

Statistiques Fondamentales

Simon Coste

2024-01-09

Table des matières

Organisation	4
Utiliser ce site	4
1 Introduction	5
1.1 Un exemple pour fixer les idées	5
1.2 Qu'est-ce qu'un problème statistique ?	6
1.3 Qu'est-ce qu'un estimateur ?	8
1.4 Points de vue	8
2 Estimation de paramètre	10
2.1 Précision d'un estimateur	10
2.2 Consistances	10
2.3 Normalité asymptotique	11
2.4 Deux outils sur la normalité asymptotique	11
2.5 La méthode des moments	13
2.6 Deux estimateurs importants	14
3 Exercices	15
3.0.1 Questions	15
3.0.2 Variance empirique	15
3.0.3 Estimation de masse	15
3.0.4 Uniforme (1)	16
3.0.5 Uniforme (2)	16
3.0.6 Gamma	16
3.0.7 Gumbel	17
3.0.8 Yule-Simon	17
4 Intervalles de confiance	18
4.1 Principe	18
4.2 Exemples gaussiens	18
4.2.1 Estimation de μ	19
4.2.2 Estimation de σ	20
4.3 Exemples asymptotiques	21
5 Outils pour construire des intervalles de confiance	22
5.1 Quantiles	22

5.2	Calculs de lois	23
5.2.1	Loi du chi-deux	23
5.3	Inégalités de concentration	24
6	Exercices	25
6.1	Questions	25
6.2	Poisson	25
6.3	Uniforme	25
7	Test d'hypothèse	26
8	Économétrie	27
9	Théorie de l'information	28
10	Estimation de densité	29
	Et après ?	30

Organisation

Bienvenue sur la page du cours de Statistiques Fondamentales (STF8) du master Mathématiques Fondamentales et Appliquées de l'Université Paris-Cité. Les notes de cours sont accessibles à tous. De nombreux auteurs s'y sont succédés ; je suis le dernier en date, mais les versions précédentes ont été travaillées par Clément Levrard, Stéphane Boucheron, Stéphane Gaïffas, Pierre Youssef.

- Les CM ont lieu les jeudi à (8h30 - 10h30), et les vendredi (10h45 - 12h45) **sauf le premier cours qui a lieu lundi 8 janvier à 10h45-12h45.**
- Les TD ont lieu lundi (13h45 - 16h45) et vendredi (13h30 - 15h30), de lundi 8 janvier à vendredi 16 février.
- Il y aura deux contrôles de 2h, le vendredi 26 janvier et lundi 12 février.
- L'examen a lieu le 1er mars de 13h30 à 16h30.
- Il y aura une interro de 5 minutes chaque semaine le jeudi.

Utiliser ce site

Chaque chapitre de ce livre contient une page dédiée au cours théorique, et contiendra dans un futur proche une page d'exercices.

La saveur du cours est essentiellement mathématique et nous n'aurons pas de TP d'info ; cependant, je vous recommande vraiment d'essayer d'appliquer tout ça via votre langage de programmation favori, c'est-à-dire ~~Python~~ **R SAS C++ Julia**. J'essaierai autant que possible de fournir des mini-jeux de données avec des petits challenges pour appliquer ce que vous apprenez en cours.

Ces notes sont mises en lignes et totalement accessibles via [Quarto](#). Si vous savez comment utiliser `git`, n'hésitez pas à corriger toutes les erreurs que vous pourriez voir (et Dieu sait qu'elles seront nombreuses) via des pull requests.

1 Introduction

Les outils des statistiques furent créés pour analyser des phénomènes quantitatifs dans lesquels la présence de *bruit* ou de *hasard* rendait l'analyse classique moins opérante. Il peut typiquement s'agir de problèmes dans lesquels de nombreuses données indépendantes ont été générées par le même phénomène. Dans cette section, nous allons développer un exemple pour bien comprendre les questions qui se posent et la façon de les résoudre, puis nous poserons quelques bases qui nous permettront d'utiliser le langage des mathématiques.

1.1 Un exemple pour fixer les idées

Une grande enseigne de distribution possède $n = 100$ magasins identiques, qui génèrent chaque année un chiffre d'affaire annuel (CA, en millions d'euros). Ce chiffre oscille autour d'une valeur de référence μ . Cette valeur n'est pas observée ; ce qui est observé, ce sont tous les chiffres d'affaires des n magasins, qui fluctuent tous autour de la vraie valeur μ . Ces fluctuations proviennent de nombreuses sources : erreurs comptables, perturbations des ventes dues aux fournisseurs ou aux prix, etc. Ce qu'on observe, c'est donc des chiffres x_1, \dots, x_n qui ne sont pas tous égaux ; comment avoir une idée de la véritable valeur de μ ?

Estimation. Évidemment, la moyenne empirique

$$\bar{x}_n = \frac{x_1 + \dots + x_n}{n}$$

vient naturellement à l'esprit. En faisant le calcul, on trouve $\bar{x}_n \approx 21,6$. Cette valeur est une *estimation* du CA moyen μ . Ce chiffre peut être utilisé par l'enseigne, par exemple pour jauger la rentabilité d'un possible plan d'ouverture de nouveaux magasins.

Précision. On pourrait se demander à quel point cette estimation est précise ou, disons, essayer de quantifier l'erreur possible qu'on fait si l'on dit que μ est égal à 21,6 millions d'euros. Cela nécessite de faire quelques hypothèses sur le hasard qui génère les fluctuations des x_i autour de μ . Ces fluctuations observées au cours de l'année proviennent de l'agrégation de toutes les fluctuations quotidiennes, lesquelles sont à peu près indépendantes, et pour cette raison on peut supposer (pour commencer) que ces fluctuations sont gaussiennes et ont à peu près la même variance, disons $\sigma^2 = 1$. Comme on a supposé que les x_i sont des réalisations d'une loi gaussienne $N(\mu, 1)$, alors on sait que \bar{x}_n est la réalisation d'une loi $N(\mu, 1/n)$, ou encore que $\bar{x}_n - \mu$ est la réalisation d'une gaussienne centrée de variance $1/n$. Les lois gaussiennes sont bien connues ; par exemple, avec probabilité supérieure à 99%, une

gaussienne $N(0, \sigma^2)$ est comprise entre les valeurs $-2,96\sigma$ et $2,96\sigma$. Autrement dit, il y a 99% de chances pour que le nombre $|\bar{x} - \mu|$, qui représente l'erreur d'estimation, soit plus petite que $2,96/\sqrt{n} = 2,96/10 \approx 0,3$.

