



UNIVERSITÉ DE NANTES

Statistiques Inférentielles

X32M010

Licence de Mathématiques, 3ème année

Erwan BRUGALLÉ

03/05/2020

Erwan Brugallé, Université de Nantes, Laboratoire de Mathématiques Jean Leray, 2 rue de la Houssinière, F-44322 Nantes Cedex 3, France
Email : erwan.brugalle@math.cnrs.fr

Table des matières

Préambule	5
1 Estimation ponctuelle	9
1.1 Définitions	9
1.2 Digression : preuve du théorème central limite	16
1.3 Méthode des moments	18
1.4 Maximum de vraisemblance	24
1.5 Exercices	34
2 Intervalles de confiance	39
2.1 Quantiles	39
2.2 Intervalles de confiance	44
2.3 Méthode du pivot	46
2.4 Intervalles de confiance asymptotiques	53
2.5 Exercices	57
3 Tests statistiques	61
3.1 Principes généraux	62
3.2 Probabilité critique, ou p -valeur	68
3.3 Construction d'un test à partir d'un intervalle de confiance	70
3.4 Test de Student	72
3.5 Test d'adéquation du χ^2	74
3.6 Test d'indépendance du χ^2	75
3.7 Exercices	76
Bibliographie	79

Préambule

Ce cours est une introduction élémentaire aux statistiques inférentielles. Commençons par un exemple simple mais emblématique de problème que l'on y traite.

On veut déterminer la proportion θ de la population nantaise qui aime le café. On interroge pour cela 1000 personnes, et on obtient comme résultat 600 *oui* et 400 *non*. On estime alors que

$$\theta \simeq \frac{600}{1000} = 0,6.$$

Mais quel sens à donner au signe \simeq ? A-t-on vraiment $\theta \simeq 0,6$? Peut-on espérer un résultat différent si on interroge 500 ou 2000 personnes? Existe-t-il des méthodes moins naïve pour “calculer” θ ? C'est ce type de questions auquel s'attaquent les statistiques inférentielles.

La première chose à faire est de *modéliser le problème*, c'est à dire traduire notre problème de la vie réelle¹ en un problème mathématique. Dans notre exemple, on interprète le sondage des 1000 personnes comme le résultat de l'observation ω de 1000 variables aléatoires X_1, \dots, X_{1000} indépendantes suivant une même loi de Bernoulli de paramètre θ . On rappelle qu'une variable aléatoire X suit une loi de Bernoulli de paramètre $\theta \in [0; 1]$, abrégé en $X \rightsquigarrow Ber(\theta)$, si

$$P(X = 0) = 1 - \theta \quad \text{et} \quad P(X = 1) = \theta.$$

Ici $X_j(\omega) = 1$ si la j -ème personne aime le café, et $X_j(\omega) = 0$ sinon. Notre estimation précédente de θ s'écrit alors

$$\theta \simeq \frac{1}{1000} \sum_{j=1}^{1000} X_j(\omega) = \frac{600}{1000}.$$

Dans ce cadre mathématique, cette approximation se justifie par la loi forte des grands nombres qui nous dit que presque sûrement,

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow[n \rightarrow +\infty]{} E(X_1) = \theta.$$

1. Pour autant que cela ait un sens.

En revenant à notre problème de départ, on peut ainsi conclure qu'approximer θ par 0,6 n'est pas totalement saugrenu.

C'est une première étape. Le bon sens et la loi forte des grands nombres nous disent que plus nous interrogeons de nantais, plus nous connaissons θ précisément. Mais quel est le bon compromis entre moins de travail et plus de précision ? Est ce qu'interroger 500 personnes aurait suffi, ou au contraire est ce que 2000 personnes est un seuil minimal pour avoir un résultat significatif ? Autrement dit, à quelle *vitesse* $\frac{1}{n} \sum_{j=1}^n X_j$ converge-t-elle vers θ ? Une piste pour répondre est donnée par un autre théorème célèbre de probabilités : le théorème central limite. N'allons pas plus loin dans les détails techniques ici, mais retenons ceci : la loi forte des grands nombres et le théorème central limite sont deux piliers des statistiques inférentielles, et ce cours constitue essentiellement une variation musicale sur la combinaison de ces deux résultats.

Le but de ce cours est de pousser (un peu) plus loin le formalisme qui nous a permis de "résoudre" le problème d'estimer la proportion de la population nantaise aimant le café. Posons pour cela dès à présent le cadre mathématique dans lequel nous nous placerons dans toute la suite.

Cadre général : On observe un vecteur (x_1, \dots, x_n) de n nombres réels, réalisation de n variables aléatoires X_1, \dots, X_n indépendantes et identiquement distribuées de loi inconnue. On cherche alors, sur la base des observations (x_1, \dots, x_n) , à déterminer certaines caractéristiques de la loi des X_j , comme par exemple l'espérance ou la variance.

Définition. Le n -uplet (X_1, \dots, X_n) est appelé un échantillon, et le vecteur (x_1, \dots, x_n) est appelé une réalisation ou observation de l'échantillon (X_1, \dots, X_n) .

Avertissement : Ce cours aborde *uniquement* les méthodes mathématiques applicables aux *modèles mathématiques* issus de certains problèmes du monde réel. Nous n'aborderons quasiment pas, et de manière extrêmement naïve lorsque nous le ferons, l'aspect *modélisation* d'un problème concret donné. Il est cependant **fondamental** de garder à l'esprit que la modélisation d'un problème n'est pas une science exacte. Elle repose au contraire fortement, et souvent inconsciemment, sur les biais (scientifiques, culturels, moraux, politiques, économiques, ...) du modélisateur concernant le problème en question et son cadre. Différencier le bon du mauvais chasseur dépend avant tout du côté du fusil duquel on se trouve.

Dans l'exemple discuté plus haut, la loi forte des grands nombres et le théorème central limite sont difficilement attaquables du point de vue d'une suite de variables aléatoires indépendantes et identiquement distribuées suivant une loi de Bernoulli. Mais que nous fussent-ils réellement sur la proportion de la population nantaise aimant le café ? C'est avant tout une question de croyance en la pertinence de notre modèle mathématique. On peut imaginer moult objections à ce modèle, comme par exemple :

- La population nantaise est finie. Quel sens cela a-t-il de lui appliquer des théorèmes limites en mathématiques portant sur un nombre infini d’observations ?
- Notre modèle suppose que les sondés disent la vérité. C’est sans doute plausible en ce qui concerne le goût pour le café², mais qu’en serait-il pour une question plus délicate ?
- Il existe des gens qui détestent catégoriquement le café, et d’autres qui ne l’aiment pas mais quand même un petit peu parfois.

Quelle est alors la légitimité concernant la population nantaise du 0,6 obtenu par notre savant calcul ? À chacun de se forger sa propre opinion.

Terminons par une parabole tirée du livre [Ell17] illustrant que des a priori différents peuvent produire des inférences différentes, à l’aide de traitements mathématiques rigoureux. Pour faire simple, supposons que vous receviez 10 jours d’affilée un email d’une personne inconnue, vous prédisant chaque jour correctement la hausse ou la baisse d’une action en bourse. Le 11ème jour, cette personne vous demande alors une rémunération en échange de sa prédiction quotidienne. Est-il raisonnable d’accepter cette proposition ?

Si la personne est un escroc qui vous a fait 10 prédictions au hasard, la chance qu’elle ait fait la bonne prédiction 10 jours consécutifs est de

$$\frac{1}{2^{10}} = \frac{1}{1024} \simeq 0,001,$$

ce qui est fort improbable. On pourrait alors être enclin à accorder une certaine confiance à notre devin. Mais on pourrait aussi imaginer que cette personne ait choisi 1024 adresses email au hasard, que le premier jour elle envoie 512 prédictions de hausse et 512 prédictions de baisse, qu’elle raye ensuite de son carnet d’adresse les 512 récipiendaires de la prédiction qui s’est avérée fautive, puis que le deuxième jour elle envoie 256 prédictions de hausse et 256 prédictions de baisse, et ainsi de suite. Au bout de 10 jours, il y aura exactement une personne pour laquelle les 10 prédictions se seront révélées justes, et par malchance cela est tombé sur vous. Avec cette vision de la situation, il serait peu avisé d’accepter la proposition du 11ème jour.

Ainsi, notre décision finale dépendra fortement du fait que notre modèle ne considère que les données réellement observées (la prédiction correcte) ou bien au contraire de sa prise en compte de données imaginées ou potentielles (les 1023 prédictions fautes). À l’heure des supercalculateurs et des métadonnées, où une interprétation “objective” de données tient facilement lieu de preuve irréfutable, il peut être utile de garder cette parabole à l’esprit.

On peut ne pas les aimer, les statistiques ont néanmoins une influence profonde et quotidienne sur notre société et son développement économique, technologique, sani-

2. Cette affirmation repose elle-même sur ma croyance en le fait qu’aimer ou non le café est une caractéristique personnelle anodine.

taire, ... Connaître leurs principes mathématiques sous-jacents ne peut être qu'un facteur d'émancipation.

Tous les thèmes abordés dans ce cours sont classiques, et ce polycopié ne prétend à aucune originalité. Je me suis en particulier largement appuyé sur les notes de cours des années précédentes de Nicolas Pétrélis et Paul Rochet, ainsi que sur les livres [Lec19, DKLM05]. Le lecteur désirant approfondir le sujet est invité à s'y reporter. Ce cours nécessite quelque familiarité avec les bases des probabilités : variable aléatoire, espérance, variance, convergence de variables aléatoires, loi forte des grands nombres, ... On pourra se reporter au livre [Lec19] pour un rafraîchissement sur le sujet. J'essayerai cependant autant que possible de rappeler définitions et énoncés nécessaires au fur et à mesure de la progression. Chacune des trois sections se termine par un recueil d'exercices donnés en TD. Ces exercices ont été rédigés en particulier par Caroline Robet, Aymeric Stamm et Jeffery Petit.

Je tiens à remercier tout particulièrement mes collègues Nicolas Pétrélis et Aymeric Stamm pour leur patience face à mes innombrables questions, ainsi que les auditeurs du cours magistral pour leur enthousiasme et leurs nombreuses remarques pertinentes en direct.

Chapitre 1

Estimation ponctuelle

1.1 Définitions

On suppose que la loi de probabilité P_{X_j} des X_j dépend d'un paramètre $\theta \in \mathbb{R}^k$ inconnu, dont on cherche à déterminer la "vraie" valeur θ_0 à partir d'un échantillon. On note (P_θ) l'ensemble des lois de probabilités possibles pour les X_j . On pourra avoir en tête les exemples suivants

- $X_j \rightsquigarrow Ber(\theta)$ avec $\theta \in [0; 1]$;
- X_j suit une loi exponentielle de paramètre $\lambda \in \mathbb{R}^+$, abrégé en $X_j \rightsquigarrow Exp(\lambda)$, i.e. X_j admet pour densité

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}};$$

- X_j suit une loi normale d'espérance m et de variance σ^2 avec $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$, abrégé en $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$, i.e. X_j admet pour densité

$$f_{m, \sigma^2} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2}.$$

Définition 1.1.1. *Un estimateur de θ (ou une statistique) est une fonction mesurable $\hat{\theta}_n = g(X_1, \dots, X_n)$ de l'échantillon.*

Soulignons que $\hat{\theta}_n$ est une variable aléatoire par définition. Il s'agit donc de trouver une fonction g donnant une valeur, ou estimation, fiable de θ , et ce pour toute observation (x_1, \dots, x_n) .

Exemple 1.1.2. Dans l'exemple du préambule, lorsque $X_j \rightsquigarrow Ber(\theta)$, nous avons considéré la moyenne empirique des X_j

$$\hat{m}_1 = \frac{1}{n} \sum_{j=1}^n X_j$$

comme estimateur de θ . Nous renvoyons à la section 1.3 pour une justification de la notation \hat{m}_1 au lieu de $\hat{\theta}_n$. Insistons sur le fait que bien que n n'apparaisse pas dans la notation \hat{m}_1 , cet estimateur dépend néanmoins bel et bien de la taille de l'échantillon. Nous sacrifions à ce léger abus traditionnel afin d'alléger les notations.

Il s'agit maintenant de quantifier l'efficacité d'un estimateur, et la première chose à faire est d'étudier sa convergence. Or il existe plusieurs notions de convergence en probabilités, produisant autant de notions de consistance pour un estimateur. On rappelle qu'étant donné une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ et une variable aléatoire X , on dit que X_n converge vers X

1. *presque sûrement*, noté $X_n \xrightarrow[ps]{} X$, si

$$X_n(x) \xrightarrow[n \rightarrow +\infty]{} X(x)$$

pour tout x en dehors d'un ensemble de mesure nulle ;

2. *dans L^p* , noté $X_n \xrightarrow[L^p]{} X$, si ¹ $X \in L^p$ et

$$E((X_n - X)^p) \xrightarrow[n \rightarrow +\infty]{} 0;$$

3. *en probabilité*, noté $X_n \xrightarrow[\mathbb{P}]{} X$, si

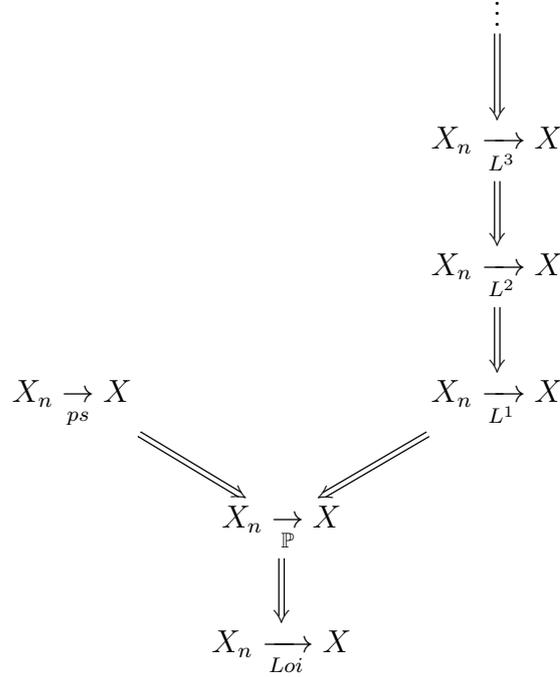
$$\forall \varepsilon > 0, P(|X_n - X| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$$

4. *en loi*, noté $X_n \xrightarrow[Loi]{} X$, si pour toute fonction continue bornée f on a

$$E(f(X_n)) \xrightarrow[n \rightarrow +\infty]{} E(f(X)).$$

Rappelons aussi qu'on a les implications suivantes entre ces différentes notions de convergence :

1. Une variable aléatoire X est dans L^p si $E(|X|^p) < +\infty$. Dans ce cas, l'espérance de X^p est bien définie.



On prendra bien garde que la convergence presque sûre n'entraîne pas la convergence dans L^p , ni réciproquement.

Définition 1.1.3. Soit $\hat{\theta}_n$ un estimateur de θ . On dit que

1. $\hat{\theta}_n$ est fortement consistant si

$$\forall \theta, P_{X_j} = P_\theta \implies \hat{\theta}_n \xrightarrow{ps} \theta;$$

2. $\hat{\theta}_n$ est consistant si

$$\forall \theta, P_{X_j} = P_\theta \implies \hat{\theta}_n \xrightarrow{\mathbb{P}} \theta;$$

3. $\hat{\theta}_n$ converge vers θ dans L^2 si

$$\forall \theta, P_{X_j} = P_\theta \implies E \left((\hat{\theta}_n - \theta)^2 \right) \xrightarrow{n \rightarrow +\infty} 0.$$

La quantité $E \left((\hat{\theta}_n - \theta)^2 \right)$ est appelée erreur quadratique moyenne et est notée $EQM(\hat{\theta}_n)$.

D'après ce qui précède, un estimateur fortement consistant est consistant.

Exemple 1.1.4. On suppose ici que $X_j \in L^1$, et on considère la moyenne empirique des X_j

$$\hat{m}_1 = \frac{1}{n} \sum_{j=1}^n X_j.$$

La loi forte des grands nombres (LFGN) nous dit exactement que $\hat{m}_1 \xrightarrow{ps} E(X_1)$. Autrement dit, dans la terminologie introduite à la définition 1.1.3, la variable aléatoire \hat{m}_1 est un estimateur fortement consistant de $E(X_1)$.

Dans le cas où $X_n \rightsquigarrow Ber(\theta)$, on a $E(|X_j|) = E(X_j) = \theta$, et \hat{m}_1 est donc un estimateur fortement consistant de θ .

Exemple 1.1.5. Supposons maintenant que $X_j \rightsquigarrow Exp(\lambda)$ avec $\lambda \in \mathbb{R}^+$. Comme X_j est une variable aléatoire positive, on a $E(|X_j|) = E(X_j)$. On calcule alors

$$\begin{aligned} E(X_j) &= \int_0^{+\infty} \lambda x e^{-\lambda x} dx \\ &= \int_0^{+\infty} e^{-\lambda x} dx \\ &= \frac{1}{\lambda}, \end{aligned}$$

la deuxième égalité étant obtenue en intégrant par parties. La LFGN nous dit alors que la moyenne empirique \hat{m}_1 des X_j converge presque sûrement vers $\frac{1}{\lambda}$. En prenant l'inverse, on obtient que

$$\hat{\lambda}_n = \frac{1}{\hat{m}_1} = \frac{n}{\sum_{j=1}^n X_j}$$

est un estimateur fortement consistant de λ .

Il peut arriver qu'on ait à notre disposition plusieurs estimateurs d'un même paramètre θ . On est alors en droit de se demander si l'un de ces estimateurs est meilleur que les autres. Il n'est peut-être pas inutile ici de rappeler qu'un classement dépend cruciallement de la fonction de classement utilisée. Autrement dit, savoir si un estimateur est meilleur qu'un autre dépend de ce que l'on attend d'un *bon* estimateur². L'erreur quadratique moyenne constitue une fonction de classement élémentaire.

Définition 1.1.6. *Étant donné deux estimateurs $\hat{\theta}_n$ et $\tilde{\theta}_n$ de θ , on dit que $\hat{\theta}_n$ est meilleur que $\tilde{\theta}_n$ si*

$$\forall \theta, EQM(\hat{\theta}_n) \leq EQM(\tilde{\theta}_n).$$

Une autre des qualités attendues d'un estimateur est d'être sans biais.

2. Pour distinguer un bon estimateur d'un mauvais, on pourra se reporter à l'étude correspondante concernant bon et mauvais chasseurs.

Définition 1.1.7. Le biais d'un estimateur $\hat{\theta}_n$ de θ est défini par

$$b(\hat{\theta}_n) = E(\hat{\theta}_n - \theta) = E(\hat{\theta}_n) - \theta.$$

On dit que $\hat{\theta}_n$ est sans biais si $b(\hat{\theta}_n) = 0$, et asymptotiquement sans biais si

$$b(\hat{\theta}_n) \xrightarrow{n \rightarrow +\infty} 0.$$

Exemple 1.1.8. On suppose ici que $X_j \in L^1$, et on considère encore une fois la moyenne empirique \hat{m}_1 des X_j . On a alors

$$E(\hat{m}_1) = \frac{1}{n} \sum_{j=1}^n E(X_j) = E(X_1).$$

Ainsi \hat{m}_1 est un estimateur sans biais de $E(X_1)$. Si $X_j \rightsquigarrow \text{Ber}(\theta)$, alors \hat{m}_1 est un estimateur fortement consistant et sans biais de θ .

L'erreur quadratique moyenne et le biais d'un estimateur sont liés.

Lemme 1.1.9. Soit $\hat{\theta}_n$ un estimateur de θ . Si X_j et $\hat{\theta}_n$ sont dans L^2 , alors³

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

Démonstration. Il s'agit d'un simple calcul :

$$\begin{aligned} EQM(\hat{\theta}_n) &= E\left((\hat{\theta}_n - \theta)^2\right) \\ &= E\left(\hat{\theta}_n^2 - 2\hat{\theta}_n \cdot \theta + \theta^2\right) \\ &= E(\hat{\theta}_n^2) - 2\theta \cdot E(\hat{\theta}_n) + \theta^2 \\ &= \text{Var}(\hat{\theta}_n) + E(\hat{\theta}_n)^2 - 2\theta \cdot E(\hat{\theta}_n) + \theta^2 \\ &= \text{Var}(\hat{\theta}_n) + b(\hat{\theta}_n)^2 \end{aligned}$$

comme annoncé. ☺

Exemple 1.1.10. Supposons que $X_j \in L^2$, et continuons l'étude de la moyenne empirique \hat{m}_1 des X_j . On a déjà vu que \hat{m}_1 est sans biais, et donc

$$EQM(\hat{m}_1) = \text{Var}(\hat{m}_1) = \frac{1}{n^2} \text{Var}\left(\sum_{j=1}^n X_j\right).$$

3. On rappelle que $\text{Var}(X) = E\left((X - E(X))^2\right) = E(X^2) - E(X)^2$.

Puisque les X_j sont indépendantes, on a⁴ $Var\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n Var(X_j)$ et donc

$$EQM(\hat{m}_1) = \frac{1}{n} Var(X_1).$$

Ainsi, la moyenne empirique est un estimateur de $E(X_1)$ fortement consistant et convergeant dans L^2 . On peut dire qu'il s'agit là d'un bon estimateur.

Remarquons qu'on peut être encore plus précis concernant la convergence dans L^2 de \hat{m}_1 : cette convergence s'effectue en $\frac{1}{n}$. Formulé autrement, l'estimateur \hat{m}_1 a une vitesse de convergence de l'ordre de n .

Plus généralement, on peut essayer d'évaluer la vitesse de convergence d'une suite réelle positive (u_n) tendant vers 0 de la manière suivante : c'est l'exposant a (ou la suite (n^a)) tel(le) que $n^a \cdot u_n$ converge vers un élément de \mathbb{R}^* . Pour $\varepsilon > 0$, on a alors

$$n^{a-\varepsilon} \cdot u_n \xrightarrow{n \rightarrow +\infty} 0 \quad \text{et} \quad n^{a+\varepsilon} \cdot u_n \xrightarrow{n \rightarrow +\infty} +\infty.$$

En d'autres termes, c'est l'exposant a pour lequel $n^a \cdot u_n$ a une limite non dégénérée. Par exemple la suite $\left(\frac{2n}{n^3+n+1}\right)$ a une vitesse de convergence de n^2 vers 0 :

- $n^2 \cdot \frac{2n}{n^3+n+1} = \frac{2n^3}{n^3+n+1} \xrightarrow{n \rightarrow +\infty} 2,$
- $\forall \varepsilon > 0, n^{2-\varepsilon} \cdot \frac{2n}{n^3+n+1} = n^{-\varepsilon} \cdot \frac{2n^3}{n^3+n+1} \xrightarrow{n \rightarrow +\infty} 0,$
- $\forall \varepsilon > 0, n^{2+\varepsilon} \cdot \frac{2n}{n^3+n+1} = n^\varepsilon \cdot \frac{2n^3}{n^3+n+1} \xrightarrow{n \rightarrow +\infty} +\infty.$

Cette définition de vitesse de convergence n'est cependant pas satisfaisante, puisque la suite $\left(\frac{1}{n \log n}\right)$ n'admet alors pas de vitesse de convergence : $\frac{n^a}{n \log n}$ tend vers 0 si $a \leq 1$ et vers $+\infty$ si $a > 1$. Il faut donc affiner quelque peu cette notion.

