Tests Statistiques

Master 1: 2020-2021

Prof. Abdelhakim Necir

Département de Mathématiques (Univ. Biskra)

Lundi 14 Décembre 2020

Estimation poncutelle

- **Estimation nonparamétrique**: aucune paramétrisation sur la loi de probabilité.
- **Estimation paramétrique**: la loi de probabilité est paramétrisée.

Estimation nonparamétrique

- Estimation de loi de probabilité F ou sa densité $f = F'$
- Estimation des paramètres statistiques: $\mu := \mathbf{E}[X]$, $\sigma^2 := \mathbf{V}[X]$, Q_α (quantile d'ordre α), le mode, l'étendue,...

- Soit X une va ayant une loi de probabilité $F(x) = P(X \leq x)$. Supposons qu'on possède un échantillon aléatoire $X_1, ..., X_n$ de X.
- Donner un estimateur de F noté $\widehat{F}_n$
- Réponse: la fonction de répartition empirique

$$\widehat{F}_n(x) = F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x).$$

Nous avons déjà évoquer que

$$\mathbf{E}[F_n(x)] = F(x).$$

En d'autre termes F_n est **estimateur sans biais** de F.

Pour tout x fixé,

$$F_n(x) \xrightarrow{P} F(x) \text{ quand } n \longrightarrow \infty.$$

En d'autre termes F_n est estimateur convergeant de F ou **estimateur consistent**.

Pour tout x fixé,

$$\frac{\sqrt{n}\left(F_n(x) - F(x)\right)}{\sqrt{F(x)\left(1 - F(x)\right)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \text{ quand } n \longrightarrow \infty.$$

En d'autre termes F_n est **estimateur asymptotiquement normal** (gaussien).

Estimateur de la moyenne μ est:

$$\widehat{\mu} = \overline{X}.$$

Nous avons:

$$\mathbf{E}\left[\widehat{\mu}\right] = \mathbf{E}\left[\overline{X}\right] = \mathbf{E}\left[X\right] = \mu.$$

Donc $\widehat{\mu}$ est estimateur sans biais de μ. En outre, d'après la loi des grands nombres

$$\widehat{\mu} \xrightarrow{p} \mu,$$

c'est à dire $\widehat{\mu}$ est **un estimateur consistent de** μ.

De plus d'après le théorème central limite

$$\sqrt{n}\frac{\widehat{\mu}-\mu}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,1\right),$$

C'est à dire que $\widehat{\mu}$ est **asymptotiquement normal**. Pour n assez grand,

$$\sqrt{n}\frac{\widehat{\mu}-\mu}{\sigma} \rightsquigarrow \mathcal{N}\left(0,1\right).$$

Estimation nonparamétrique

Rappel:

$$X \rightsquigarrow \mathcal{N}\left(a, b^2\right), b > 0 \Longleftrightarrow \frac{X-a}{b} \rightsquigarrow \mathcal{N}(0,1).$$

Donc pour n assez grand:

$$\sqrt{n}\frac{\widehat{\mu}-\mu}{\sigma} \rightsquigarrow \mathcal{N}(0,1) \Longleftrightarrow \widehat{\mu} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Pus généralement, pour une fonction g donnée, un estimateur de $\mu_g = g(X_i)$ est:

$$\widehat{\mu}_g = \frac{1}{n} \sum_{i=1}^{n} g(X_i).$$

En particulier un estimateur de le moment d'ordre 2 $\mu^{(2)} = \mathbf{E}\left[X^2\right]$ est

$$\widehat{\mu}^{(2)} = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Estimation de la variance:

Nous avons

$$\sigma^2 = \mathbf{Var}\left[X\right] = \mathbf{E}\left[X - \mu\right]^2 = \mu^{(2)} - \mu^2.$$

Donc une estimateur σ^2 est

$$\begin{aligned}\widehat{\sigma}^2 &= \widehat{\mu}^{(2)} - \widehat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 =: \widetilde{S}_n^2.\end{aligned}$$

Estimation de la variance:
On démontre que

$$\mathbf{E}\left[\widehat{\sigma}^2\right] = \mathbf{E}\left[\widetilde{S}_n^2\right] = \sigma^2 - \frac{1}{n}\sigma^2.$$

En d'autre termes $\widetilde{S}_n^2$ est un **estimateur biaisé** de σ^2.

Correction de biais: on remarque que

$$\mathbf{E}\left[\widetilde{S}_n^2\right] = \left(1 - \frac{1}{n}\right)\sigma^2 = \frac{n-1}{n}\sigma^2,$$

ce qui implique

$$\mathbf{E}\left[\frac{n}{n-1}\widetilde{S}_n\right] = \sigma^2.$$

On pose

$$S_n^2 =: \frac{n}{n-1}\widetilde{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2,$$

avec $\mathbf{E}\left[S_n^2\right] = \sigma^2$. En d'autre termes S_n^2 est estimateur sans biais de σ^2.

On pose

$$S_n^2 = \frac{n-1}{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\right],$$

Nous avons $\frac{n-1}{n} \to 1$, et d'apres la loi des grands nombres

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mathbf{E}\left[X^2\right] = \mu_2 \text{ et } \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbf{E}\left[X\right] = \mu$$

et par conséquent

$$S_n^2 \xrightarrow{P} \mu_2 - \mu^2 = \sigma^2.$$

Donc S_n^2 **un estimateur consistent** de σ^2.

