrer que pour (X_n) i.i.d. selon une loi de X telle que $E[X] = \mu$ et $var(X) = \sigma^2$, on a pour tout n :

$$E(\bar{X}_n) = \mu \quad \text{et} \quad var(\bar{X}_n) = \frac{\sigma^2}{n} \quad \text{avec} \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Dans ce cadre, la moyenne d'échantillon \bar{X}_n est un estimateur sans biais de μ . Illustrer ceci pour $X \sim \mathcal{E}(\lambda)$.

Exercice 31. (Estimateur sans biais de la variance)

On a évoqué à l'Exercice 14, deux estimateurs convergents vers σ^2 :

$$S'_n{}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Vérifier que S'^2 est un estimateur sans biais de σ^2 . Pour le montrer, considérer $X_i - \bar{X}_n = X_i - \mu + \mu - \bar{X}_n$.

On observe en Figure 2.3 que si l'estimateur représenté en noir est manifestement moins biaisé que le rouge, il est en revanche beaucoup plus variable puisque sa dispersion est beaucoup plus importante. Il est naturel de se demander lequel de ces deux estimateurs est finalement le plus satisfaisant : a-t-on plus de chances d'être proche de la valeur cible en tirant aléatoirement un point noir ou un point rouge ? Quelle est en moyenne l'estimateur le plus proche de la valeur cible ? Formellement, et en considérant l'erreur quadratique, il s'agit de savoir quel estimateur $\hat{\theta}_n$ réalise le minimum de $\mathbb{E}[(\hat{\theta}_n - \theta)^2]$ où θ est le paramètre à estimer. Cette erreur quadratique moyenne peut être écrite en fonction du biais et de la variance de l'estimateur. Plus précisément, c'est la somme du carré du biais et de la variance :

$$\mathbb{E}((\hat{\theta}_n - \theta)^2) = (\mathbb{E}(\hat{\theta}_n) - \theta)^2 + \mathbb{E}((\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))^2) = b_n^2(\hat{\theta}_n) + var(\hat{\theta}_n). \quad (2.8.1)$$

En pratique cette quantité est inconnue (tout comme le biais d'ailleurs) puisque θ est inconnu. Cependant, c'est souvent cette erreur qui est retenue comme mesure de la qualité d'un estimateur dans les études théoriques. Dans une étude simulée, les quatités en présence dans l'équation (2.8.1) peuvent être approchées. C'est l'objet de la section suivante.

2.8.2 Comparaison des estimateurs du paramètre de position

Soit X_1, \dots, X_n des variables aléatoires i.i.d. à densité. On suppose que cette densité, notée f , est symétrique par rapport à une valeur réelle θ , c'est-à-dire que

$$f(\theta + x) = f(\theta - x), \forall x \in \mathbb{R}.$$

Le but de cet exercice est d'étudier les propriétés de trois estimateurs de θ : la moyenne, la médiane et le "mid-range", définis par

$$\hat{\theta}_1 = \frac{X_1 + \dots + X_n}{n}, \quad \hat{\theta}_2 = Med_n \quad \text{et} \quad \hat{\theta}_3 = \frac{\min_i X_i + \max_i X_i}{2}.$$

1. On commence par poser $n = 1000$ et par générer n réalisations indépendantes d'une v.a. de loi gaussienne $\mathcal{N}(5, 4)$:

```
n=1000 ;
X=randn(n,mean=5,sd=2) ;
```

2. On vérifie que la répartition des éléments de X est proche de la loi gaussienne :

```
par(bg="cornsilk",lwd=2,col="darkblue")
hist(X,breaks=20,freq=F,col="cyan")
curve(dnorm(x,mean=5,sd=2),add=T)
```

Question : Quelle est la valeur de θ dans ce cas ?

3. On veut étudier les comportements des estimateurs $\hat{\theta}_1, \hat{\theta}_2$ et $\hat{\theta}_3$ lorsque $n \rightarrow \infty$.

- (a) Pour cela, on écrit la fonction suivante :

```
location_estimator=function(U,theta)
{
  n=length(U)
```

```

theta1=cumsum(U)/(1:n)
theta2=1:n
for (i in 1:n)
  theta2[i]=median(U[1:i])
end
par(mfrow=c(1,2),bg="cornsilk",lwd=2,col="darkblue")
plot(1:n,theta1,type="l",main="moyenne")
abline(h=theta,col="darkred")
plot(1:n,theta2,type="l",main="mediane")
abline(h=theta,col="darkred")
return(matrix(c(theta1,theta2),n,2))
}

```

Questions :

- Si U est un vecteur dont les coordonnées sont (u_1, \dots, u_n) , que représente le vecteur `theta1`? le vecteur `theta2`?
- A quoi sert l'option `type="l"` dans la commande `plot`?

- (b) On appelle cette fonction à l'aide des commandes

```
X=rnorm(n,mean=5,sd=2)
est=location_estimator(X,5)
```

Question : Que remarque-t-on? A quoi correspond la matrice `est`?

- (c) Compléter la fonction `location_estimator` pour qu'elle renvoie également la valeur de $\hat{\theta}_3$ (on pourra utiliser les commandes `cummax` et `cummin`). L'estimateur $\hat{\theta}_3$ est-il convergent?

4. On veut maintenant comparer les 3 estimateurs.

- (a) Pour cela, on écrit une fonction qui génère N échantillons dont chacun est de taille n et est distribué suivant la loi normale $\mathcal{N}(5, 4)$.

Pour chaque échantillon, on calcule les 3 estimateurs. Cela nous donne 3 séries numériques de N valeurs $\hat{\theta}_1^j$, $j = 1, \dots, N$, $\hat{\theta}_2^j$, $j = 1, \dots, N$ et $\hat{\theta}_3^j$, $j = 1, \dots, N$.

On trace ensuite les boxplots de ces 3 séries numériques.

```
compare_estimators=function(N,n)
{
  X=matrix(rnorm(N*n,mean=5,sd=2),N,n)
  theta1=apply(X,1,mean)
  theta2=apply(X,1,median)
  theta3=(apply(X,1,min)+apply(X,1,max))/2
  par(bg="cornsilk",lwd=2,col="darkblue")
  boxplot(theta1,theta2,theta3,col="cyan")
}
```

Question : Que fait la commande `apply`?

- (b) On appelle cette fonction : `compare_estimators(200,1000)`

Au vue de ce résultat, lequel des 3 estimateurs préfère-t-on?

- (c) On modifie la fonction `compare_estimators` en remplaçant la loi normale par la loi uniforme sur $[0, 10]$: `rnorm(N*n, 0, 10)`. Lequel des 3 estimateurs préfère-t-on dans ce cas-là ?
- (d) On modifie encore la fonction `compare_estimators` en remplaçant la loi uniforme par la loi de Cauchy : `rcauchy(N*n, location=5, scale=2)`.
- i. On observe d'abord que le troisième estimateur est incontestablement le moins bon des trois (le graphe correspondant doit être inclus dans le compte-rendu.)
 - ii. On supprime le calcul de $\hat{\theta}_3$ ainsi que l'affichage de son boxplot de la fonction `compare_estimators` et on relance la commande : `compare_estimators(200, 1000)`. Le résultat obtenu est-il en faveur de $\hat{\theta}_1$ ou $\hat{\theta}_2$?
- Résumer les réponses obtenues dans le tableau suivant :

Loi	$\mathcal{N}(5, 4)$	$\mathcal{C}(5, 2)$	$\mathcal{U}_{[0, 10]}$
Meilleur estimateur parmi $\hat{\theta}_1$, $\hat{\theta}_2$ et $\hat{\theta}_3$			

- (e) Si l'on dispose de vraies données, de loi inconnue, réparties de façon symétrique par rapport à une valeur θ , que fait-on pour estimer cette valeur ?