Ce dernier raisonnement peut être vu d'une autre façon. Dire que \bar{x}_n et μ ne diffèrent pas de plus de 0,3, c'est équivalent à dire que μ appartient à l'intervalle $[\bar{x} - 0,3, \bar{x} + 0,3]$. En d'autres termes, avec une probabilité supérieure à 99%, le vrai CA μ de chaque magasin se situe entre 21,3 et 21,9. Cela laisse tout de même une chance de 1% que le paramètre μ ne soit pas dans cette région.

Tests. Il existe encore un autre point de vue sur ce problème. Par exemple, le conseil d'administration de la firme veut s'assurer que le dirigeant a bien tenu sa promesse selon laquelle le CA de chaque magasin était supérieur à 21 millions d'euros. La valeur exacte de μ n'est pas le plus important : ce qui nous intéresse maintenant, c'est plutôt d'être sûrs que μ n'est pas inférieur au seuil de 21. Le dirigeant, fin statisticien, effectue alors un raisonnement par l'absurde *en probabilité*. Supposons que le CA μ soit effectivement égal à 21 (ou même, inférieur). Alors, par les mêmes calculs que ci-dessus, cela voudrait dire qu'avec 99% de chances, \bar{x}_n et 21 ne devraient pas différer de plus de 0,3 ; autrement dit, que \bar{x}_n devrait se situer entre 20,7 et 21,3. Ce n'est pas le cas, puisque $\bar{x}_n = 21,6$. Si μ est réellement plus petit que 21, alors ce qu'on a observé est extrêmement peu probable. Par contraposée probabiliste, il est raisonnable de rejeter l'hypothèse selon laquelle μ est inférieur à 21.

Les trois points de vue donnés ci-dessus sont en quelque sorte les piliers de l'analyse statistique. L'estimation consiste à deviner une valeur cachée dans du bruit ; les intervalles de confiance consistent à donner une région dans laquelle se trouve cette valeur ; les tests d'hypothèse permettent de raisonner de façon logique sur cette valeur.

L'objectif du cours de statistiques de quantifier l'incertitude liée au hasard dans chacun de ces objectifs. Comme dans les exemples donnés ci-dessus, c'est un ensemble de méthodes scientifiques qui s'appuient sur la théorie des probabilités ; dans ce cours, on fera des *hypothèses* sur le hasard qui est en jeu, et on en tirera des conséquences *probables* sur le modèle sous-jacent. En théorie des probabilités, le jeu est plutôt inverse : partant d'un modèle probabiliste fixé, on essaie de déterminer quel sera le comportement des réalisations de ce modèle. Il semble difficile de faire l'un sans l'autre.

1.2 Qu'est-ce qu'un problème statistique ?

Il n'y aurait pas de statistiques s'il n'y avait pas de monde réel, et comme chacun sait, le monde réel est principalement composé de quantités aléatoires.

Un problème statistique tire donc toujours sa source d'un ensemble d'observations, disons n observations notées x_1, \dots, x_n ; cet ensemble d'observations est appelé un *échantillon*. L'hypothèse de base de tout travail statistique consiste à supposer que cet échantillon suit une certaine loi de probabilité ; l'objectif est de trouver laquelle. Évidemment, on ne va pas partir de rien : il faut bien faire des hypothèses minimales sur cette loi. Ce qu'on appelle un *modèle statistique* est le choix d'une famille de lois de probabilités que l'on suppose pertinentes.

Définition 1.1. Formellement, choisir un modèle statistique revient à choisir trois choses :

- \mathcal{X} , l'espace dans lequel vit notre échantillon ;
- \mathcal{F} , une tribu sur \mathcal{X} , pour donner du sens à ce qui est observable ou non ;
- $(P_\theta)_{\theta \in \Theta}$, une famille de mesures de probabilités sur \mathcal{X} indexée par $\theta \in \Theta$, où Θ est appelé espace des paramètres. On écrira fréquemment \mathbb{E}_θ ou Var_θ pour désigner des espérances, variances, etc., calculées avec la loi P_θ .

En pratique, dans ce cours, on aura toujours un échantillon (x_1, \dots, x_n) où les x_i vivent dans un même espace, disons \mathbb{R}^d pour simplifier. On devrait donc écrire $\mathcal{X} = \mathbb{R}^{d \times n}$; et l'on fera toujours l'hypothèse que ces observations sont indépendantes les unes des autres, et que ces observations ont la même loi de probabilité. Autrement dit, on se donnera toujours une mesure p_θ sur \mathbb{R}^d et on supposera que la loi de notre échantillon est $P_\theta = p_\theta^{\otimes n}$. Dans ce cadre, les observations x_i sont des *réalisations* de variables aléatoires X_i iid de loi p_θ .

Il faut prendre garde à distinguer les variables aléatoires X_i , qui sont des objets théoriques, de leurs réalisations x_i , qui, elles, sont bel et bien observées.

Définition 1.2. On dit qu'un modèle statistique est identifiable si $\theta \neq \theta'$ entraîne $P_\theta \neq P_{\theta'}$.

Si l'on a bien choisi notre modèle statistique, alors il existe un « vrai » paramètre, disons θ_* , tel que les observations x_1, \dots, x_n sont des réalisations de loi p_{θ_*} . L'objectif est alors de trouver θ_* ou quelque information que ce soit le concernant.