Définition 1.1.11. Soit $(u_n)_{n \in \mathbb{N}}$ une suite de réels positifs qui converge vers 0, et $(a_n)_{n \in \mathbb{N}}$ une suite de réels positifs telle que $\lim_{n \rightarrow +\infty} a_n = +\infty$. On dit que la suite (u_n) converge à vitesse (a_n) si

$$\lim_{n \rightarrow +\infty} a_n \cdot u_n \in \mathbb{R}_{>0}.$$

Exemple 1.1.12. La suite $\left(\frac{2n}{n^3+n+1}\right)$ a une vitesse de convergence de n^2 , et les suites $\left(\frac{1}{n \log n}\right)$ et $\left(\frac{1}{n \log n+n}\right)$ ont une vitesse de convergence de $n \log n$.

Nous venons de développer la notion de vitesse de convergence d'une suite réelle positive, mais cette notion se généralise à d'autres situations. Nous sommes bien évidemment intéressés ici à définir et étudier la vitesse de convergence d'un estimateur. On dit d'une variable aléatoire réelle qu'elle est non dégénérée si elle n'est pas nulle presque sûrement.

4. Le lecteur est chaudement invité à redémontrer par lui-même ce fait élémentaire.

Définition 1.1.13. Soit $\hat{\theta}_n$ un estimateur consistant de θ , et $(a_n)_{n \in \mathbb{N}}$ une suite de réels positifs telle que $\lim_{n \rightarrow +\infty} a_n = +\infty$. On dit que la suite $\hat{\theta}_n$ converge à vitesse (a_n) si

$$\forall \theta, P_{X_j} = P_\theta \implies a_n \cdot (\hat{\theta}_n - \theta) \xrightarrow[\text{Loi}]{} Y,$$

où Y est une variable aléatoire non dégénérée.

Le choix de la convergence en loi à la définition 1.1.13 a plusieurs origines. Tout d'abord il s'agit de la notion la plus faible de convergence comme nous l'avons vu plus haut. Mais plus important encore, le théorème central limite permet de calculer la vitesse de convergence d'un estimateur dans bien des cas.

Théorème 1.1.14 (Théorème central limite – TCL). Soit $(X_j)_{j \in \mathbb{N}}$ une suite de variables aléatoires dans L^2 , indépendantes et identiquement distribuées. On note $m_1 = E(X_1)$ et $\sigma^2 = \text{Var}(X_1)$, et on considère la moyenne empirique \hat{m}_1 des X_j . Alors

$$\sqrt{n} \cdot (\hat{m}_1 - m_1) \xrightarrow[\text{Loi}]{} \mathcal{N}(0, \sigma^2).$$

Lorsque⁵ $\sigma^2 \neq 0$, l'énoncé se reformule

$$\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma} \xrightarrow[\text{Loi}]{} \mathcal{N}(0, 1).$$

Le TCL est un résultat absolument fondamental en probabilités et statistiques, compagnon de la LFGN. On sait par cette dernière que \hat{m}_1 converge presque sûrement vers m_1 . Le TCL nous assure alors non seulement que \hat{m}_1 converge vers m_1 à la vitesse \sqrt{n} , mais qu'en plus la loi non dégénérée vers laquelle converge $\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma}$ est indépendante de la loi des X_j ! Ainsi, la loi normale centrée réduite $\mathcal{N}(0, 1)$ apparaît comme une loi limite universelle, ce que personnellement je trouve assez fascinant.

Lorsqu'un estimateur $\hat{\theta}_n$ de θ converge dans L^2 , sa vitesse de convergence est liée à la vitesse de convergence de la suite $EQM(\hat{\theta}_n)$. En fait la première est bien souvent la racine carrée de la deuxième⁶.

Exemple 1.1.15. Supposons encore une fois que $X_j \in L^2$, et étudions encore et toujours la moyenne empirique \hat{m}_1 des X_j . Nous avons vu à l'exemple 1.1.10 que $EQM(\hat{m}_1)$ une vitesse de convergence de n , et le TCL nous assure que \hat{m}_1 converge vers m_1 à la vitesse \sqrt{n} .

Tout cela est bien joli, mais nous n'avons essentiellement étudié qu'un seul estimateur jusqu'à présent : la moyenne empirique. Nous allons voir dans les sections 1.3 et 1.4 quelques méthodes de construction d'estimateurs. Mais avant cela, donnons une preuve du TCL.

5. La variance d'une variable aléatoire X est nulle si et seulement si X est constante presque sûrement. Dans ce cas, le TCL est trivial.

6. La présence de la racine carrée s'explique par le fait que la norme dans L^2 est $\sqrt{E(X^2)}$ et non $E(X^2)$. On pourra se reporter à son cours de topologie préférée à ce sujet.

1.2 Digression : preuve du théorème central limite

La démonstration présentée ici utilise les fonctions caractéristiques, ce qui sera l'occasion de quelques rappels à leur sujet. Étant donné X une variable aléatoire réelle, on définit sa *fonction caractéristique* par

$$\begin{aligned} \phi_X : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto E(e^{itX}) . \end{aligned}$$

Attention, l'ensemble d'arrivée de ϕ_X est bien \mathbb{C} et non \mathbb{R} , puisque e^{itX} est une variable aléatoire à valeur dans \mathbb{C}^7 . Puisque X est à valeur réelle, on a $|e^{itX}| = 1$ et donc la fonction e^{itX} est bien intégrable pour tout $t \in \mathbb{R}$. On rappelle sans démonstration les propriétés principales de ϕ_X .

Théorème 1.2.1. *Les fonctions caractéristiques de variables aléatoires réelles vérifient les propriétés suivantes :*

1. $\forall t \in \mathbb{R}, \phi_X(t) = \phi_Y(t) \implies X$ et Y ont même loi ;
2. X et Y indépendantes $\implies \forall t \in \mathbb{R}, \phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$;
3. $X \in L^p \implies \phi_X$ est dérivable p fois en 0 et

$$\phi_X^{(p)}(0) = \frac{1}{i^p} E(X^p);$$

4. *Théorème de Levy : Si (X_n) est une suite de variables aléatoires réelles, alors*

$$X_n \xrightarrow[\text{Loi}]{} X \iff \forall t \in \mathbb{R}, \phi_{X_n}(t) \xrightarrow[n \rightarrow +\infty]{} \phi_X(t);$$

5. $\phi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}}$.

Ce rappel étant fait, on peut maintenant démontrer le TCL.

Preuve du théorème 1.1.14. On suppose que $\sigma \neq 0$, la preuve étant triviale sinon. On pose

$$Y_j = \frac{X_j - m_1}{\sigma}.$$

On a donc $E(Y_j) = 0$ et $Var(Y_j) = 1$, et

$$\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma} = \sum_{j=1}^n \frac{Y_j}{\sqrt{n}}.$$

7. On rappelle en passant que $E(e^{itX}) = E(Re(e^{itX})) + iE(Im(e^{itX}))$.

Par le théorème 1.2.1(3), nous avons

$$\phi_{Y_j}(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Le théorème 1.2.1(2) implique alors

$$\begin{aligned} \phi_{\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma}}(t) &= \phi_{\sum_{j=1}^n \frac{Y_j}{\sqrt{n}}}(t) \\ &= \left(\phi_{\frac{Y_1}{\sqrt{n}}}(t) \right)^n \\ &= \left(\phi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) \right)^n \\ &= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right)^n. \end{aligned}$$

L'étude de la limite de cette suite, à t fixé, lorsque n tend vers l'infini est standard. On se rappelle que

$$\log(1+x) = x + \underset{x \rightarrow 0}{o}(x),$$

et on écrit⁸

$$\begin{aligned} \log \left(\phi_{\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma}}(t) \right) &= n \log \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right) \\ &= n \left(-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right) \\ &= -\frac{t^2}{2} + o(1). \end{aligned}$$

Ainsi

$$\log \left(\phi_{\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma}}(t) \right) \xrightarrow{n \rightarrow +\infty} -\frac{t^2}{2},$$

et

$$\phi_{\sqrt{n} \cdot \frac{\hat{m}_1 - m_1}{\sigma}}(t) \xrightarrow{n \rightarrow +\infty} e^{-\frac{t^2}{2}}.$$

Le théorème central limite se déduit maintenant du Théorème 1.2.1(4) et (5). ⊙

8. On fera attention à ce qu'on a besoin d'une détermination *complexe* du logarithme ici, puisqu'une fonction caractéristique est à valeur dans \mathbb{C} .

1.3 Méthode des moments

La première idée pour construire un estimateur est de généraliser la moyenne empirique en considérant les moments d'ordres supérieurs.

Définition 1.3.1. *On dit qu'une variable aléatoire X admet un moment d'ordre p si $X \in L^p$, c'est à dire si $E(|X|^p) < +\infty$. Dans ce cas, on définit le moment d'ordre p de X par*

$$m_p = E(X^p).$$

Soulignons que le nombre réel p n'est pas nécessairement entier dans la définition précédente.

Lemme 1.3.2. *Si $X \in L^p$, alors $X \in L^q$ pour tout $q \in]0; p]$.*

Démonstration. Comme $q \leq p$, la fonction $x \mapsto x^{\frac{p}{q}}$ est convexe sur \mathbb{R}^+ . On obtient donc par l'inégalité de Jensen

$$E(|X|^p) = E\left((|X|^q)^{\frac{p}{q}}\right) \geq E(|X|^q)^{\frac{p}{q}}.$$

Donc $E(|X|^q) < +\infty$ et X est bien dans L^q . ⊙

Profitons de la preuve précédente pour rappeler l'inégalité de Jensen, bien utile dans la pratique.

Théorème 1.3.3 (Inégalité de Jensen). *Soit ϕ une fonction convexe mesurable. Alors pour toute variable aléatoire X dans L^1 ou positive, on a*

$$\phi(E(X)) \leq E(\phi(X)).$$

De plus si ϕ est strictement convexe, alors $\phi(E(X)) = E(\phi(X))$ si et seulement si X est presque sûrement constante.

Remarque 1.3.4. Par définition une fonction réelle ϕ est convexe si pour tout x et y , on a (voir la figure 1.1)

$$\phi\left(\frac{x+y}{2}\right) \leq \frac{\phi(x) + \phi(y)}{2}.$$

En considérant la mesure de probabilité μ de support $\{x, y\}$ définie par $\mu(x) = \mu(y) = \frac{1}{2}$, cette inégalité se réécrit sous la forme

$$\phi(E(X)) \leq E(\phi(X)).$$

Le théorème 1.3.3 nous dit donc que l'on peut déduire l'inégalité de Jensen générale uniquement à partir de l'hypothèse qu'elle est vérifiée pour les mesures très simples que sont les mesures equidistribuées dont le support est réduit à deux points. C'est manifestement un joli résultat.

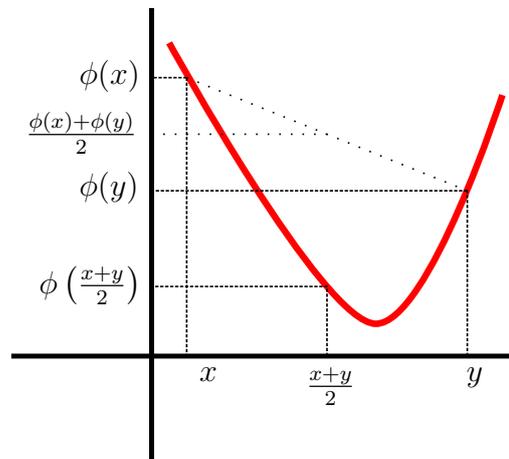


FIGURE 1.1 – Fonction convexe.

Remarque 1.3.5. Pour revenir au lemme 1.3.2, ne perdons pas de vue que ce résultat est *faux* si la mesure n'est pas une mesure de probabilité (ou de masse finie). Par exemple, la fonction $x \mapsto \frac{1}{x^2}$ est intégrable sur $[1; +\infty[$, mais sa racine carrée $x \mapsto \frac{1}{x}$ ne l'est pas.

Après ces quelques rappels de probabilités et d'intégration, revenons aux statistiques proprement dites.

Définition 1.3.6. Si $X_j \in L^p$, on appelle le moment empirique d'ordre p l'estimateur

$$\hat{m}_p = \frac{1}{n} \sum_{j=1}^n X_j^p.$$

Par la LFGN, \hat{m}_p est un estimateur sans biais et fortement consistant de m_p . En appliquant le TCL à la variable aléatoire X_j^p , on peut en particulier calculer la vitesse de convergence de \hat{m}_p .

Lemme 1.3.7. Si $X_j \in L^{2p}$, alors

$$\sqrt{n} \cdot (\hat{m}_p - m_p) \xrightarrow{\text{Loi}} \mathcal{N}(0, m_{2p} - m_p^2).$$

En première approximation, la construction d'estimateurs par la méthode des moments peut se résumer ainsi : si les moments d'ordre p s'expriment en fonction de θ , on cherche à inverser ce système et à exprimer θ en fonction des moments, et on estime ces derniers par les moments empiriques. Quelques exemples illustreront sans doute mieux la méthode que cette phrase absconse.

Exemple 1.3.8. On a déjà vu que \hat{m}_1 est un estimateur sans biais et fortement consistant de $E(X_1)$ si $X_1 \in L^1$.

Exemple 1.3.9. On a vu à l'exemple 1.1.5 que $\hat{\lambda}_n = \frac{1}{\hat{m}_1}$ est un estimateur fortement consistant de λ si $X_j \rightsquigarrow \text{Exp}(\lambda)$ avec $\lambda \in \mathbb{R}^+$.

Exemple 1.3.10. Si $X_j \in L^2$, alors la variance de X_j est définie par

$$\text{Var}(X_j) = m_2 - m_1^2.$$

On en déduit donc que $\hat{m}_2 - \hat{m}_1^2$ est un estimateur sans biais et fortement consistant de $\text{Var}(X_1)$. Calculons son biais :

$$\begin{aligned} E(\hat{m}_2 - \hat{m}_1^2) &= E(\hat{m}_2) - E(\hat{m}_1^2) \\ &= m_2 - \frac{1}{n^2} E \left(\left(\sum_{j=1}^n X_j \right)^2 \right) \\ &= m_2 - \frac{1}{n^2} E \left(\sum_{j=1}^n X_j^2 + \sum_{j \neq k} X_j X_k \right) \\ &= m_2 - \frac{1}{n} m_2 - \frac{1}{n^2} \sum_{j \neq k} E(X_j X_k). \end{aligned}$$

Comme les X_j sont indépendantes, on a $E(X_j X_k) = E(X_j) \cdot E(X_k) = m_1^2$. On obtient donc

$$\begin{aligned} E(\hat{m}_2 - \hat{m}_1^2) &= \frac{n-1}{n} m_2 - \frac{n^2-n}{n^2} m_1^2 \\ &= \frac{n-1}{n} (m_2 - m_1^2) \\ &= \frac{n-1}{n} \text{Var}(X_1). \end{aligned}$$

Ainsi, l'estimateur $\hat{m}_2 - \hat{m}_1^2$ est biaisé. Il est cependant facile de construire un estimateur sans biais de $\text{Var}(X_1)$ en considérant

$$\hat{S}_n^2 = \frac{n}{n-1} (\hat{m}_2 - \hat{m}_1^2),$$

qui est aussi fortement consistant. Un calcul sans difficulté montre que \hat{S}_n s'écrit aussi

$$\hat{S}_n^2 = \frac{1}{n-1} \left(\sum_{j=1}^n X_j^2 - \left(\sum_{j=1}^n X_j \right)^2 \right) = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{m}_1)^2.$$

Exemple 1.3.11. On prendra bien garde qu'il n'y a pas unicité dans la manière de construire un estimateur par la méthode des moments. Revenons à un échantillon de variables aléatoires $X_j \rightsquigarrow \text{Exp}(\lambda)$ avec $\lambda \in \mathbb{R}^+$. On a déjà vu à l'exemple 1.3.9 que $\hat{\lambda}_n = \frac{1}{\hat{m}_1}$ est un estimateur fortement consistant de λ . On calcule aisément par intégration par parties que

$$\begin{aligned} m_2 &= E(X_1^2) \\ &= \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx \\ &= 2 \int_0^{+\infty} x e^{-\lambda x} dx \\ &= \frac{2}{\lambda}. \end{aligned}$$

Ainsi $\lambda = \sqrt{\frac{2}{m_2}}$, et donc $\tilde{\lambda}_n = \sqrt{\frac{2}{\hat{m}_2}}$ est aussi un estimateur fortement consistant de λ .

On verra à l'exemple 1.3.16 une méthode pour comparer ces deux estimateurs. On peut déjà sentir intuitivement que $\hat{\lambda}_n$ sera meilleur pour des petites valeurs de λ , et que $\tilde{\lambda}_n$ sera meilleur pour des grandes valeurs du paramètre. En effet plus λ est petit, plus X_j a une probabilité importante de prendre de grandes valeurs puisque

$$P(X_j \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t} \xrightarrow{\lambda \rightarrow 0} 0.$$

La fonction $x \mapsto x^2$ est beaucoup plus sensible aux erreurs d'imprécision que la fonction $x \mapsto x$ lorsque x est grand, alors qu'elle l'est beaucoup moins quand x est petit. Ainsi \hat{m}_1 (et donc $\hat{\lambda}_n$) sera moins sensible aux erreurs d'imprécision pour de petites valeurs de λ , alors que \hat{m}_2 (et donc $\tilde{\lambda}_n$) le sera moins pour de grandes valeurs de λ .

On observe dans les exemples traités que θ s'exprime en général comme une fonction des moments, plutôt que comme un moment lui-même. On a déjà utilisé la proposition suivante implicitement, il est temps de la formuler et d'en donner une preuve.

Proposition 1.3.12. Soit $X_j \in L^p$, et $g : \mathbb{R}^p \rightarrow \mathbb{R}^k$ une fonction continue. Alors $\hat{\theta}_n = g(\hat{m}_1, \dots, \hat{m}_p)$ est un estimateur fortement consistant de $\theta = g(m_1, \dots, m_p)$.

Démonstration. Par la LFGN, on sait que \hat{m}_i converge presque sûrement vers m_i . Une union finie (et même dénombrable) d'ensembles de mesure nulle est de mesure nulle, donc le p -uplet $(\hat{m}_1, \dots, \hat{m}_p)$ converge presque sûrement vers (m_1, \dots, m_p) . Par continuité de la fonction g , l'estimateur $\hat{\theta}_n$ converge presque sûrement vers θ . ☺

Il n'est cependant pas vrai que $g(\hat{m}_1, \dots, \hat{m}_p)$ est sans biais même si les \hat{m}_i le sont. La variance empirique étudiée à l'exemple 1.3.10 donne un contre-exemple. On peut en revanche raisonnablement espérer une absence asymptotique de biais.

Proposition 1.3.13. *Si $X_j \in L^p$ et $g : \mathbb{R}^p \rightarrow \mathbb{R}^k$ est une fonction continue bornée, alors $\hat{\theta}_n = g(\hat{m}_1, \dots, \hat{m}_p)$ est un estimateur asymptotiquement sans biais de $\theta = g(m_1, \dots, m_p)$.*

Démonstration. On a déjà vu à la proposition 1.3.12 que $\hat{\theta}_n \xrightarrow[ps]{} \theta$. De plus, on a

$$|\hat{\theta}_n| = |g(\hat{m}_1, \dots, \hat{m}_p)| \leq \|g\|_\infty.$$

Comme la fonction constante $\|g\|_\infty$ est intégrable puisque la mesure est de masse finie, le théorème de convergence dominée nous assure que $E(\hat{\theta}_n)$ converge vers $E(\theta)$. \odot

On peut aussi formuler un TCL pour un estimateur $\hat{\theta}_n = g(\hat{m}_1, \dots, \hat{m}_p)$. Par soucis de clarté, nous nous bornerons au cas $p = 1$. Le cas général se traite à l'aide d'un peu de calcul différentiel.

Proposition 1.3.14 (Δ -méthode). *On suppose que $X_j \in L^2$, et que $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction dérivable en m_1 . On pose*

$$\sigma = \text{Var}(X_j), \quad \hat{\theta}_n = g(\hat{m}_1), \quad \text{et} \quad \theta = g(m_1).$$

Alors

$$\sqrt{n} \cdot (\hat{\theta}_n - \theta) \xrightarrow[Loi]{} \mathcal{N}(0, g'(m_1)^2 \sigma^2).$$

En particulier, si $g'(m_1) \neq 0$, alors l'estimateur $\hat{\theta}_n$ de θ est fortement consistant de vitesse de convergence \sqrt{n} .

Démonstration. On écrit un développement limité de la fonction g en m_1 :

$$\begin{aligned} \hat{\theta}_n &= g(\hat{m}_1) \\ &= g(m_1) + (\hat{m}_1 - m_1)g'(m_1) + (\hat{m}_1 - m_1)\phi(\hat{m}_1 - m_1), \end{aligned}$$

avec $\phi(t) \xrightarrow[t \rightarrow 0]{} 0$. On en déduit donc

$$\sqrt{n} (\hat{\theta}_n - \theta) = \sqrt{n} (\hat{\theta}_n - \theta)g'(m_1) + \sqrt{n} (\hat{\theta}_n - \theta)\phi(\hat{m}_1 - m_1).$$

Le TCL assure que $\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[Loi]{} \mathcal{N}(0, \sigma^2)$, d'où l'on déduit

$$\sqrt{n} (\hat{\theta}_n - \theta)g'(m_1) \xrightarrow[Loi]{} \mathcal{N}(0, g'(m_1)^2 \sigma^2).$$

Comme $\hat{m}_1 \xrightarrow[ps]{} m_1$, on a $\phi(\hat{m}_1 - m_1) \xrightarrow[ps]{} 0$, et donc $\phi(\hat{m}_1 - m_1) \xrightarrow[\mathbb{P}]{} 0$. Une première application du théorème de Slutsky nous donne alors

$$\sqrt{n} (\hat{\theta}_n - \theta)\phi(\hat{m}_1 - m_1) \xrightarrow[Loi]{} 0.$$

Une deuxième application du théorème de Slutsky nous donne alors le résultat. \odot

Rappelons que la convergence en loi ne se comporte pas forcément très bien avec les sommes et les produits. Par exemple, si $X \rightsquigarrow Ber(\frac{1}{2})$ alors $Y = 1 - X \rightsquigarrow Ber(\frac{1}{2})$; on a donc $X_n = X \xrightarrow[Loi]{} X$ et $Y_n = Y \xrightarrow[Loi]{} X$, en revanche $X + Y$ suit une loi uniforme sur $[0; 1]$ et n'a donc pas la même loi que $2X$. Le théorème de Slutsky utilisé dans la preuve ci-dessus permet d'assurer dans certains cas que la limite de la somme ou du produit est bien la somme ou le produit des limites. C'est un énoncé très important en pratique, et nous le rappelons ci-dessous.