Exercice 32. (Régression linéaire simple)

On vérifierait dans un exercice ultérieur que l'estimateur des moindres carrés ordinaires (MCO) de la pente de la droite de régression est sans biais. Ici, on se propose d'étudier l'estimateur de cette pente de la droite défini en prenant la pente de la droite obtenue en joignant le point (x_i, y_i) associé à la plus petite valeur de X à celui associé à la plus grande.

1. Créer une fonction prenant comme paramètres des vecteurs d'observations $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ et qui donne la pente de la droite ajustée au nuage des points (x_i, y_i) , $i = 1, \dots, n$ de la manière qui est décrite ci-dessus.
2. En générant aléatoirement des points sur le même modèle qu'en MCO (notion à voir !), calculer une valeur approchée du biais de cette estimateur.
3. Comparer la variance de cet estimateur à celle de l'estimateur des moindres carrés.
4. Comparer les erreurs quadratiques moyennes de ces deux estimateurs.

2.8.3 Intervalles de confiance (IC)

IC pour la moyenne d'un échantillon gaussien : on observe un n -échantillon, x_1, \dots, x_n , distribué suivant la loi gaussien $\mathcal{N}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma > 0$ sont des paramètres inconnus. Le but est de déterminer un IC de niveau 95% (par exemple) pour μ et d'étudier ses propriétés statistiques. On sait que μ et σ^2 sont bien estimées respectivement par la moyenne empirique

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

et par la variance empirique sans biais

$$s_n'{}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

1. La théorie dit que la variable aléatoire $\hat{\theta}_n = \frac{(\bar{X} - \mu)\sqrt{n}}{S_n'}$ suit la loi de Student $t_{(n-1)}$ à $(n-1)$ degrés de liberté (cf. polycopié du cours). On cherche à avoir une confirmation empirique de ce résultat. Pour cela, on génère N échantillons de taille n de loi $\mathcal{N}(\mu, \sigma^2)$, pour des valeurs de μ et de σ^2 données. Pour chacun de ces N échantillons, on calcule la valeur de $\hat{\theta}_n$. Cela nous donne une série numérique t_n^1, \dots, t_n^N .

```
check_student=function(N,n,mu,sigma)
{
  X=matrix(rnorm(N*n,mean=mu,sd=sigma),N,n)
  t=sqrt(n)*(apply(X,1,mean)-mu)/apply(X,1,sd)
  return(t)
}
```

Si cette série est vraiment distribuée suivant la loi de Student t_{n-1} , alors son histogramme devrait être proche de la densité de la loi t_{n-1} . Pour vérifier cela, on exécute les commandes :

```
N=5000 ; n=30 ;
t=check_student(N,n,2,3)
par(bg="cornsilk",lwd=2,col="darkblue")
hist(t,breaks=30,freq=F,col="cyan")
curve(dt(x,n-1),add=T,lwd=2)
```

2. Questions :

- Déterminer les quantiles d'ordre 2.5% et 97.5% de la série numérique t_n^1, \dots, t_n^N . (On lira l'aide en ligne de la commande `quantile`.)
- Faire varier les valeurs de μ et de σ . Cela influence t-il le résultat ?
- Augmenter N jusqu'à 40000 (si les capacités de la machine le permettent) et refaire l'expérience. Quelles sont les valeurs obtenues pour les deux quantiles ?
- Soient a et b les valeurs obtenues dans la question précédente. Si $\hat{\theta}_n$ est une v.a. distribuée suivant la loi t_{n-1} , quelle est la probabilité $\mathbb{P}(\hat{\theta}_n \in [a, b])$?
- Vérifier par le calcul que $\hat{\theta}_n \in [a, b]$ équivaut à

$$\mu \in \left[\bar{X}_n - \frac{S_n}{\sqrt{n}} b, \bar{X}_n - \frac{S_n}{\sqrt{n}} a \right] \quad (2.8.2)$$

Quel est le pourcentage des échantillons (parmi les N échantillons générés) pour lesquels (2.8.2) est satisfaite ?

3. *Application* : On mesure la force de compression d'un ciment en moulant de petits cylindres et en mesurant la pression X , mesurée en kg/cm², à partir de laquelle ils se cassent. Pour 10 cylindres utilisés on observe les pressions suivantes :

19.6 19.9 20.4 19.8 20.5 21.0 18.5 19.7 18.4 19.4

et on suppose que la variable aléatoire X obéit à une loi gaussienne. On veut déterminer un IC de niveau 95% pour la moyenne de X . Pour cela,

- (a) Rentrer les données :

$$X = c(19.6, 19.9, 20.4, 19.8, 20.5, 21.0, 18.5, 19.7, 18.4, 19.4)$$

- (b) Déterminer la valeur de n et les valeurs de a et de b correspondant à cette valeur de n .
(c) En déduire l'IC demandé par la formule (2.8.2).

2.8.4 Calcul de l'EMV à l'aide de la commande `mle`

La méthode du MLE est une méthode pour trouver un estimateur (on ne sait pas s'il vérifera de bonnes propriétés, en particulier il sera souvent biaisé, mais c'est déjà un bon départ). Elle consiste à regarder pour quelle valeur du paramètre on a le plus de chance d'obtenir le résultat qu'on a observé, i.e., à maximiser la vraisemblance.

Par exemple, cherchons, avec cette méthode, un estimateur de la moyenne d'une variable suivant une loi normale de variance 1 à l'aide d'un échantillon de 5 individus (dans cet exemple, on obtient en fait la moyenne empirique).

```
# Choix de la moyenne
m <- runif(1, min=-1, max=1)
# Les n individus
n <- 5
v <- rnorm(n, mean=m)
# Calcul de la densité
N <- 1000
l <- seq(-2,2, length=N)
y <- vector()
for (i in l) {
  y <- append(y, prod(dnorm(v,mean=i)))
}
plot(y ~ l, type='l')
# Moyenne réelle
points(m, prod(dnorm(v,mean=m)), lwd=3)
# Moyenne empirique
points(mean(v), prod(dnorm(v,mean=mean(v))), col='red', lwd=3)
```

- La commande `optimize` permet de trouver un tel maximum, en dimension 1.

```
optimize(dnorm, c(-1,2), maximum=T)
$maximum
$objective
```