Dans un modèle identifiable, la statistique inférentielle (classique) permet de faire trois choses :

- Trouver une valeur approchée du vrai paramètre θ_* (estimation ponctuelle).
- Donner une zone de Θ dans laquelle le vrai paramètre θ_* a des chances de se trouver (intervalle de confiance).
- Répondre à des questions binaires sur θ_* , par exemple « θ_* est-il positif ? ».

1.3 Qu'est-ce qu'un estimateur ?

Définition 1.3. Une *statistique* est une fonction mesurable des observations. Plus formellement, si le modèle statistique fixé est $(\mathcal{X}, \mathcal{F}, P)$, alors une statistique est n'importe quelle fonction mesurable de $(\mathcal{X}, \mathcal{F})$.

- 1) Le premier point important est qu'une statistique ne peut pas prendre θ en argument. Ses valeurs ne doivent dépendre du paramètre θ qu'au travers de P_θ .
- 2) Le second point important est que, si X est une variable aléatoire et T une statistique, alors $T(X)$ est une variable aléatoire. On peut donc définir des quantités théoriques liées à T : typiquement, si X a pour loi P_θ , on peut définir la valeur moyenne de T sous le modèle P_θ comme

$$\mathbb{E}_\theta[T(X)] = \int_{\mathcal{X}} T(x) P_\theta(dx)$$

ou encore sa variance $\mathbb{E}_\theta[T(X)^2] - (\mathbb{E}_\theta[T(X)])^2$, etc. On peut aussi calculer la valeur de cette statistique sur l'échantillon dont on dispose, c'est-à-dire $T(x_1, \dots, x_n)$. Par exemple, la moyenne empirique d'un n -échantillon réel est la fonction $T : (a_1, \dots, a_n) \rightarrow n^{-1}(a_1 + \dots + a_n)$. Si les x_i sont des réalisations des variables aléatoires X_i , alors $T(x_1, \dots, x_n)$ est une réalisation de la variable aléatoire $T(X_1, \dots, X_n)$.

- 3) Ce qui ne se voit pas dans la définition, c'est qu'une bonne statistique devrait être facilement calculable ; à la place de *statistique*, on peut penser à *algorithme* : une bonne statistique doit pouvoir être calculée facilement par un algorithme ne prenant en entrée que les échantillons x_i .

Si le but est de deviner la valeur de θ à partir des observations, il est naturel de considérer des statistiques à valeurs dans Θ . C'est précisément la définition d'un estimateur.

Définition 1.4. Dans le modèle $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$, un estimateur de θ est une statistique à valeurs dans Θ .

En fait, on n'est pas obligés de vouloir estimer précisément θ . Peut-être qu'on veut estimer quelque chose qui dépend de θ , mais qui n'est pas θ ; disons, une fonction $\varphi(\theta)$. Dans ce cas, un estimateur de $\varphi(\theta)$ sera simplement une statistique à valeurs dans l'espace où vit $\varphi(\theta)$.

1.4 Points de vue

Inférence paramétrique. La plupart des expériences/modèles statistiques que nous rencontrerons dans ce cours, seront de nature dite *paramétrique*, autrement dit indexés par des parties de \mathbb{R}^d . Le mot “paramètre” est en lui-même trompeur : on parle souvent de paramètre d'une distribution pour désigner ce qui devrait plutôt s'appeler une fonctionnelle. Par exemple,

la moyenne, la covariance d'une distribution sur \mathbb{R}^d sont des paramètres de cette distribution. Les quantiles, l'asymétrie, la kurtosis sont d'autres paramètres.

Statistique non paramétrique. Tous les modèles ne sont pas paramétriques au sens ci-dessus : dans de nombreux développements des statistiques, par exemple en estimation de densité, on travaille sur des modèles plus riches qui n'admettent pas de paramétrisation naturelle par une partie d'un espace euclidien de dimension finie. C'est ce qu'on appelle l' *estimation non-paramétrique*. Nous y reviendrons au dernier chapitre.

Statistique bayésienne. En statistique paramétrique, les paramètres θ déterminent le hasard qui génère les observations x_i . La statistique bayésienne consiste à renverser le point de vue, et à rendre le paramètre θ lui-même aléatoire ; sa loi, appelée *prior*, mesure “le degré de connaissance a priori” qu'on en a. La règle de Bayes explique comment cette loi est modifiée par les observations. C'est un point de vue qui ne sera pas abordé dans ce cours.

2 Estimation de paramètre

On fixe un modèle statistique $(\mathcal{X}, \mathcal{F}, (P_\theta))$, et l'on cherche à estimer le paramètre θ ou un autre paramètre qui dépend de θ . Dans ce chapitre, on explique comment juger la qualité d'un estimateur, et l'on donne une technique générale pour construire de bons estimateurs dans des situations assez naturelles.

2.1 Précision d'un estimateur

Définition 2.1 (Biais, risque quadratique).

- Le biais de $\hat{\theta}$ est la quantité $\mathbb{E}_\theta[\hat{\theta} - \theta]$. L'estimateur est dit *sans biais* s'il est de biais nul.
- Le risque quadratique de $\hat{\theta}$ est la quantité $\mathbb{E}_\theta[|\hat{\theta} - \theta|^2]$.

En pratique, on peut vouloir estimer non pas θ lui-même, mais un paramètre $\psi = \psi_\theta$ qui dépend de θ , comme $\cos(\theta)$ ou $|\theta|$ par exemple. Dans ce cas, si $\hat{\psi}$ est un estimateur de ψ alors le biais est défini par $\mathbb{E}_\theta[\hat{\psi} - \psi_\theta]$ et le risque quadratique par $\mathbb{E}_\theta[|\hat{\psi} - \psi_\theta|^2]$.

La dépendance du risque quadratique vis à vis de la taille de l'échantillon est une question importante en statistique mathématique. Elle concerne la vitesse d'estimation (pour une suite d'expériences donnée, quelles sont les meilleures vitesses envisageables, et comment les obtenir ?).