Théorème 1.3.15 (Théorème de Slutsky). *Soit (X_n) et (Y_n) deux suites de variables aléatoires réelles telles que*

$$X_n \xrightarrow[Loi]{} X \quad \text{et} \quad Y_n \xrightarrow[Loi]{} a \in \mathbb{R}.$$

Alors

$$X_n + Y_n \xrightarrow[Loi]{} X + a \quad \text{et} \quad Y_n X_n \xrightarrow[Loi]{} aX.$$

Exemple 1.3.16. On a vu à l'exemple 1.1.5 que $\hat{\lambda}_n = \frac{1}{\hat{m}_1}$ est un estimateur fortement consistant de λ si $X_j \rightsquigarrow Exp(\lambda)$ avec $\lambda \in \mathbb{R}^+$. On a ici $\hat{\lambda}_n = g(\hat{m}_1)$ avec $g(t) = \frac{1}{t}$. En particulier $g'(t) = -\frac{1}{t^2}$ et

$$g'(m_1) = g'\left(\frac{1}{\lambda}\right) = -\lambda^2.$$

On a aussi calculé à l'exemple 1.3.11 que $m_2 = \frac{2}{\lambda^2}$, ce qui nous donne

$$\sigma^2 = m_2 - m_1^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

On a donc $g'(m_1)^2 \sigma^2 = \frac{\lambda^4}{\lambda^2} = \lambda^2$, et la Δ -méthode nous donne alors

$$\sqrt{n} (\hat{\lambda}_n - \lambda) \xrightarrow[Loi]{} \mathcal{N}(0, \lambda^2).$$

Nous avons vu à l'exemple 1.3.11 que $\tilde{\lambda}_n = \sqrt{\frac{2}{\hat{m}_2}}$ est aussi un estimateur fortement consistant de λ . On a donc ici $\tilde{\lambda}_n = h(\hat{m}_2)$ avec $h(t) = \sqrt{\frac{2}{t}}$. En particulier $h'(t) = -\frac{1}{\sqrt{2}t^{3/2}}$ et

$$h'(m_2) = h'\left(\sqrt{\frac{2}{\lambda}}\right) = -\frac{\lambda^{3/2}}{2^{5/4}}.$$

On va appliquer ici la Δ -méthode à la variable aléatoire X_j^2 . On calcule facilement par intégration par parties que

$$E(X_j^p) = \int_0^{+\infty} \lambda x^p e^{-\lambda x} dx = \frac{p!}{\lambda^p},$$

ce qui nous donne

$$\begin{aligned} \text{Var}(X_j^2) &= E(X_j^4) - E(X_j^2)^2 \\ &= \frac{24}{\lambda^4} - \frac{4}{\lambda^4} \\ &= \frac{20}{\lambda^4}. \end{aligned}$$

La Δ -méthode nous donne alors

$$\sqrt{n} (\tilde{\lambda}_n - \lambda) \xrightarrow[\text{Loi}]{} \mathcal{N}\left(0, \frac{5}{\sqrt{2}\lambda}\right).$$

D'un point de vue TCL, l'écart type limite est donc meilleur pour l'estimateur $\hat{\lambda}_n$ lorsque λ est petit, et est meilleur pour $\tilde{\lambda}_n$ lorsque λ est grand. Ceci formalise le raisonnement intuitif que nous avons tenu à la fin de l'exemple 1.3.11.

1.4 Maximum de vraisemblance

Une autre idée pour estimer le paramètre θ dont dépend la loi de probabilité P_θ de X_j , est de considérer la probabilité de faire l'observation (x_1, \dots, x_n) , puis de choisir la valeur de θ qui maximise cette probabilité. En d'autres termes, on considère la valeur de θ qui rend l'observation dont nous disposons la plus probable.

Pour simplifier, nous supposons dans cette partie que le paramètre θ à déterminer est dans \mathbb{R} .

Définition 1.4.1. La vraisemblance de l'observation (x_1, \dots, x_n) est la fonction de $\theta \mapsto V(\theta)$ suivante

1. Si P_θ est une loi discrète :

$$V(\theta) = \prod_{j=1}^n P_\theta(X_j = x_j).$$

2. Si P_θ est une loi continue de densité f_θ :

$$V(\theta) = \prod_{j=1}^n f_\theta(x_j).$$

Un estimateur du maximum de vraisemblance de θ , noté MV_θ est une valeur de θ pour laquelle la fonction $\theta \mapsto V(\theta)$ est maximum.

Soulignons bien que c'est la valeur de θ qui rend $V(\theta)$ maximum qui nous intéresse, *et non pas* la valeur de ce maximum. Dans le cas où P_θ est une loi discrète, on a par indépendance des X_j

$$V(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n).$$

Un maximum de vraisemblance est donc effectivement une valeur de θ qui rend l'observation (x_1, \dots, x_n) la plus probable. Un raisonnement analogue combiné avec un soupçon de discrétisation justifient de manière analogue le maximum de vraisemblance dans le cas où P_θ est une loi continue à densité.

Exemple 1.4.2. Supposons que $X_j \rightsquigarrow Ber(\theta)$. Rappelons que dans ce cas $x_j = 0$ ou $x_j = 1$ et

$$P_\theta(X_j = 1) = \theta \quad \text{et} \quad P_\theta(X_j = 0) = 1 - \theta.$$

Puisque le nombre de x_j valant 1 est précisément $\sum_{j=1}^n x_j$, on obtient

$$V(\theta) = \theta^{\sum_{j=1}^n x_j} (1 - \theta)^{n - \sum_{j=1}^n x_j}.$$

Pour chercher le maximum de $V(\theta)$, il est de bon goût de calculer sa dérivée. Or calculer la dérivée d'un produit est un peu fastidieux, alors que calculer la dérivée d'une somme est beaucoup plus aisé. Puisqu'un logarithme transforme justement les produits en sommes, considérons la *log-vraisemblance*

$$\begin{aligned} v(\theta) &= \log(V(\theta)) \\ &= \sum_{j=1}^n x_j \cdot \log \theta + \left(n - \sum_{j=1}^n x_j \right) \cdot \log(1 - \theta). \end{aligned}$$

La fonction $x \mapsto \log(x)$ étant croissante, la fonction $\theta \mapsto V(\theta)$ admet un maximum en $\tilde{\theta}$ si et seulement si la fonction $\theta \mapsto v(\theta)$ admet un maximum en $\tilde{\theta}$. Le maximum de ces deux fonctions n'ont pas de raison a priori d'être égaux, mais cela n'a pas d'importance ici, comme nous l'avons déjà souligné plus haut. On a

$$v'(\theta) = \frac{\sum_{j=1}^n x_j}{\theta} - \frac{n - \sum_{j=1}^n x_j}{1 - \theta},$$

et donc

$$\begin{aligned} v'(\theta) = 0 &\iff (1 - \theta) \cdot \sum_{j=1}^n x_j = \theta \cdot \left(n - \sum_{j=1}^n x_j \right) \\ &\iff \theta = \frac{1}{n} \sum_{j=1}^n x_j \\ &\iff \theta = \hat{n}_1. \end{aligned}$$

Pour s'assurer que \hat{m}_1 est bien un maximum de $v(\theta)$, calculons sa dérivée seconde :

$$v''(\theta) = -\frac{\sum_{j=1}^n x_j}{\theta^2} - \frac{n - \sum_{j=1}^n x_j}{(1-\theta)^2} < 0,$$

et la fonction $v(\theta)$ atteint bien son maximum pour $\theta = \hat{m}_1$. Dans le cas de variables aléatoires suivant une loi de Bernoulli, on obtient donc

$$MV_\theta = \hat{m}_1.$$

On pourrait se dire “tout ça pour ça!”. Au contraire, il est plutôt rassurant que dans la situation la plus simple où l'on voit un estimateur “évident”, différentes méthodes donnent précisément cet estimateur. C'est une sorte de preuve par neuf de l'inférence, nous indiquant qu'il est possible que nous ne fassions pas n'importe quoi.

Exemple 1.4.3. Le deuxième exemple d'estimation que nous avons vu concerne la valeur de λ pour un échantillon suivant une loi $Exp(\lambda)$ avec $\lambda \in \mathbb{R}^+$. Nous avons déjà vu aux exemples 1.1.5 et 1.3.11 deux estimateurs fortement consistants de λ :

$$\hat{\lambda}_n = \frac{1}{\hat{m}_1} \quad \text{et} \quad \tilde{\lambda}_n = \sqrt{\frac{2}{\hat{m}_2}}.$$

Pour calculer la vraisemblance, rappelons d'abord qu'une variable aléatoire de loi exponentielle est positive. On peut donc supposer que $x_j \geq 0$, sinon il ne fait aucun sens de supposer que X_j est positive. La vraisemblance s'écrit donc

$$\begin{aligned} V(\lambda) &= \prod_{j=1}^n \lambda e^{-\lambda x_j} \\ &= \lambda^n \cdot e^{-\lambda \sum_{j=1}^n x_j}, \end{aligned}$$

et la log-vraisemblance

$$v(\lambda) = n \log \lambda - \lambda \sum_{j=1}^n x_j.$$

On a donc

$$v'(\lambda) = \frac{n}{\lambda} - \sum_{j=1}^n x_j,$$

et

$$\begin{aligned} v'(\lambda) = 0 &\iff \lambda = \frac{n}{\sum_{j=1}^n x_j} \\ &\iff \lambda = \hat{\lambda}_n. \end{aligned}$$

Vérifions que $\hat{\lambda}_n$ est bien un maximum de $v(\lambda)$:

$$v''(\lambda) = -\frac{n}{\lambda^2} < 0,$$

et on a alors

$$MV_\lambda = \hat{\lambda}_n.$$

Cela peut être une raison de préférer l'estimateur $\hat{\lambda}_n$ à $\tilde{\lambda}_n$.

Exemple 1.4.4. Voici un exemple où l'estimateur du maximum de vraisemblance ne s'obtient pas de manière évidente par la méthode des moments. Supposons que $X_j \rightsquigarrow U([0; \theta])$ suive une loi uniforme sur l'intervalle $[0; \theta]$, c'est à dire admette pour densité

$$f_\theta = \frac{1}{\theta} \mathbb{1}_{[0; \theta]}.$$

En particulier la variable aléatoire X_j prend ses valeurs presque sûrement dans $[0; \theta]$. De même qu'à l'exemple précédent, on peut supposer que $x_j \in [0; \theta]$, sinon il serait idiot de supposer que X_j prend ses valeurs presque sûrement dans $[0; \theta]$. La vraisemblance s'écrit donc

$$V(\theta) = \theta^{-n}.$$

La fonction $\theta \mapsto \theta^{-n}$ est décroissante, donc son maximum est atteint pour la plus petite valeur de θ possible. Nous savons par hypothèse que $\theta \geq x_j$ pour tout j , soit $\theta \geq \max(x_1, \dots, x_n)$. Donc la plus petite valeur possible de θ est $\max(x_1, \dots, x_n)$, c'est à dire

$$MV_\theta = \max(x_1, \dots, x_n).$$

On peut aussi construire un estimateur de θ par la méthode des moments. Pour cela on calcule

$$E(X_j) = \frac{1}{\theta} \int_0^\theta x dx = \frac{\theta}{2}.$$

Ainsi, la variable aléatoire

$$\hat{\theta}_n = 2\hat{m}_1$$

est un estimateur fortement consistant de θ . On sait de plus par le TCL⁹ que la vitesse de convergence de $\hat{\theta}_n$ est \sqrt{n} . D'après l'exercice 1.3, l'estimateur MV_θ a une vitesse de convergence de n , ce qui en fait un estimateur beaucoup plus performant que $\hat{\theta}_n$.

La proposition suivante permet de faciliter parfois le calcul d'un estimateur du maximum de vraisemblance.

9. On vérifie facilement que X_j est dans L^2 .

Proposition 1.4.5. 1. *Transformation des observations : Si $g : \mathbb{R} \rightarrow \mathbb{R}$ est un C^1 -difféomorphisme et $Y_j = g(X_j)$, alors les estimateurs du maximum de vraisemblance de Y_j et de X_j coïncident.*

2. *Transformation du paramètre : Si g est une bijection et $\eta = g(\theta)$, alors*

$$MV_\eta = g(MV_\theta).$$

Démonstration. Nous donnons une démonstration dans le cas où X_j admet une densité f_θ . Le cas d'une variable aléatoire discrète est laissé au lecteur.

1. On note h_θ la densité de Y_j . Étant donné une fonction borélienne ϕ quelconque, on a

$$\begin{aligned} E(\phi(Y_j)) &= E(\phi \circ g(X_j)) \\ &= \int_{-\infty}^{+\infty} \phi \circ g(x) \cdot f_\theta(x) dx. \end{aligned}$$

En appliquant le changement de variable $y = g(x)$, on obtient

$$E(\phi(Y_j)) = \int_{-\infty}^{+\infty} \phi(y) \cdot \frac{f_\theta \circ g^{-1}(y)}{|g'(g^{-1}(y))|} dy,$$

d'où l'on déduit que Y_j a pour densité

$$h_\theta(x) = \frac{f_\theta \circ g^{-1}(y)}{|g'(g^{-1}(y))|}.$$

La vraisemblance de l'observation (y_1, \dots, y_n) est donc

$$\begin{aligned} V_Y(\theta) &= \prod_{j=1}^n \frac{f_\theta \circ g^{-1}(y_j)}{|g'(g^{-1}(y_j))|} \\ &= \frac{1}{\prod_{j=1}^n |g'(g^{-1}(y_j))|} \cdot \prod_{j=1}^n f_\theta \circ g^{-1}(y_j) \\ &= \frac{1}{\prod_{j=1}^n |g'(g^{-1}(y_j))|} \cdot \prod_{j=1}^n f_\theta(x_j) \\ &= \frac{1}{\prod_{j=1}^n |g'(g^{-1}(y_j))|} \cdot V_X(\theta). \end{aligned}$$

La vraisemblance de (y_1, \dots, y_n) est donc égale à une constante (relativement à θ) fois la vraisemblance de (x_1, \dots, x_n) . En particulier, les deux fonctions atteignent leur maximum pour les mêmes valeurs de θ .

2. C'est une tautologie : la vraisemblance de X_i pour le paramètre $\eta = g(\theta)$ est égale à $V \circ g^{-1}(\eta)$, où $V(\theta)$ est la vraisemblance de X_i pour le paramètre θ .

☺

Exemple 1.4.6. Supposons que $X_j \rightsquigarrow \text{Exp}(\lambda)$ avec $\lambda \in \mathbb{R}^+$. On a vu à l'exemple 1.4.3 que l'estimateur du maximum de vraisemblance de X_j est donné par

$$MV_\lambda = \frac{1}{\hat{m}_1} = \frac{n}{\sum_{j=1}^n X_j}.$$

On déduit alors de la proposition 1.4.5 que l'estimateur du maximum de vraisemblance de $Y_j = \frac{1}{X_j}$ est donné par

$$MV_\lambda = \frac{n}{\sum_{j=1}^n X_j} = \frac{n}{\sum_{j=1}^n \frac{1}{Y_j}},$$

et que l'estimateur du maximum de vraisemblance de X_j si $X_j \rightsquigarrow \text{Exp}(\frac{1}{\eta})$ est

$$MV_\eta = \frac{1}{MV_\lambda} = \hat{m}_1.$$

Nous avons vu à la section 1.1 qu'un estimateur doit au minimum être consistant pour être intéressant. Dans ce cours et les exercices, il est en général facile de montrer à la main que l'estimateur du maximum de vraisemblance est consistant. Il existe néanmoins des conditions générales permettant d'assurer la consistance forte de l'estimateur du maximum de vraisemblance en présence d'une densité continue f_θ . Pour énoncer ce résultat, nous avons besoin d'introduire quelques notations supplémentaires :

- θ_0 désigne une valeur particulière du paramètre θ ;
- $l_\theta(x) = \log f_\theta(x)$;
- $g(\theta) = \int l_\theta(x) f_{\theta_0}(x) dx = E_{\theta_0}(l_\theta)$.

À la dernière ligne, la notation $E_{\theta_0}(l_\theta)$ indique que l'on calcule l'espérance de la fonction l_θ pour la mesure de probabilité P_{θ_0} .

Théorème 1.4.7. *On suppose que X_j admet une densité f_θ , et que pour tout θ_0 ,*

1. $\forall \theta, E_{\theta_0}(|l_\theta|) = \int |l_\theta(x)| f_{\theta_0}(x) dx < +\infty$ (la fonction g est donc bien définie) ;
2. Si $(\theta_n)_{n \in \mathbb{N}}$ est une suite réelle telle que $g(\theta_n) \xrightarrow{n \rightarrow +\infty} g(\theta_0)$, alors $\theta_n \xrightarrow{n \rightarrow +\infty} \theta_0$;
3. $\sup_\theta \left| \frac{1}{n} \sum_{j=1}^n l_\theta(X_j) - g(\theta) \right| \xrightarrow{n \rightarrow +\infty} 0$ presque sûrement pour la mesure de probabilité P_{θ_0} .

Alors un estimateur du maximum de vraisemblance de X_j est fortement consistant.

Démonstration. On notera que l'hypothèse 2 implique en particulier que

$$f_\theta \stackrel{ps}{=} f_{\theta_0} \implies \theta = \theta_0.$$

Étape 1. On montre d'abord que θ_0 est l'unique valeur pour laquelle la fonction $g(\theta)$ atteint son maximum. On a pour tout θ

$$\begin{aligned} g(\theta) - g(\theta_0) &= E_{\theta_0}(l_\theta - l_{\theta_0}) \\ &= E_{\theta_0}(\log f_\theta - \log f_{\theta_0}) \\ &= E_{\theta_0}\left(\log \frac{f_\theta}{f_{\theta_0}}\right). \end{aligned}$$

La fonction $x \mapsto \log x$ est concave, donc l'inégalité de Jensen concave donne

$$g(\theta) - g(\theta_0) \leq \log E_{\theta_0}\left(\frac{f_\theta}{f_{\theta_0}}\right).$$

Or

$$\begin{aligned} E_{\theta_0}\left(\frac{f_\theta}{f_{\theta_0}}\right) &= \int \frac{f_\theta(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) dx \\ &= \int f_\theta(x) dx \\ &= 1, \end{aligned}$$

puisque f_θ est une densité de probabilité. On en déduit alors

$$g(\theta) - g(\theta_0) \leq 0,$$

i.e. la fonction g atteint son maximum en θ_0 . Si de plus $g(\theta) = g(\theta_0)$, alors l'inégalité de Jensen ci-dessus est une égalité, c'est à dire

$$E_{\theta_0}\left(\log \frac{f_\theta}{f_{\theta_0}}\right) = \log E_{\theta_0}\left(\frac{f_\theta}{f_{\theta_0}}\right).$$

La condition d'égalité dans l'inégalité de Jensen nous assure alors qu'il existe une constante $c \in \mathbb{R}$ telle que $f_\theta = cf_{\theta_0}$ presque sûrement. En intégrant les deux côtés de cette égalité et en utilisant le fait que f_θ et f_{θ_0} sont des densités de probabilités, on obtient $c = 1$. Donc $f_\theta = f_{\theta_0}$ presque sûrement, et $\theta = \theta_0$ par l'hypothèse 2.

Étape 2. On montre qu'un estimateur du maximum de vraisemblance MV_θ converge presque sûrement vers θ_0 si P_{θ_0} est la loi de probabilité de X_j . Toujours d'après l'hypothèse 2, il suffit de montrer pour cela que $g(MV_\theta)$ converge presque sûrement vers $g(\theta_0)$. Par l'étape 1, on sait que $g(\theta_0) - g(\theta) \geq 0$.

Remarquons que la vraisemblance logarithmique est par définition

$$v(\theta) = \sum_{j=1}^n l_{\theta}(X_j),$$

en particulier

$$\forall \theta, \sum_{j=1}^n l_{MV_{\theta}}(X_j) \geq \sum_{j=1}^n l_{\theta}(X_j).$$

Cette inégalité nous permet d'écrire

$$\begin{aligned} g(\theta_0) - g(MV_{\theta_0}) &= g(\theta_0) - \frac{1}{n} \sum_{j=1}^n l_{MV_{\theta_0}}(X_j) + \frac{1}{n} \sum_{j=1}^n l_{MV_{\theta_0}}(X_j) - g(MV_{\theta_0}) \\ &\leq g(\theta_0) - \frac{1}{n} \sum_{j=1}^n l_{\theta_0}(X_j) + \frac{1}{n} \sum_{j=1}^n l_{MV_{\theta_0}}(X_j) - g(MV_{\theta_0}) \\ &\leq \left| g(\theta_0) - \frac{1}{n} \sum_{j=1}^n l_{\theta_0}(X_j) \right| + \left| \frac{1}{n} \sum_{j=1}^n l_{MV_{\theta_0}}(X_j) - g(MV_{\theta_0}) \right| \\ &\leq 2 \cdot \sup_{\theta} \left| \frac{1}{n} \sum_{j=1}^n l_{\theta}(X_j) - g(\theta) \right|. \end{aligned}$$

Puisque $g(\theta_0) - g(MV_{\theta_0}) \geq 0$, on obtient donc

$$|g(\theta_0) - g(MV_{\theta_0})| \leq 2 \cdot \sup_{\theta} \left| \frac{1}{n} \sum_{j=1}^n l_{\theta}(X_j) - g(\theta) \right| \xrightarrow[ps]{} 0.$$

On a ainsi montré que $g(MV_{\theta_0})$ converge presque sûrement vers $g(\theta_0)$, impliquant que MV_{θ_0} converge presque sûrement vers θ_0 . \odot

Sous certaines hypothèses techniques supplémentaires, il est possible de donner un énoncé de type TCL pour l'estimateur du maximum de vraisemblance. Nous n'explicitons pas ces conditions ici afin de ne pas obscurcir la signification du théorème. Précisons simplement que le rôle de ces hypothèses manquantes est justifier chaque étape de la preuve du théorème¹⁰. Le lecteur intéressé est chaudement invité à les expliciter.

