Problèmes inverses — partie 2
Estimation par maximum de vraisemblance
Chapitre 2: estimation paramétrique

Nicolas Champagnat

Ecole des Mines de Nancy, 28/04/2020

Introduction (1)

- On a vu au cours précédent la notion de modèle statistique, et en particulier de modèle paramétrique
- L'objet de ce cours est l'estimation paramétrique, c-à-d l'estimation de certains paramètres d'intérêt du modèle à partir d'observations
- Notre intérêt principal portera sur la méthode du maximum de vraisemblance
- Avant de s'y intéresser dans les cours suivants, nous allons placer aujourd'hui le cadre général de l'estimation paramétrique

Introduction (2)

- On se place sur un espace probabilisable $(\Omega, \mathcal{F})$
- On considère le modèle statistique paramétrique

$$\left(\mathcal{H}^n, \{\mathbb{P}_\theta\}_{\theta \in \Theta}\right),$$

  où $\mathcal{H} \subset \mathbb{R}^k$ et $\Theta \subset \mathbb{R}^d$.
- On suppose que le paramètre d'intérêt est $g(\theta)$ avec $g : \Theta \to \mathbb{R}^p$.
- Un exemple typique est le cas où g est la projection des p premières coordonnées de θ

Exemple du modèle gaussien

Par exemple, dans le modèle gaussien

$$\left(\mathbb{R}^n, \{\mathcal{N}(m, \sigma^2)^{\otimes n}\}_{m\in\mathbb{R},\sigma\in[0,+\infty)}\right),$$

le paramètre θ est le couple (m, σ), et l'ensemble Θ est $\mathbb{R} \times \mathbb{R}_+$.

Le paramètre d'intérêt peut être par exemple la moyenne m de la loi inconnue, c'est-à-dire la première composante du paramètre.

Question statistique typique que se pose l'expérimentateur

Détermination de la taille n de l'échantillon nécessaire à réaliser l'objectif de l'expérimentateur,
où l'objectif est formulé en termes des paramètres du modèle.

Dans l'exemple précédent, l'expérimentateur peut souhaiter par exemple

- estimer le paramètre m à une erreur $0,01$ près
- tester si le paramètre m appartient à un intervalle $[a,b]$ donné, qui pourrait avoir été obtenu indépendamment par une autre expérience, par exemple afin de valider le modèle

Échantillons (1)

La notion mathématique d'échantillon est différente de celle vue au dernier cours.

$\rightsquigarrow$ formalisation nécessaire pour exploiter la théorie des probabilités (LGN, TCL) afin de quantifier la qualité d'une estimation.

Définition

- *Un échantillon $(X_1, \ldots, X_n)$ est une v.a. à valeurs dans $\mathcal{H}^n$ dont la loi est $\mathbb{P}_\theta$ pour un certain $\theta \in \Theta$ (à estimer).*
- *L'échantillon de loi $\mathbb{P}_\theta$ est une v.a. $(X_1, \ldots, X_n)$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ de loi $\mathbb{P}_\theta$.*

Échantillons (2)

- Cas particulier important : si $\mathbb{P}_\theta = \mathbb{Q}_\theta^{\otimes n}$ pour tout θ, cela signifie que les échantillons sont i.i.d.
  C'est une hypothèse fréquente lorsque le processus de sélection est un tirage aléatoire avec remise dans la population, ou sans remise mais avec un population beaucoup plus grande que la taille n de l'échantillon.
- Attention ! Ne pas confondre échantillon et observations : un échantillon est une variable aléatoire $(X_1, \ldots, X_n)$ et les observations $(x_1, \ldots, x_n)$ sont une réalisation de l'échantillon :

$$\exists \omega \in \Omega, \quad (x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)).$$

Échantillons (3)

Définition

- *Une statistique S est une fonction mesurable de l'échantillon $(X_1, \dots, X_n)$: $S = S(X_1, \dots, X_n)$.*
- *Un estimateur est une statistique à valeurs dans $g(\Theta)$ (l'ensemble des valeurs possibles du paramètre d'intérêt).*

- Un estimateurs est classiquement notés avec des « chapeaux » : $\hat{g}$ ou $\hat{\theta}$. On écrira $\hat{g}_n$ et $\hat{\theta}_n$ si on veut rendre compte de la dépendance en la taille de l'échantillon.
- Une statistique est un cas particulier de v.a.
- Pour le moment, pas de notion de qualité d'estimation.
- Exemple du jeu de pile-ou-face du cours précédent : les v.a.

$$X_1 \quad \text{et} \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

sont toutes les deux des estimateurs $\rightsquigarrow$ lequel est le meilleur ?