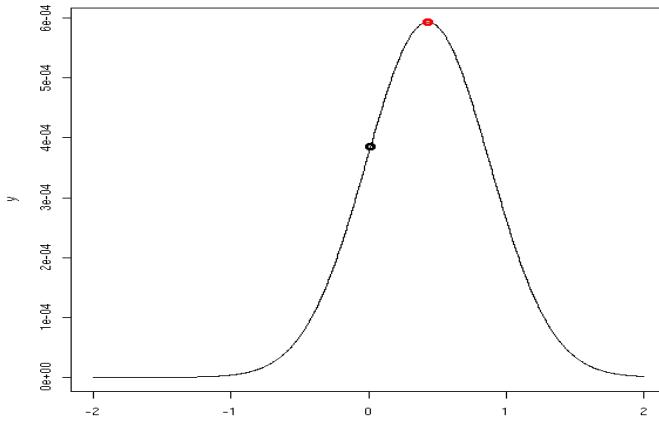


FIGURE 2.4 – Estimateur MLE

- En dimensions supérieures, on peut utiliser la commande `optim` (pour trouver un minimum).

```
f <- function (arg) {
  x <- arg[1]
  y <- arg[2]
  (x-1)^2 + (y-3.14)^2
} optim(c(0,0), f)
```

- ou la commande `nlm` : A FAIRE

On obtient :

```
$par
$value
$counts
function gradient
 59 NA
$convergence
$message
NULL
```

- On peut exiger une méthode d'optimisation particulière, par exemple le recuit simulé (Simulated Annealing, en anglais) – qui donne en fait un résultat très imprécis.

```
optim(c(0,0), f, method="SANN")
```

Exercice 33. Nous souhaitons modéliser la durée de vie d'un lave-linge (en années). Pour cela nous disposons d'un échantillon de 50 valeurs :

```
10.273 1.035 63.952 33.153 21.708 1.612 22.216 17.682 4.730 2.236 0.746 22.762 23.455
30.993 30.138 3.140 12.884 40.893 35.560 5.999 5.962 7.069 23.743 30.881 6.647 2.393
3.736 0.866 46.587 3.728 4.173 10.127 40.502 15.275 0.030 2.517 3.134 15.530 15.156
```

```
7.919 32.617 18.881 1.221 0.360 8.641 36.070 16.217 10.520 21.354 1.366
```

On commence par saisir les données dans une variable qu'on appelle X :

```
X <- c(10.273, 1.035, 63.952, 33.153, 21.708, 1.612, 22.216, 17.682, 4.730, 2.236,
0.746, 22.762, 23.455, 30.993, 30.138, 3.140, 12.884, 40.893, 35.560, 5.999, 5.962,
7.069, 23.743, 30.881, 6.647, 2.393, 3.736, 0.866, 46.587, 3.728, 4.173, 10.127,
40.502, 15.275, 0.030, 2.517, 3.134, 15.530, 15.156, 7.919, 32.617, 18.881, 1.221,
0.360, 8.641, 36.070, 16.217, 10.520, 21.354, 1.366)
```

- (a) Représenter graphiquement cette distribution.
- (b) Le graphique précédent nous suggère de modéliser la durée de vie du lave-linge par une loi exponentielle $\mathcal{E}(\theta)$. On cherche maintenant estimer le paramètre θ de cette loi en utilisant la méthode du maximum de vraisemblance.
 - i. Pour cela, on définit une fonction qui calcule la log-vraisemblance négative de la loi exponentielle : `log.vrais.neg=function(theta=1){
 if(theta>0)
 -sum(dexp(X, theta, log=TRUE))
 else NA }`
 - ii. Que fait la commande `dexp` appelée avec l'option `log=TRUE` ?
 - iii. Faut-il minimiser ou maximiser la fonction `log.vrais.neg` pour obtenir l'EMV ?
 - iv. La fonction `mle` (Maximum Likelihood Estimator) de la librairie `stats4` permet de calculer l'EMV dans un modèle paramétrique spécifié. Il faut d'abord télécharger le package `stats4` puis le charger :


```
install.packages("stats4")
library(stats4)
fit = mle(log.vrais.neg)
summary(fit)
```

 Quelle est l'estimation obtenue pour θ ?
 - v. Nous avons vu dans le cours que $\hat{\theta}_n = 1/\bar{X}_n$ est l'EMV de θ . En déduire une estimation pour θ puis la comparer l'estimation précédente.
- (c) Superposer sur le graphique de la question 1 la densité de probabilité de la loi $\mathcal{E}(\hat{\theta}_n)$.

Exercice 34. *Propriétés asymptotiques*

Un estimateur de maximum de vraisemblance n'est pas forcément unique (la vraisemblance peut avoir plusieurs maxima), ni sans biais, ni de variance minimale, ni efficace. Mais il possède d'excellentes propriétés asymptotiques, pour peu que la loi des observations vérifie les conditions de régularité déjà évoquées pour la quantité d'information.

Théorème 2.8.1. *Si les X_i sont indépendantes et de même loi dépendant d'un paramètre réel θ , cette loi vérifiant des conditions de régularité (cf. cours correspondant), on a :*

$$\sqrt{I_n(\theta)}(\hat{\theta}_n - \theta) \rightarrow \mathcal{N}(0, 1) \text{ en loi}$$

Ceci signifie que, $\hat{\theta}_n$ est asymptotiquement gaussien, sans biais et efficace.

Illustrer par des simulations la convergence en loi de l'EMV dans chacun des cas suivants :

- un échantillon aléatoire de loi $\mathcal{E}(\frac{1}{\theta})$ avec $\theta = 50$.
- un échantillon aléatoire de loi $\mathcal{P}(\lambda)$ avec $\lambda = 5$.

Exercice 35. *EMM versus EMV*

Nous avons vu que la méthode des moments et la méthode du maximum de vraisemblance donnent les mêmes résultats dans le cas de certaines lois de probabilité parmi les plus élémentaires (Poisson, exponentielle, normale, ...). En fait, dans la plupart des cas, les deux méthodes fournissent des estimateurs différents. C'est le cas par exemple de la loi uniforme ou aussi de la loi gamma. On a deux estimateurs différents pour chaque paramètre. On doit donc se demander quel est le meilleur d'entre eux.

- (a) *Loi uniforme $\mathcal{U}[0, \theta]$.* Pour ce modèle, l'estimateur obtenu par la méthode des moments est :

$$\tilde{\theta}_n = 2\bar{X}_n$$

et l'estimateur du maximum de vraisemblance est :

$$\hat{\theta}_n = \max\{X_1, \dots, X_n\}$$

Au vu des résultats obtenus, quelle méthode d'estimation préfère-t-on utiliser ?

- (b) *Loi gamma $\gamma(p, \theta)$.* Rappelons que si X_1, \dots, X_n est un échantillon aléatoire de $\gamma(p, \theta)$, alors les estimateurs de p et θ , par la méthode des moments, sont :

$$\tilde{p}_n = \frac{\bar{X}_n^2}{S_n^2} \text{ et } \tilde{\theta}_n = \frac{\bar{X}_n}{S_n^2}$$

D'autre part, l'EMV \hat{p}_n de p est solution de l'équation implicite :

$$n \ln p - n \ln \bar{X}_n - n \frac{\Gamma'(p)}{\Gamma(p)} \sum_{i=1}^n \ln(X_i) = 0$$

Il n'y a pas d'expression explicite de \hat{p}_n . Cette équation est à résoudre par des méthodes numériques. Une fois \hat{p}_n déterminé, on en déduit que $\hat{\theta}_n = \hat{p}_n / \bar{X}_n$.

