# Statistiques Fondamentales

Simon Coste

2024-01-08

# Table des matières

# 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 $\mu$. 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 $\mu$. 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 $\mu$ ?

**Estimation.** Évidemment, la moyenne empirique

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

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 $\mu$. 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 $\mu$ 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 $\mu$. 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 $\mu$ ne diffèrent pas de plus de $0,3$, c'est équivalent à dire que $\mu$ 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 $\mu$ de chaque magasin se situe entre $21,3$ et $21,9$. Cela laisse tout de même une chance de 1% que le paramètre $\mu$ 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 $\mu$ n'est pas le plus important : ce qui nous intéresse maintenant, c'est plutôt d'être sûrs que $\mu$ 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 $\mu$ 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 $\mu$ 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 $\mu$ 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 hypothèse 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 $\Theta$ est appelé espace des paramètres. On écrira fréquemment $\mathbb{E}_\theta$ ou $\mathrm{Var}_\theta$ pour désigner des espérances, variances, etc., calculées avec la loi $P_\theta$.

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 $p_\theta$ 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 $p_\theta$.

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 $\theta_\star$, tel que les observations $x_1, \dots, x_n$ sont des réalisations de loi $p_{\theta_\star}$. L'objectif est alors de trouver $\theta_\star$ 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 $\theta_\star$ (estimation ponctuelle).
- Donner une zone de $\Theta$ dans laquelle le vrai paramètre $\theta_\star$ a des chances de se trouver (intervalle de confiance).
- Répondre à des questions binaires sur $\theta_\star$, par exemple « $\theta_\star$ 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 $\theta$ en argument. Ses valeurs ne doivent dépendre du paramètre $\theta$ qu'au travers de $P_\theta$.

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 $P_\theta$, on peut définir la valeur moyenne de $T$ sous le modèle $P_\theta$ 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 $T : (a_1, \dots, a_n) \to n^{-1}(a_1 + \cdots + a_n)$. 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 $\theta$ à partir des observations, il est naturel de considérer des statistiques à valeurs dans $\Theta$. 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 $\theta$ est une statistique à valeurs dans $\Theta$.

En fait, on n'est pas obligés de vouloir estimer précisément $\theta$. Peut-être qu'on veut estimer quelque chose qui dépend de $\theta$, mais qui n'est pas $\theta$ ; 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 $\theta$ déterminent le hasard qui génère les observations $x_i$. La statistique bayésienne consiste à renverser le point de vue, et à rendre le paramètre $\theta$ 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 $\theta$ ou un autre paramètre qui dépend de $\theta$. 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 $\widehat{\theta}$ est la quantité $\mathbb{E}_\theta[\widehat{\theta} - \theta]$. L'estimateur est dit *sans biais* s'il est de biais nul.
- Le risque quadratique de $\widehat{\theta}$ est la quantité $\mathbb{E}_\theta[|\widehat{\theta} - \theta|^2]$.

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

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)**.**

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

## 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 $\epsilon > 0$

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

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

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

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

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

## 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 gaussiens est relativement fréquent, et beaucoup d'estimateurs sont des sommes de réalisations de variables indépendantes.

**Définition 2.4** (normalité asymtotique)**.** Soit $\theta$ un paramètre à estimer, et $\hat{\theta}_n$ une suite d'estimateurs de $\theta$. 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

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

où $\Sigma$ est une matrice de covariance qui dépend peut-être de $\theta$ — pour éviter les cas dégénérés, on demande à ce que $\Sigma$ 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 tache 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 $\varphi$,*

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

*Démonstration.* Fixons une fonction test $\varphi$ 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 $t \to \varphi(t, c)$ 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, $\varphi$ 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 $\{|Y_n - c| \leqslant \delta\}$. 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

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

- 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 $|\mathbb{E}A| \leqslant \varepsilon$ dès que $n$ est assez grand, ce qui montre bien que $\mathbb{E}A \to 0$.

□

**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 $g : \mathbb{R} \to \mathbb{R}$ dérivable en $\alpha$ (de dérivée non nulle en $\alpha$), on a*

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

Plus généralement, si les $X_n$ sont à valeurs dans $\mathbb{R}^d$ et que $\sqrt{n}(X_n - \alpha) \to N(0, \Sigma)$, alors pour toute application $g : \mathbb{R}^d \to \mathbb{R}^k$, 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 statistiques intégrable $T$ (pas forcément réelle), dont la moyenne n'est pas le paramètre $\theta$ lui-même, mais plutôt une *fonction* de $\theta$ :

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

C'est cette fonction $\varphi$ qu'on appelle *moment*. Typiquement,

- la moyenne d'une loi $\mathcal{E}(\theta)$ n'est pas $\theta$ 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,

