Chapitre Un

Performances

Sommaire

Dans ce chapitre nous allons étudier comment comparer des estimateurs. Un estimateur T_n de θ sera un bon estimateur s'il s'approche suffisamment du paramètre dans un sens que nous allons préciser.

1.1 Qualité d'un estimateur

1.1.1 Risque quadratique, biais et variance

Définition : *risque quadratique.*
Le risque quadratique ou erreur quadratique moyenne d'un estimateur T_n est donné par l'expression suivante :

$$EQM(T_n) = \mathbb{E}[(T_n - \theta)^2]$$

Le risque quadratique permet de comparer deux estimateurs, c'est une mesure assez brute de la qualité d'un estimateur qui peut être raffiné avec les notions de biais et de variance.

Définition : *bias d'un estimateur*
On appelle biais de T_n pour θ la valeur $b_\theta(T_n) := \mathbb{E}(T_n) - \theta$
Par conséquent on dira d'un estimateur T_n de θ qu'il est sans biais si et seulement si $\mathbb{E}(T_n) = \theta$, dans le cas contraire, on dira que T_n est un estimateur biaisé. En d'autres termes un estimateur est sans biais si son espérance est égale au paramètre recherché.

Exercice : Montrer que $EQM(T_n) = Var(T_n) + b_\theta(T_n)^2$

Le risque quadratique d'un estimateur est égal à sa variance plus le carré de son biais. Lorsqu'on évalue plusieurs estimateurs le meilleur sera sans biais et de variance minimale. Cependant, le biais et la variance ne sont pas les seuls propriété caractérisant la qualité d'un estimateur. En effet, augmenter la taille de l'échantillon est aussi à prendre en considération quand on compare des estimateurs. D'ailleurs le succès du big data revient à utiliser d'énormes échantillons, ce qui conduit à bénéficier des performances asymptotiques des estimateurs.

Exercice : On considère une variable aléatoire X telle que $\mathbb{E}[X] = \mu$ et $Var(X) = \sigma^2$.

- Montrer que les estimateurs de la moyenne et de la variance d'une loi Gaussienne $\mathcal{N}(\mu, \sigma^2)$ que vous avez trouvé à l'aide du maximum de vraisemblance sont sans biais.
- Montrer que l'estimateur de la variance empirique $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ est biaisé. Pour cela commencez par développer et simplifier cette somme.
- Déduisez en un estimateur de la variance non biaisé et retrouvez l'estimateur de la variance empirique corrigé présenté plus haut $\hat{S}_n^2$.

1.1.2 Convergence des estimateurs

Définition : *convergence d'un estimateur.*
Un estimateur T_n de θ est dit convergeant s'il converge en probabilité vers θ lorsque n tend vers l'infini. On notera $T_n \xrightarrow{\mathcal{P}} \theta$

Définition : *convergence en moyenne quadratique.*
L'estimateur T_n converge en moyenne quadratique vers θ si et seulement si son erreur quadratique moyenne tend vers 0 quand n tend vers l'infini :

$$T_n \xrightarrow{L^2} \theta \Leftrightarrow \lim_{n\to\infty} \mathbb{E}[(T_n - \theta)^2] = 0$$

Remarque : Lorsqu'un estimateur est convergent, son biais converge vers 0. On dit aussi qu'il est asymptotiquement sans biais.

Théorème *loi faible des grands nombres*
Soit $X_1, \cdots, X_n$ un ensemble de variables aléatoires $i.i.d$ d'espérance finie $\mathbb{E}[X_i] = \mu$, alors l'estimateur de la moyenne empirique converge en probabilité vers la moyenne μ

$$\bar{X}_n \xrightarrow{\mathcal{P}} \mu$$

Théorème *loi forte des grands nombres*
Sous les mêmes hypothèses que le théorème précédent

$$\bar{X}_n \xrightarrow{p.s} \mu$$

Remarque :

- Les deux lois des grands nombres justifient l'interprétation fréquentiste des probabilités. C'est l'intuition selon laquelle, l'observation d'un grand nombre de données issues d'un phénomène aléatoire permet de déterminer les paramètres de la loi sous jacente. L'exemple le plus direct est le lancé d'une pièce équilibré pour obtenir pile ou face. À mesure que vous augmentez le nombre de lancés, vous obtiendrez une probabilité convergent vers 0.5. Ces théorèmes justifient théoriquement les big data pour l'estimation paramétrique.

— Vous noterez cependant que ces théorèmes concernent l'estimateur de la moyenne empirique. Cela laisse suggérer que si vous construisez votre propre estimateur à l'avenir, il faudra peut-être prouver que celui ci converge bien. Enfin, vous observerez que la seule différence entre ces deux théorèmes et la nature de la convergence sous jacente.

— Bien que la loi faible des grands nombres nous indique que la distribution de $\bar{X}_n$ se concentre autour de $\mathbb{E}[X]$, elle ne nous dit rien sur la loi que suit cet estimateur. Le théorème de la limite centrale (TCL) dit que $\bar{X}_n$ a une distribution qui est approximativement normale avec une moyenne $\mathbb{E}[X]$ et une variance $Var(X)$. Étonnamment, rien n'est supposé au sujet de la loi de X, sauf l'existence de la moyenne et de la variance.

Théorème *théorème central limite*
Soit (X_i) une suite de v.a i.i.d et $\bar{X}_n$ l'estimateur de la moyenne empirique, alors on a le résultat suivant :

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Var(X))$$

Le point clé du théorème central limite est de donner la loi de l'estimateur
Exercice :

— Montrez que les estimateurs de la moyenne empirique et de la variance empirique sont convergent en moyenne quadratique.

— Montrez l'équivalence entre les deux formulations suivantes du TCL :

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Var(X)) \Leftrightarrow \sqrt{n}(\frac{\bar{X}_n - \mathbb{E}(X)}{\sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

— Donnez un diagramme d'implication comparant les 4 types de convergence suivants : convergence en loi, convergence en probabilité, convergence presque sûre, et convergence quadratique. Le but de cet exercice est de vous donner un aperçu de haut niveau des propriétés de convergence.