- i. Simuler 1000 échantillons de taille 100 selon la loi $\gamma(10, 3)$, puis calculer pour chacun de ces échantillons la valeur prise par les estimateurs \tilde{p}_n , $\tilde{\theta}_n$, \hat{p}_n et $\hat{\theta}_n$. Pour les EMV on pourra utiliser la fonction `mle`.
- ii. Après avoir effectué les comparaisons numériques et graphiques nécessaires, commenter la pertinence de chacune de deux méthodes d'estimation : laquelle préfère-t-on utiliser ?

2.8.5 Calcul de l'EMV à l'aide de la commande `mle` pour les deux paramètres de la loi gamma

Le but de cet exercice est d'apprendre à se servir de la commande `mle` afin de calculer l'estimateur du maximum de vraisemblance (EMV) dans un modèle paramétrique spécifié. On va illustrer l'utilisation de cette commande sur l'exemple de la loi Gamma.

Rappelons qu'une variable aléatoire X suit une loi Gamma de paramètre de forme $a > 0$ et de paramètre d'échelle $\sigma > 0$ si X admet la densité de probabilité :

$$f(x, a, \sigma) = \frac{x^{a-1}}{\sigma^a \Gamma(a)} e^{-x/\sigma} \mathbb{1}_{]0, +\infty[}(x),$$

où $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ désigne la fonction Gamma.

Les 50 valeurs suivantes ont été générées par une loi gamma :

```
39.878 29.137 57.321 39.140 66.776 48.028 42.325 31.200 38.632 42.914 60.969
22.076 52.446 45.257 42.626 62.504 22.684 69.196 42.383 61.339 45.803 74.707
33.048 72.423 43.670 65.279 42.714 59.785 101.742 59.641 44.749 44.161 58.488
46.448 25.280 67.619 66.846 80.208 98.492 41.149 40.395 22.220 34.628 77.768
48.161 48.909 66.267
```

On cherche

- à estimer les paramètres a et σ qui ont été utilisés pour générer ces valeurs,
- et à fournir un IC pour chacun de ces paramètres.

(a) Pour cela, on commence par saisir les données dans une variable qu'on appelle X :

```
X = c(77.551, 45.195, 50.626, 39.878, 29.137, 57.321, 39.140, 66.776, 48.028, 42.325, 31.200,
38.632, 42.914, 60.969, 22.076, 52.446, 45.257, 42.626, 62.504, 22.684, 69.196, 42.383, 61.339,
45.803, 74.707, 33.048, 72.423, 43.670, 65.279, 42.714, 59.785, 101.742, 59.641, 44.749, 44.161,
58.488, 46.448, 25.280, 67.619, 66.846, 80.208, 98.492, 41.149, 40.395, 22.220, 34.628, 77.768,
48.161, 48.909, 66.267)
```

(b) On définit une fonction qui calcule la log-vraisemblance négative de la loi Gamma :

```
ll=function(a=1, sigma=1)
```

```
{
  if(a > 0 && sigma > 0)
    -sum(dgamma(X, shape=a, scale=sigma, log=TRUE))
  else
    NA
}
```

- i. Que fait la commande `dgamma` appelée avec l'option `log=TRUE` ?
 - ii. Etant donné un vecteur $X = (x_1, \dots, x_n)$ et deux valeurs réelles a et σ , quelle est la valeur renournée par `ll(a,sigma)` ?
 - iii. Faut-il minimiser ou maximiser la fonction `ll` pour obtenir l'EMV ?
- (c) Afin d'obtenir l'EMV à l'aide de la commande `mle` (Maximum Likelihood Estimator), il faut d'abord charger le package `stats4` :
- ```
library(stats4)
fit = mle(ll)
summary(fit)
```
- Quelles sont les estimations obtenues pour  $a$  et  $\sigma$  ?
- (d) La commande `mle` ne fait pas que calculer l'EMV, elle peut aussi être utilisée pour déterminer des intervalles de confiance pour les paramètres ainsi que pour calculer la matrice de covariance limite :
- ```
vcov(fit)
par(mfrow=c(2,1),bg="cornsilk",col="blue",lwd=2)
plot(profile(fit), absVal=FALSE)
confint(fit,level=0.95)
```
- Quel est l'IC de niveau 95% pour le paramètre a ? 85% pour le paramètre σ ?
- (e) On peut même afficher la surface représentée par la log-vraisemblance négative. Cela peut se faire en utilisant les commandes `persp`, `image` et `contour` comme dans l'exemple suivant :

```

# Calcul des valeurs de la log-vraisemblance
K=80
x=(1 :K)/4; y=(1 :K)/4; z=c();
for (i in 1 :length(x))
  for (j in 1 :length(y))
    z=c(z,ll(x[i],y[j]))
# Transformation des valeurs calculées
z=matrix(z,length(x),length(y))
z=log(0.001+((z-min(z))/(max(z)-min(z))))
# Le contenu des 7 lignes suivantes peut etre utilisé comme une boite noire
nrz <- nrow(z)
ncz <- ncol(z)
jet.colors <- colorRampPalette( c("blue", "green") )
nbcoll <- 100
color <- jet.colors(nbcoll)
zfacet <- z[-1, -1] + z[-1, -ncz] + z[-nrz, -1] + z[-nrz, -ncz]
facetcol <- cut(zfacet, nbcoll)
# Visualisation des résultats
par(bg="cornsilk",lwd=1,mfrow=c(1,2))
image(x,y,z,col = cm.colors(50))
contour(x,y,z,add=T,col="darkred")
persp(x,y,z,ticktype="detailed",expand=0.5,col=color[facetcol],shade=0.4)

```

La surface affichée représente-t-elle la log-vraisemblance négative ou une transformation de celle-ci ? Donner la formule de la transformation le cas échéant et justifier son utilisation (on essaiera d'exécuter les commandes ci-dessus sans utiliser la transformation).

- (f) Afin de vérifier que les commandes ci-dessus fournissent vraiment une estimation des paramètres (a, σ) de la loi Gamma, effectuer le test suivant.
 - i. Générer $n = 1000$ valeurs aléatoires de la loi Gamma de paramètres $a = 5$ et $\sigma = 3$.
 - ii. Déterminer une estimation de a et de σ .
 - iii. Déterminer un IC de niveau 95% pour a et pour σ .
 - iv. Vérifier que les estimations obtenues sont proches des vraies valeurs.
- (g) Lors de la définition de la fonction `ll` dans la question 2, nous avons utilisé les valeurs initiales $a = 1$ et $\sigma = 1$. La nécessité de donner des valeurs initiales aux paramètres vient du fait que le calcul (approximatif) de l'EMV est fait par une méthode d'optimisation du type “descente de gradient”. La log-vraisemblance négative de la loi Gamma n'étant pas convexe, le point de minimum trouvé pourrait dépendre de la valeur initiale. Etudier la dépendance des valeurs initiales de l'EMV calculé ci-dessus en faisant varier les valeurs initiales et en observerant ce qui se passe avec l'EMV.
- (h) *Question pour le compte-rendu :* nous avons un échantillon de $n = 99$ valeurs réelles qui semblent être distribuées suivant une loi de Cauchy de paramètre de position a et de