Théorème 1.4.8. *Sous les hypothèses que la preuve présentée ci-dessous est correcte, on a*

$$P_{X_j} = P_{\theta_0} \implies \sqrt{n}(MV_{\theta_0} - \theta_0) \xrightarrow[Loi]{} \mathcal{N}(0, I(\theta_0)^{-1}),$$

10. Écrire la démonstration avant le théorème est une pratique courante en mathématiques. Les cours racontant en général l'histoire à l'envers, cet aspect de la recherche est rarement évoqué.

où

$$I(\theta) = - \int \frac{\partial^2 l_\theta}{\partial^2 \theta}(x) f_{\theta_0}(x) dx = -E_{\theta_0} \left(\frac{\partial^2 l_\theta}{\partial^2 \theta} \right)$$

Démonstration. On note $s_\theta = \frac{\partial l_\theta}{\partial \theta}$, et on considère la fonction $s : \theta \mapsto s_\theta$. Un développement limité à l'ordre 1 en θ_0 de la fonction s nous donne

$$s_{MV_\theta} = s_{\theta_0} + (MV_\theta - \theta_0) \frac{ds_\theta}{d\theta} \Big|_{\theta=\theta_0} + (MV_\theta - \theta_0) \cdot \phi(MV_\theta - \theta_0),$$

avec ϕ une fonction vérifiant $\lim_{t \rightarrow 0} \phi(t) = 0$. Par définition de s , on a aussi

$$\frac{ds_\theta}{d\theta} = \frac{\partial^2 l_\theta}{\partial^2 \theta},$$

et donc

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=MV_\theta}(X_j) &= \frac{1}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0}(X_j) \\ &+ (MV_\theta - \theta_0) \cdot \left(\frac{1}{n} \sum_{j=1}^n \frac{\partial^2 l_\theta}{\partial^2 \theta} \Big|_{\theta=\theta_0}(X_j) + \phi(MV_\theta - \theta_0) \right). \end{aligned}$$

Or MV_θ est la valeur de θ pour laquelle la fonction $v(\theta) = \sum_{j=1}^n l_\theta(X_j)$ est maximum, ce qui nous donne

$$\sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=MV_\theta}(X_j) = \frac{\partial}{\partial \theta} \left(\sum_{j=1}^n l_\theta(X_j) \right) \Big|_{\theta=MV_\theta} = 0.$$

On peut donc écrire

$$MV_\theta - \theta_0 = \frac{1}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0}(X_j) \times \frac{-1}{\frac{1}{n} \sum_{j=1}^n \frac{\partial^2 l_\theta}{\partial^2 \theta} \Big|_{\theta=\theta_0}(X_j) + \phi(MV_\theta - \theta_0)},$$

ou encore

$$\sqrt{n}(MV_\theta - \theta_0) = \frac{\sqrt{n}}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0}(X_j) \times \frac{-1}{\frac{1}{n} \sum_{j=1}^n \frac{\partial^2 l_\theta}{\partial^2 \theta} \Big|_{\theta=\theta_0}(X_j) + \phi(MV_\theta - \theta_0)}.$$

Il ne nous reste plus qu'à calculer la limite en loi du terme de gauche. On sait d'après le théorème 1.4.7 que MV_θ converge presque sûrement vers θ_0 , ce qui implique

$$\phi(MV_\theta - \theta_0) \xrightarrow{ps} 0.$$

La LFGN nous dit que

$$\frac{1}{n} \sum_{j=1}^n \frac{\partial^2 l_\theta}{\partial^2 \theta} \Big|_{\theta=\theta_0}(X_j) \xrightarrow{ps} E_{\theta_0} \left(\frac{\partial^2 l_\theta}{\partial^2 \theta} \Big|_{\theta=\theta_0} \right) = -I(\theta_0),$$

d'où l'on déduit

$$\frac{-1}{\frac{1}{n} \sum_{j=1}^n \frac{\partial^2 l_\theta}{\partial^2 \theta} \Big|_{\theta=\theta_0}(X_j) + \phi(MV_\theta - \theta_0)} \xrightarrow{ps} I(\theta_0)^{-1}.$$

Toujours par la LFGN, on a

$$\frac{1}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0}(X_j) \xrightarrow{ps} E_{\theta_0} \left(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0} \right).$$

Rappelons que g désigne toujours la fonction définie avant l'énoncé du théorème 1.4.7.

On peut alors calculer explicitement la limite précédente :

$$\begin{aligned} E_{\theta_0} \left(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0} \right) &= \frac{\partial}{\partial \theta} E_{\theta_0}(l_\theta) \Big|_{\theta=\theta_0} \\ &= g'(\theta_0) \\ &= 0, \end{aligned}$$

puisque θ_0 est le maximum de la fonction g . Le TCL nous assure maintenant que

$$\frac{\sqrt{n}}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0}(X_j) \xrightarrow{Loi} \mathcal{N}(0, \text{Var}(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0})),$$

et il ne nous reste donc plus qu'à calculer

$$\begin{aligned} \text{Var}(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0}) &= E_{\theta_0} \left(\left(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 \right) - E_{\theta_0} \left(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 \\ &= E_{\theta_0} \left(\left(\frac{\partial l_\theta}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 \right). \end{aligned}$$

Par définition de $l_\theta = \log f_\theta$, on a

$$\frac{\partial l_\theta}{\partial \theta} = \frac{\partial f_\theta}{\partial \theta} \times \frac{1}{f_\theta},$$

et

$$\begin{aligned} \frac{\partial^2 l_\theta}{\partial \theta^2} &= \frac{\partial^2 f_\theta}{\partial \theta^2} \times \frac{1}{f_\theta} - \left(\frac{\partial f_\theta}{\partial \theta} \times \frac{1}{f_\theta} \right)^2 \\ &= \frac{\partial^2 f_\theta}{\partial \theta^2} \times \frac{1}{f_\theta} - \left(\frac{\partial l_\theta}{\partial \theta} \right)^2. \end{aligned}$$

On obtient alors enfin

$$\begin{aligned}
 \text{Var}\left(\frac{\partial l_\theta}{\partial \theta}\Big|_{\theta=\theta_0}\right) &= I(\theta_0) + E_{\theta_0}\left(\frac{\partial^2 f_\theta}{\partial \theta^2}\Big|_{\theta=\theta_0} \times \frac{1}{f_{\theta_0}}\right) \\
 &= I(\theta_0) + \int \frac{\partial^2 f_\theta}{\partial \theta^2}\Big|_{\theta=\theta_0}(x) dx \\
 &= I(\theta_0) + \frac{\partial^2}{\partial \theta^2} \int f_\theta(x) dx \Big|_{\theta=\theta_0} \\
 &= I(\theta_0) + \frac{\partial^2}{\partial \theta^2}(1)\Big|_{\theta=\theta_0} \\
 &= I(\theta_0).
 \end{aligned}$$

En résumé, nous avons montré

$$\frac{\sqrt{n}}{n} \sum_{j=1}^n \frac{\partial l_\theta}{\partial \theta}\Big|_{\theta=\theta_0}(X_j) \xrightarrow[\text{Loi}]{} \mathcal{N}(0, I(\theta_0)),$$

et le théorème de Slutsky nous assure enfin que

$$\sqrt{n}(MV_\theta - \theta_0) \xrightarrow[\text{Loi}]{} I(\theta_0)^{-1} \mathcal{N}(0, I(\theta_0)) = \mathcal{N}(0, I(\theta_0)^{-1}).$$

Ouf!

☺

1.5 Exercices

Exercice 1.1. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. telles que $\text{Var}(X_1) = \sigma^2 < \infty$. On pose $E(X_1) = \theta$ et $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

1. Montrer que $\hat{\theta}_n$ est un estimateur sans biais de θ .
2. Montrer qu'il est consistant en précisant sa vitesse de convergence.
3. Montrer qu'il converge également en moyenne quadratique.

Exercice 1.2. Soit X la variable aléatoire réelle égale au nombre de pannes que subit un certain type d'appareil électroménager. On suppose que X suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}^+$ i.e.

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

Soit $n \in \mathbb{N}^*$ et (X_1, \dots, X_n) un n -échantillon de X . Ici $\lambda > 0$ est un réel inconnu que l'on souhaite estimer à l'aide de (X_1, \dots, X_n) . On pose $S_n = \sum_{i=1}^n X_i$ et $\bar{X}_n = \frac{1}{n} S_n$.

1. Déterminer la loi de S_n , en déduire que \bar{X}_n est un estimateur sans biais de λ .

2. Dorénavant, on cherche à estimer la probabilité qu'il n'y ait aucune panne. Cette probabilité est notée θ .
 - (a) Exprimer θ en fonction de λ .
 - (b) On pose $T_n = \exp(-\bar{X}_n)$. Calculer $E(T_n)$ et montrer que T_n n'est pas un estimateur sans biais de θ . Est-il asymptotiquement sans biais ?
 - (c) On pose $\hat{\theta}_n = (1 - \frac{1}{n})^{S_n}$. Montrer que $\hat{\theta}_n$ est un estimateur sans biais de θ
 - (d) Calculer la variance de $\hat{\theta}_n$. Est-ce un estimateur consistant ?

Exercice 1.3. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. de loi uniforme sur $[0, \theta]$.

1. Déterminer la loi de $M_n = \max\{X_1, \dots, X_n\}$.
2. Montrer que M_n converge en probabilité vers θ .
3. Montrer que M_n converge presque sûrement vers θ .
4. Déterminer la limite en loi de $n(M_n - \theta)$ et conclure sur la vitesse de convergence de M_n .
5. Proposer une correction \widehat{M}_n de M_n qui soit sans biais. Comparer les erreurs quadratiques moyennes des deux estimateurs.

Exercice 1.4. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles indépendantes et de même loi dont la densité f_θ est définie pour $\theta \in \mathbb{R}_+^*$ par :

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer la loi de $m_n = \min\{X_1, \dots, X_n\}$.
2. Montrer que m_n converge en moyenne quadratique vers θ .
3. Étudier la convergence en loi de $n(m_n - \theta)$.

Exercice 1.5. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. de loi uniforme sur $[\theta, 2\theta]$ où $\theta > 0$. On note $m_n = \min\{X_1, \dots, X_n\}$ et $M_n = \max\{X_1, \dots, X_n\}$.

1. Déterminer les densités de probabilité de m_n et M_n .
2. En déduire $E(m_n)$ et $E(M_n)$.
On admet dans la suite que :

$$\text{Var}(m_n) = \text{Var}(M_n) = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

3. Afin d'estimer θ on propose $\hat{\theta}_n = M_n/2$. Étudier son biais, son erreur quadratique moyenne et déduire sa convergence dans L^2 .

4. Quel estimateur de θ peut-on construire à partir de m_n ? Est-il sans biais? Si non, le modifier afin qu'il le devienne.
5. Comparer l'erreur quadratique moyenne de $\hat{\theta}_n$ et de l'estimateur sans biais construit ci-dessus.

Exercice 1.6. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles dont la loi est définie pour $n \in \mathbb{N}$ par :

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = n) = \frac{1}{n}.$$

1. Montrer que $(X_n)_{n \geq 1}$ converge en probabilité, mais pas en moyenne quadratique, vers zéro quand n tend vers l'infini.
2. Soit $(Y_n)_{n \geq 1}$ une suite de v.a. de même loi $\mathcal{N}(0, 1)$ et indépendantes des v.a. (X_n) . Étudier la convergence en loi de $Z_n = X_n + Y_n$ et la limite de $Var(Z_n)$ quand n tend vers l'infini.

Exercice 1.7. Soit X une variable aléatoire réelle de loi

$$P(X = 0) = \frac{a}{a+1}, \quad P(X = 1) = \frac{1}{a+1},$$

$n \in \mathbb{N}^*$ et (X_1, \dots, X_n) un n -échantillon de même loi que X . Ici $a > 0$ est un réel inconnu que l'on souhaite estimer à l'aide de (X_1, \dots, X_n) . Déterminer un estimateur de a par la méthode des moments.

Exercice 1.8. Soit X une variable aléatoire réelle de densité

$$f(x) = p \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + (1-p) \frac{1}{\sqrt{4\pi}} e^{-\frac{x^2}{4}},$$

$n \in \mathbb{N}^*$ et (X_1, \dots, X_n) un n -échantillon de même loi que X . Ici $p \in]0, 1[$ est un réel inconnu que l'on souhaite estimer à l'aide de (X_1, \dots, X_n) . Déterminer un estimateur de p par la méthode des moments.

Exercice 1.9. Soit X une variable aléatoire réelle de loi

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

$n \in \mathbb{N}^*$ et (X_1, \dots, X_n) un n -échantillon de X . Ici $\lambda > 0$ est un réel inconnu que l'on souhaite estimer à l'aide de (X_1, \dots, X_n) . Déterminer l'estimateur du maximum de vraisemblance de λ .

Exercice 1.10. Soit n variables aléatoires réelles X_1, \dots, X_n indépendantes et identiquement distribuées de loi $\mathcal{N}(0, 1)$. On définit Y_j par

$$Y_j = \frac{\theta}{j^\alpha} + \sigma X_j,$$

où $\alpha \geq 0$ et $\sigma > 0$ sont des réels connus. On veut estimer le réel $\theta > 0$.

1. Déterminer la vraisemblance et la log-vraisemblance de (Y_1, \dots, Y_n) .
2. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ .
3. Est-il sans biais? Déterminer les valeurs de α pour lesquelles l'estimateur $\hat{\theta}_n$ converge.

Chapitre 2

Intervalles de confiance

Au chapitre précédent, nous avons calculé une valeur supposée approchée du paramètre θ . Un intervalle de confiance est un intervalle où la valeur de θ a de bonnes chances de se trouver, permettant ainsi “d’épaissir” un peu cette valeur approchée. Un tel intervalle offre donc une réponse à la question de la signification du signe “ \simeq ” dans l’affirmation “ $\theta \simeq 0,6$ ” de l’exemple du préambule.

2.1 Quantiles

Commençons par quelques rappels sur les fonctions de répartition et les quantiles d’une variable aléatoire.

Définition 2.1.1. Soit X une variable aléatoire réelle. La fonction de répartition de X est la fonction

$$F_X : \mathbb{R} \longrightarrow [0; 1] \\ x \longmapsto P(X \leq x) .$$

On rappelle sans démonstration quelques propriétés des fonctions de répartition.

Proposition 2.1.2. Soit X une variable aléatoire réelle.

1. La fonction F_X est croissante, continue à droite (i.e. $\lim_{x \rightarrow x_0^+} F_X(x) = F_X(x_0)$) et limité à gauche (plus précisément $\lim_{x \rightarrow x_0^-} F_X(x) = P(X < x_0)$).
2. La fonction F_X a pour limite en $\pm\infty$

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

3. Si F_X est continue et C^1 par morceaux, alors X admet pour densité $f_X = F'_X$, i.e.

$$P(X \leq x) = \int_{-\infty}^x f_X(x) dx.$$

Il découle de la propriété 1. ci-dessus que F_X est continue en x_0 si et seulement si $P(X = x_0) = 0$.

Définition 2.1.3. Soit X une variable aléatoire réelle et $\alpha \in]0; 1[$. Un quantile d'ordre α de X est un réel q_α satisfaisant

$$P(X \leq q_\alpha) \geq \alpha \quad \text{et} \quad P(X \geq q_\alpha) \geq 1 - \alpha.$$

Le fait que q_α est un quantile d'ordre α de X admet une interprétation graphique. Considérons le *graphe complété* $\tilde{\Gamma}(F_X)$ de la fonction de répartition F_X , c'est à dire le graphe de F_X auquel on a ajouté un segment vertical à chaque point de discontinuité de F_X afin d'obtenir un ensemble $\tilde{\Gamma}(F_X)$ homéomorphe à un segment (voir figures 2.1 et 2.2). Le réel q_α est alors un quantile d'ordre α de X si et seulement si le point (q_α, α) est sur $\tilde{\Gamma}(F_X)$. Avec cette interprétation graphique, on déduit facilement de la proposition 2.1.2 l'existence d'un quantile d'ordre α pour tout α dans $]0; 1[$.

Si $P(X = q_\alpha) = 0$, alors

$$P(X < q_\alpha) = P(X \leq q_\alpha) \geq \alpha \quad \text{et} \quad P(X > q_\alpha) = P(X \geq q_\alpha) \geq 1 - \alpha.$$

Puisque l'on a aussi

$$1 = P(X \neq x_0) = P(X < q_\alpha) + P(X > q_\alpha),$$

on a alors nécessairement

$$P(X \leq q_\alpha) = \alpha \quad \text{et} \quad P(X \geq q_\alpha) = 1 - \alpha.$$

Ainsi, si X est une fonction définie sur une population, alors q_α est la valeur de X vérifiant que $X \leq x_0$ pour une proportion α de la population, et $X \geq x_0$ pour une proportion $1 - \alpha$ de la population.

Exemple 2.1.4. Si $X \rightsquigarrow U([0; 1])$, alors

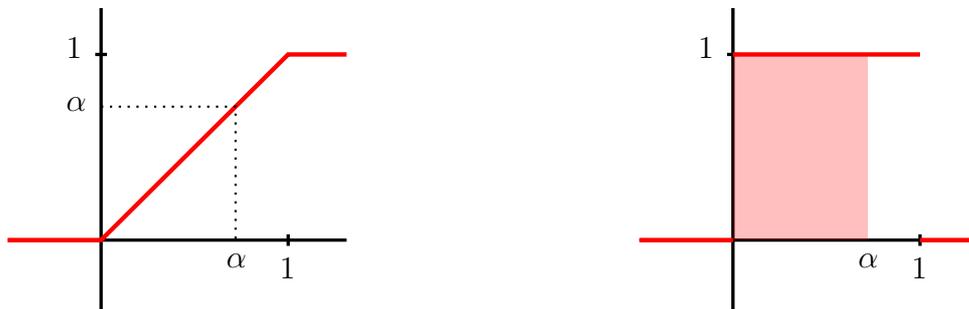
$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in [0; 1] \\ 1 & \text{si } x \geq 1 \end{cases} \quad \text{et} \quad f_X = F'_X = \mathbb{1}_{[0;1]}.$$

Le graphe de ces deux fonctions sont représentés à la figure 2.1. On a alors que α est un quantile d'ordre α de X . On a de plus unicité de ce quantile. Remarquons que par définition de f_X , les quantiles d'ordre α sont caractérisés par

$$\int_{-\infty}^{q_\alpha} f_X(x) dx \geq \alpha \quad \text{et} \quad \int_{q_\alpha}^{+\infty} f_X(x) dx \geq 1 - \alpha,$$

ce qui se traduit dans le cas présent par l'équation

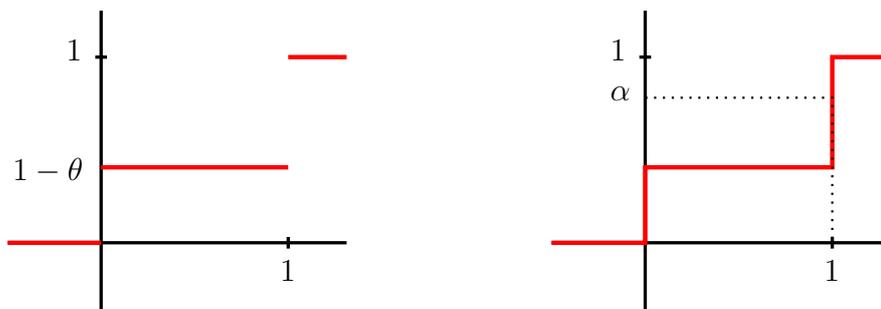
$$\int_{-\infty}^{q_\alpha} f_X(x) dx = \alpha.$$

a) Graphe de $F_{U([0;1])}$, égal ici à $\tilde{\Gamma}(F_{U([0;1])})$.b) Graphe de $f_{U([0;1])}$, la partie coloriée est d'aire α .FIGURE 2.1 – Quantiles de la loi uniforme $U([0; 1])$.

Exemple 2.1.5. Gardons bien en tête qu'il n'y a pas nécessairement unicité d'un quantile lorsque F_X n'est pas strictement croissante. Prenons un exemple concret et supposons que $X \rightsquigarrow \text{Ber}(\theta)$. On a dans ce cas la fonction de répartition

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - \theta & \text{si } x \in [0; 1[\\ 1 & \text{si } x \geq 1 \end{cases} .$$

Le graphe de la fonction F_X et l'ensemble $\tilde{\Gamma}(F_X)$ sont représentés à la figure 2.2. On

a) Graphe de $F_{\text{Ber}(\theta)}$.b) L'ensemble $\tilde{\Gamma}(F_{\text{Ber}(\theta)})$.FIGURE 2.2 – Quantiles de la loi uniforme $\text{Ber}(\theta)$.

voit alors graphiquement que

- tout réel $q \in [0; 1]$ est un quantile d'ordre $1 - \theta$;
- si $\alpha \in]0; 1 - \theta[$, alors 0 est l'unique quantile d'ordre α ;

— si $\alpha \in]1 - \theta, 1[$, alors 1 est l'unique quantile d'ordre α .

Vérifions cette observation graphique par un calcul explicite. Pour tout $q \in [0; 1[$, on a

$$P(X \leq q) = P(X = 0) = 1 - \theta \quad \text{et} \quad P(X \geq q) = P(X = 1) = \theta,$$

d'où l'on déduit que q est un quantile d'ordre $1 - \theta$. Pour tout $\alpha \in]1 - \theta; 1[$, on a

$$P(X \leq 1) = 1 \geq \alpha \quad \text{et} \quad P(X \geq 1) = \theta \geq 1 - \alpha.$$

Ainsi 1 est aussi un quantile d'ordre α , et c'est le seul puisque $P(X \leq q) = 1 - \theta < \alpha$ si $q \in [0; 1[$. De même pour tout $\alpha \in]0; 1 - \theta[$, on a

$$P(X \leq 0) = 1 - \theta \geq \alpha \quad \text{et} \quad P(X \geq 0) = 1 \geq 1 - \alpha.$$

Ainsi 0 est un quantile d'ordre α , et c'est le seul puisque $P(X \geq q) = \theta < 1 - \alpha$ si $q \in]0; 1[$.

On peut assurer l'unicité des quantiles des variables aléatoires à densité continue.

Lemme 2.1.6. Soit X une variable aléatoire réelle admettant une densité f_X vérifiant :

- $f_X^{-1}(]0; +\infty[)$ est un intervalle d'extrémités $a < b$;
- f_X est continue sur $]a; b[$.

Alors X admet un unique quantile d'ordre α pour tout $\alpha \in]0; 1[$.