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

ce qui permet d'estimer $\varphi(\theta)$. Peut-on alors estimer $\theta$ ? Si la fonction $\varphi$ est inversible et si $\bar{T}_n$ appartient presque sûrement à l'ensemble de définition de $\varphi^{-1}$, alors $\varphi^{-1}(\bar{T}_n)$ est bien définie. Pour qu'en plus cette quantité converge presque sûrement vers $\theta$, il faut s'assurer que $\varphi^{-1}$ est continue. C'est par exemple le cas lorsque l'ensemble des paramètres $\Theta$ est un ouvert, et que $\varphi$ 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 $\varphi$ 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 $\theta$. 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*

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

*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 $N(0, \mathrm{Cov}_\theta(T))$ par le TCL. Une simple application de la delta-méthode (Théorème 2.3) donne alors le résultat, puisque la matrice jacobienne de $\varphi^{-1}$ en $\varphi(\theta)$ n'est autre que l'inverse de la matrice jacobienne de $\varphi$ en $\theta$.

□

# 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 $\lambda$, 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 $\mu$ et variance $\sigma^2$.

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

### 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 $\alpha$. %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 $n \to \infty$.
3. Étudier le comportement en loi de $n(1 - R_n)$ quand $n \to \infty$.

### 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 $\theta$.

1. Déterminer un estimateur de $\theta$ à partir de $\bar{X}_n$.
2. On considère l'estimateur $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Déterminer les propriétés asymotptiques 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 chercche à estimer les paramètres.

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

### 3.0.7 Gumbel

La loi de Gumbel (centrale) de paramètre $\beta$ 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 $\beta$.

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 $n \geqslant 1$ et $B$ est la fonction beta.

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

# 4 Intervalles de confiance

## 4.1 Principe

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

Pour simplifier, on supposera d'abord que $\theta$ 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

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

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 $n$,

$$P_\theta(\theta \in I_n) \geqslant 1 - \alpha.$$

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 Trois exemples

### 4.2.1 $\sigma$ est connu

On se place dans un modèle gaussien où $X_1, \dots, X_n$ sont indépendantes de loi $N(\mu, \sigma^2)$ et on cherche à estimer $\mu$. Nous avons déjà vu que la moyenne empirique $\bar{X}_n$ est un estimateur convergent de $\mu$. Or, nous savons aussi la loi *exacte* de $\bar{X}_n$, qui est $N(\mu, \sigma^2/n)$.

Le moment est idéal pour rappeler l'existence et le calcul des *quantiles* d'une loi — voir ci dessous. Si l'on se donne un niveau de confiance $\alpha = 99$, alors

$$\mathbb{P}(\sqrt{n}|\bar{X}_n - \mu| > z_{0.99}) = 99\%.$$

et $z_{0.99} \approx 2.32$. Or,

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

est équivalent à

$$\bar{X}_n - \frac{z_{0.99}\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X}_n + \frac{z_{0.99}\sigma}{\sqrt{n}}.$$

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 $\sigma$ est connu. De plus, nous venons de montrer que $P_\mu(\mu \in [A, B]) = 99\%$.

### 4.2.2 $\sigma$ est inconnu

### 4.2.3 La loi est inconnue

## 4.3 Outils pour construire des intervalles de confiance

### 4.3.1 Quantiles

Si $X$ est une variable aléatoire sur $\mathbb{R}$, un quantile d'ordre $\alpha \in ]0, 1[$, noté $q_\alpha$, est un nombre tel que $\mathbb{P}(X \leqslant q_\alpha) = \alpha$. Lorsque $X$ est continue, un tel nombre existe forcément (pourquoi ?). Les quantiles symétriques $z_\alpha$ sont, eux, définis par $\mathbb{P}(|X| \leqslant z_\alpha) = \alpha$.

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

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

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

En règle générale, 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.

| $\alpha$ | 90% | 95% | 98% | 99% | 99.9% | 99.99999% |
|---|---|---|---|---|---|---|
| $z_\alpha$ | 1.64 | 1.96 | 2.32 | 2.57 | 3.2 | 5.32 |

**Théorème 4.1** (Queues de distribution de la gaussienne)**.** *Si $x$ est plus grand que 1,*

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

*En particulier, si $x$ est grand, $\mathbb{P}(X \geqslant x) \sim e^{-x^2/2}/x\sqrt{2\pi}$ 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.

□

### 4.3.2 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

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

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 4.2.** *Soit $X$ une variable aléatoire de carré intégrable. Alors,*

$$\mathbb{P}(|X-\mathbb{E}[X]| \geqslant x) \leqslant \frac{\mathrm{Var}(X)}{x^2}.$$

*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 $\mathrm{Var}(X)$.

□

**Théorème 4.3** (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 $S_n = X_1 + \dots + X_n$.*

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

*Démonstration.* À écrire.

□

# 5 Test d'hypothèse

Cours 3-4

# 6 Économétrie

Cours 5-6-7

# 7 Théorie de l'information

Cours 8-9-10

# 8 Estimation de densité

Cours 11-12

# Et après ?

nasuitenasute