paramètre d'échelle s inconnus. Ces valeurs sont

−7.54, 82.51, 14.27, 3.96, 189.98, 17.20, −20.07, 52.66, 93.47, −33.57, 13.13,
 −1.26, 12.69, 53.33, 2.85, −7.25, 13.30, −5.67, −38.99, 24.24, 4.17, 12.30,
 21.59, −6.70, 1.24, 13.91, 30.24, 3.35, 6.45, −26.22, 72.65, 10.12, −1.64,
 21.49, 391.11, 26.53, 146.60, 2.11, 5.84, 14.25, 7.17, 4.96, −9.55, 7.89,
 −2.31, 91.11, 8.39, 6.23, 25.45, 9.36, 102.44, −7.28, −40.02, −8.86, 14.11,
 6.84, −11.15, −6.67, −84.82, −241.41, −0.14, −72.95, 21.09, 53.47, −3.80,
 −10.64, 19.71, 45.89, −124.30, −2.02, −1.67, 7.81, −9.76, 6.25, 16.68, 8.88,
 32.14, 1.29, −10.00, −5.03, −66.77, 12.85, 15.32, 31.27, 6.59, 3.92, 8.61,
 15.38, −1.34, 14.11, 10.53, 2.35, −94.19, 16.45, 2.97, 12.26, 4.15, 10.63, 5.47

- i. Ecrire une fonction qui calcule l'EMV pour un échantillon distribué selon la loi de Cauchy $\mathcal{C}(a, s)$.
- ii. Donner une estimation des paramètres a et s , ainsi que des IC de niveau 95% pour ces deux paramètres.
- iii. Les valeurs obtenues, dépendent-elles des valeurs initiales données à la log-vraisemblance ? Expliquer le résultat.
- iv. Afficher la surface de la log-vraisemblance (éventuellement transformée par une transformation croissante qui améliore le résultat visuel).

Chapitre 3

Annexes : Lois utiles en statistique

Les notions présentées dans ces annexes doivent être acquises. En effet, celles-ci font partie du programme de L2 MIASHS. Pour les étudiants qui arrivent d'autres formations ou qui ne sont pas très familiers avec ces notions, sont invités à travailler sérieusement ces annexes. Bien entendu, cela ne fera pas l'objet d'une séance de travaux pratiques.

1. Dans le répertoire TP_L3MIASHS_Stat2 qui a été créé sur le bureau.
2. Lancer ensuite R et modifier le répertoire de travail en allant dans **Fichier -> Changer le Répertoire Courant** et en choisissant le répertoire Bureau/TP_L3MIASHS_Stat2 qui a été créé.
3. Ouvrir une fenêtre d'éditeur **Fichier -> Nouveau Script**.
4. Sauver le fichier dans le répertoire courant sous le nom TP0.R : **Fichier -> Sauver sous**
5. Pour les différentes questions, on peut utiliser un “copier-coller” à partir de ce document. *Il est fortement recommandé de saisir toutes les commandes dans la fenêtre ouverte de l'éditeur.* Pour exécuter les commandes saisies, il suffit de les sélectionner avec la souris et d'appuyer simultanément sur les touches **Ctrl et R**.
6. Pour inclure des commentaires dans le programme, ce qui est fortement recommandé, utiliser le caractère **#**. Tout ce qui suit le caractère **#** sera négligé lors de l'exécution.
7. Penser à sauvegarder régulièrement le contenu du fichier TP3.R en appuyant sur les touches **Ctrl et S**.

Le TCL affirme que, quelque soit la loi du caractère étudié dans la population, la moyenne d'une suite de variables aléatoires i.i.d. tirées dans cette population suit approximativement une loi normale pour n assez grand.

Pour des variables i.i.d. selon une loi normale, la moyenne d'échantillon suit exactement une loi normale, et ce quelque soit la valeur de n . Plus précisément, si X_1, \dots, X_n suit une loi $\mathcal{N}(\mu, \sigma^2)$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

De nombreux estimateurs sont fondés sur des calculs de moyenne. Le fait de supposer la normalité de la loi, du caractère étudié, dans la population facilite donc les calculs et permet de travailler avec des égalités et non plus avec des approximations. Cette hypothèse n'est pas trop gênante, en pratique, dès que les tailles d'échantillons sont suffisantes. La loi normale est donc très utile aussi bien dans la pratique que pour l'étude des propriétés théoriques des estimateurs. D'autant plus que d'autres lois, comme la loi de χ^2 ou la loi de Student (et par transitivité la loi de Fisher) se déduisent de la loi normale et sont utiles en pratique.

3.1 Loi normale

Exercice 36. (Autour de la loi normale, gaussienne ou de Laplace-Gauss)

1. Représenter la densité de la loi normale de paramètres (μ, σ^2) en utilisant les fonctions pré-définies dans R.
2. Retrouver la courbe de la densité en représentant la fonction

$$x \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

3. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. En utilisant le logiciel R, déterminer

$$\mathbb{P}(X \in [\mu - \sigma, \mu + \sigma]), \mathbb{P}(X \in [\mu - 2\sigma, \mu + 2\sigma]) \text{ et } \mathbb{P}(X \in [\mu - 3\sigma, \mu + 3\sigma]).$$

Corrigé.

1. Via les codes suivants, on obtient les courbes de la fonction de densité de probabilité et de la fonction de répartition (cf. Figure 3.1) :

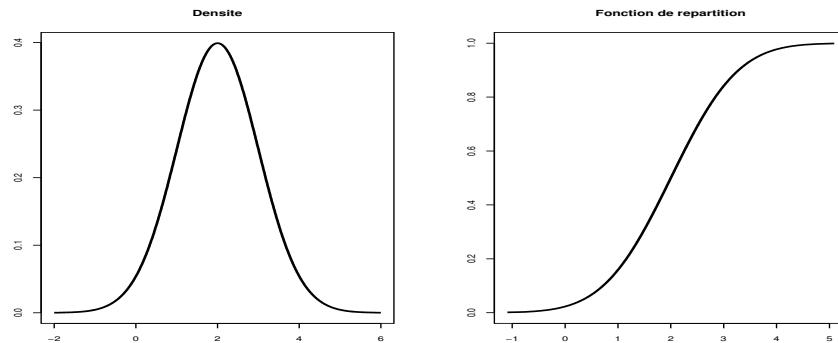


FIGURE 3.1 – Loi normale : densité de probabilité (à gauche) et fonction de répartition (à droite)

```
mu <- 2
sigma <- 1
```

```

par(mfrow=c(1,2))
grille.x <- seq(mu-4*sigma,mu+4*sigma,length=1000)
plot(grille.x,dnorm(grille.x,mean=mu,sd=sigma),type="l",lwd=2,main="Densite")
grille.q <- seq(0,1,length=1000)
plot(qnorm(grille.q,mean=mu,sd=sigma,lower.tail = TRUE),
+grille.q,type="l",lwd=2,main="Fonction de repartition")

```

2. On retrouve la partie de gauche de la Figure 3.1 en définissant et représentant la fonction proposée :

```

N <- function(x,mu,sigma) {
  (1/sqrt(2*pi*sigma^2))*exp(-0.5*(x-mu)^2/sigma^2)
}
plot(grille.x,N(grille.x,mu,sigma),type="l",lwd=2)