Théorème 2.1 (Décomposition biais-variance).

$$\mathbb{E}_\theta[|\hat{\theta} - \theta|^2] = \underbrace{\text{Var}_\theta(\hat{\theta})}_{\text{variance}} + \underbrace{\mathbb{E}_\theta[\hat{\theta} - \theta]^2}_{\text{carré du biais}} .$$

2.2 Consistances

Pour introduire la notion de consistance d'une suite d'estimateurs, nous aurons besoin des notions de convergence en probabilité et de convergence presque sûre.

Définition 2.2. Une suite de variables aléatoires X_n à valeurs dans \mathbb{R}^k converge en probabilité vers une variable aléatoire X à valeurs dans \mathbb{R}^k , vivant sur cet espace probabilisé si et seulement si, pour tout $\epsilon > 0$

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

Définition 2.3 (consistance d'un estimateur). Une suite d'estimateurs $(\hat{\theta}_n)$ est consistante pour l'estimation de θ lorsque, pour tout $\theta \in \Theta$,

$$\forall \varepsilon > 0, \quad \lim_n P_\theta(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

La suite est fortement consistante si, pour tout $\theta \in \Theta$,

$$\hat{\theta}_n \rightarrow \theta \quad P_\theta - \text{p.s.}$$

2.3 Normalité asymptotique

Lorsqu'un estimateur est consistant, on peut se demander à quoi ressemblent ses fluctuations autour de sa valeur limite. Le théorème central limite indique que le comportement asymptotique gaussien est relativement fréquent, et beaucoup d'estimateurs sont des sommes de réalisations de variables indépendantes.

Définition 2.4 (normalité asymptotique). Soit θ un paramètre à estimer, et $\hat{\theta}_n$ une suite d'estimateurs de θ . On dit que ces estimateurs sont *asymptotiquement gaussiens* (ou *normaux*) si, après les avoir renormalisés convenablement, ils convergent en loi vers une loi gaussienne. Autrement dit, s'il existe une suite a_n de nombres réels tels que

$$a_n(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\text{loi}} N(0, \Sigma)$$

où Σ est une matrice de covariance qui dépend peut-être de θ — pour éviter les cas dégénérés, on demande à ce que Σ soit non-nulle.

2.4 Deux outils sur la normalité asymptotique

La normalité asymptotique n'est pas intéressante en elle-même ; l'idée est plutôt de chercher le comportement asymptotique de la statistique recentrée pour pouvoir en déduire des garanties en terme de risque asymptotique ou d'intervalle de confiance. Nous utiliserons cela de nombreuses fois dans la suite ; la normalité asymptotique sera par exemple la clé de la construction de nombreux intervalles de confiance. Aussi, prouver que des estimateurs sont asymptotiquement normaux est une tâche importante, qui est grandement simplifiée par les deux outils suivants.

Théorème 2.2 (Lemme de Slutsky). *Soit (X_n) une suite de variables aléatoire qui converge en loi vers X et (Y_n) une suite de variables aléatoires qui converge en probabilité (ou en loi) vers une constante c . Alors, le couple (X_n, Y_n) converge en loi vers (X, c) ; autrement dit, pour toute fonction continue bornée φ ,*

$$\mathbb{E}[\varphi(X_n, Y_n)] \rightarrow \mathbb{E}[\varphi(X, c)].$$

Démonstration. Fixons une fonction test φ continue à support compact, donc bornée par un certain M . Il faut montrer que $\mathbb{E}[\varphi(X_n, Y_n) - \varphi(X, c)]$ tend vers zéro. L'intégrande est égal à la somme de $A = \varphi(X_n, Y_n) - \varphi(X_n, c)$ et de $B = \varphi(X_n, c) - \varphi(X, c)$.

Comme X_n tend en loi vers X et que $t \rightarrow \varphi(t, c)$ est continue bornée, l'espérance de B tend vers zéro. Il faut donc montrer que l'espérance de A tend vers zéro. On fixe un $\varepsilon > 0$.

- Par le [théorème de Heine](#), φ est uniformément continue : il existe $\delta > 0$ tel que $|(x, y) - (x', y')| < \delta$ entraîne que $|\varphi(x, y) - \varphi(x', y')| < \varepsilon/2$.
- On introduit l'événement $\{|Y_n - c| \leq \delta\}$. Par le point précédent, sur cet événement on a $|A| < \varepsilon/2$. Hors de cet événement, on peut toujours borner $|A|$ par $2M$. On a donc

$$|\mathbb{E}A| \leq \mathbb{P}(|Y_n - c| \leq \delta)\varepsilon/2 + \mathbb{P}(|Y_n - c| > \delta)2M.$$

- Comme Y_n converge en probabilité vers c , lorsque n est assez grand on a $\mathbb{P}(|Y_n - c| > \delta) < \varepsilon/4M$.
- En regroupant tout ce qui a été dit, on obtient bien $|\mathbb{E}A| \leq \varepsilon$ dès que n est assez grand, ce qui montre bien que $\mathbb{E}A \rightarrow 0$.

□

Théorème 2.3 (Delta-méthode). *Soit (X_n) une suite de variables aléatoires réelles telle que $\sqrt{n}(X_n - \alpha)$ converge en loi vers $N(0, \sigma^2)$. Pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ dérivable en α (de dérivée non nulle en α), on a*

$$\sqrt{n}(g(X_n) - g(\alpha)) \xrightarrow[n \rightarrow \infty]{loi} N(0, g'(\alpha)^2 \sigma^2).$$

Plus généralement, si les X_n sont à valeurs dans \mathbb{R}^d et que $\sqrt{n}(X_n - \alpha) \rightarrow N(0, \Sigma)$, alors pour toute application $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$, la suite $\sqrt{n}(g(X_n) - g(\alpha))$ converge en loi vers

$$N(0, Dg(\alpha)\Sigma Dg(\alpha)^\top)$$

où $Dg(x)$ est la [matrice jacobienne](#) de g en x .