Comment caractériser la qualité d'un estimateur ?

- L'estimateur $\hat{g} = \hat{g}(X_1, \dots, X_n)$ est bon si la quantité $\hat{g}(x_1, \dots, x_n)$ est proche de la vraie valeur $g(\theta_0)$ lorsque $x_1, \dots, x_n$ est la réalisation d'un échantillon de loi $\mathbb{P}_{\theta_0}$
- θ_0 inconnu $\rightsquigarrow$ cette proximité doit être vraie pour tout $\theta \in \Theta$
- rien n'empêche la réalisation $(x_1, \dots, x_n)$ d'être identique pour différentes valeurs de $\theta \rightsquigarrow$ l'approximation ne peut être bonne qu'avec une proba < 1
- $\rightsquigarrow$ deux seuils pour mesurer la qualité d'un estimateur : seuil d'erreur $\varepsilon > 0$ aussi petit que possible et niveau de confiance $1 - \eta$ aussi proche de 1 que possible, tels que

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta(|\hat{g} - g(\theta)| \leq \varepsilon) \geq 1 - \eta$$

Estimation par insertion

Principe :

- considérons le modèle statistique i.i.d. $(\mathcal{H}^n, \{\mathbb{Q}_\theta^{\otimes n}\}_{\theta\in\Theta})$, où $\mathbb{Q}_\theta$ est une probabilité sur $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$
- supposons que l'on peut écrire le paramètre d'intérêt $g(\theta)$ comme

$$g(\theta) = \varphi(\mathbb{Q}_\theta), \quad \forall\theta \in \Theta,$$

  pour une certaine fonction φ définie sur les mesures de probabilités
- estimation par insertion : remplacer la mesure $\mathbb{Q}_\theta$ par son approximation Q_n (mesure empirique) dans l'équation précédente : on obtient l'estimateur

$$\hat{g}_n = \varphi(Q_n).$$

- Si la fonction φ est suffisamment régulière, le théorème de Varadarajan implique que $\hat{g}_n$ converge p.s. vers la valeur réelle du paramètre d'intérêt $g(\theta)$.

Méthode des moments

La méthode des moments est un cas particulier de la méthode d'estimation par insertion, lorsque le paramètre d'intérêt $g(\theta)$ s'écrit comme une fonction d'un (vecteur de) moment(s) de la loi $\mathbb{Q}_\theta$.

Exemples de fonctions des moments : Supposons que $\mathcal{H} \subset \mathbb{R}$.

- Si $g(\theta) = \int_{\mathbb{R}} x\mathbb{Q}_\theta(dx)$ (premier moment), alors l'estimateur par la méthode des moments est la moyenne empirique, notée $\bar{X}_n$, de l'échantillon :

$$\hat{g}_n = \int_{\mathbb{R}} x\, Q_n(dx) = \frac{1}{n}\int_{\mathbb{R}} x \sum_{i=1}^{n} \delta_{X_i}(dx) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n.$$

- Si $g(\theta)$ est la variance de $\mathbb{Q}_\theta$, c'est-à-dire $g(\theta) = \int_{\mathbb{R}} x^2\, \mathbb{Q}_\theta(dx) - \left(\int_{\mathbb{R}} x\, \mathbb{Q}_\theta(dx)\right)^2$, $g(\theta)$ est une fonction régulière des moments d'ordre 1 et 2. La méthode des moments donne pour estimateur la variance empirique, notée $\hat{\sigma}_n^2$, de l'échantillon :

$$\hat{g}_n = \int_{\mathbb{R}} x^2 Q_n(dx) - \left(\int_{\mathbb{R}} x Q_n(dx)\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \hat{\sigma}_n^2.$$

Exemples (1)

Par ex. l'estimateur $\bar{X}_n$ obtenu dans le jeu de pile-ou-face peut être obtenu par la méthode des moments.