```

3. On retrouve les valeurs classiques :

$$\begin{aligned}\mathbb{P}(X \in [\mu - \sigma, \mu + \sigma]) &\approx 0.68, \\ \mathbb{P}(X \in [\mu - 2\sigma, \mu + 2\sigma]) &\approx 0.95 \\ \mathbb{P}(X \in [\mu - 3\sigma, \mu + 3\sigma]) &\approx 0.99\end{aligned}$$

de la manière suivante :

```

pnorm(mu+sigma,mean=mu,sd=sigma)-pnorm(mu-sigma,mean=mu,sd=sigma)
[1] 0.6826895
pnorm(mu+sigma,mean=mu,sd=sigma)-pnorm(mu-sigma,mean=mu,sd=sigma)
[1] 0.6826895
pnorm(mu+2*sigma,mean=mu,sd=sigma)-pnorm(mu-2*sigma,mean=mu,sd=sigma)
[1] 0.9544997
pnorm(mu+3*sigma,mean=mu,sd=sigma)-pnorm(mu-3*sigma,mean=mu,sd=sigma)
[1] 0.9973002

```

3.2 Loi de chi-deux : χ^2

Sous l'hypothèse de normalité, la variable aléatoire \bar{X}_n est distribuée selon une loi normale. Il est naturel de se demander, que sous cette hypothèse, quelle loi suivrait l'estimateur de la variance S_n' . Celle-ci dérive de la loi du χ^2 , qui est la loi de la somme des carrés de lois normales indépendantes, et qui est associée à la loi d'une variance en général.

Définition 3.2.1. Soit U_1, \dots, U_k des v. a. i.i.d. à $\mathcal{N}(0, 1)$. On appelle la loi du chi-deux à k degrés de liberté χ_k^2 , la loi de la v. a.

$$\sum_{i=1}^k U_i^2$$

et on a

$$\mathbb{E}(\chi_k^2) = k \quad \text{et} \quad \text{var}(\chi_k^2) = 2k.$$

Définition 3.2.2. La densité de probabilité de la loi de chi-deux à k degrés de liberté χ_k^2 :

$$f(x) = \frac{(1/2)^{\frac{k}{2}}}{\Gamma(\frac{k}{2})} e^{-x/2} x^{\frac{k}{2}-1} \text{ pour } x \geq 0$$

Une illustration de cette loi fait l'objet de l'Exercice 37.

Quelle est alors la loi de $S_n'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ sous l'hypothèse $X \sim \mathcal{N}(\mu, \sigma^2)$?

Il est clair d'après la définition que

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2$$

Cependant, les termes $(X_i - \bar{X}_n)^2$ constitutifs de $S_n'^2$ ne sont pas indépendants et ils ne suivent pas une loi normale centrée-réduite. En revanche, on peut montrer (cf. Saporta (2006), par exemple) que

$$(n-1) \frac{S_n'^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 \sim \chi_{n-1}^2$$

Cela traduit le fait que les n termes $\frac{X_i - \bar{X}_n}{\sigma}$ "équivalent" à $(n-1)$ termes indépendants distribués selon une loi normale centrée-réduite.

Exercice 37. (Loi de χ^2)

Représenter la courbe de la densité, de la fonction de répartition ainsi que la distribution de nombres simulés pour la loi χ_n^2 .

Corrigé. Nous représentons les choses avec $p = 10$ pour obtenir la Figure 3.2 :

```
par(mfrow=c(1,3))
ddl <- 10
grille.x <- seq(0,3*ddl,by=0.01)
plot(grille.x,dchisq(grille.x,ddl),type="l",main=paste(c("df=",ddl)))
plot(grille.x,pchisq(grille.x,ddl),type="l",main=paste(c("df=",ddl)))
n <- 1000
hist(rchisq(n,ddl),nclass=30,freq=FALSE,main=paste(c("df=",ddl)))
```

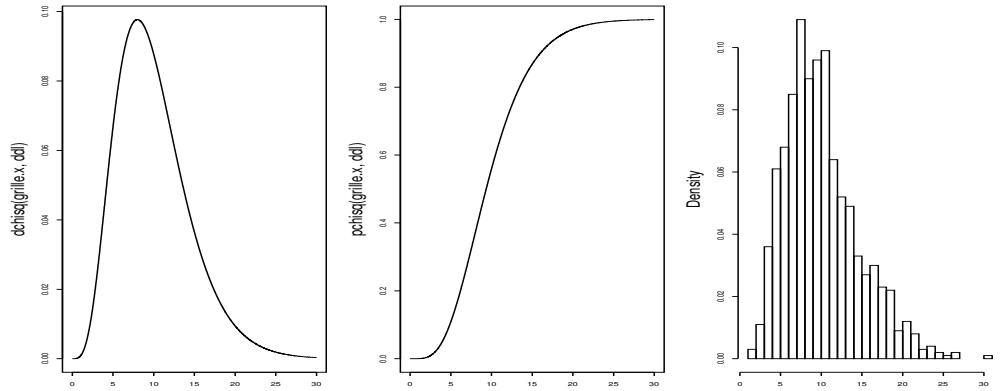
Exercice 38. (Distribution de la variance échantillonnale)

Générer des échantillons aléatoires X_1, \dots, X_n de taille n , les X_i étant i.i.d. selon une loi $\mathcal{N}(\mu, \sigma^2)$.

Représenter les distributions de $S_n'^2$ et de $(n-1) \frac{S_n'^2}{\sigma^2}$.

Corrigé. On obtient la figure Figure 3.3 avec les commandes suivantes :

```
K <- 10000 # nombre d'échantillons
n <- 10 # taille des échantillons
mu <- 0
```



```

sigma2 <- 1
S2<-apply(matrix(rnorm(n*K,mean=mu,sd=sqrt(sigma2)),ncol=n),FUN=var,MARGIN=1)
hist(S2,nclass=30,freq=FALSE)

hist((n-1)*S2/sigma2,nclass=30,freq=FALSE)
abline(v=(n-1),col="red",lwd=2)
abline(v=mean((n-1)*S2/sigma2),col="green")

```

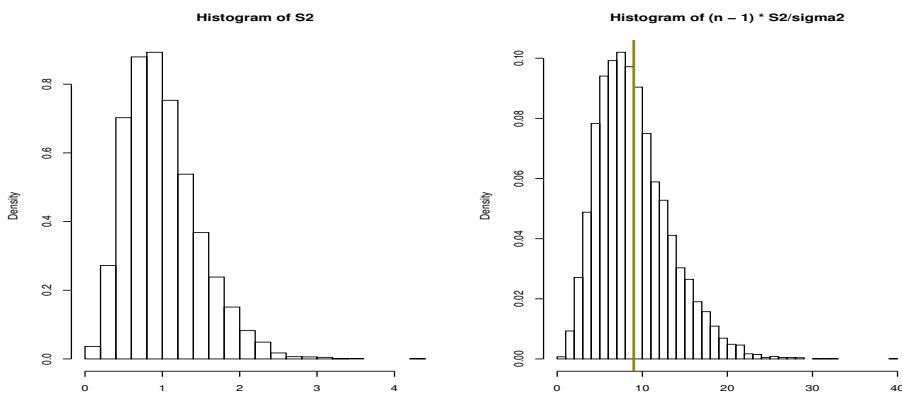


FIGURE 3.3 – Distribution de $S_n^{2'}$ et de $(n - 1) S_n^{2'}/\sigma^2$

3.3 Loi de Student

On a vu que la moyenne d'échantillon \bar{X} est un estimateur sans biais de la moyenne μ inconnue dans la population. On sait que si les X_i sont i.i.d. selon une loi $\mathcal{N}(\mu, \sigma^2)$ alors $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$. Ceci est équivaut à

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1). \quad (3.3.1)$$

Pour estimer μ , cette variable aléatoire est peu utile en pratique car σ est bien souvent inconnue également. Dès lors, il est naturel de travailler avec la variable aléatoire obtenue en remplaçant dans (3.3.1), σ par S'_n qui en est un estimateur :

$$T_n = \frac{\bar{X} - \mu}{S'_n / \sqrt{n}} \quad (3.3.2)$$

Le dénominateur étant cette fois lui aussi aléatoire. Il est clair que cette variable ne suit pas une loi normale centrée-réduite. La variable T_n suit une loi dite de Student.