Démonstration. L'application

$$\begin{array}{ccc} F_X|_{]a; b[} &]a; b[& \longrightarrow &]0; 1[\\ x & \longmapsto & F_X(x) = \int_a^x f_X(y) dy \end{array}$$

est continue et strictement croissante. Elle vérifie de plus

$$\lim_{x \rightarrow a} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow b} F_X(x) = 1,$$

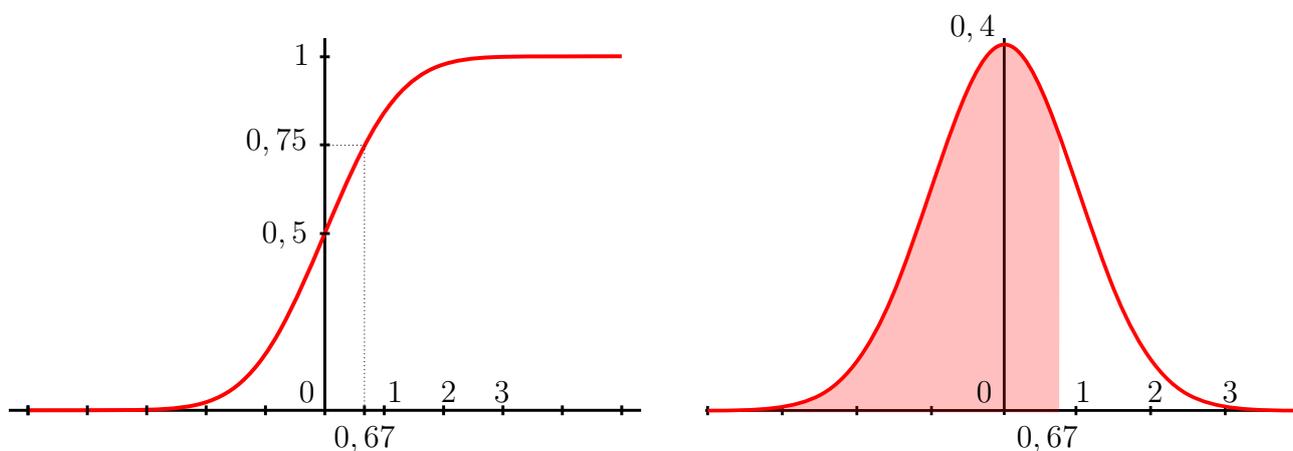
c'est donc une bijection de $]a; b[$ sur $]0; 1[$. ⊙

Exemple 2.1.7. Un exemple **fondamental** de quantiles concerne la loi normale centrée réduite $\mathcal{N}(0, 1)$. Rappelons qu'une variable aléatoire X est de loi $\mathcal{N}(0, 1)$ si sa fonction de répartition est

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Dans toute la suite du texte, on notera $q_\alpha^{\mathcal{N}}$ le quantile d'ordre α de la loi $\mathcal{N}(0, 1)$. À part $q_{0,5}^{\mathcal{N}} = 0$, les quantiles de $\mathcal{N}(0, 1)$ ne se calculent en général pas explicitement. Un ordinateur peut cependant en calculer de très bonnes approximations très facilement. En pratique, on utilisera les approximations données dans le tableau suivant. On remarque qu'on a la relation

$$q_{1-\alpha}^{\mathcal{N}} = -q_\alpha^{\mathcal{N}}.$$

a) Graphe de $F_{\mathcal{N}(0,1)}$.b) Graphe de $f_{\mathcal{N}(0,1)}$,
la partie coloriée est d'aire 0,75.FIGURE 2.3 – Quantiles de la loi normale centrée réduite $\mathcal{N}(0, 1)$.

α	0,025	0,05	0,25	0,5	0,75	0,95	0,975
$q_\alpha^{\mathcal{N}}$	-1,96	-1,64	-0,67	0	0,67	1,64	1,96

TABLE 2.1 – Quantiles approchés de la loi normale centrée réduite $\mathcal{N}(0, 1)$.

La relation précédente est vraie pour toute loi symétrique.

Proposition 2.1.8. *Si X est une variable aléatoire réelle à densité continue symétrique (i.e. $f_X(-x) = f_X(x)$), alors $-q_\alpha$ est un quantile de X d'ordre $1 - \alpha$ si et seulement si q_α est un quantile de X d'ordre α .*

Démonstration. Il s'agit d'un simple calcul :

$$\begin{aligned}
 P(X \leq -q_\alpha) &= \int_{-\infty}^{-q_\alpha} f_X(y) dy \\
 &= \int_{q_\alpha}^{+\infty} f_X(u) du \\
 &= 1 - \int_{-\infty}^{q_\alpha} f_X(u) du \\
 &= 1 - \alpha,
 \end{aligned}$$

où la deuxième ligne est obtenue par le changement de variables $u = -y$.

☺

2.2 Intervalles de confiance

Rappelons que l'on se donne un échantillon X_1, \dots, X_n de n variables aléatoires indépendantes et identiquement distribuées de loi dépendant d'un paramètre $\theta \in \mathbb{R}^k$ inconnu. On cherche ici à donner une marge d'erreur à l'estimation ponctuelle de θ vue au chapitre précédent. Pour simplifier, nous supposons dans ce chapitre que le paramètre θ à déterminer est dans \mathbb{R} .

Définition 2.2.1. Soit $\alpha \in]0; 1[$. Un intervalle aléatoire $I(X_1, \dots, X_n) \subset \mathbb{R}$ est un intervalle de confiance de niveau $1 - \alpha$ si

1. $I(X_1, \dots, X_n)$ ne dépend pas de θ , mais seulement de X_1, \dots, X_n et α ;
2. $\forall \theta, P_\theta(\theta \in I(X_1, \dots, X_n)) \geq 1 - \alpha$.

Heuristiquement, si on dispose de plusieurs observations de l'échantillon X_1, \dots, X_n , on s'attend à ce qu'une proportion $1 - \alpha$ des intervalles de confiance contienne θ , et qu'une proportion α ne le contienne pas. On ne sera donc *jamais* sûr qu'un intervalle de confiance contient bien la valeur de θ cherchée. On pourra seulement affirmer que cela est bien le cas avec une certaine probabilité fixée à l'avance. Le nombre α correspond ainsi à la marge d'erreur que le statisticien se fixe pour donner un encadrement de θ .

Par définition, un intervalle de confiance est un intervalle *aléatoire* qui varie en fonction de l'observation (x_1, \dots, x_n) . Afin d'alléger les notations, nous noterons un intervalle de confiance I plutôt que $I(X_1, \dots, X_n)$, mais la dépendance de I en les X_j est à garder en tête.

Remarque 2.2.2. On peut étendre sans difficulté la définition 2.2.1 aux cas où le paramètre θ est dans \mathbb{R}^k avec $k \geq 2$. On parlera alors de *région de confiance*.

Exemple 2.2.3. Supposons que $X_j \rightsquigarrow \mathcal{N}(\theta, 1)$ avec $\theta \in \mathbb{R}$ inconnu. Puisque $m_1 = \theta$, on sait que la moyenne empirique

$$\hat{m}_1 = \frac{1}{n} \sum_{j=1}^n X_j$$

est un estimateur fortement consistant de θ . On rappelle qu'étant donné deux variables aléatoires *indépendantes* X et Y , on a

$$X \rightsquigarrow \mathcal{N}(m, \sigma^2) \text{ et } Y \rightsquigarrow \mathcal{N}(m', \sigma'^2) \implies X + Y \rightsquigarrow \mathcal{N}(m + m', \sigma^2 + \sigma'^2).$$

On en déduit que

$$\hat{m}_1 \rightsquigarrow \mathcal{N}\left(\theta, \frac{1}{n}\right) \quad \text{et} \quad \sqrt{n} (\hat{m}_1 - \theta) \rightsquigarrow \mathcal{N}(0, 1).$$

Le point important ici est que le membre de gauche dans cette dernière identité ne dépend pas de θ . On aura donc pour tout θ

$$P_\theta(\sqrt{n} (\hat{m}_1 - \theta) \leq q_{1-\alpha}^{\mathcal{N}}) = 1 - \alpha,$$

c'est à dire

$$P_\theta(\theta \geq \hat{m}_1 - \frac{q_{1-\alpha}^{\mathcal{N}}}{\sqrt{n}}) = 1 - \alpha.$$

Autrement dit, un intervalle de confiance pour θ de niveau $1 - \alpha$ est donné par

$$I = [\hat{m}_1 - \frac{q_{1-\alpha}^{\mathcal{N}}}{\sqrt{n}}; +\infty[.$$

On est donc passé d'une estimation ponctuelle \hat{m}_1 de θ à l'intervalle I non borné. Cela nous laisse une sacré marge d'indétermination dans l'estimation de θ ... On peut cependant obtenir un intervalle de confiance borné en remarquant qu'il suffit de choisir un intervalle $[a; b]$ au dessus duquel l'aire de la densité $f_{\mathcal{N}(0,1)}$ soit égale à $1 - \alpha$, c'est à dire

$$\int_a^b f_{\mathcal{N}(0,1)}(y) dy = 1 - \alpha.$$

On voit alors graphiquement à la figure 2.3b que l'intervalle $[a; b]$ a une longueur minimale si et seulement si il est symétrique par rapport 0, c'est à dire si

$$a = -q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \quad \text{et} \quad b = q_{1-\frac{\alpha}{2}}^{\mathcal{N}}.$$

Vérifions qu'on obtient bien ainsi un intervalle de confiance pour θ de niveau $1 - \alpha$:

$$\begin{aligned} P_\theta(-q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \leq \sqrt{n} (\hat{m}_1 - \theta) \leq q_{1-\frac{\alpha}{2}}^{\mathcal{N}}) &= P_\theta(\sqrt{n} (\hat{m}_1 - \theta) \leq q_{1-\frac{\alpha}{2}}^{\mathcal{N}}) - P_\theta(\sqrt{n} (\hat{m}_1 - \theta) \leq -q_{1-\frac{\alpha}{2}}^{\mathcal{N}}) \\ &= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} \\ &= 1 - \alpha, \end{aligned}$$

la deuxième ligne étant une conséquence de la proposition 2.1.8. On a donc montré que

$$P_\theta \left(\theta \in \left[\hat{m}_1 - \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}}; \hat{m}_1 + \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \right] \right) = 1 - \alpha,$$

c'est à dire qu'un intervalle de confiance pour θ de niveau $1 - \alpha$ est aussi donné par

$$I' = \left[\hat{m}_1 - \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}}; \hat{m}_1 + \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \right].$$

L'intervalle I' a une longueur beaucoup plus petite que l'intervalle I qui est non borné. Sans information supplémentaire, on préférera donc l'intervalle I' à I .

Il ressort de l'exemple précédent qu'il n'y a en général pas unicité dans la construction d'un intervalle de confiance d'un niveau donné. En pratique lorsqu'on a le choix, on choisit un intervalle de confiance de longueur minimale.

2.3 Méthode du pivot

Construire un intervalle de confiance n'est en général pas une tâche facile. En étudiant de plus près l'exemple 2.2.3, on voit que la clef qui nous a permis de construire les intervalles I et I' est une fonction des X_j et de θ , en l'occurrence $\sqrt{n}(\hat{m}_1 - \theta)$, dont la loi $\mathcal{N}(0, 1)$ ne dépend plus de θ . Nous généralisons cette méthode dans cette section.

Définition 2.3.1. Une fonction $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ est dite pivotale si la loi de $g(\theta, X_1, \dots, X_n)$ ne dépend pas de θ .

Une fonction pivotale permet la construction d'intervalles de confiance.

Proposition 2.3.2. Soit $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction pivotale pour θ, X_1, \dots, X_n , et soit $U \subset \mathbb{R}$ tel que $P(Z \in U) = 1 - \alpha$, où Z est une variable aléatoire réelle de même loi que $g(\theta, X_1, \dots, X_n)$. Alors pour

$$I = \{\theta \mid g(\theta, X_1, \dots, X_n) \in U\},$$

on a pour tout θ

$$P_\theta(\theta \in I) = 1 - \alpha.$$

Démonstration. C'est une tautologie :

$$\begin{aligned} P_\theta(\theta \in I) &= P(g(\theta, X_1, \dots, X_n) \in U) \\ &= P(Z \in U) \\ &= 1 - \alpha. \end{aligned}$$

☺

Comme souvent avec une tautologie, il est sans plus judicieux d'illustrer la proposition 2.3.2 à travers quelques exemples élémentaires que de la justifier par une heuristique obscure. Les quatre exemples qui suivent portent sur la construction d'intervalles de confiance pour l'espérance ou la variance de la loi normale $\mathcal{N}(m, \sigma^2)$ suivant que l'autre paramètre est connu ou non. Ces exemples sont **fondamentaux** pour le reste du cours. On y verra en particulier apparaître deux lois centrales en statistiques : la loi du χ^2 et la loi de Student.

Exemple 2.3.3. Supposons que $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$ avec σ connu mais m inconnu. De même qu'à l'exemple 2.2.3, nous avons

$$\hat{m}_1 \rightsquigarrow \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \quad \text{et} \quad \sqrt{n} \frac{\hat{m}_1 - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1).$$

Autrement dit, la fonction $g(\theta, X_1, \dots, X_n) = \sqrt{n} \frac{\hat{m}_1 - m}{\sigma}$ est pivotale. Le même raisonnement qu'à l'exemple 2.2.3 nous donne alors

$$\left[\hat{m}_1 - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; \hat{m}_1 + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right].$$

comme intervalle de confiance de niveau $1 - \alpha$ pour m .

Exemple 2.3.4. Supposons que $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$ avec m connu mais σ inconnu. On sait par la LFGN que

$$\hat{\sigma}_n^2 = \hat{m}_2 - m^2 = \frac{1}{n} \sum_{j=1}^n (X_j - m)^2$$

est un estimateur fortement consistant de σ^2 . En divisant par σ^2 , on obtient

$$\begin{aligned} \frac{\hat{\sigma}_n^2}{\sigma^2} &= \frac{1}{n} \sum_{j=1}^n \left(\frac{X_j - m}{\sigma} \right)^2 \\ &= \frac{1}{n} \sum_{j=1}^n Y_j^2, \end{aligned}$$

avec $Y_j = \frac{X_j - m}{\sigma}$. Puisque $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$, on a $Y_j \rightsquigarrow \mathcal{N}(0, 1)$. Ainsi $\frac{\hat{\sigma}_n^2}{\sigma^2}$ est la moyenne arithmétique des carrés de n variables aléatoires indépendantes suivant une loi $\mathcal{N}(0, 1)$. La loi de $\frac{\hat{\sigma}_n^2}{\sigma^2}$ ne dépend donc que de n , et en particulier pas de σ .

Définition 2.3.5. Si Y_1, \dots, Y_n sont n variables aléatoires indépendantes de même loi $\mathcal{N}(0, 1)$, alors la loi de $\sum_{j=1}^n Y_j^2$ est appelée loi du χ^2 à n degrés de liberté, et est notée $\chi^2(n)$.

La loi du χ^2 à n degrés de liberté est centrale en statistiques. Prenons le temps de la regarder d'un peu plus près. On rappelle que la fonction gamma est définie sur $\mathbb{R}_{>0}$ par

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt,$$

et vérifie

$$\Gamma(n+1) = n! \quad \forall n \in \mathbb{N}.$$

Théorème 2.3.6. La loi du χ^2 à n degrés de liberté est une loi gamma $\Gamma(\frac{n}{2}, \frac{n}{2})$, c'est à dire admet pour densité la fonction

$$f_{\chi^2(n)}(x) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{\{x \geq 0\}}.$$

Peu importe la tête exacte de la densité $f_{\chi^2(n)}(x)$, ce qui compte est qu'un ordinateur peu facilement calculer une approximation aussi fine que l'on veut de n'importe lequel de ses quantiles.

Démonstration. D'après l'exercice 2.1, il suffit de montrer le théorème dans le cas $n = 1$. Puisque $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, il s'agit donc de montrer que

$$f_{\chi^2(1)}(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-\frac{x}{2}}}{\sqrt{x}} \mathbb{1}_{\{x \geq 0\}}.$$

On applique pour cela la méthode de la variable muette. Soit donc h une fonction borélienne bornée. On a alors

$$\begin{aligned} E(h(Y_1^2)) &= \int_{-\infty}^{+\infty} h(x^2) f_{\mathcal{N}(0,1)}(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} h(x^2) e^{-\frac{x^2}{2}} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} h(x^2) e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} h(y) \frac{e^{-\frac{y}{2}}}{\sqrt{y}} dy, \end{aligned}$$

où la dernière ligne est obtenue en effectuant le changement de variables $y = x^2$. \odot

Les graphes des premières fonctions $f_{\chi^2(n)}$ sont représentés à la figure 2.4. Dans toute la suite du texte, on notera $q_{\alpha}^{\chi^2(n)}$ le quantile d'ordre α de la loi $\chi^2(n)$. Si Z est une variable aléatoire suivant une loi $\chi^2(n)$, on a

$$P(Z \in [q_{\frac{\alpha}{2}}^{\chi^2(n)}; q_{1-\frac{\alpha}{2}}^{\chi^2(n)}]) = P(Z \leq q_{1-\frac{\alpha}{2}}^{\chi^2(n)}) - P(Z \leq q_{\frac{\alpha}{2}}^{\chi^2(n)}) = 1 - \alpha.$$

En revenant à la construction d'un intervalle de confiance pour σ , on obtient donc

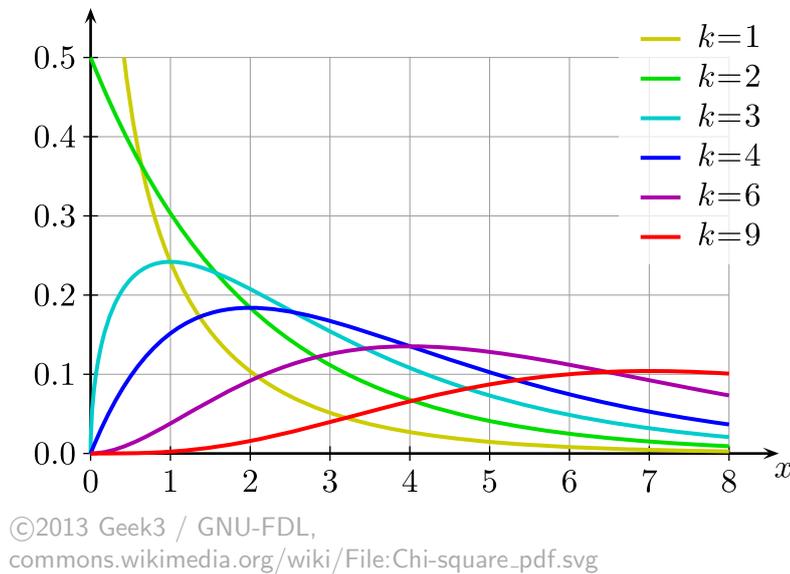
$$P\left(\frac{n \hat{\sigma}_n^2}{\sigma^2} \in [q_{\frac{\alpha}{2}}^{\chi^2(n)}; q_{1-\frac{\alpha}{2}}^{\chi^2(n)}]\right) = 1 - \alpha,$$

c'est à dire

$$P\left(\sigma^2 \in \left[\frac{n}{q_{1-\frac{\alpha}{2}}^{\chi^2(n)}} \hat{\sigma}_n^2; \frac{n}{q_{\frac{\alpha}{2}}^{\chi^2(n)}} \hat{\sigma}_n^2\right]\right) = 1 - \alpha.$$

En conclusion, nous avons montré que

$$I = \left[\frac{n}{q_{1-\frac{\alpha}{2}}^{\chi^2(n)}} \hat{\sigma}_n^2; \frac{n}{q_{\frac{\alpha}{2}}^{\chi^2(n)}} \hat{\sigma}_n^2\right]$$

FIGURE 2.4 – Densité $f_{\chi^2(k)}$ de la loi $\chi^2(k)$.

est un intervalle de confiance de niveau $1 - \alpha$ pour σ^2 .

Pour $n = 100$ et $\alpha = 0,05$, on a $q_{0,025}^{\chi^2(100)} \simeq 74,22$ et $q_{0,975}^{\chi^2(100)} \simeq 129,56$, et l'intervalle I est alors

$$I \simeq [0,76 \hat{\sigma}_n^2; 1,34 \hat{\sigma}_n^2].$$

Terminons cet exemple par une remarque asymptotique sur les lois du χ^2 à n degrés de liberté. Soit (Z_j) une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\chi^2(1)$. On calcule aisément $E(Z_j) = 1$ et $Var(Z_j) = 2$. Par la LFGN et le TCL, on a alors

$$\frac{1}{n} \sum_{j=1}^n Z_j \xrightarrow[ps]{} 1 \quad \text{et} \quad \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n Z_j - 1 \right) \xrightarrow[Loi]{} \mathcal{N}(0, 2).$$

Ainsi lorsque n est grand (en pratique si $n > 30$), on peut approximer la loi $\chi^2(n)$ par la loi $\mathcal{N}(n, 2n)$, et utiliser l'approximation

$$q_{\alpha}^{\chi^2(n)} \simeq \sqrt{2n} q_{\alpha}^{\mathcal{N}} + n.$$

Pour $n = 100$, cette approximation donne $q_{0,025}^{\chi^2(100)} \simeq 72,3$ et $q_{0,975}^{\chi^2(100)} \simeq 127,7$.

Exemple 2.3.7. Supposons que $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$ avec m et σ inconnus, et que nous voulions déterminer un intervalle de confiance pour σ . L'idée de départ est la même qu'à

l'exemple 2.3.4 en prenant en compte que m , le moment d'ordre 1, ne peut plus être explicitement utilisé puisqu'inconnu. On remplace donc l'estimateur $\hat{\sigma}_n^2$ par la variance empirique \tilde{S}_n^2 :

$$\tilde{S}_n^2 = \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \hat{m}_1)^2.$$

Après division par σ^2 , il vient

$$\begin{aligned} \frac{\tilde{S}_n^2}{\sigma^2} &= \frac{1}{n} \sum_{j=1}^n \left(\frac{X_j - \hat{m}_1}{\sigma} \right)^2 \\ &= \frac{1}{n} \sum_{j=1}^n \tilde{Y}_j^2, \end{aligned}$$

avec $\tilde{Y}_j = \frac{X_j - \hat{m}_1}{\sigma}$. La situation ressemble fort à celle de l'exemple 2.3.4, mais les \tilde{Y}_j ne sont maintenant plus indépendantes puisqu'elles vérifient la relation

$$\sum_{j=1}^n \tilde{Y}_j = 0.$$

Il est facile de voir qu'il s'agit de la seule¹ relation vérifiée par les \tilde{Y}_j . La variable aléatoire “perd” donc un degré de liberté (ou d'indépendance) par rapport au cas où les \tilde{Y}_j sont indépendantes. Cette heureuse terminologie de “degré de liberté” offre une explication heuristique² au théorème suivant, dont une démonstration est proposée à l'exercice 2.3.