Autre exemple : Considérons le modèle statistique

$$\left(\mathbb{R}_+^n, \{\mathcal{U}([0,\theta])^{\otimes n}\}_{\theta>0}\right).$$

- Ici, le paramètre à estimer θ est la borne supérieure des valeurs possibles des observations, qui n'est pas un moment de la loi $\mathbb{Q}_\theta = \mathcal{U}([0,\theta])$.
- En revanche, son premier moment s'exprime comme une fonction simple du paramètre θ :

$$\int_{\mathbb{R}} x\,\mathcal{U}([0,\theta])(dx) = \int_0^\theta x\,\frac{dx}{\theta} = \frac{\theta}{2}.$$

- La méthode des moments donne donc l'estimateur $\hat{\theta}_n = 2\bar{X}_n$

Exemples (2)

- La méthode d'inférence par insertion s'applique également, puisque

$$\theta = \inf\{x \in \mathbb{R}_+ \text{ tels que } \mathbb{Q}_\theta(]x, +\infty[) = 0\}.$$

- On obtient le second estimateur

$$\begin{aligned}\hat{\theta}'_n &= \inf\{x \in \mathbb{R}_+ \text{ tels que } Q_n(]x, +\infty[) = 0\} \\ &= \inf\left\{x \in \mathbb{R}_+ \text{ tels que } x \geq X_i,\ \forall i \in \{1, \ldots, n\}\right\},\end{aligned}$$

c'est-à-dire $\hat{\theta}'_n = \max_{1 \leq i \leq n} X_i$.

Comment savoir lequel des deux estimateurs est le meilleur ?

Biais d'un estimateur

On note $\mathbb{E}_\theta$ (resp. $\mathbb{V}_\theta$) l'espérance (resp. la variance) par rapport à la loi $\mathbb{P}_\theta$.

$$\mathbb{V}_\theta(Z) = \mathbb{E}_\theta\left[\left[(Z - \mathbb{E}_\theta Z)(Z - \mathbb{E}_\theta Z)^T\right]\right].$$

Définition

Une statistique S est d'ordre $p \in [1, +\infty)$ si $S \in \mathbb{L}^p(\mathbb{P}_\theta)$ pour tout $\theta \in \Theta$.

Définition

Un estimateur $\hat{g}_n$ de $g(\theta)$ est dit :

- *sans biais lorsque $\mathbb{E}_\theta \hat{g}_n = g(\theta)$ pour tout $\theta \in \Theta$*
- *asymptotiquement sans biais lorsque*

$$\lim_{n\to+\infty} \mathbb{E}_\theta \hat{g}_n = g(\theta), \quad \forall\theta \in \Theta.$$

$\mathbb{E}_\theta \hat{g} - g(\theta)$ est le biais de l'estimateur $\hat{g}$.

Exemple 1

Si $\mathbb{P}_\theta = \mathbb{Q}_\theta^{\otimes n}$ et $g(\theta)$ est la moyenne de la loi $\mathbb{Q}_\theta$, la méthode de moments donne l'estimateur de moyenne empirique

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

Il est sans biais, puisque

$$\mathbb{E}_\theta \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\theta X_i = \mathbb{E}_\theta X_1 = \int_{\mathcal{H}} x\,\mathbb{Q}_\theta(dx).$$

Exemple 2

Si $g(\theta)$ est la variance de la loi $\mathbb{Q}_\theta$: variance empirique

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

Il est biaisé mais asymptotiquement sans biais :

$$\begin{aligned}
\mathbb{E}_\theta \hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i^2) - \frac{1}{n^2}\sum_{1\leq i,j\leq n}\mathbb{E}_\theta(X_iX_j) \\
&= \mathbb{E}_\theta(X_1^2) - \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i^2) - \frac{1}{n^2}\sum_{i\neq j}\mathbb{E}_\theta X_i \mathbb{E}_\theta X_j \\
&= \left(1 - \frac{1}{n}\right)\mathbb{E}_\theta(X_1^2) - \frac{n(n-1)}{n^2}(\mathbb{E}_\theta X_1)^2 = \frac{n-1}{n}\mathbb{V}_\theta(X_1).
\end{aligned}$$

On préfère à la variance empirique $\hat{\sigma}_n^2$ sa version sans biais

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

Risque quadratique

Afin de mesurer la qualité d'un estimateur, on définit la notion de risque.