Démonstration. À écrire.

□

2.5 La méthode des moments

Il existe plusieurs techniques générales pour *construire* des estimateurs. La méthode des moments est la plus naturelle, et donne beaucoup des estimateurs avec de bonnes propriétés.

Dans un modèle statistique, supposons qu'on dispose d'une statistique intégrable T (pas forcément réelle), dont la moyenne n'est pas le paramètre θ lui-même, mais plutôt une *fonction* de θ :

$$\mathbb{E}_\theta[T(X)] = \varphi(\theta).$$

C'est cette fonction φ qu'on appelle *moment*. Typiquement,

- la moyenne d'une loi $\mathcal{E}(\theta)$ n'est pas θ mais $1/\theta$.
- la moyenne d'une loi log-normale de paramètres $(0, \sigma^2)$ est $e^{\sigma^2/2}$.

Prenons la moyenne empirique associée à cet estimateur, \bar{T}_n . Par la loi des grands nombres,

$$\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T(X_i) \rightarrow \varphi(\theta) \quad P_\theta - ps,$$

ce qui permet d'estimer $\varphi(\theta)$. Peut-on alors estimer θ ? Si la fonction φ est inversible et si \bar{T}_n appartient presque sûrement à l'ensemble de définition de φ^{-1} , alors $\varphi^{-1}(\bar{T}_n)$ est bien définie. Pour qu'en plus cette quantité converge presque sûrement vers θ , il faut s'assurer que φ^{-1} est continue. C'est par exemple le cas lorsque l'ensemble des paramètres Θ est un ouvert, et que φ est un difféomorphisme sur son image — une situation si fréquente qu'elle mérite son propre théorème, et si agréable qu'elle garantit que l'estimateur associé est asymptotiquement normal.

Théorème 2.4 (Estimation par moments). *Sous l'hypothèse mentionnée ci-dessus (la fonction φ est un difféomorphisme), l'estimateur*

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

est presque sûrement bien défini pour tout n suffisamment grand ; il est également consistant pour l'estimation de θ . En outre, si T est de carré intégrable, cet estimateur est asymptotiquement normal, au sens où $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une gaussienne centrée de matrice de covariance

$$(D\varphi(\theta))^{-1} \text{Cov}_\theta(T) (D\varphi(\theta))^\top)^{-1}.$$

Démonstration. La première partie a essentiellement été démontrée un peu plus haut. Pour la seconde, il faut d'abord remarquer que si T est de carré intégrable, alors $\sqrt{n}(\bar{T}_n - \varphi(\theta))$ converge vers une loi $N(0, \text{Cov}_\theta(T))$ par le TCL. Une simple application de la delta-méthode (Théorème 2.3) donne alors le résultat, puisque la matrice jacobienne de φ^{-1} en $\varphi(\theta)$ n'est autre que l'inverse de la matrice jacobienne de φ en θ .

□

2.6 Deux estimateurs importants

Deux estimateurs sont omniprésents en statistique : la moyenne empirique et la variance empirique. Ils sont pertinents dans n'importe quel modèle où les observations sont des réalisations de variables iid possédant une moyenne μ et une variance σ^2 .

La moyenne empirique est définie par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Il est évident que $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X] = \mu$. Cet estimateur est donc toujours sans biais, et son risque quadratique est égal à sa variance, c'est-à-dire $\frac{\sigma^2}{n}$.

L'estimateur de la variance empirique est défini comme

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

Théorème 2.5. *L'estimateur $\hat{\sigma}_n^2$ est sans biais.*

Démonstration. À écrire.

□

3 Exercices

3.0.1 Questions

- Montrer que la convergence en loi vers une constante implique la convergence en probabilité.
- Montrer que, si un modèle statistique n'est pas identifiable, alors il ne peut exister aucun estimateur convergent.
- Trouver un couple de variables aléatoires (X_n, Y_n) tel que X_n converge en loi et Y_n converge en loi, mais le couple ne converge pas en loi.
- On observe un échantillon de lois de Poisson de paramètre λ , que l'on estime par la moyenne empirique. Calculer le risque quadratique de cet estimateur.
- Quelle est la loi d'une somme de lois de Bernoulli indépendantes ? L'écart-type ?

3.0.2 Variance empirique

On se donne Y_1, \dots, Y_n , i.i.d de moyenne μ et variance σ^2 .

1. On suppose μ connu. Donner un estimateur non biaisé de σ^2 .
2. On suppose μ inconnu. Calculer l'espérance de $\sum_{i=1}^n (Y_i - \bar{Y}_n)^2$. En déduire un estimateur non biaisé de σ^2 .

3.0.3 Estimation de masse

Au cours de la seconde guerre mondiale, l'armée alliée notait les numéros de série X_1, \dots, X_n de tous les tanks nazis capturés ou détruits, afin d'obtenir un estimateur du nombre total N de tanks produits.

1. Proposer un modèle pour le tirage de X_1, \dots, X_n .
2. Calculer l'espérance de \bar{X}_n . En déduire un estimateur non biaisé de N . Indication: la loi de n tirages sans remise est échangeable.
3. Étudier la loi de $X_{(n)}$ et en déduire un estimateur non biaisé de N .
4. Proposer deux intervalles de confiance de niveau α . %de la forme $[aS, bS]$ avec $a, b \in \mathbb{R}$ et S une statistique résumée. On pourra utiliser le fait que l'inégalité de Hoeffding s'applique également aux tirages sans remise.

Selon Ruggles et Broodie (1947, JASA), la méthode statistique a fourni comme estimation une production moyenne de 246 tanks/mois entre juin 1940 et septembre 1942. Des méthodes d'espionnage traditionnelles donnaient une estimation de 1400 tanks/mois. Les chiffres officiels du ministère nazi des Armements ont montré après la guerre que la production moyenne était de 245 tanks/mois.

3.0.4 Uniforme (1)

On considère (X_1, \dots, X_n) un échantillon de loi uniforme sur $] \theta, \theta + 1[$.