Supposons que $\mu_4 := \mathbf{E}[X-\mu]^4 < \infty$, alors

$$\sqrt{n}\left(S_n^2 - \sigma^2\right) \xrightarrow{D} \mathcal{N}\left(0, \mu_4 - \sigma^4\right).$$

En d'autre termes S_n^2 est un **asymptotiquement normal** de σ^2.

Théorème central limite multidimentionnel

Soit $Y_1, ..., Y_n$ une suite de vecteurs aléatoires de dimension $d \geq 1$ iid. Supposons que $\mathbf{E}\left[Y_i^2\right] < \infty$, alors

$$\frac{1}{\sqrt{n}}\left(Y_1 + ... + Y_n - n\mathbf{E}\left[Y_i\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sum\right),$$

où $\sum$ est la matrice de variance-covariance de Y_1.

Théorème de Slutsky

Soit $Z, Z_1, Z_2..., W_1, W_2, ...$ des va définies dans un même espace de probabilité. Supposons que $Z_n \xrightarrow{\mathcal{D}} Z$ et $W_n \xrightarrow{p} c$ où c est une constante finie. Alors

$$Z_n + W_n \xrightarrow{\mathcal{D}} Z + c.$$

Application: Nous avons

$$\begin{aligned}\widetilde{S}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \left(\overline{X} - \mu\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)\right)^2.\end{aligned}$$

Soit $Y_i = \left(X_i - \mu, \left(X_i - \mu\right)^2\right)$, $i = 1, 2, ..., n$. Par définition $\mathbf{E}\left[\left(V, W\right)\right] := \left(\mathbf{E}\left[V\right], \mathbf{E}\left[W\right]\right)$. Nous avons

$$\mathbf{E}\left[Y_i\right] = \left(0, \sigma^2\right) \text{ et } \mathbf{E}\left[Y_i^2\right] = \left(\sigma^2, \mathbf{E}\left[\left(X_i - \mu\right)^4\right]\right).$$

D'après le TCLM, si $\mu_4 := \mathbf{E}\left[(X_i - \mu)^4\right] < \infty$ on a

$$\frac{1}{\sqrt{n}}\left(\sum_{\mathbf{i=1}}^{\mathbf{n}}\left(X_i - \mu, (X_i - \mu)^2\right) - n\left(0, \sigma^2\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sum\right),$$

où $\sum$ est la matrice de variance-covriance de Y_i :

$$\sum = \begin{pmatrix} \mathbf{Var}\left[X_i - \mu\right] & \mathbf{Cov}\left[X_i - \mu, (X_i - \mu)^2\right] \\ \mathbf{Cov}\left[(X_i - \mu)^2, X_i - \mu\right] & \mathbf{Var}\left[X_i - \mu\right]^2 \end{pmatrix}.$$

Nous avons

$$\mathbf{Var}\left[X_i - \mu\right] = \sigma^2,$$

$$\mathbf{Cov}\left[X_i - \mu, (X_i - \mu)^2\right] = \mathbf{E}\left[X_i - \mu\right]^3,$$

et

$$\mathbf{Var}\left[X_i - \mu\right]^2 = \mu_4 - \sigma^4.$$

Donc en particulier

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}(X_i - \mu)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\right) \tag{1}$$

et

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}(X_i - \mu)^2 - n\sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mu_4 - \sigma^4\right). \tag{2}$$

Les limites (1) et (2) peuvent être réecrites comme suit:

$$\sqrt{n}\left(\overline{X}-\mu\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\sigma^2\right)$$

et

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2-\sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\mu_4-\sigma^4\right). \tag{3}$$

On en déduit que $\sqrt{n}\left(\overline{X}_n-\mu\right)$ est bornée en probabilité et que $\overline{X}_n-\mu \xrightarrow{P} 0$. Donc

$$\sqrt{n}\left(\overline{X}_n-\mu\right)^2 \xrightarrow{P} 0.$$

En ultilisant le théorème **Slutsky,** avec

$$Z_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-\sigma^2\right) \text{ et } W_n = \sqrt{n}\left(\overline{X}_n-\mu\right)^2 \xrightarrow{P} 0,$$

on obtient

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-\sigma^2\right)-\sqrt{n}\left(\overline{X}_n-\mu\right)^2 \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{E}\left[X_i-\mu\right]^4-\sigma^4\right).$$

En d'autres termes

$$\sqrt{n}\left(\widetilde{S}_n^2-\sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{E}\left[X_i-\mu\right]^4-\sigma^4\right).$$

Remarquons que

$$\widetilde{S}_n^2 = \frac{n-1}{n} S_n^2.$$

Alors

$$\sqrt{n}\left(\frac{n-1}{n} S_n^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{E}\left[X_i - \mu\right]^4 - \sigma^4\right).$$

Ce qui implique que

$$\frac{n-1}{n}\sqrt{n}\left(S_n^2 - \frac{n}{n-1}\sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{E}\left[X_i - \mu\right]^4 - \sigma^4\right).$$

En d'autres termes

$$\sqrt{n}\left(S_n^2 - \sigma^2 - \frac{1}{n-1}\sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{E}\left[X_i - \mu\right]^4 - \sigma^4\right).$$

Théorème central limite multidimentionnel (TCLM)

En appliquant encore une fois le théorème **Slutsky**, avec

$$Z_n = \sqrt{n}\left(S_n^2 - \sigma^2\right) \text{ et } W_n = \frac{\sqrt{n}}{n-1}\sigma^2 \to 0,$$

on obient

$$\sqrt{n}\left(S_n^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{E}\left[X_i - \mu\right]^4 - \sigma^4\right).$$