Définition

Soit $\hat{g}$ un estimateur de $g(\theta)$ d'ordre de 2. Pour tout $\theta \in \Theta$, on définit le risque quadratique de $\hat{g}$ sous $\mathbb{P}_\theta$ comme

$$\mathcal{R}(\theta, \hat{g}) = \mathbb{E}_\theta \left(|\hat{g} - g(\theta)|^2 \right).$$

Proposition (décomposition biais-variance)

Soit $\hat{g}$ un estimateur de $g(\theta)$ d'ordre 2. Alors

$$\mathcal{R}(\theta, \hat{g}) = |\mathbb{E}_\theta \hat{g} - g(\theta)|^2 + \mathbb{E}_\theta |\hat{g} - \mathbb{E}_\theta \hat{g}|^2 = |\mathbb{E}_\theta \hat{g} - g(\theta)|^2 + \mathbb{V}_\theta(\hat{g}).$$

Preuve de la décomposition biais-variance

Rappelons que $|\cdot|$ est la norme euclidienne.
En utilisant que $\hat{g} - g(\theta) = (\hat{g} - \mathbb{E}_\theta \hat{g}) + (\mathbb{E}_\theta \hat{g} - g(\theta))$ et que pour tout vecteur $u, v \in \mathbb{R}^p$, $|u+v|^2 = |u|^2 + 2u^T v + |v|^2$, on obtient

$$\begin{aligned}\mathcal{R}(\theta, \hat{g}) &= \mathbb{E}_\theta |\hat{g} - \mathbb{E}_\theta \hat{g}|^2 + 2\mathbb{E}_\theta (\hat{g} - \mathbb{E}_\theta \hat{g})^T (\mathbb{E}_\theta \hat{g} - g(\theta)) + \mathbb{E}_\theta |\mathbb{E}_\theta \hat{g} - g(\theta)|^2 \\ &= \mathbb{E}_\theta |\hat{g} - \mathbb{E}_\theta \hat{g}|^2 + |\mathbb{E}_\theta \hat{g} - g(\theta)|^2\end{aligned}$$

puisque $\mathbb{E}_\theta (\hat{g} - \mathbb{E}_\theta \hat{g}) = 0$.

Consistance

Nous introduisons maintenant les concepts utiles pour pouvoir relier la taille d'un échantillon à une précision donnée.

Définition

Un estimateur $\hat{g}_n$ est dit consistant lorsque

$$\hat{g}_n \xrightarrow[n\to+\infty]{\mathbb{P}_\theta} g(\theta), \quad \forall\theta\in\Theta.$$

Attention ! la loi $\mathbb{P}_\theta$ dépend de n.

Consistance forte

Dans le cas d'un échantillon i.i.d., $\mathbb{P}_\theta = \mathbb{Q}_\theta^{\otimes n}$ et la consistance signifie que

$$\forall \varepsilon > 0,\ \forall \theta \in \Theta, \qquad \lim_{n\to+\infty} \mathbb{Q}_\theta^{\otimes n}\left(|\hat{g}_n(X_1,\ldots,X_n) - g(\theta)| > \varepsilon\right) = 0.$$

Si on remplace $\mathbb{P}_\theta$ par $\mathbb{P}'_\theta$ la loi d'une suite i.i.d. $(X_i)_{i\geq 1}$ de loi commune $\mathbb{Q}_\theta$, alors la consistance s'écrit

$$\lim_{n\to+\infty} \mathbb{P}'_\theta\left(|\hat{g}_n(X_1,\ldots,X_n) - g(\theta)| > \varepsilon\right) = 0.$$

avec $\mathbb{P}'_\theta$ indépendant de n.
Dans ce cas, on peut définir la notion de consistance forte de l'estimateur comme

$$\lim_{n\to+\infty} \hat{g}_n = g(\theta), \quad \mathbb{P}'_\theta\text{-presque sûrement}, \quad \forall \theta \in \Theta.$$

Rq. La consistance forte implique la consistance.
Dans la suite, on fera toujours l'abus de notation d'écrire $\mathbb{P}_\theta$ à la place de $\mathbb{P}'_\theta$.

Exemple

Dans le jeu de pile-ou-face, l'estimateur du paramètre θ par la méthode des moments est la moyenne empirique

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

C'est un estimateur consistant, et même fortement consistant.