1. Donner la densité de la loi de la variable $R_n = X_{(n)} - X_{(1)}$, où $X_{(1)} = \min(X_1, \dots, X_n)$ et $X_{(n)} = \max(X_1, \dots, X_n)$.
2. Étudier les différents modes de convergence de R_n quand $n \rightarrow \infty$.
3. Étudier le comportement en loi de $n(1 - R_n)$ quand $n \rightarrow \infty$.

3.0.5 Uniforme (2)

Soit X_1, \dots, X_n un échantillon de loi $\mathcal{U}([0, \theta])$, avec $\theta > 0$. On veut estimer θ .

1. Déterminer un estimateur de θ à partir de \bar{X}_n .
2. On considère l'estimateur $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Déterminer les propriétés asymptotiques de ces estimateurs.
3. Comparer les performances des deux estimateurs.

3.0.6 Gamma

La loi Gamma $\Gamma(\alpha, \beta)$ de paramètres $\alpha, \beta > 0$ a pour densité

$$x \mapsto \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0.$$

On se donne un échantillon (X_1, \dots, X_n) de loi $\Gamma(\alpha, \beta)$ et on cherche à estimer les paramètres.

1. On suppose le paramètre β connu. Proposer un estimateur de α par la méthode des moments.
2. On suppose à présent que les deux paramètres α, β sont inconnus. Proposer un estimateur de (α, β) par la méthode des moments.

3.0.7 Gumbel

La loi de Gumbel (centrale) de paramètre β a pour fonction de répartition $F(x) = e^{-e^{-x/\beta}}$. On observe un échantillon de lois de Gumbel et l'on cherche à estimer β .

1. Calculer la densité des lois de Gumbel, ainsi que leur moyenne et variance [indice : 0.57721...]
2. En déduire un estimateur convergent dont on calculera le risque quadratique et les propriétés asymptotiques.

3.0.8 Yule-Simon

Une variable aléatoire X suit la loi de Yule-Simon de paramètre $\rho > 0$ lorsque $\mathbb{P}(X = n) = \rho B(n, 1 + \rho)$, où $n \geq 1$ et B est la [fonction beta](#).

1. Montrer que si $\rho > 1$, alors $\mathbb{E}[X] = \rho/(\rho - 1)$.
2. Trouver un estimateur de ρ et donner ses propriétés.

4 Intervalles de confiance

4.1 Principe

Dans un modèle statistique, l'estimation du paramètre d'intérêt θ par intervalles de confiance consiste à spécifier un intervalle calculable à partir des données, et qui contient θ avec grande probabilité : en d'autres termes, une *région de confiance* pour θ .

Pour simplifier, on supposera d'abord que θ est un paramètre réel.

Définition 4.1 (intervalle de confiance). Un intervalle de confiance de niveau $1 - \alpha$ est un intervalle $I = [A, B]$ dont les bornes A, B sont des statistiques, et tel que pour tout θ ,

$$P_{\theta}(\theta \in I) \geq 1 - \alpha.$$

Un intervalle de confiance de niveau asymptotique $1 - \alpha$ est une *suite* d'intervalles $I_n = [A_n, B_n]$ dont les bornes A_n, B_n sont des statistiques, et tels que pour tout n ,

$$P_{\theta}(\theta \in I_n) \geq 1 - \alpha.$$

Le terme « niveau » désigne $1 - \alpha$; la vocation de ce nombre est d'être proche de 1, typiquement 99%. Le nombre α est parfois appelé « erreur » ou « marge d'erreur » ; la vocation de ce nombre est d'être proche de zéro, typiquement 1%.

Il n'y a rien d'autre à savoir sur les intervalles de confiance ; tout l'art de la chose consiste à savoir les construire. Commençons par trois exemples essentiels à plusieurs titres.

4.2 Exemples gaussiens

On dispose de variables aléatoires X_1, \dots, X_n gaussiennes de loi $N(\mu, \sigma^2)$. On va donner des intervalles de confiance pour l'estimation des paramètres μ et σ dans plusieurs cas de figure. Le moment est idéal pour rappeler l'existence et le calcul des *quantiles* d'une loi — voir ci dessous.

4.2.1 Estimation de μ

Lorsque σ est connue.

Nous avons déjà vu que la moyenne empirique \bar{X}_n est un estimateur sans biais de μ . Or, nous savons aussi la loi *exacte* de \bar{X}_n , qui est $N(\mu, \sigma^2/n)$. Autrement dit,

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \sim N(0, 1). \quad (4.1)$$

Si l'on se donne une marge d'erreur $\alpha = 1\%$, alors

$$\mathbb{P}((\sqrt{n}/\sigma)|\bar{X}_n - \mu| > z_{0.99}) = 1\%$$

où $z_{0.99} \approx 2.32$. Or,

$$\frac{\sqrt{n}}{\sigma}|\bar{X}_n - \mu| > z_{0.99} \quad (4.2)$$

est équivalent à

$$\bar{X}_n - \frac{z_{0.99}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{0.99}\sigma}{\sqrt{n}}. \quad (4.3)$$

Le passage de Équation 4.2 à Équation 4.3 est souvent appelé *pivot*.

Nous avons donc les deux bornes de notre intervalle de confiance :

$$A = \bar{X}_n - \frac{z_{0.99}\sigma}{\sqrt{n}}$$
$$B = \bar{X}_n + \frac{z_{0.99}\sigma}{\sqrt{n}}.$$

Ces deux quantités sont bien des statistiques, car σ est connu. De plus, nous venons de montrer que $P_\mu(\mu \in [A, B]) = 99\%$. Ici, le choix de la marge d'erreur $\alpha = 1\%$ ne jouait aucun rôle particulier ; ainsi, un intervalle de confiance de niveau $1 - \alpha$ pour l'estimation de μ est donné par

$$\left[\bar{X}_n - \frac{z_{1-\alpha}\sigma}{\sqrt{n}} \ ; \ \bar{X}_n + \frac{z_{1-\alpha}\sigma}{\sqrt{n}} \right].$$

Exercice. Montrer que cet intervalle de confiance est plus grand possible.