Théorème 2.3.8. *Soit X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées suivant une loi $\mathcal{N}(m, \sigma^2)$, et posons*

$$\tilde{Y}_j = \frac{X_j - \hat{m}_1}{\sigma}.$$

Alors $\sum_{j=1}^n \tilde{Y}_j^2$ est de loi $\chi^2(n-1)$.

On applique maintenant exactement la même méthode qu'à l'exemple 2.3.4 pour montrer que

$$I = \left[\frac{n}{q_{1-\frac{\alpha}{2}}^{\chi^2(n-1)}} \tilde{S}_n^2, \frac{n}{q_{\frac{\alpha}{2}}^{\chi^2(n-1)}} \tilde{S}_n^2 \right]$$

1. À multiplication par un scalaire près, bien entendu.

2. Ce degré de liberté correspond en fait à la dimension de l'espace vectoriel engendré par les variables aléatoires \tilde{Y}_j , c'est à dire au rang de la famille $(\tilde{Y}_1, \dots, \tilde{Y}_n)$. C'est sans doute le point de vue le plus significatif pour démontrer le théorème 2.3.8, en utilisant les vecteurs gaussiens. La méthode proposée à l'exercice 2.3 est plus ad hoc.

est un intervalle de confiance de niveau $1 - \alpha$ pour σ^2 .

Remarquons qu'on pourrait tout aussi bien considérer la variance empirique débiaisée vue à l'exemple 1.3.10

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{m}_1)^2$$

à la place de la variance empirique. L'intervalle de confiance obtenu est le même, et s'écrit alors

$$I = \left[\frac{n-1}{q_{1-\frac{\alpha}{2}}^{\chi^2(n-1)}} \hat{S}_n^2, \frac{n-1}{q_{\frac{\alpha}{2}}^{\chi^2(n-1)}} \hat{S}_n^2 \right].$$

Exemple 2.3.9. Supposons que $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$ avec m et σ inconnus, et que nous voulions cette fois déterminer un intervalle de confiance pour m . L'idée de départ est d'utiliser encore une fois la formule

$$\sqrt{n} \frac{\hat{m}_1 - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1).$$

La valeur de σ n'étant pas connue ici, on la remplace par la variance empirique débiaisée³. On considère donc la variable aléatoire

$$\begin{aligned} \sqrt{n} \frac{\hat{m}_1 - m}{\hat{S}_n} &= \frac{\sqrt{n} \frac{\hat{m}_1 - m}{\sigma}}{\sqrt{\frac{\hat{S}_n^2}{\sigma^2}}} \\ &= \frac{\sqrt{n} \frac{\hat{m}_1 - m}{\sigma}}{\sqrt{\frac{1}{n-1} \frac{(n-1)\hat{S}_n^2}{\sigma^2}}} \\ &= \frac{Z_1}{\sqrt{\frac{1}{n-1} Z_2}}, \end{aligned}$$

où $Z_1 = \sqrt{n} \frac{\hat{m}_1 - m}{\sigma}$ et $Z_2 = \frac{(n-1)\hat{S}_n^2}{\sigma^2}$. On sait que $Z_1 \rightsquigarrow \mathcal{N}(0, 1)$, et le théorème 2.3.8 nous assure que $Z_2 \rightsquigarrow \chi^2(n-1)$. Ainsi, les lois de Z_1 et Z_2 ne dépendent plus de m . Si ces variables aléatoires étaient indépendantes, cela assurerait alors que la fonction $g(m, X_1, \dots, X_n) = \sqrt{n} \frac{\hat{m}_1 - m}{\hat{S}_n}$ est pivotale. Cette indépendance est assurée par la proposition suivante, que nous admettrons.

Proposition 2.3.10. *Les variables aléatoires Z_1 et Z_2 sont indépendantes. En particulier, la loi de $\sqrt{n} \frac{\hat{m}_1 - m}{\hat{S}_n}$ ne dépend que de n .*

3. On choisit ici la variance empirique débiaisée au lieu de la variance empirique afin de rendre plus transparent l'apparition des termes $n-1$ dans les formules. On peut faire les calculs avec la variance empirique, et le résultat est le même.

Définition 2.3.11. On appelle loi de Student à n degrés de liberté, notée $St(n)$, la loi de $\frac{X}{\sqrt{\frac{1}{n}Y}}$ où X et Y sont deux variables aléatoires indépendantes vérifiant

$$X \rightsquigarrow \mathcal{N}(0, 1) \quad \text{et} \quad Y \rightsquigarrow \chi^2(n).$$

Comme la loi $\chi^2(n)$, la loi de Student est centrale en statistiques. On a donc montré que

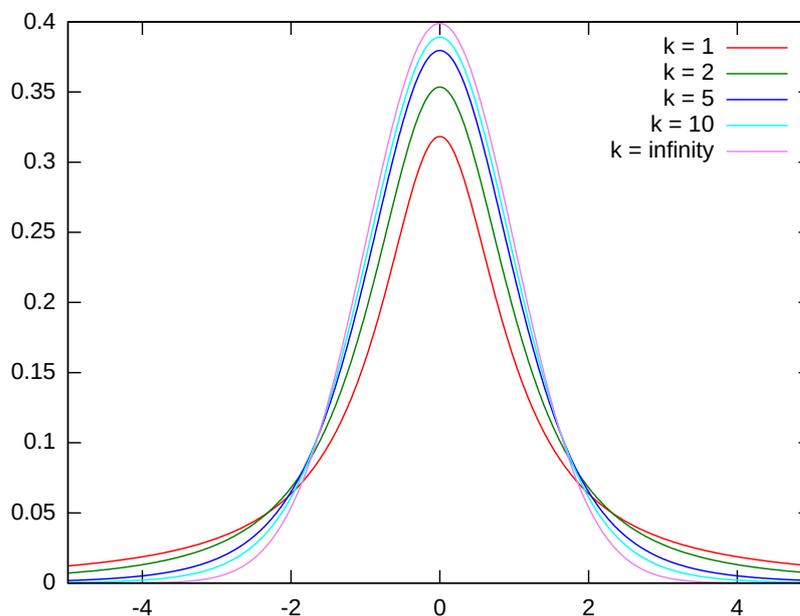
$$\sqrt{n} \frac{\hat{m}_1 - m}{\hat{S}_n} \rightsquigarrow St(n - 1).$$

Pour construire un intervalle de confiance pour m , il faut donc regarder d'un peu plus près les lois de Student et leurs quantiles. On admettra la proposition suivante.

Proposition 2.3.12. La loi $St(n)$ admet pour densité

$$f_{St(n)}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Encore une fois, peu importe l'expression exacte de $f_{St(n)}$, retenons simplement qu'un ordinateur peut facilement calculer une valeur approchée de ses quantiles. Les graphes des premières fonctions $f_{\chi^2(n)}$ sont représentés à la figure 2.5. Dans toute la suite du



commons.wikimedia.org/wiki/File:Student_density.svg

FIGURE 2.5 – Densité $f_{St(k)}$ de la loi $St(k)$.

texte, on notera $q_\alpha^{St(n)}$ le quantile d'ordre α de la loi $St(n)$. D'après la figure 2.5, la suite de fonction $(f_{St(n)})$ semble converger vers la fonction $f_{\mathcal{N}(0,1)}$. C'est effectivement le cas, et nous laissons la preuve de la propriété suivante en exercice au lecteur.

Proposition 2.3.13. *Si (X_n) est une suite de variables aléatoires telle que $X_n \rightsquigarrow St(n)$, alors*

$$X_n \xrightarrow[\text{Loi}]{} \mathcal{N}(0, 1).$$

Pour revenir à la construction d'un intervalle de confiance pour m , on obtient par le même calcul qu'à l'exemple 2.3.3 l'intervalle

$$I = \left[\hat{m}_1 - \frac{\hat{S}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{St(n-1)}; \hat{m}_1 + \frac{\hat{S}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{St(n-1)} \right]$$

comme intervalle de confiance de niveau $1 - \alpha$ pour m .

Pour la petite histoire, la loi de Student a été exhibée par William Gosset en 1908. Alors employé à la brasserie Guinness de Dublin et tenu à la confidentialité, il publia sous le nom de Student. C'est ce nom qui passa à la postérité.

2.4 Intervalles de confiance asymptotiques (on fait ce qu'on peut avec ce qu'on a)

Il est en général difficile de construire des fonctions pivotales et plus généralement des intervalles de confiance. Les théorèmes limites type TCL mettent en lumière la notion, plus faible mais plus facile à manier, d'intervalle de confiance asymptotique.

Définition 2.4.1. *Soit $\alpha \in]0; 1[$. Une suite d'intervalles aléatoires $I(X_1, \dots, X_n) \subset \mathbb{R}$ est appelée un intervalle de confiance asymptotique de niveau $1 - \alpha$ si*

1. $I(X_1, \dots, X_n)$ ne dépend pas de θ , mais seulement de X_1, \dots, X_n et α ;
2. $\forall \theta, \lim_{n \rightarrow \infty} P_\theta(\theta \in I(X_1, \dots, X_n)) \geq 1 - \alpha$.

Les fonctions asymptotiquement pivotales sont aux intervalles de confiance asymptotiques ce que les fonctions pivotales sont aux intervalles de confiance.

Définition 2.4.2. *Une fonction⁴ $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ est dite asymptotiquement pivotale si $g(\theta, X_1, \dots, X_n)$ converge en loi, lorsque n tend vers $+\infty$, vers une variable aléatoire dont la loi ne dépend pas de θ .*

4. Il s'agit plutôt d'une suite de fonctions $g_n : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$, mais passons sur ce point.

Un exemple de méthode de construction d'intervalle de confiance asymptotique est la suivante. Supposons que l'on dispose d'un estimateur consistant $\hat{\theta}_n$ de θ qui vérifie un énoncé du type TCL

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sigma} \xrightarrow{\text{Loi}} \mathcal{N}(0, 1),$$

avec $\sigma^2 = \text{Var}(X_j)$. La fonction

$$g(\theta, X_1, \dots, X_n) = \sqrt{n} \frac{\hat{\theta}_n - \theta}{\sigma}$$

est donc asymptotiquement pivotale. On applique ensuite la même méthode qu'avec les fonctions pivotales pour construire un intervalle de confiance, maintenant asymptotique. En général, la variance σ^2 est inconnue, et la proposition suivante est alors bien utile.

Proposition 2.4.3. *Supposons $X_j \in L^2$, et soit $\hat{\theta}_n$ un estimateur consistant de θ vérifiant*

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sigma} \xrightarrow{\text{Loi}} \mathcal{N}(0, 1),$$

avec $\sigma^2 = \text{Var}(X_j)$. Si $\hat{\sigma}_n$ est un estimateur fortement consistant de σ , alors

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \xrightarrow{\text{Loi}} \mathcal{N}(0, 1).$$

En particulier, l'intervalle

$$\left[\hat{\theta}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; \hat{\theta}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right].$$

est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ .

Démonstration. On a

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} = \sqrt{n} \frac{\hat{\theta}_n - \theta}{\sigma} \cdot \frac{\sigma}{\hat{\sigma}_n}.$$

Par hypothèse $\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sigma}$ converge en loi vers $\mathcal{N}(0, 1)$, et $\hat{\sigma}_n$ converge en probabilité vers σ . Donc $\frac{\sigma}{\hat{\sigma}_n}$ converge en probabilité vers 1, et le théorème de Slutsky nous assure que

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \xrightarrow{\text{Loi}} \mathcal{N}(0, 1).$$

Autrement dit, la fonction

$$g(\theta, X_1, \dots, X_n) = \sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$$

est asymptotiquement pivotale. En particulier,

$$\lim_{n \rightarrow \infty} P_{\theta} \left(\sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \in \left[-q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right] \right) = 1 - \alpha,$$

et donc

$$\lim_{n \rightarrow \infty} P_{\theta} \left(\theta \in \left[\hat{\theta}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; \hat{\theta}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right] \right) = 1 - \alpha,$$

ce qui démontre la deuxième assertion. \odot

On connaît déjà la variance empirique et la variance empirique débiaisée

$$\tilde{S}_n^2 = \hat{m}_2 - \hat{m}_1^2 \quad \text{et} \quad \hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{m}_1)^2$$

comme estimateurs fortement consistants de σ^2 . Suivant la situation, d'autres estimateurs consistants de σ peuvent se présenter.

Exemple 2.4.4. Si $X_j \rightsquigarrow Ber(\theta)$, on a alors $\sigma^2 = \theta(1 - \theta)$. L'estimateur \hat{m}_1 de θ est fortement consistant et le TCL nous dit que

$$\sqrt{n} \frac{\hat{m}_1 - \theta}{\sqrt{\theta(1 - \theta)}} \xrightarrow{Loi} \mathcal{N}(0, 1).$$

Puisque $X_j^2 = X_j$, la variance empirique est égale à $\hat{m}_1(1 - \hat{m}_1)$. La proposition 2.4.3 nous assure alors que

$$\sqrt{n} \frac{\hat{m}_1 - \theta}{\sqrt{\hat{m}_1(1 - \hat{m}_1)}} \xrightarrow{Loi} \mathcal{N}(0, 1)$$

et que

$$\left[\hat{m}_1 - \sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; \hat{m}_1 + \sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right]$$

est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ .

En revenant à l'exemple de l'estimation de la proportion de la population nantaise aimant le café, nous pouvons enfin donner un sens au signe " \simeq " dans " $\theta \simeq 0,6$ ". On a $n = 1000$, et en prenant pour $\alpha = 0,05$ on obtient

$$\sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} = \sqrt{\frac{0,6 \times 0,4}{1000}} \times 1,96 \simeq 0,03.$$

Il est donc raisonnable d'affirmer, avec un degré de confiance de 95%, que la proportion de la population nantaise aimant le café se trouve dans l'intervalle $[0,57; 0,63]$. On notera bien que les 5% d'incertitude ne portent pas sur les bornes de l'intervalle, mais sur la présence ou non de θ dans cet intervalle.

Exemple 2.4.5. Si $X_j \rightsquigarrow \text{Exp}(\lambda)$, on a vu aux exemples 1.1.5 et 1.3.11 que

$$E(X_j) = \frac{1}{\lambda} \quad \text{et} \quad \sigma^2 = \frac{2}{\lambda^2},$$

et que $\hat{\lambda}_n = \frac{1}{\bar{m}_1}$ est un estimateur fortement consistant de λ . Comme application de la Δ -méthode, nous avons aussi montré à l'exemple 1.3.16 que

$$\sqrt{n} \frac{\hat{\lambda}_n - \lambda}{\lambda} \xrightarrow{\text{Loi}} \mathcal{N}(0, 1).$$

À partir de là, deux voies possibles s'ouvrent pour construire un intervalle de confiance asymptotique pour λ .

1. On écrit

$$\sqrt{n} \frac{\hat{\lambda}_n - \lambda}{\lambda} = \sqrt{n} \left(\frac{\hat{\lambda}_n}{\lambda} - 1 \right).$$

On prend donc $g(\lambda, X_1, \dots, X_n) = \sqrt{n} \left(\frac{\hat{\lambda}_n}{\lambda} - 1 \right)$ comme fonction asymptotiquement pivotale. Puisque

$$-q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \leq \sqrt{n} \left(\frac{\hat{\lambda}_n}{\lambda} - 1 \right) \leq q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \iff \frac{\sqrt{n} \hat{\lambda}_n}{\sqrt{n} + q_{1-\frac{\alpha}{2}}^{\mathcal{N}}} \leq \lambda \leq \frac{\sqrt{n} \hat{\lambda}_n}{\sqrt{n} - q_{1-\frac{\alpha}{2}}^{\mathcal{N}}},$$

on en déduit que

$$I_1 = \left[\frac{\sqrt{n} \hat{\lambda}_n}{\sqrt{n} + q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}; \frac{\sqrt{n} \hat{\lambda}_n}{\sqrt{n} - q_{1-\frac{\alpha}{2}}^{\mathcal{N}}} \right]$$

est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour λ .

2. On utilise l'estimation fortement consistante $\hat{\lambda}_n$ de λ pour écrire

$$\sqrt{n} \frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} = \sqrt{n} \left(1 - \frac{\lambda}{\hat{\lambda}_n} \right) \xrightarrow{\text{Loi}} \mathcal{N}(0, 1).$$

Puisque

$$-q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \leq \sqrt{n} \left(1 - \frac{\lambda}{\hat{\lambda}_n} \right) \leq q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \iff \hat{\lambda}_n - \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \hat{\lambda}_n \leq \lambda \leq \hat{\lambda}_n + \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \hat{\lambda}_n,$$

on en déduit que

$$I_1 = \left[\hat{\lambda}_n - \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \hat{\lambda}_n; \hat{\lambda}_n + \frac{q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \hat{\lambda}_n \right]$$

est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour λ .

Y a-t-il un meilleur intervalle de confiance entre I_1 et I_2 ? Pour répondre à cette question, calculons la longueur de chacun de ces deux intervalles. On a

$$\begin{aligned} l(I_1) &= \sqrt{n} \hat{\lambda}_n \left(\frac{1}{\sqrt{n} - q_{1-\frac{\alpha}{2}}^{\mathcal{N}}} - \frac{1}{\sqrt{n} + q_{1-\frac{\alpha}{2}}^{\mathcal{N}}} \right) \\ &= \hat{\lambda}_n \frac{2\sqrt{n} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{n - (q_{1-\frac{\alpha}{2}}^{\mathcal{N}})^2} \\ &= \hat{\lambda}_n \frac{2 q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n} - \frac{(q_{1-\frac{\alpha}{2}}^{\mathcal{N}})^2}{\sqrt{n}}} \end{aligned}$$

et

$$l(I_2) = \hat{\lambda}_n \frac{2 q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}}.$$

Donc l'intervalle I_2 est meilleur que I_1 . Ils sont cependant asymptotiquement égaux.

2.5 Exercices

Exercice 2.1. Soit X une variable aléatoire suivant une loi Gamma $\Gamma(\alpha, \beta)$ avec $\alpha > 0$ et $\beta > 0$. On rappelle que la densité de X s'écrit :

$$f_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x \in \mathbb{R}^+,$$

où Γ est la fonction Gamma définie par :

$$\begin{aligned} \Gamma : \mathbb{R}^{*+} &\rightarrow \mathbb{R}^{*+} \\ x &\mapsto \int_0^\infty e^{-y} y^{x-1} dy. \end{aligned}$$

On admet, à toutes fins utiles, la relation de récurrence suivante :

$$\Gamma(x+1) = x\Gamma(x), \quad \text{pour tout } x > 0.$$

En outre, la fonction caractéristique de X est donnée par :

$$\varphi_X(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}, \quad \forall t \in \mathbb{R}.$$

1. Montrer que $X \in L^2$ et que $E(X) = \alpha/\beta$ et $Var(X) = \alpha/\beta^2$.

2. Soit $X_1 \rightsquigarrow \Gamma(\alpha_1, \beta)$ et $X_2 \rightsquigarrow \Gamma(\alpha_2, \beta)$ deux variables aléatoires indépendantes, avec $\alpha_1 > 0$, $\alpha_2 > 0$ et $\beta > 0$. Montrer que $X_1 + X_2 \rightsquigarrow \Gamma(\alpha_1 + \alpha_2, \beta)$.
3. Soit $Z \rightsquigarrow \mathcal{N}(0, 1)$. Montrer que $Z^2 \rightsquigarrow \mathcal{G}(1/2, 1/2)$. On rappelle que $\Gamma(1/2) = \sqrt{\pi}$.
4. Dédurre des deux questions précédentes que si Z_1, \dots, Z_n sont indépendantes et identiquement distribuées de loi $\mathcal{N}(0, 1)$, alors $\sum_{j=1}^n Z_j^2 \rightsquigarrow \chi^2(n)$.

Exercice 2.2. Soit X une variable aléatoire suivant une loi de Student $St(m)$ avec $m \in \mathbb{N}^*$. On rappelle que la densité de X s'écrit :

$$f_X(x) = \frac{1}{\sqrt{m\pi}} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}.$$

Montrer que $X \in L^1$ si et seulement si $m > 1$, que $X \in L^2$ si et seulement si $m > 2$, et que $E(X) = 0$ pour $m > 1$.

Exercice 2.3. Soit X_1, \dots, X_n indépendantes et identiquement distribuées de loi $\mathcal{N}(\mu, \sigma^2)$. On définit les estimateurs empiriques pour les deux paramètres comme suit :

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{S}_{n-1}^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{m}_1)^2.$$

1. Montrer que $\hat{m}_1 \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n)$.
2. Montrer que \hat{m}_1 et \hat{S}_{n-1}^2 sont deux variables aléatoires indépendantes.
3. Montrer que $\frac{(n-1)\hat{S}_{n-1}^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$.

Exercice 2.4. Donner un exemple de variable aléatoire ayant des quantiles non uniques.

Exercice 2.5. On a mesuré la consommation quotidienne en eau dans une ville sur une année. On suppose les 365 observations x_1, \dots, x_n issues de variables aléatoires X_1, \dots, X_n indépendantes et de loi normale $\mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnues. On rappelle que, sous l'hypothèse de normalité des données, les estimateurs empiriques sans biais de la moyenne μ et de la variance σ^2 , qui sont respectivement :

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{m}_1)^2,$$

sont indépendants et que :

$$\hat{m}_1 \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{et} \quad \frac{(n-1)\hat{S}_n^2}{\sigma^2} \rightsquigarrow \chi^2(n-1).$$

On mesure sur l'échantillon une moyenne $\hat{m}_1 = 18$ (en milliers de m³) et un écart type $\hat{S}_n = 2$.

1. Construire un intervalle de confiance à 95% pour estimer σ
Indication : prendre respectivement 313 et 419 comme valeurs des quantiles d'ordre 0.025 et 0.975 de la loi $\chi^2(364)$.
2. Montrer que $g(\mu; X_1, \dots, X_n) = \sqrt{n} (\bar{X}_n - \mu) / \hat{S}_n$ est une fonction pivotale pour le paramètre μ et donner sa loi.
3. Construire un intervalle de confiance à 95% pour estimer μ .
4. Dédire un intervalle de confiance à 95% sur la consommation mensuelle moyenne pour un mois de 30 jours.
5. On veut évaluer la consommation maximale attendue du 1^{er} janvier prochain, que l'on note X_{n+1} . Montrer que

$$\sqrt{\frac{n}{n+1}} \frac{X_{n+1} - \hat{m}_1}{\hat{S}_n}$$

suit une loi de Student $St(n-1)$.