En effet, sous $\mathbb{P}_\theta$, pour $\theta \in [0, 1]$, $(X_i)_{i \geq 1}$ est une suite i.i.d. de loi $\mathrm{Bern}(\theta)$. Ainsi, la loi des grands nombres assure que, sous $\mathbb{P}_\theta$,

$$\bar{X}_n \xrightarrow[n \to +\infty]{} \mathbb{E}_\theta(X_1) = \theta \quad \text{presque sûrement.}$$

Vitesse et loi limite d'un estimateur

Définition

Soit $(v_n)_{n\geq 1}$ une suite de réels positifs telle que $v_n \to +\infty$ quand $n \to +\infty$. On dit que $\hat{g}_n$ est un estimateur de $g(\theta)$ de vitesse v_n si, pour tout $\theta \in \Theta$, il existe une loi $\ell(\theta)$ sur $\mathbb{R}^p$ différente de δ_0, appelée loi limite de $\hat{g}_n$, telle que

$$v_n(\hat{g}_n - g(\theta)) \xrightarrow[n\to+\infty]{\text{loi sous } \mathbb{P}_\theta} \ell(\theta), \quad \forall \theta \in \Theta.$$

Si toutes les lois $\ell(\theta)$ sont gaussiennes, on dit que $\hat{g}_n$ est un estimateur asymptotiquement normal (ou asymptotiquement gaussien).

Exemple (suite)

Dans le jeu de pile-ou-face, la vitesse de l'estimateur $\bar{X}_n$ peut se déduire du théorème central limite.
Soit $\theta \in [0, 1]$ fixé. Sous $\mathbb{P}_\theta$, les v.a. X_i sont i.i.d. de loi $\text{Bern}(\theta)$, et donc de moyenne θ et variance $\theta(1-\theta)$. On obtient donc

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow[n \to +\infty]{\text{loi sous } \mathbb{P}_\theta} \mathcal{N}(0, \theta(1-\theta)).$$

Ainsi, l'estimateur $\bar{X}_n$ est asymptotiquement normal, de vitesse $\sqrt{n}$.

Lien avec la consistance

Un estimateur $\hat{g}$ de vitesse $\nu_n \to +\infty$ est toujours consistant.

En effet,

$$\hat{g}_n - g(\theta) = \frac{1}{\nu_n} \times \nu_n(\hat{g}_n - g(\theta)).$$

- Le premier facteur converge en probabilité vers 0.
- Le second facteur converge en loi vers $\ell(\theta)$.
- Lemme de Slutsky : le couple

$$(1/\nu_n \, , \, \nu_n(\hat{g}_n - g(\theta)))$$

  converge en loi vers $(0, \ell(\theta))$.
- Donc le produit des deux éléments du couple converge en loi vers le produit des limites :
  $\hat{g}_n - g(\theta)$ converge en loi vers 0, donc en probabilité

V.a. pivotale

Définition

Une v.a. G définie sur $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est dite pivotale si sa loi sous $\mathbb{P}_\theta$ est indépendante du paramètre θ.

- Supposons que l'estimateur $\hat{g}_n$ est de vitesse $\nu_n \to +\infty$ et que sa loi limite $\ell(\theta)$ est t.q.

$$\sigma(\theta)G \sim \ell(\theta) \quad \text{où } \sigma(\theta) > 0 \text{ et } G \text{ est pivotale.}$$

- Par exemple si $\ell(\theta) = \mathcal{N}(0, \sigma(\theta)^2)$, et dans ce cas $G \sim \mathcal{N}(0,1)$.
- Supposons que $\hat{\sigma}_n$ est un estimateur consistant de $\sigma(\theta)$. D'après le lemme de Slutsky,

$$\left(\nu_n(\hat{g}_n - g(\theta))\,,\,\hat{\sigma}_n\right) \xrightarrow[n\to+\infty]{\text{loi sous } \mathbb{P}_\theta} (\sigma(\theta)G\,,\,\sigma(\theta)).$$

- Puisque $\sigma(\theta) > 0$, on en déduit la convergence du quotient :

$$\frac{\nu_n}{\hat{\sigma}_n}(\hat{g}_n - g(\theta)) \xrightarrow[n\to+\infty]{\text{loi sous } \mathbb{P}_\theta} G.$$

Exemple du jeu de pile-ou-face

- $\bar{X}_n$ est un estimateur de θ asymptotiquement normal, de variance asymptotique $\theta(1-\theta)$.
- On peut donc appliquer le raisonnement précédent avec $\sigma(\theta) = \sqrt{\theta(1-\theta)}$ et $G \sim \mathcal{N}(0,1)$:
- Puisque $\bar{X}_n$ est un estimateur consistant de θ, on en déduit que

$$\hat{\sigma}_n = \sqrt{\bar{X}_n(1-\bar{X}_n)}$$

  est un estimateur consistant de $\sigma(\theta)$.
- On en déduit que

$$\sqrt{\frac{n}{\bar{X}_n(1-\bar{X}_n)}}\,(\bar{X}_n - \theta) \xrightarrow[n\to+\infty]{\text{loi sous } \mathbb{P}_\theta} \mathcal{N}(0,1). \tag{1}$$