Lorsque σ est inconnue.

Lorsque σ n'est pas connu, les bornes A, B ci-dessus ne sont pas des statistiques, car elles dépendent de σ . Heureusement, on peut estimer σ sans biais via l'estimateur

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

Que se passe-t-il si, dans la statistique Équation 4.1, on remplace σ par son estimation $\hat{\sigma}_n^2$? La loi n'est plus gaussienne. Il se trouve que la loi de

$$\frac{\sqrt{n}}{\hat{\sigma}_n^2}(\bar{X}_n - \mu) \quad (4.4)$$

s'appelle *loi de Student* à $n - 1$ paramètres de liberté. Cette loi de référence est bien connue ; sa densité est

$$\rho_k(x) = \frac{1}{Z_k} \left(\frac{1}{1 + \frac{x^2}{k}} \right)^{\frac{k+1}{2}},$$

où Z_k est la constante de normalisation donnée par

$$Z_k = \sqrt{k\pi} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)}.$$

Ses quantiles ont également été calculés avec une grande précision. On notera $t_{k,\alpha}$ le quantile symétrique de niveau α d'une loi de Student de paramètre k . Alors,

$$P_{\mu,\sigma^2}(|T_n| > t_{n-1,\alpha}) \leq \alpha.$$

Par le même raisonnement que tout à l'heure, l'inégalité

$$\left| \frac{\sqrt{n}}{\hat{\sigma}_n^2}(\bar{X}_n - \mu) \right| > t_{n-1,\alpha}$$

est équivalente à

$$\bar{X}_n - \frac{t_{n-1,\alpha}\hat{\sigma}_n^2}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{t_{n-1,\alpha}\hat{\sigma}_n^2}{\sqrt{n}}.$$

et les deux côtés de ces inégalités sont des statistiques; en les notant A, B , on a bien trouvé un intervalle de confiance de niveau α , c'est-à-dire tel que $P_{\mu,\sigma^2}(\mu \in [A, B]) = 1 - \alpha$.

Théorème 4.1 (Intervalle de Student). *Un intervalle de confiance de niveau $1 - \alpha$ pour l'estimation de μ lorsque σ n'est pas connue est donné par*

$$\left[\bar{X}_n - \frac{t_{1-\alpha}\hat{\sigma}_n}{\sqrt{n}} \ ; \ \bar{X}_n + \frac{t_{1-\alpha}\hat{\sigma}_n}{\sqrt{n}} \right].$$

4.2.2 Estimation de σ

Supposons maintenant qu'on désire estimer la variance σ^2 .

Lorsque μ est connue.

En supposant que μ est connue, l'estimateur des moments le plus naturel pour estimer σ^2 est évidemment

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

Comme les $(X_i - \mu)/\sigma$ sont des variables aléatoires gaussiennes centrées réduites, l'estimateur $\tilde{\sigma}_n^2 \times (n/\sigma^2)$ est une somme de n gaussiennes standard indépendantes. La loi de cette statistique est connue : c'est une [loi du chi-deux](#) à n paramètres de liberté. Cette loi n'est pas symétrique, puisqu'elle est supportée sur $[0, \infty[$. On note souvent $k_{n,\alpha}^-$ et $k_{n,\alpha}^+$ les nombres les plus éloignées possible (ils existent) tels que $\mathbb{P}(k_{n,\alpha}^- < \chi^2(n) < k_{n,\alpha}^+) = 1 - \alpha$. Ainsi,

$$P_{\sigma^2}(k_{n,\alpha}^- < \frac{n\tilde{\sigma}_n^2}{\sigma^2} < k_{n,\alpha}^+) = \alpha.$$

En pivotant comme dans les exemples précédents, on obtient que l'intervalle

$$\left[\frac{n\tilde{\sigma}_n^2}{k_{n,\alpha}^+} ; \frac{n\tilde{\sigma}_n^2}{k_{n,\alpha}^-} \right]$$

est un intervalle de confiance de niveau α pour σ^2 .

Lorsque μ est inconnue.

Cette fois, on utilise l'estimateur déjà évoqué plus tôt, à savoir

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

La loi de $(n-1)\hat{\sigma}_n^2/\sigma^2$ est encore une loi du chi-deux, mais à $n-1$ paramètres de liberté. Ainsi, le même raisonnement que ci-dessus donne l'intervalle de confiance de niveau α suivant :

$$\left[\frac{(n-1)\tilde{\sigma}_n^2}{k_{n-1,\alpha}^+} ; \frac{(n-1)\tilde{\sigma}_n^2}{k_{n-1,\alpha}^-} \right].$$

4.3 Exemples asymptotiques

- Estimation du paramètre p dans un modèle de Bernoulli.
- Estimation de moyenne dans un modèle non-gaussien.

5 Outils pour construire des intervalles de confiance

5.1 Quantiles

Si X est une variable aléatoire sur \mathbb{R} , un quantile d'ordre $\alpha \in]0, 1[$, noté q_α , est un nombre tel que $\mathbb{P}(X \leq q_\alpha) = \alpha$. Lorsque X est continue, un tel nombre existe forcément, car la fonction de répartition $F(x) = \mathbb{P}(X \leq x)$ est une surjection continue. Les quantiles symétriques z_α sont, eux, définis par $\mathbb{P}(|X| \leq z_\alpha) = \alpha$.

Si la loi de X est de surcroît symétrique, les quantiles symétriques s'expriment facilement en fonction des quantiles classiques. En effet, $\mathbb{P}(|X| \leq z)$ est égal à $\mathbb{P}(X \leq z) - \mathbb{P}(X \leq -z)$. Or, si la loi de X est symétrique, alors $\mathbb{P}(X \leq -z) = 1 - \mathbb{P}(X \leq z)$, et donc

$$\mathbb{P}(|X| \leq z) = 2\mathbb{P}(X \leq z) - 1.$$

Il suffit alors de choisir pour z le quantile $q_{\frac{1+\alpha}{2}}$ pour obtenir $\mathbb{P}(|X| \leq z) = \alpha$.