Définition 3.3.1. La loi de Student à p degrés de liberté est la loi du rapport, indépendant, d'une loi normale centrée-réduite et de la racine d'une loi de χ^2 divisé par son degré de liberté p .

Cette définition peut sembler alambiquée. Il faut comprendre qu'elle se formule de cette manière de sorte à correspondre, précisément, à des v.a. comme celles définies par l'équation (3.3.2), qui sont des variables/lois utiles en pratique. En effet, on peut écrire T_n sous la forme suivante

$$T_n = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \cdot \frac{1}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2}} \quad (3.3.3)$$

On fait ainsi apparaître $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ qui suit une loi normale centrée-réduite et $\sqrt{\frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2}$ qui suit une loi de χ^2 à $(n-1)$ degrés de liberté. On admet l'indépendance (cf. Cornillon & Matzner-Lober (2011), par exemple) et on est bien dans le cadre de la définition.

Exercice 39. (Simulation d'une variable suivant une loi de Student)

1. En revenant à la définition, simuler une variable aléatoire suivant une loi de Student.
2. Représenter cette distribution.

Corrigé. Les codes suivants donnent la Figure 3.4.

```
K <- 10000 #nombre de valeurs simulées
p <- 10 #ddl
N <- rnorm(K,mean=0,sd=1) #numérateur
D <- sqrt(rchisq(K, df=p)/p) #dénominateur
T <- N/D
hist(T,freq=FALSE,nclass=20)
```

Exercice 40. (Comparaison avec la loi normale)

En utilisant les fonctions pré-définies de R, représenter sur un même graphe les densités d'une loi de Student à 2, 5, 10 et 50 degrés de liberté ainsi que celle de la loi normale centrée-réduite.

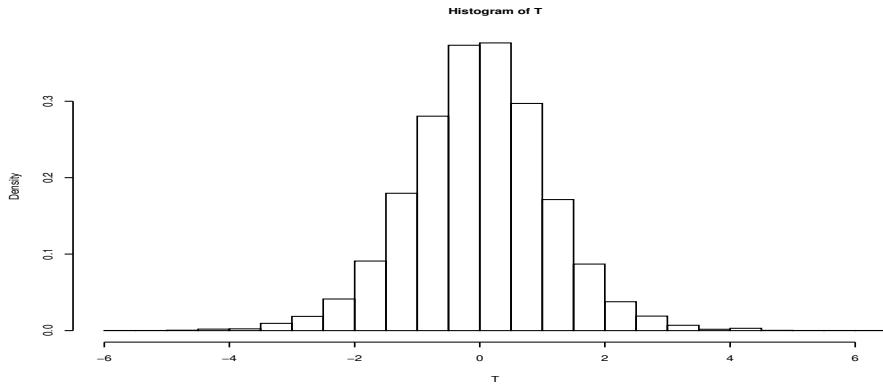


FIGURE 3.4 – Simulation d'une loi de Student

Corrigé. On observe (cf. Figure 3.5) que la densité d'une loi de Student se rapproche de celle d'une loi normale centrée-réduite. L'approximation étant d'autant meilleure que le degré de liberté est grand.

```
grille.x <- seq(-4,4,by=0.01)
plot(grille.x,dnorm(grille.x,mean=0,sd=1),type="l",lwd=2)
lines(grille.x,dt(grille.x,df=2),col="brown")
lines(grille.x,dt(grille.x,df=5),col="blue")
lines(grille.x,dt(grille.x,df=10),col="green")
lines(grille.x,dt(grille.x,df=50),col="red")
```

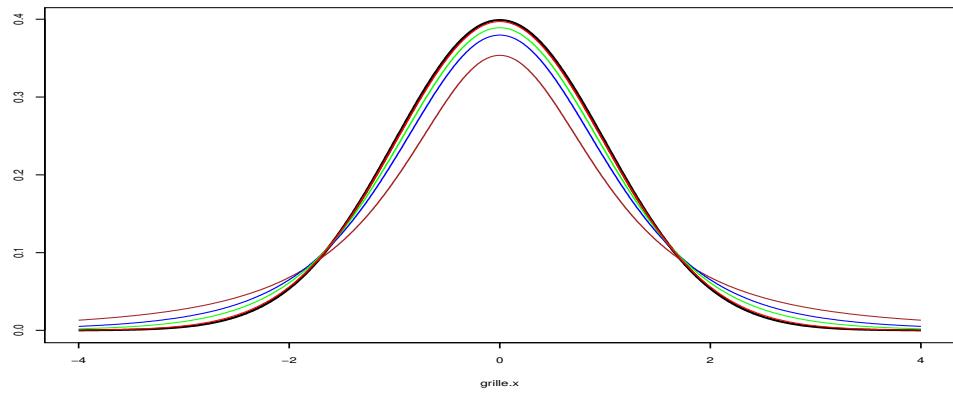


FIGURE 3.5 – Loi normale versus loi de Student

Exercice 41. (Comparaison avec la loi normale (suite))

Comparer des quantiles classiques pour la loi normale centrée-réduite et la loi de Student.

Corrigé. Notons u_α le quantile d'ordre α de la loi normale centrée-réduite $\mathbb{P}(X < u_\alpha) = \alpha$. On a par exemple les quantiles classiques : $u_{0.025} = -1.96$, $u_{0.5} = 0$ et $u_{0.975} = 1.96$. On les retrouve ainsi :

```
qnorm(p=0.025,mean=0,sd=1)
[1] -1.959964
qnorm(p=0.5,mean=0,sd=1)
[1] 0
qnorm(p=0.975,mean=0,sd=1)
[1] 1.959964
```

Du fait de la symétrie de la loi, on peut écrire :

$$\mathbb{P}(-u_{1-\alpha/2} < X < u_{1-\alpha/2}) = 1 - \alpha.$$

Cela correspond au schéma de la Figure 3.6.