Intervalles de confiance

Définition

Soit $\alpha \in]0,1[$ un niveau de confiance fixé. Un intervalle de confiance pour $g(\theta) \in \mathbb{R}$ de niveau de confiance $1-\alpha$ est une statistique I à valeur dans les intervalles de $\mathbb{R}$, c'est-à-dire de la forme

$$I = [a(X_1, \ldots, X_n)\ ,\ b(X_1, \ldots, X_n)]$$

pour des statistiques $a < b$, telle que

$$\mathbb{P}_\theta(g(\theta) \in I) = 1 - \alpha, \quad \forall \theta \in \Theta.$$

- Un intervalle de confiance est un compromis entre l'erreur d'estimation (la longueur de l'intervalle) et la confiance dans le résultat (le paramètre $1-\alpha$).
- Si α est petit, l'intervalle de confiance sera grand ; si l'intervalle de confiance est petit, la confiance $1-\alpha$ sera faible.

Exemple

- Modèle statistique gaussien :

$$\left(\mathbb{R}^n, \{\mathcal{N}(\theta,1)^{\otimes n}\}_{\theta\in\mathbb{R}}\right).$$

- Soit $\alpha \in]0,1[$ et q_α le quantile gaussien d'ordre $1-\frac{\alpha}{2}$: si $G\sim\mathcal{N}(0,1)$,

$$\mathbb{P}(G\le q_\alpha) = \int_{-\infty}^{q_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = 1-\frac{\alpha}{2}.$$

- $\sqrt{n}(\bar{X}_n-\theta)$ est une v.a. gaussienne de moyenne nulle et de variance $n\frac{1}{n^2}\sum_{i=1}^n \mathbb{V}_\theta(X_i) = 1$, donc

$$\mathbb{P}_\theta(\sqrt{n}|\bar{X}_n-\theta| > q_\alpha) = \mathbb{P}(G>q_\alpha) + \mathbb{P}(-G>q_\alpha) = \alpha.$$

- Donc

$$\mathbb{P}_\theta\left(\theta\in\left[\bar{X}_n - \frac{q_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{q_\alpha}{\sqrt{n}}\right]\right) = \mathbb{P}_\theta(\sqrt{n}|\bar{X}_n-\theta|\le q_\alpha) = 1-\alpha.$$

- Ainsi, $\left[\bar{X}_n - \frac{q_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{q_\alpha}{\sqrt{n}}\right]$ est un intervalle de confiance pour θ de niveau de confiance $1-\alpha$.

Intervalles de confiance asymptotiques

Définition

Soit $\alpha \in]0,1[$ *un niveau de confiance fixé. Une statistique* I *à valeur dans les intervalles de* $\mathbb{R}$ *est un* *intervalle de confiance par excès* *pour* $g(\theta) \in \mathbb{R}$ *de niveau de confiance* $1-\alpha$ *si*

$$\mathbb{P}_\theta(g(\theta) \in I) \geq 1-\alpha, \quad \forall \theta \in \Theta.$$

Définition

Soit $\alpha \in]0,1[$ *un niveau de confiance fixé. Une statistique* I_n *à valeur dans les intervalles de* $\mathbb{R}$ *est un* *intervalle de confiance asymptotique* *pour* $g(\theta) \in \mathbb{R}$ *de niveau de confiance* $1-\alpha$ *si*

$$\lim_{n \to +\infty} \mathbb{P}_\theta(g(\theta) \in I_n) = 1-\alpha, \quad \forall \theta \in \Theta.$$

Exemple du jeu de pile-ou-face

Dans le jeu de pile-ou-face, la convergence dans l'équation (1) permet d'obtenir un intervalle de confiance asymptotique :

Fixons un niveau de confiance $\alpha \in]0, 1[$. On a

$$\begin{aligned}\lim_{n\to+\infty} \mathbb{P}_\theta\left(\theta \in \left[\bar{X}_n - q_\alpha\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n - q_\alpha\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]\right) \\ = \lim_{n\to+\infty} \mathbb{P}_\theta\left(\sqrt{\frac{n}{\bar{X}_n(1-\bar{X}_n)}}(\bar{X}_n - \theta) \in [-q_\alpha, q_\alpha]\right) \\ = \mathbb{P}(G \in [-q_\alpha, q_\alpha]) = 1 - \alpha,\end{aligned}$$

où $G \sim \mathcal{N}(0, 1)$.