Les quantiles s'obtiennent en inversant la fonction de répartition : lorsque celle-ci est une bijection sur $]0, 1[$, alors $q_\alpha = F^{-1}(\alpha)$. En règle générale, il n'y a pas de forme fermée. Par exemple, pour une loi gaussienne standard,

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$$

qui elle-même n'a pas d'écriture plus simple. Fort heureusement, les outils de calcul numérique permettent d'effectuer ces calculs avec une grande précision. La table suivante donne les quantiles symétriques de la gaussienne.

α	90%	95%	98%	99%	99.9%	99.99999%
z_α	1.64	1.96	2.32	2.57	3.2	5.32

Voir aussi la [règle 1-2-3](#).

Théorème 5.1 (Queues de distribution de la gaussienne). *Si x est plus grand que 1,*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \leq \mathbb{P}(X > x) \leq \frac{1}{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

En particulier, si x est grand, $\mathbb{P}(X \geq x) \sim e^{-x^2/2}/x\sqrt{2\pi}$ avec une erreur d'ordre $O(e^{-x^2/2}/x^3)$.

À titre d'exemple, pour $x = 2.32$ cette approximation donne 98.83%, ce qui est remarquablement proche de 98%. Pour $x = 2.57$ on trouve 99.42%.

Démonstration. À écrire.

□

5.2 Calculs de lois

- calcul de la densité de la loi de Student.
- calcul de la densité de la loi du chi-deux.
- approximation binomiale et théorème de Berry-Esséen.

5.2.1 Loi du chi-deux

Soit X une loi gaussienne standard. Calculons la densité de X^2 ; pour toute fonction-test φ , $\mathbb{E}[\varphi(X^2)]$ est donné par

$$\frac{1}{\sqrt{2\pi}} \int e^{-x^2/2} \varphi(x^2) dx.$$

Cette intégrale est symétrique, donc on peut ajouter un facteur 2 et intégrer sur $[0, \infty[$. En posant $u = x^2$, on obtient alors la valeur

$$\frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-u/2} \varphi(u) \frac{1}{2\sqrt{u}} du.$$

On reconnaît la densité d'une [loi Gamma](#) de paramètres $(1/2, 1/2)$. Cette loi s'appelle *loi du chi-deux* et on la note $\chi_2(1)$. Sa transformée de Fourier est donnée par

$$\mathbb{E}[e^{itX^2}] = \frac{1}{\sqrt{1-2it}}.$$

Soient maintenant X_1, \dots, X_n des lois gaussiennes standard indépendantes. Chaque X_i^2 est une $\chi_2(1)$; leur somme S_n a donc pour loi la convolée n fois de χ_2 . Calculons sa transformée de Fourier :

$$\mathbb{E}[e^{itS_n}] = \mathbb{E}[e^{itX^2}]^n = \left(\frac{1}{\sqrt{1-2it}}\right)^n = \left(\frac{1}{1-2it}\right)^{\frac{n}{2}}.$$

On reconnaît la transformée de Fourier d'une loi $\Gamma(n/2, 1/2)$; cette loi s'appelle loi du chi-deux à n paramètres de liberté et elle est notée $\chi_2(n)$. Sa densité est

$$\rho_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-x/2} x^{n/2-1} \mathbf{1}_{x>0}.$$

5.3 Inégalités de concentration

Les outils de base pour construire des intervalles de confiance sont les inégalités de concentration. Une inégalité de concentration pour une variable aléatoire intégrable X est une inégalité de type

$$\mathbb{P}(|X - \mathbb{E}[X]| > x) \leq (\text{quelque chose de petit quand } x \text{ est grand}),$$

c'est-à-dire une inégalité qui contrôle la probabilité pour que les réalisations d'une variable aléatoire X soient éloignées de leur valeur moyenne.

Théorème 5.2. *Soit X une variable aléatoire de carré intégrable. Alors,*

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

Démonstration. Élever au carré les deux membres de l'inégalité, puis appliquer l'inégalité de Markov à la variable aléatoire positive $|X - \mathbb{E}[X]|^2$ dont l'espérance est $\text{Var}(X)$.

□

Théorème 5.3 (Inégalité de Hoeffding). *Soient X_1, \dots, X_n des variables aléatoires indépendantes, pas forcément de même loi. On suppose que chaque X_i est à valeurs dans un intervalle borné $[a_i, b_i]$ et on pose $S_n = X_1 + \dots + X_n$.*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq e^{-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

Démonstration. À écrire.

□

6 Exercices

6.1 Questions

- Soit X_n une variable aléatoire de loi de Student de paramètre n . Montrer que X_n converge en loi vers $N(0, 1)$.
- Construire un intervalle de confiance pour le paramètre d'une loi de Poisson.
- Donner un intervalle de confiance de la forme $[A, +\infty[$ pour la moyenne d'un échantillon gaussien.
- Même question pour la variance dans un modèle gaussien centré.

6.2 Poisson

On suppose que l'on observe X_1, \dots, X_n i.i.d de loi $\mathcal{P}(\theta)$.

1. Étudier \bar{X}_n .
2. Montrer que $\sqrt{\bar{X}_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sqrt{\theta}$.
3. Donner deux intervalles de confiance au niveau 98% pour $\sqrt{\theta}$, et les comparer.

6.3 Uniforme

Soit X_1, \dots, X_n un échantillon iid de loi $\mathcal{U}[0, \theta]$. Donner un intervalle de confiance non asymptotique pour θ .

7 Test d'hypothèse

Cours 3-4

8 Économétrie

Cours 5-6-7

9 Théorie de l'information

Cours 8-9-10

10 Estimation de densité

Cours 11-12

Et après ?

nasuitenasute