6. Dédire une majoration de X_{n+1} de niveau de significativité $\alpha = 0.05$.

Exercice 2.6. Soit (X_1, \dots, X_n) un n -échantillon de la loi uniforme $\mathcal{U}[0, \theta]$ avec $\theta > 0$. On pose $M_n = \max(X_1, \dots, X_n)$.

1. Montrer que θ/M_n est pivotale pour le paramètre θ .
2. Déterminer la densité de probabilité de θ/M_n .
3. Soit $0 < \alpha < 1$. Montrer que $[M_n, \alpha^{-1/n} M_n]$ est un intervalle de confiance pour θ de niveau $1 - \alpha$.

Exercice 2.7. Un biochimiste étudie un type de moisissure qui attaque les cultures de blé. La toxine contenue dans cette moisissure est obtenue sous forme d'une solution organique. La quantité de substance toxique par gramme de solution est une variable aléatoire réelle X que l'on suppose suivre une loi normale. L'unité est le milligramme. On mesure la quantité de substance toxique par gramme de solution. Sur 9 extraits, on a obtenu les mesures suivantes :

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1.2	0.8	0.6	1.1	1.2	0.9	1.5	0.9	1.0

1. Déterminer une estimation ponctuelle de la moyenne et de l'écart type de la quantité de substance toxique par gramme de solution.
2. Déterminer un intervalle de confiance pour la quantité moyenne de substance toxique par gramme de solution de niveau 95%.

3. Déterminer un intervalle de confiance pour l'écart type de la quantité de substance toxique par gramme de solution de niveau 95%.

Exercice 2.8. Une usine fabrique des câbles. La masse maximale (en tonnes) supportée par un câble est une variable aléatoire réelle X que l'on suppose suivre une loi normale de moyenne μ inconnue et d'écart type $\sigma = 0.5$. Une étude portant sur un échantillon de 50 câbles a donné une moyenne des charges maximales supportées égales à 12.2 tonnes.

1. Déterminer un intervalle de confiance pour μ de niveau 99%.
2. Déterminer la taille d'échantillon minimale pour que la longueur de l'intervalle de confiance pour μ de niveau 99% soit inférieure ou égale à 0.2.

Exercice 2.9. On s'intéresse à la proportion p de la population touchée par le syndrome de dépression hivernale. Sur un échantillon de $n = 400$ personnes interrogées, 16 affirment en souffrir.

1. Décrire le modèle statistique.
2. Construire un intervalle de confiance asymptotique de niveau 95% pour p .

Chapitre 3

Tests statistiques

Nous avons vu jusqu'à présent comment estimer un paramètre à partir d'un échantillon, de manière ponctuelle ou moins brutalement par un intervalle. En pratique, on peut aussi être confronté à un problème non pas d'estimation, mais de choix entre deux hypothèses contradictoires sur le paramètre. Un test statistique est un moyen de décider en faveur de l'une ou l'autre de ces hypothèses lorsque les informations disponibles ne permettent pas un choix certain. Attention à ne pas comprendre cette dernière phrase de travers ! Un test statistique n'est pas, loin de là, un oracle omniscient départageant le vrai du faux. Il s'agit uniquement d'un algorithme proposant un choix lorsque l'on est dans l'indécision, sans garantir la véracité de l'hypothèse retenue. Il s'agit en quelque sorte d'un conseiller. La décision de suivre un conseil dépend avant tout de la confiance placée en la personne qui le donne, il en est de même avec un test statistique. Tout le sel de l'exercice consiste ainsi à élaborer des tests dont les conclusions ne peuvent pas être immédiatement qualifiées de farfelues.

Nous développerons les deux exemples élémentaires suivants tout au long de ce chapitre.

Exemple 3.0.1. On souhaite déterminer si une pièce de monnaie est équilibrée. La probabilité d'obtenir "pile" en lançant cette pièce suit une loi $Ber(\theta)$, et le problème est donc de choisir entre les deux hypothèses

$$H_0 : \theta = \frac{1}{2} \qquad H_1 : \theta \neq \frac{1}{2}.$$

Pour cela, on effectue 100 lancers et on obtient une proportion $\hat{m}_1 = 0,58$ de "piles". Est-il raisonnable de penser que cette pièce est équilibrée ?

Exemple 3.0.2. Un producteur de champignons cultivés en grotte sait d'expérience que sa production mensuelle suit une loi normale de moyenne 1000 et d'écart type 100. Il essaie gratuitement un nouveau système d'arrosage pendant 12 mois et obtient les résultats suivants :

Mois	1	2	3	4	5	6	7	8	9	10	11	12
Production	1081	1108	947	1019	1100	1107	1004	994	976	1000	1123	1101

TABLE 3.1 – $\hat{m}_1 = 1046,67$

Après ces 12 mois d’essai, notre cultivateur doit décider de l’acquisition éventuelle de ce système d’arrosage. En particulier, si m est la nouvelle moyenne de production mensuelle, il doit décider entre les hypothèses

$$H_0 : m = 1000$$

$$H_1 : m > 1000.$$

Comme évoqué dans le préambule, la modélisation, c’est à dire le passage d’une question “concrète” à un problème mathématique, constitue une étape cruciale dans la mise en place d’un test statistique. Cette étape est néanmoins sujette à bien des biais, conscients et inconscients, de la part du modélisateur. On pourra remarquer dans le cas du cultivateur de champignons que ce dernier n’imagine pas que le nouveau système d’arrosage puisse faire baisser sa production, ce qui révèle un certain optimiste quand au progrès et à l’honnêteté du constructeur dudit système. Ainsi, comme dans l’exemple du devin boursier, la réponse à une question dépendra en grande partie de sa modélisation mathématique. Au delà de l’aspect modélisation, la réponse à une question dépend également de la façon dont on la pose. Nous verrons dans la suite qu’un test statistique ne donnera pas forcément la même réponse aux questions “A ou B ?” et “B ou A ?”.

3.1 Principes généraux

Rappelons que $(X_j)_{j \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes et identiquement distribuées dont la loi de probabilité P_θ dépend d’un paramètre θ inconnu. Nous supposons ici que $\theta \in \Theta \subset \mathbb{R}$. Étant donné une partition $\Theta = \Theta_0 \sqcup \Theta_1$, on cherche à décider si $\theta \in \Theta_0$ ou $\theta \in \Theta_1$.

Définition 3.1.1. *On appelle l’hypothèse nulle, ou hypothèse neutre, l’alternative*

$$H_0 : \theta \in \Theta_0.$$

On appelle l’hypothèse alternative

$$H_1 : \theta \in \Theta_1.$$

Cette différence de dénomination introduit une brisure de symétrie entre H_0 et H_1 qui ne paraît pas forcément pertinente à ce stade. Nous verrons dans la suite que les traitements de H_0 et H_1 seront bien différents : H_0 sera acceptée comme vraie par défaut, alors que H_1 ne sera acceptée que si les données sont à charge contre H_0 . On voit alors

bien pourquoi les questions “ H_0 ou H_1 ?” et “ H_1 ou H_0 ?” ne sont pas les mêmes, tout comme un système judiciaire basé sur la présomption d’innocence est fondamentalement différent d’un système judiciaire basé sur la présomption de culpabilité.

Il s’agit donc de trancher en faveur de H_0 ou H_1 en fonction d’une observation (x_1, \dots, x_n) de (X_1, \dots, X_n) . Pour cela, on construit une *région critique* RC pour laquelle on décide que

- si $(x_1, \dots, x_n) \in RC$, alors on rejette H_0 pour H_1 ;
- si $(x_1, \dots, x_n) \notin RC$, alors on ne rejette pas H_0 .

Exemple 3.1.2. Concernant la pièce de monnaie de l’exemple 3.0.1, on a

$$\Theta = [0; 1], \quad \Theta_0 = \left\{ \frac{1}{2} \right\}, \quad \text{et} \quad \Theta_1 = \left[0; \frac{1}{2} \cup \left[\frac{1}{2}; 1 \right] \right].$$

On prendra¹

$$RC = \left\{ \left| \hat{m}_1 - \frac{1}{2} \right| > C \right\}$$

comme région critique. La région RC est dite *bilatérale*, c’est à dire s’étend de part et d’autre² de $\frac{1}{2}$. Toute cette formalisation ne fait pour l’instant que déplacer le problème, qui devient maintenant : comment choisir C de manière optimale ?

Exemple 3.1.3. Dans l’exemple 3.0.2 du cultivateur de champignons, on a

$$\Theta = [1000; +\infty[, \quad \Theta_0 = \{1000\}, \quad \text{et} \quad \Theta_1 =]1000; +\infty[.$$

On prendra

$$RC = \{ \hat{m}_1 > 1000 + C \}$$

comme région critique. La région RC est dite ici *unilatérale*, c’est à dire que 1000 est sur la frontière de RC . Encore une fois, on est ramené à résoudre le problème du choix optimal de C .

Définition 3.1.4. Il y a deux types d’erreur possibles dans le choix de H_0 ou de H_1 :

- L’erreur de première espèce : on rejette H_0 alors que H_0 est vraie ;
- L’erreur de deuxième espèce : on ne rejette pas H_0 alors que H_0 est fausse.

Le risque de première espèce est la fonction

$$\begin{aligned} \tilde{\alpha} : \Theta_0 &\longrightarrow [0; 1] \\ \theta &\longmapsto P_\theta((X_1, \dots, X_n) \in RC), \end{aligned}$$

1. Pour être correct, il faudrait écrire $RC = \{(x_1, \dots, x_n) \mid |\hat{m}_1 - \frac{1}{2}| > C\}$. Nous ferons fréquemment cet abus afin d’alléger un peu les notations.

2. Ou encore $\frac{1}{2}$ est contenue dans l’intérieur, dans \mathbb{R} , de RC .

et le *niveau* d'un test est le réel

$$\alpha = \sup_{\theta \in \Theta_0} \tilde{\alpha}(\theta).$$

Il est fréquent qu'on ne puisse calculer en pratique que le *niveau asymptotique* d'un test, défini par $\lim_{n \rightarrow +\infty} \alpha$. Le *risque de deuxième espèce* est la fonction

$$\begin{aligned} \beta : \Theta_1 &\longrightarrow [0; 1] \\ \theta &\longmapsto P_\theta((X_1, \dots, X_n) \notin RC), \end{aligned}$$

et la puissance d'un test est la fonction

$$\begin{aligned} \pi : \Theta_1 &\longrightarrow [0; 1] \\ \theta &\longmapsto P_\theta((X_1, \dots, X_n) \in RC) = 1 - \beta(\theta). \end{aligned}$$

Définition 3.1.5. On dit qu'un test est consistant si

$$\forall \theta \in \Theta_1, \quad \pi(\theta) \xrightarrow{n \rightarrow +\infty} 1.$$

Autrement dit, un test est consistant si il ne se trompe asymptotiquement pas. On peut résumer ces définitions dans le tableau suivant.

		Vérité	
		H_0	H_1
Décision	H_0	OK $1 - \alpha$	2ème espèce β
	H_1	1ère espèce α	OK π

TABLE 3.2 – Erreurs de première et deuxième espèces

Exemple 3.1.6. Revenons au cas de la pièce de monnaie des exemples 3.0.1 et 3.1.6, où

$$\Theta = [0; 1], \quad \Theta_0 = \left\{\frac{1}{2}\right\}, \quad \Theta_1 = \left[0; \frac{1}{2} \cup \left[\frac{1}{2}; 1\right], \quad \text{et} \quad RC = \left\{|\hat{m}_1 - \frac{1}{2}| > C\right\}.$$

On a donc

$$\alpha = P_{\frac{1}{2}}\left(|\hat{m}_1 - \frac{1}{2}| > C\right) \quad \text{et} \quad \beta(\theta) = P_\theta\left(|\hat{m}_1 - \frac{1}{2}| \leq C\right).$$

On observe alors un phénomène troublant : minimiser le risque de première espèce α maximise le risque de deuxième espèce β , et réciproquement ! En effet minimiser α revient ici à augmenter C . Par exemple puisque $\hat{m}_1 \in [0; 1]$, on aura $\alpha = 0$ si $C \geq \frac{1}{2}$. Mais d'un autre côté, si C augmente alors β aussi : on aura $\beta(\theta) = 1$ dès que $C \geq \frac{1}{2}$.

Cette impossibilité de minimiser à la fois les risques de première et deuxième espèces n'est pas spécifique à l'exemple ci-dessus. Il s'agit d'un phénomène général, une sorte de principe d'incertitude statistique. Pour reprendre notre parallèle juridique, l'existence d'un système de procès qui éviterait assurément les innocents condamnés et les coupables relâchés semble peu probable. Il est cependant facile d'éviter l'un ou l'autre de ces inconvénients : si on pense que rien ne justifie un innocent en prison, alors il suffit de relâcher tout le monde ; si au contraire on pense que tout coupable doit être châtié, alors il suffit d'enfermer tout le monde. Dans les deux cas, si une des deux situations problématiques est évitée à coup sûr, l'autre sera tout autant à coup sûr inévitable. Tout système judiciaire doit donc se situer entre ces deux extrêmes. Déterminer où placer le curseur est à présent un affaire de sensibilité, probablement pas de mathématiques.

À la lumière des généralités qui précèdent, il se dégage le principe suivant pour la construction d'un test statistique. On fixe d'abord le niveau α du test à une valeur petite (0,05 ou 0,01 par exemple) et on cherche une région critique de niveau (éventuellement asymptotique) α . Si on a le choix entre plusieurs régions critiques à α fixé, on essaiera alors de minimiser le risque de deuxième espèce β .

En d'autres termes, on ne rejette l'hypothèse H_0 que si les observations sont aberrantes en supposant H_0 vraie. Cette dernière est ainsi considérée vraie par défaut : si les observations ne sont en désaccord ni avec H_1 ni avec H_0 , l'hypothèse H_0 sera toujours préférée. Une grande partie de la difficulté consiste donc à bien choisir les hypothèses H_0 et H_1 . Faute de quoi, un test statistique ne sera rien d'autre qu'une justification pseudo-scientifique d'une opinion préconçue.

Exemple 3.1.7. On suppose que $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$ avec σ^2 connue, et on veut tester

$$H_0 : m = m_0 \qquad H_1 : m \neq m_0$$

Si H_0 est vraie, alors

$$\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} \rightsquigarrow \mathcal{N}(0, 1).$$

On rejettera donc H_0 si la valeur de $\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma}$ est trop grande. Pour avoir un test de niveau α , on pourra prendre comme région critique

$$RC = \left\{ \hat{m}_1 \notin \left[m_0 - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right] \right\}$$

puisque

$$P_{m_0} \left(\left| \sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} \right| > q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right) = \alpha.$$

Vérifions que cette région critique définit un test consistant. Si $m \neq m_0$, alors

$$\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} = \sqrt{n} \frac{\hat{m}_1 - m}{\sigma} + \sqrt{n} \frac{m - m_0}{\sigma}.$$

On a

$$\sqrt{n} \frac{\hat{m}_1 - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1),$$

et

$$\begin{aligned} \beta(m) &= P_m \left(\left| \sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} \right| \leq q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right) \\ &= P_m \left(-q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma} \leq \sqrt{n} \frac{\hat{m}_1 - m}{\sigma} \leq q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma} \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma}}^{q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma}} e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Puisque

$$\lim_{n \rightarrow +\infty} -q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma} = \lim_{n \rightarrow +\infty} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma} = \pm\infty$$

suivant le signe de $m - m_0$, on en déduit que $\lim_{n \rightarrow +\infty} \beta(m) = 0$. Le test est donc bien consistant.

Exemple 3.1.8. Reprenons l'exemple précédent, mais supposons que l'on sache maintenant que $m \geq m_0$. On cherche alors à décider entre les deux hypothèses

$$H_0 : m = m_0 \qquad H_1 : m > m_0$$

La région critique de l'exemple 3.1.7 définit toujours un test consistant de niveau α . Mais en utilisant l'information supplémentaire que nous avons sur m , nous pouvons faire mieux en considérant la région critique

$$\widetilde{RC} = \left\{ \hat{m}_1 > m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^{\mathcal{N}} \right\}.$$

La région \widetilde{RC} définit aussi un test de niveau α puisque

$$P_{m_0} \left(\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} > q_{1-\alpha}^{\mathcal{N}} \right) = \alpha,$$

et consistant puisque, si $m > m_0$,

$$\begin{aligned} \widetilde{\beta}(m) &= P_m \left(\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} \leq q_{1-\alpha}^{\mathcal{N}} \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{q_{1-\alpha}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma}} e^{-\frac{x^2}{2}} dx \xrightarrow[n \rightarrow +\infty]{} 0. \end{aligned}$$

L'inégalité $1 - \frac{\alpha}{2} > 1 - \alpha$ entraîne l'inégalité $q_{1-\frac{\alpha}{2}}^{\mathcal{N}} > q_{1-\alpha}^{\mathcal{N}}$ sur les quantiles. Les erreurs de deuxième espèce pour les deux régions critiques se comparent ainsi

$$\beta(m) - \tilde{\beta}(m) = \frac{1}{\sqrt{2\pi}} \left(\int_{q_{1-\alpha}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma}}^{q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma}} e^{-\frac{x^2}{2}} dx - \int_{-\infty}^{-q_{1-\frac{\alpha}{2}}^{\mathcal{N}} + \sqrt{n} \frac{m_0 - m}{\sigma}} e^{-\frac{x^2}{2}} dx \right),$$

et on peut voir avec un peu d'effort supplémentaires que

$$\beta(m) - \tilde{\beta}(m) \geq 0.$$

Ainsi pour un même niveau de test, la région critique \widetilde{RC} minimise le risque de deuxième espèce. Le deuxième test est donc meilleur. On peut aussi voir que \widetilde{RC} est meilleure que RC en observant que

$$RC \cap \{m > m_0\} \subset \widetilde{RC}.$$

Autrement dit, la région \widetilde{RC} est plus sensible aux observations en contradiction avec H_0 . Plus on utilise d'informations pour construire un test, plus la région critique est susceptible d'être large.

Revenons à notre producteur de champignons des exemples 3.0.2 et 3.1.3. On suppose que la variance $\sigma^2 = 100^2$ de la production mensuelle ne change pas avec le nouveau système d'arrosage. Pour $\alpha = 0,05$, on a alors

$$q_{0,95}^{\mathcal{N}} = 1,64, \quad \hat{m}_1 = 1046,67, \quad n = 12,$$

et

$$m_0 + \frac{\sigma}{\sqrt{n}} q_{0,95}^{\mathcal{N}} = 1000 + \frac{100}{\sqrt{12}} \times 1,64 \simeq 1047,49.$$

Notre producteur décide alors en faveur de H_0 , et de ne pas investir dans ce système d'arrosage.

Exemple 3.1.9. Il est sans doute plus réaliste de ne pas supposer comme à l'exemple précédent, que la variance $\sigma^2 = 100^2$ de la production mensuelle ne change pas avec le nouveau système d'arrosage. On estime alors la nouvelle variance par la variance empirique débiaisée

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{m}_1)^2.$$

On a vu à l'exemple 2.3.9 que

$$\sqrt{n} \frac{\hat{m}_1 - m}{\sigma} \rightsquigarrow St(n-1),$$

on montre donc avec le même raisonnement qu'à l'exemple 3.1.8 que la région critique

$$RC = \left\{ \hat{m}_1 > m_0 + \frac{\hat{S}_n}{\sqrt{n}} q_{1-\alpha}^{St(n-1)} \right\}.$$

définit un test consistant de niveau α . Pour $\alpha = 0,05$, on a alors

$$q_{0,95}^{St(11)} = 1,8, \quad \hat{m}_1 = 1046,67, \quad \hat{S}_n^2 = 3949,73, \quad n = 12,$$

et

$$m_0 + \frac{\hat{S}_n}{\sqrt{n}} q_{0,95}^{\mathcal{N}} = 1000 + \frac{\sqrt{3949,73}}{\sqrt{12}} \times 1,8 \simeq 1032,58.$$

Notre producteur décide cette fois de rejeter H_0 en faveur de H_1 .

3.2 Probabilité critique, ou p -valeur

Le concept de probabilité critique est une reformulation de ce que nous avons vu à la section précédente, avec un point de vue légèrement différent où le niveau α d'un test est maintenant une variable.

Définition 3.2.1. *Soit un test de niveau α défini par une région critique $RC(\alpha)$ dépendant de α . La probabilité critique, ou p -valeur, du test est la plus petite valeur α_c de α telle que l'échantillon appartienne à $RC(\alpha)$:*

$$\alpha_c = \inf \{ \alpha \mid (x_1, \dots, x_n) \in RC(\alpha) \}.$$

On a donc l'alternative suivante :

1. $\alpha_c < \alpha$ et on rejette H_0 pour H_1 ;
2. $\alpha_c \geq \alpha$ et on ne rejette pas H_0 .

Revisitons les exemples 3.1.7, 3.1.8 et 3.1.9 sous l'angle de la probabilité critique. Rappelons que dans ces trois exemples, on suppose que $X_j \rightsquigarrow \mathcal{N}(m, \sigma^2)$.

Exemple 3.2.2. On suppose de plus σ^2 connue, et on veut tester

$$H_0 : m = m_0 \qquad H_1 : m \neq m_0$$

avec la région critique

$$RC = \left\{ \hat{m}_1 \notin \left[m_0 - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right] \right\} = \left\{ \sqrt{n} \frac{|\hat{m}_1 - m_0|}{\sigma} > q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right\}.$$

La probabilité critique α_c est alors définie par

$$\sqrt{n} \frac{|\hat{m}_1 - m_0|}{\sigma} = q_{1-\frac{\alpha_c}{2}}^{\mathcal{N}}.$$

En prenant la valeur de la fonction de répartition de la loi $\mathcal{N}(0, 1)$ pour chacun deux membres, on obtient

$$F_{\mathcal{N}(0,1)} \left(\sqrt{n} \frac{|\hat{m}_1 - m_0|}{\sigma} \right) = 1 - \frac{\alpha_c}{2},$$

soit

$$\alpha_c = 2 \left(1 - F_{\mathcal{N}(0,1)} \left(\sqrt{n} \frac{|\hat{m}_1 - m_0|}{\sigma} \right) \right).$$

Exemple 3.2.3. On suppose encore σ^2 connue, et on veut tester

$$H_0 : m = m_0 \qquad H_1 : m > m_0$$

avec la région critique

$$RC = \left\{ \hat{m}_1 > m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^{\mathcal{N}} \right\} = \left\{ \sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} > q_{1-\alpha}^{\mathcal{N}} \right\}.$$

La probabilité critique α_c est alors définie par

$$\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} = q_{1-\alpha_c}^{\mathcal{N}},$$

soit

$$\alpha_c = 1 - F_{\mathcal{N}(0,1)} \left(\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} \right).$$

Les valeurs numériques provenant de l'exemple du producteur de champignons sont

$$\sqrt{12} \frac{\hat{m}_1 - m_0}{\sigma} = 1,6 \quad \text{et} \quad 1 - F_{\mathcal{N}(0,1)}(1,6) = 0,053.$$

Au niveau $\alpha = 0,05 < 0,053$, le producteur ne rejette donc pas H_0 .