δ-méthode

On a vu comment déduire du TCL la loi limite et un intervalle de confiance quand le paramètre à estimer est θ.
Le résultat suivant permet de traiter le cas d'un estimateur de $g(\theta)$.

Proposition (δ-méthode)

Soit Y_n une suite de v.a. à valeurs dans $\mathbb{R}^d$ et $y \in \mathbb{R}^d$. Supposons que $\nu_n(Y_n - y)$ converge en loi vers une v.a. Z, pour une certaine suite $\nu_n \to +\infty$. Soit $g : \mathbb{R}^d \to \mathbb{R}^p$ une fonction $\mathcal{C}^1$, et soit $J_g(y)$ sa matrice jacobienne au point $y \in \mathbb{R}^d$. Alors

$$\nu_n \left(g(Y_n) - g(y)\right) \xrightarrow[n\to+\infty]{\text{loi sous } \mathbb{P}_\theta} J_g(y)Z.$$

Ce résultat permet de déduire, à partir de la vitesse et la loi limite d'un estimateur $\hat{\theta}_n$ de θ, la vitesse et la loi limite de l'estimateur $g(\hat{\theta}_n)$ de $g(\theta)$.

Régression linéaire multiple (cf. exercice 8)

- On cherche à expliquer des observations $(x_1, \ldots, x_n)$, par exemple la consommation électrique de différents foyers, à l'aide d'un certain nombre de facteurs quantitatifs connus, par exemple l'âge moyen des personnes du foyer, la catégorie socio-professionnelle ou le cours du pétrole.
- Ces facteurs sont appelés régresseurs, représentés par un vecteur $R_i \in \mathbb{R}^k$ pour chaque observation x_i.
  On note R la matrice formée des vecteurs lignes $R_1, \ldots, R_n$.
- Modèle de régression linéaire multiple : modèle i.i.d. où la loi de chaque élément X d'un échantillon $(X_1, \ldots, X_n)$ est donnée par

$$X = R\theta + \varepsilon, \quad \text{où } \varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathrm{Id}),$$

où les paramètres inconnus sont $\theta \in \mathbb{R}^k$ le vecteur des poids décrivant l'influence de chaque régresseur sur les observations, et $\sigma > 0$ l'écart-type du bruit.

Régression linéaire multiple (suite)

On considère donc le modèle statistique gaussien

$$\left(\mathbb{R}^n, \{\mathcal{N}_n(R\theta, \sigma^2 \mathrm{Id})\}_{\theta\in\mathbb{R}^k,\ \sigma>0}\right).$$

On peut toujours supposer (quitte à réduire le nombre de régresseurs si certains sont redondants) que la matrice R est de rang k (en particulier, $k \leq n$). Dans ce cas, on a le résultat suivant, où la notion d'estimateur du maximum de vraisemblance est définie dans le chapitre suivant.

Régression linéaire multiple (suite et fin)

Théorème

Sous les hypothèses précédentes, l'estimateur du maximum de vraisemblance de (θ, σ) est donné par $(\hat{\theta}_n, \hat{s}_n)$ où

$$\hat{\theta}_n = (R^T R)^{-1} R^T X \quad \text{et} \quad \hat{s}_n^2 = \frac{1}{n} |X - R\hat{\theta}_n|^2,$$

où $|\cdot|$ désigne la norme euclidienne. De plus, sous la loi $\mathbb{P}_{(\theta,\sigma)}$, ces estimateurs ont pour loi

$$\hat{\theta}_n \sim \mathcal{N}_k(\theta, \sigma^2 (R^T R)^{-1}) \quad \text{et} \quad \frac{n}{\sigma^2}\hat{s}_n^2 \sim \chi^2(n-k),$$

où $\chi^2(n-k)$ est la loi du khi-2 à $n-k$ degrés de liberté. En particulier,

$$\hat{\theta}_n \text{ est un estimateur sans biais de } \theta$$
$$\text{et } \frac{n}{n-k}\hat{s}_n^2 \text{ est un estimateur sans biais de } \sigma^2.$$