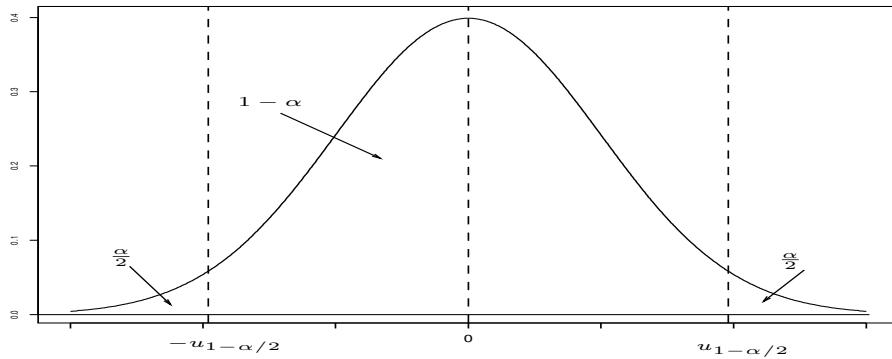


FIGURE 3.6 – Quantiles de la loi normale

3.4 Loi de Fisher

Le loi de Fisher se déduit des lois de chi-deux et de la loi de Student. Elle est très utile dans les tests statistiques. On la rencontre quand on compare des variances (en particulier dans l'ANOVA).

```
curve(df(x,1,1), xlim=c(0,2), ylim=c(0,.8), lty=2)
curve(df(x,3,1), add=T)
curve(df(x,6,1), add=T, lwd=3)
curve(df(x,3,3), add=T, col='red')
curve(df(x,6,3), add=T, lwd=3, col='red')
curve(df(x,3,6), add=T, col='blue')
curve(df(x,6,6), add=T, lwd=3, col='blue')
title(main="Densit de la loi de Fisher")
legend(par('usr')[2], par('usr')[4], xjust=1, c('df=(1,1)', 'df=(3,1)', 'df=(6,1)',
```

```
'df=(3,3)', 'df=(6,3)', 'df=(3,6)', 'df=(6,6)'), lwd=c(1,1,3,1,3,1,3),
lty=c(2,1,1,1,1,1), col=c(par("fg"), par("fg"), par("fg"),
'red', 'red', 'blue',
'blue')))
```

Pour la définir on se donne d'abord X et Y deux v.a.r. indépendantes. Illustrer les résultats ci-dessous avec la commande `qqplot`.

1. Caractérisation de la loi du chi-deux χ_2^2 : Si $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$, alors

$$X^2 + Y^2 \sim \chi_2^2$$

2. Caractérisation de la loi de Student t_ν : Si $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi_p^2$, alors

$$\frac{X}{\sqrt{Y}} \sim t_\nu$$

Prendre $\nu = 3.9$.

3. Caractérisation de la loi de Fisher $\mathcal{F}(\nu_1, \nu_2)$: Si $X \sim \chi_{\nu_1}^2$ et $Y \sim \chi_{\nu_2}^2$, alors

$$\frac{X/\nu_1}{Y/\nu_2} \sim \mathcal{F}(\nu_1, \nu_2)$$

Prendre $(\nu_1, \nu_2) = (2.1, 8.3)$.

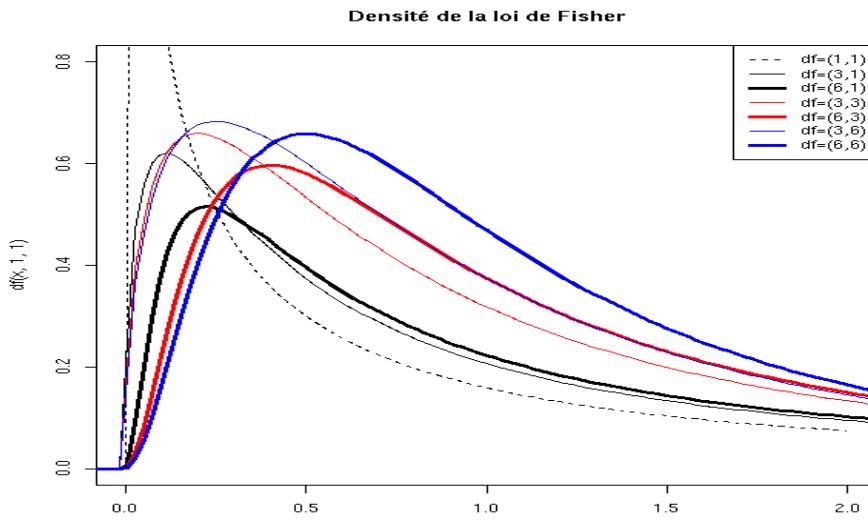


FIGURE 3.7 – Densité de la loi de Fisher en fonction de (ν_1, ν_2)

3.5 La loi gamma

Soient $\theta > 0$ et $p > 0$. On dit qu'une variable aléatoire X suit la loi gamma de paramètres θ et p si, sa densité de probabilité est de la forme :

$$f(x) = \frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1} \text{ pour } x \geq 0$$

Donc $E(X) = p/\theta$ et $var(X) = p/\theta^2$.

Nous exposons ci-dessous quelques propriétés de cette loi.

1. Si $X \sim \gamma(p, \theta)$ alors $Y = \theta X \sim \gamma(p, 1)$ qu'on note $\gamma(p)$.
2. Si les deux variables aléatoires X et Y sont indépendantes et $X \sim \gamma(p, \theta)$ et $Y \sim \gamma(q, \theta)$ alors $X + Y \sim \gamma(p+q, \theta)$ qu'on note $\gamma(p+q, \theta)$.
3. Si $p = 1$ alors $\gamma(1, \theta) = \mathcal{E}(\theta)$.
4. $\gamma(n/2, 1/2) = \chi_n^2$.

Bibliographie

- [1] Genon-Catalot V. Duhamel C. et Meyre T. Cottrell, M., *Exercices de probabilités : Licence, maîtrise, écoles d'ingénieurs.*
- [2] Y (1993) Dodge, *Statistique, dictionnaire encyclopédique*, Dunod, Paris.
- [3] J.J. (1997) Drosbeke, *Elément de statistique*, Ellipses, Paris.
- [4] D. Ghorbanzadeh, *Probabilités : exercices corrigés*, Edt Technip.
- [5] C. (2000) Goldfarb, B. et Pardoux, *Introduction à la méthode statistique*, Dunod, Paris.
- [6] V. Hurlin, C. et Mignon, *Statistique et probabilités : économie-gestion*, Edt Dunod.
- [7] Maubourguet Tournieret Lacaze, Mailhes, *Probabilités et statistique appliquées : résumé de cours et illustration*, Edt CEPADUES, collection Polytechnique.
- [8] J.P. Lecoutre, *Probabilités : exercices corrigés avec rappels de cours*, Edt Masson.
- [9] _____, *Statistique et probabilités - 6e éd. cours et exercices corrigés*, Eco Sup, Dunod.
- [10] _____, *Td statistique et probabilités : Qcm et exercices corrigés*, Edt Dunod.
- [11] J.P. (2002) Lecoutre, *Statistique et probabilités. travaux dirigés*, Dunod, Paris.
- [12] A. Ruegg, *Statistiques et probabilités*, Edt Presses Polytechniques et universitaires romandes.
- [13] M. (1994) Tenehaus, *Méthodes statistiques en gestion*, Dunod, Paris.
- [14] R. Veysseyre, *Statistiques et probabilités pour l'ingénieur*, Edt Dunod.

- [15] E. Vigneron, C. et Elisabeth Logak, *Probabilités et statistiques 1.*, Edt Diderot Editeur, Arts et sciences.