Exemple 3.2.4. On ne suppose plus σ^2 connue, et on veut tester

$$H_0 : m = m_0 \qquad H_1 : m > m_0$$

avec la région critique

$$RC = \left\{ \hat{m}_1 > m_0 + \frac{\hat{S}_n}{\sqrt{n}} q_{1-\alpha}^{St(n-1)} \right\} = \left\{ \sqrt{n} \frac{\hat{m}_1 - m_0}{\hat{S}_n} > q_{1-\alpha}^{St(n-1)} \right\}.$$

La probabilité critique α_c est alors définie par

$$\sqrt{n} \frac{\hat{m}_1 - m_0}{\hat{S}_n} = q_{1-\alpha_c}^{St(n-1)},$$

soit

$$\alpha_c = 1 - F_{St(n-1)} \left(\sqrt{n} \frac{|\hat{m}_1 - m_0|}{\hat{S}_n} \right).$$

Avec les valeurs numériques provenant de l'exemple du producteur de champignons, on trouve $\alpha_c = 0,014$. Au niveau $\alpha = 0,05 > 0,014$, le producteur rejette donc maintenant H_0 pour H_1 .

3.3 Construction d'un test à partir d'un intervalle de confiance

Comme on peut le constater dans les exemples précédents, les intervalles de confiance sont utiles dans la construction de tests statistiques. La proposition suivante formalise cette observation.

Proposition 3.3.1. *Si $I(x_1, \dots, x_n)$ est un intervalle de confiance de niveau $1 - \alpha$ pour θ , alors*

$$RC = \{(x_1, \dots, x_n) \mid I(x_1, \dots, x_n) \cap \Theta_0 = \emptyset\}$$

est une région critique pour un test de niveau au plus α . Si $I(x_1, \dots, x_n)$ est un intervalle de confiance de niveau asymptotique $1 - \alpha$ pour θ , alors RC est une région critique pour un test de niveau asymptotique au plus α .

Démonstration. On ne donne la preuve que dans le cas $I(x_1, \dots, x_n)$ de niveau $1 - \alpha$, laissant au lecteur le soin de rédiger le cas asymptotique. On veut montrer que

$$\sup_{\theta \in \Theta_0} P_\theta((x_1, \dots, x_n) \in RC) \leq \alpha.$$

Soit donc $\theta \in \Theta_0$. Par définition, l'intervalle $I(x_1, \dots, x_n)$ contient θ avec probabilité $1 - \alpha$, c'est à dire

$$P_\theta(\theta \notin I(x_1, \dots, x_n)) = \alpha.$$

Or,

$$I(x_1, \dots, x_n) \cap \Theta_0 = \emptyset \implies \theta \notin I(x_1, \dots, x_n),$$

ce qui se réécrit

$$\{I(x_1, \dots, x_n) \cap \Theta_0 = \emptyset\} \subset \{\theta \notin I(x_1, \dots, x_n)\}.$$

3.3. CONSTRUCTION D'UN TEST À PARTIR D'UN INTERVALLE DE CONFIANCE 71

On en déduit alors

$$P_\theta((x_1, \dots, x_n) \in RC) = P_\theta(\{I(x_1, \dots, x_n) \cap \Theta_0 = \emptyset\}) \leq P_\theta(\{\theta \notin I(x_1, \dots, x_n)\}) = \alpha,$$

ce qui est précisément ce qu'on voulait démontrer. \odot

Exemple 3.3.2. Supposons que X_i suive une loi $\mathcal{N}(m, \sigma^2)$ avec σ connue. On considère l'alternative

$$H_0 : m = m_0 \qquad H_1 : m \neq m_0,$$

c'est à dire $\Theta_0 = \{m_0\}$. D'après l'exemple 2.3.3, un intervalle de confiance de niveau $1 - \alpha$ pour m est donné par

$$I = \left[\hat{m}_1 - \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} ; \hat{m}_1 + \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \right],$$

ce qui fournit la région critique

$$RC = \left\{ m_0 \notin \left[\hat{m}_1 - \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} ; \hat{m}_1 + \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \right] \right\} = \left\{ \hat{m}_1 \notin \left[m_0 - \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} ; m_0 + \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \right] \right\}.$$

On retrouve donc la région critique construite à l'exemple 3.1.7.

Si on considère maintenant l'alternative

$$H_0 : m \leq m_0 \qquad H_1 : m > m_0$$

on a alors $\Theta_0 =] - \infty ; m_0]$. Le même intervalle de confiance donne la région critique

$$RC = \left\{ \hat{m}_1 > m_0 + \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}}}{\sqrt{n}} \right\}.$$

On peut cependant aussi utiliser ici l'intervalle de confiance

$$\tilde{I} = \left[\hat{m}_1 - \frac{\sigma \cdot q_{1-\alpha}^{\mathcal{N}}}{\sqrt{n}} ; +\infty \right],$$

qui nous donne la région critique

$$\widetilde{RC} = \left\{ \hat{m}_1 > m_0 + \frac{\sigma \cdot q_{1-\alpha}^{\mathcal{N}}}{\sqrt{n}} \right\}.$$

Comme $q_{1-\alpha}^{\mathcal{N}} < q_{1-\frac{\alpha}{2}}^{\mathcal{N}}$, il est préférable d'utiliser la région \widetilde{RC} plutôt que RC pour cette deuxième alternative.

Exemple 3.3.3. Supposons que X_i suive une loi $Ber(\theta)$. On considère l'alternative

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta \neq \theta_0,$$

c'est à dire $\Theta_0 = \{\theta_0\}$. D'après l'exemple 2.4.4, un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ est donné par

$$I = \left[\hat{m}_1 - \sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}} ; \hat{m}_1 + \sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right].$$

On trouve donc la région critique

$$RC = \left\{ \hat{m}_1 \notin \left[\theta_0 - \sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}} ; \theta_0 + \sqrt{\frac{\hat{m}_1(1 - \hat{m}_1)}{n}} \cdot q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right] \right\}.$$

Revenons à l'exemple de la pièce de monnaie de l'exemple 3.0.1. Pour $\alpha = 0,05$, on a les valeurs numériques

$$\theta_0 = \frac{1}{2}, \quad n = 100, \quad \hat{m}_1 = 0,58, \quad q_{0,975}^{\mathcal{N}} = 1,96,$$

ce qui donne approximativement

$$RC = \{\hat{m}_1 \notin [0,4 ; 0,6]\}$$

On conserve donc l'hypothèse H_0 . Si on avait obtenu la même valeur $\hat{m}_1 = 0,58$ mais avec cette fois $n = 1000$, alors on aurait $RC = \{\hat{m}_1 \notin [0,47 ; 0,53]\}$ et on rejeterait H_0 pour H_1 .

3.4 Test de Student

Le test de Student est un cas particulier de test construit à partir d'un intervalle de confiance asymptotique. On suppose ici que $X_j \in L^2$ est d'espérance et de variance inconnues. On pose $m = E[X_i]$ et $\sigma^2 = Var(X_i)$, et on considère l'alternative

$$H_0 : m = m_0 \qquad H_1 : m \neq m_0,$$

On considère un estimateur consistant $\hat{\sigma}_n$ de σ , et on pose

$$T_n = \sqrt{n} \frac{\hat{m}_1 - m_0}{\hat{\sigma}_n}.$$

Proposition 3.4.1. *La région*

$$RC = \left\{ |T_n| > q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right\}$$

fournit une région critique de niveau asymptotique α . De plus, ce test est consistant.

Démonstration. Sous l'hypothèse H_0 , on a par le TCL

$$\sqrt{n} \frac{\hat{m}_1 - m_0}{\sigma} \xrightarrow[\text{Loi}]{} \mathcal{N}(0, 1).$$

Puisque $\hat{\sigma}_n$ est un estimateur fortement consistant de σ , le théorème de Slutsky nous donne

$$T_n = \frac{\sigma}{\hat{\sigma}_n} \frac{\sqrt{n}(\hat{m}_1 - m_0)}{\sigma} \xrightarrow[\text{Loi}]{} \mathcal{N}(0, 1).$$

Donc

$$\left[\hat{m}_1 - \frac{\hat{\sigma}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}}; \hat{m}_1 + \frac{\hat{\sigma}_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}} \right]$$

est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour m , et la première assertion découle maintenant de la proposition 3.3.1.

Supposons maintenant que $m \neq m_0$. On a alors

$$T_n = \sqrt{n} \frac{\hat{m}_1 - m}{\hat{\sigma}_n} + \sqrt{n} \frac{m - m_0}{\hat{\sigma}_n}.$$

Le TCL implique que

$$\sqrt{n} \frac{\hat{m}_1 - m}{\hat{\sigma}_n} \xrightarrow[\text{Loi}]{} \mathcal{N}(0, 1).$$

La suite $\frac{m-m_0}{\hat{\sigma}_n}$ converge presque sûrement vers $\frac{m-m_0}{\sigma}$, et puisque $m \neq m_0$, on a

$$\sqrt{n} \frac{|m - m_0|}{\hat{\sigma}_n} \xrightarrow[\text{p.s.}]{} +\infty.$$

On obtient en particulier

$$P_m(|T_n| > q_{1-\frac{\alpha}{2}}^{\mathcal{N}}) \xrightarrow[n \rightarrow +\infty]{} 1,$$

ce qui signifie précisément que le test est consistant. ☺

Exemple 3.4.2. Revenons à la pièce de monnaie de l'exemple 3.0.1. En prenant $\hat{\sigma}_n = \sqrt{\hat{m}_1(1 - \hat{m}_1)}$, la test de Student est exactement celui construit à l'exemple 3.3.3 à partir d'un intervalle de confiance.

Remarque 3.4.3. On a traité ici le cas du test bilatéral, on laisse le soin au lecteur d'adapter la méthode au cas unilatéral.

3.5 Test d'adéquation du χ^2

On veut tester ici si il n'est pas déraisonnable de penser qu'un échantillon suit une loi discrète donnée. On considère ici une suite (X_n) de variables aléatoires indépendantes et identiquement distribuées à valeur dans $\{1, \dots, k\}$. Étant donné p_1, \dots, p_k des nombres positifs tels que

$$p_1 + \dots + p_k = 1,$$

on considère l'alternative

$$H_0 : \forall j, P(X_i = j) = p_j \qquad H_1 : \exists j, P(X_i = j) \neq p_j,$$

On pose

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i=j\}} \quad \text{et} \quad Q_n = n \sum_{j=1}^k \frac{(\hat{p}_j - p_j)^2}{p_j} = n \sum_{j=1}^k \frac{\hat{p}_j^2}{p_j} - n.$$

On admet le résultat suivant.

Théorème 3.5.1. *Sous H_0 , on a*

$$Q_n \xrightarrow[\text{Loi}]{} \chi^2(k-1).$$

Sous H_1 , on a

$$Q_n \xrightarrow[\mathbb{P}]{} +\infty.$$

En particulier, l'ensemble

$$RC = \left\{ Q_n > q_{1-\alpha}^{\chi^2(k-1)} \right\}$$

fournit une région critique pour un test consistant de niveau asymptotique α .

Exemple 3.5.2. On demande à 100 personnes de choisir un chiffre entre 0 et 9, et on cherche à savoir si les 100 chiffres obtenus forment bien une distribution uniforme sur $\{0, 1, \dots, 9\}$. On a les résultats suivants :

j	0	1	2	3	4	5	6	7	8	9
$100\hat{p}_j$	10	8	9	14	8	9	11	9	12	10
p_j	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

On calcule alors $Q_{100} = 3,2$. Comme $q_{0,95}^{\chi^2(9)} = 16,9 \gg Q_{100}$, on accepte H_0 .

Remarque 3.5.3. Il existe une version continue du test d'adéquation du χ^2 , qui s'appelle le test de Kolmogorov-Smirnov.

3.6 Test d'indépendance du χ^2

On veut tester ici l'indépendance de deux caractéristiques, par exemple la couleur des cheveux et des yeux. On considère donc une suite (X_n) de variables aléatoire indépendantes et identiquement distribuées à valeurs dans un ensemble fini J , et une suite (Y_n) de variables aléatoire indépendantes et identiquement distribuées à valeurs dans un ensemble fini K . On considère alors l'alternative

$$H_0 : (X_n) \text{ et } (Y_n) \text{ sont indépendante} \quad H_1 : (X_n) \text{ et } (Y_n) \text{ sont dépendante.}$$

On rappelle que les variables aléatoires X_i et Y_i sont indépendantes si et seulement si

$$\forall j \in J, \forall k \in K, P(X_i = j, Y_i = k) = P(X_i = j) \cdot P(Y_i = k).$$

On pose

$$N_{j,k} = \text{Card} \{i \in \{1, \dots, n\} \mid X_i = j \text{ et } Y_i = k\},$$

$$N_{j,\bullet} = \text{Card} \{i \in \{1, \dots, n\} \mid X_i = j\}, \quad N_{\bullet,k} = \text{Card} \{i \in \{1, \dots, n\} \mid Y_i = k\}.$$

On a donc

$$N_{j,\bullet} = \sum_{k \in K} N_{j,k} \quad \text{et} \quad N_{\bullet,k} = \sum_{j \in J} N_{j,k}.$$

On pose enfin

$$Q_n = \sum_{(j,k) \in J \times K} \frac{\left(N_{j,k} - \frac{N_{j,\bullet} \cdot N_{\bullet,k}}{n} \right)^2}{\frac{N_{j,\bullet} \cdot N_{\bullet,k}}{n}} = n \cdot \sum_{(j,k) \in J \times K} \frac{N_{j,k}^2}{N_{j,\bullet} \cdot N_{\bullet,k}} - n$$

On admet encore une fois le résultat suivant.

Théorème 3.6.1. *On note a le cardinal de J et b le cardinal de K . Sous H_0 , on a*

$$Q_n \xrightarrow[\text{Loi}]{} \chi^2((a-1)(b-1)).$$

Sous H_1 , on a

$$Q_n \xrightarrow[\mathbb{P}]{} +\infty.$$

En particulier, l'ensemble

$$RC = \left\{ Q_n > q_{1-\alpha}^{\chi^2((a-1)(b-1))} \right\}$$

fournit une région critique pour un test consistant de niveau asymptotique α .

Exemple 3.6.2. On veut comparer l'efficacité de deux médicaments semblables mais de prix différents. On teste pour cela un échantillon de 250 patients. Le prix du médicament est modélisé par une variable aléatoire X à valeurs dans $\{0 = \text{bon marché}, 1 = \text{cher}\}$, tandis que son efficacité est modélisée par une variable aléatoire Y à valeurs dans $\{0 = \text{non-guérison}, 1 = \text{guérison}\}$. On observe les résultats suivants

	médicament bon marché	médicament cher	$N_{\bullet,k}$
non-guérison	156	44	200
guérison	44	6	50
$N_{j,\bullet}$	200	50	250

On calcule alors

$$Q_{250} = 250 \times \left(\frac{156^2}{200 \times 200} + \frac{44^2}{50 \times 200} + \frac{44^2}{50 \times 200} + \frac{6^2}{50 \times 50} \right) - 250 = 2,5.$$

Comme $q_{0,95}^{\chi^2(1)} = 3,84 > Q_{250}$, on accepte H_0 . En d'autres termes, l'efficacité du médicament en question ne semble pas significativement liée à son prix. Autant acheter le moins cher !

3.7 Exercices

Exercice 3.1. On cherche à connaître la température d'ébullition μ , en degrés Celsius, d'un certain liquide. On effectue 16 expériences indépendantes pour mesurer cette température d'ébullition. On suppose que les mesures sont distribuées selon une variable aléatoire suivant une loi normale de moyenne μ et d'écart type σ inconnus. La moyenne empirique relevée est $\hat{m}_1 = 94,32$ degrés Celsius et l'écart type empirique $s = 1,2$. On souhaite déterminer si μ vaut 95 degrés Celsius ou non.

1. Décrire précisément les hypothèses à tester, proposer une règle de décision au niveau $\alpha = 5\%$ et conclure.
2. Que vaut la p -valeur de l'échantillon observé pour le test précédent ? Conclure au test en conséquence.

Exercice 3.2. Une usine fabrique des billes métalliques. Le diamètre en millimètres de ces billes est une variable aléatoire suivant une loi normale. L'usine s'est engagée à fournir à un client des billes dont le diamètre moyen est de 25 millimètres. Le client réceptionne sa commande. Dans le lot reçu, il prélève un échantillon de 20 billes choisies au hasard et mesure les diamètres suivants :

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
24,7	24,9	25,0	25,0	25,1	25,1	25,1	25,2	25,3	25,4
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
24,8	24,9	25,0	25,0	25,1	25,1	25,2	25,3	25,3	25,5

Peut-on dire que l'usine a respecté ses engagements ? Faire un test d'hypothèses pour y répondre.

Exercice 3.3. On désire tester si un médicament a une influence sur le comportement psychomoteur. On choisit au hasard 20 sujets qu'on répartit au hasard en deux groupes : le groupe témoin (loi X) et le groupe expérimental (loi Y). On leur fait subir la même expérience psychomotrice. On a administré auparavant le médicament aux sujets du groupe expérimental et un placebo au groupe témoin. Les résultats sont les suivants :

Groupe témoin (X)	166	167	169	170	174	173	172	170	166	173
Groupe expérimental (Y)	167	162	165	168	162	160	164	158	165	169

On a alors $\hat{m}_{1,X} = 170$, $s_X = 2.98$, $\hat{m}_{1,Y} = 164$ et $s_Y = 2.98$. On suppose que dans chaque groupe les résultats sont distribués selon une loi Gaussienne, que la variance est la même pour les deux groupes et que les performances des sujets sont indépendantes. Tester au niveau 0.05 l'hypothèse selon laquelle le médicament n'a aucun effet sur le comportement psychomoteur.

Exercice 3.4. Le revenu annuel en euros d'un individu d'une certaine population est une variable aléatoire réelle X de densité : $f_X(x) = \theta x^{-(\theta+1)} \mathbb{1}_{[1,+\infty]}(x)$ où $\theta > 1$ est un réel inconnu. Soit $n \in \mathbb{N}^*$ et (X_1, \dots, X_n) un n -échantillon de X . On pose $\hat{T}_n = \frac{1}{n} \sum_{i=1}^n \ln(X_i)$.

1. Étudier la convergence en loi de $\sqrt{n}(\theta \hat{T}_n - 1)$.
2. En déduire une région critique de niveau asymptotique $1 - \alpha$ pour tester

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta \neq \theta_0.$$

Exercice 3.5. Le nombre d'absences par semaine dans une entreprise est une variable aléatoire réelle X suivant la loi de Poisson de paramètre λ . Ici, $\lambda > 0$ est un réel inconnu. Soit $n \in \mathbb{N}^*$ et (X_1, \dots, X_n) un n -échantillon de X . On considère le test d'hypothèses :

$$H_0 : \lambda = 1 \qquad H_1 : \lambda \neq 1.$$

1. Déterminer l'estimateur par maximum de vraisemblance $\hat{\lambda}$ de λ .
2. Quel est la loi de $n\hat{\lambda}$ lorsque $\lambda = 1$?
3. En déduire une région critique de niveau $\alpha \in [0, 1]$ pour le test précédent.
4. Que vaut la p -valeur du test précédent pour un échantillon (X_1, \dots, X_n) ?
5. On dispose des données suivantes, où n_i est le nombre de semaines pour lesquelles on a relevé i absences :

i	0	1	2	3
n_i	5	6	2	3

Calculer la p -valeur du test avec ces données et conclure.

Exercice 3.6. La marque Smorties produit des bonbons au chocolat de six couleurs. Le responsable de la communication affirme que la proportion de chaque couleur est de 30% pour le brun (B), 20% pour le jaune (J), 20% pour le rouge (R), 10% pour l'orange (O), 10% pour le vert (V) et 10% pour le doré (D) dans tout échantillon de grande taille. Une expérience réalisée sur un échantillon de 370 bonbons donne les comptages suivants :

Couleur	B	J	R	O	V	D
Nombre de bonbons	84	79	75	49	36	47

Peut-on conclure, au niveau 5%, que le responsable de la communication a tort ?

Exercice 3.7. Un cabinet de communication est chargé de mettre en place un dispositif d'information juridique portant sur le harcèlement sexuel des femmes au travail. Avant de lancer une campagne nationale, ce cabinet décide de tester différents formats de communication et souhaite évaluer les éléments retenus par les femmes à qui ils sont présentés. Deux cent dix-sept femmes acceptent d'y prendre part. Comme il est important de bien contrôler le niveau des participantes à ce pré-test, le cabinet souhaite vérifier si l'échantillon possède les mêmes caractéristiques de formation scolaire que celles que l'on trouve dans la population féminine générale. À l'aide d'une classification en six catégories de diplômes, on observe les répartitions suivantes :

	% national	Nb de femmes
Aucun diplôme ou certificat d'étude primaire	20,56 %	51
BEPC seul	7,65 %	8
CAP, BEP, ou autre diplôme de ce niveau	24,05 %	60
Bac, BP, ou autre diplôme de ce niveau	17,61 %	35
Bac + 2	15,57 %	38
Diplôme supérieur à Bac +2	14,56 %	25

La répartition des femmes de l'échantillon respecte-t-elle la répartition nationale ?

Exercice 3.8. Le tableau suivant indique le classement de 124 personnes selon la couleur de leurs yeux et la couleur de leurs cheveux.

Yeux \ Cheveux	Blonds	Bruns	Noirs	Roux	Total
	Bleus	25	9	3	
Gris ou verts	13	17	10	7	47
Marrons ou noirs	7	13	8	5	33
Total	45	39	21	19	124

Peut-on conclure à l'indépendance de ces deux caractères ?

Exercice 3.9. Le tableau ci-dessous donne le nombre d'étudiants qui ont été brillants ou médiocres devant trois examinateurs A, B et C.

Résultat	Examineur		
	A	B	C
Brillants	50	47	56
Médiocres	5	14	8

Peut-on affirmer, au niveau 5%, que le résultat d'un individu dépend de l'examineur ?

Bibliographie

- [DKLM05] F.M. DEKKING, C. KRAAIKAMP, H.P. LOPUHAÄ et L.E. MEESTER : *A Modern Introduction to Probability and Statistics, Understanding Why and How*. Springer, 2005.
- [Ell17] J. ELLENBERG : *L'art de ne pas dire n'importe quoi*. Cassini, 2017.
- [Lec19] J.-P. LECOUTRE : *Statistique et Probabilités*. Dunod, 2019.