MAP 311 - Aléatoire

Leçon 7

2016-2017

Rappel : Théorème de la limite centrale

Théorème

Soit $(X_n)_{n\in\mathbb{N}^*}$ *une suite de variables aléatoires réelles indépendantes et de même loi, d'espérance* m *et de variance* $\sigma^2 > 0$.

- *La suite* $\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - m)_{n\in\mathbb{N}^*}$ *converge en loi vers une variable aléatoire de loi* $\mathcal{N}(0,\sigma^2)$.
- *Pour tous* $a \leq b$,

$$\lim_{n\to+\infty} \mathbb{P}\Big(a \leq \frac{\sqrt{n}}{\sigma}\big(\frac{1}{n}\sum_{i=1}^{n} X_i - m\big) \leq b\Big) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\Big(-\frac{x^2}{2}\Big)dx$$

Méthode Delta

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d., d'espérance m et de variance σ^2. On note

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

On connaît le comportement de $\overline{X}_n$ quand $n \to +\infty$.
Peut-on en déduire le comportement de $g(\overline{X}_n)$?

Proposition

- *Pour toute fonction continue g, $g(\overline{X}_n) \overset{n\to+\infty}{\longrightarrow} g(m)$ p.s.*
- *Pour toute fonction g de classe $\mathcal{C}^1$,*

$$\sqrt{n}(g(\overline{X}_n) - g(m)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (g'(m))^2\sigma^2)$$

- *Si $g'(m) = 0$ et g est de classe $\mathcal{C}^2$, alors*

$$n(g(\overline{X}_n) - g(m)) \xrightarrow{\mathcal{L}} \frac{1}{2}\sigma^2 g''(m)\chi^2_1$$

Exemple : Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. de loi de Bernoulli de paramètre $p \in]0,1[$.
Quel est le comportement asymptotique de $T_n = \overline{X}_n(1-\overline{X}_n)$?

On applique la méthode Delta avec $g(x) = x(1-x)$. On a $g'(x) = 1-2x$.
Si $p \neq 1/2$, $g'(p) \neq 0$ et on a alors

$$\sqrt{n}\big(T_n - p(1-p)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, p(1-p)(1-2p)^2\big)$$

Si $p = 1/2$, $g'(1/2) = 0$ et $g''(1/2) = -2$. On applique la dernière partie de la proposition :

$$n\big(T_n - \frac{1}{4}\big) \xrightarrow{\mathcal{L}} -\frac{1}{4}\chi^2_1$$

Preuve de la méthode Delta.

• Si g de classe $\mathcal{C}^1$.

Formule des accroissements finis : Pour chaque n,

$$g(\overline{X}_n) = g(m) + g'(\theta_n)(\overline{X}_n - m)$$

où θ_n se trouve entre m et $\overline{X}_n$.

Par la loi des grands nombres, $(\overline{X}_n)_n$ converge p.s. vers m et donc il en est de même pour $(\theta_n)_n$. Comme g est de classe $\mathcal{C}^1$, la suite $(g'(\theta_n))_n$ converge p.s. vers $g'(m)$. Or

$$\sqrt{n}\big(g(\overline{X}_n) - g(m)\big) = \sqrt{n}(\overline{X}_n - m)g'(\theta_n)$$

Par le TCL, $\sqrt{n}(\overline{X}_n - m)$ converge en loi vers $\mathcal{N}(0, \sigma^2)$. Comme $(g'(\theta_n))_n$ converge p.s. vers $g'(m)$, donc en probabilité, on peut appliquer le théorème de Slutsky.

• Si $g'(m) = 0$ et g de classe $\mathcal{C}^2$.
Par la formule de Taylor-Lagrange :

$$g(\overline{X}_n) = g(m) + g'(m)(\overline{X}_n - m) + \frac{1}{2}g''(\theta_n)(\overline{X}_n - m)^2,$$

où θ_n se trouve entre m et $\overline{X}_n$.
Par le TCL, $\sqrt{n}(\overline{X}_n - m)$ converge en loi vers $\mathcal{N}(0, \sigma^2)$ et donc

$$n\big(\overline{X}_n - m)^2 \xrightarrow{\mathcal{L}} \sigma^2\chi^2_1$$

Comme précedemment, $(\theta_n)_n$ converge p.s. vers m et par continuité $(g''(\theta_n))_n$ converge p.s. vers $g''(m)$.
Par le théorème de Slutsky, on conclut que

$$n\big(g(\overline{X}_n) - g(m)\big) = \frac{1}{2}g''(\theta_n)n(\overline{X}_n - m)^2 \xrightarrow{\mathcal{L}} \frac{1}{2}g''(m)\sigma^2\chi^2_1$$

TCL multi-dimensionnel

Soit $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$, de même loi, avec $\boldsymbol{m} = \mathbb{E}(\boldsymbol{X}_1)$ et $\mathbf{C}_{\boldsymbol{X}}$ la matrice de covariance de $\boldsymbol{X}_1$.

Théorème

Les vecteurs aléatoires

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{m}\right)$$

convergent en loi vers un vecteur aléatoire gaussien centré de matrice de covariance $\mathbf{C}_{\boldsymbol{X}}$.

La vitesse de convergence de $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$ vers $\boldsymbol{m}$ est toujours en $1/\sqrt{n}$, indépendante de la dimension d.

Exemple : Soit $(X_n)_n$ des variables aléatoires réelles i.i.d. centrées réduites. Soit $S_n = \sum_{i=1}^n X_i$. On sait que $S_n/\sqrt{n}$ converge en loi. Montrons que $S_n/\sqrt{n}$ ne peut pas converger en probabilité.
Si c'était le cas, la suite $(S_{2n}/\sqrt{2n} - S_n/\sqrt{n})_n$ devrait tendre en probabilité, et donc en loi, vers 0.
Or

$$\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}} = \frac{1}{\sqrt{2n}}\sum_{i=1}^{2n} X_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \frac{1-\sqrt{2}}{\sqrt{2n}}\sum_{i=1}^{n} X_i + \frac{1}{\sqrt{2n}}\sum_{i=n+1}^{2n} X_i$$

On a (en introduisant une autre suite i.i.d. X_i' de même loi que X_i)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} X_i \\ X_i' \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

Donc

$$\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}} \overset{\text{en loi}}{=} \frac{1-\sqrt{2}}{\sqrt{2}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i + \frac{1}{\sqrt{2}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

avec $\sigma^2 = \left(\frac{1-\sqrt{2}}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2$.

Statistique

Science des données.
Un statisticien travaille sur des données (résultats d'un sondage, données météorologiques,...) pour :

1. effectuer une prévision (exemple : résultat d'une élection qui aura lieu prochainement à partir d'un sondage),
2. quantifier la certitude liée à une prévision (exemple : fourchette dans le cas d'un sondage),
3. répondre à une question comme "le candidat X sera-t-il élu ?", "le réchauffement climatique est-il réel ?", pour aider à la prise de décision.

Réponses :

1. estimation statistique,
2. théorie des intervalles (ou des régions) de confiance,
3. tests de décision.

Estimation statistique

On observe des réalisations d'un phénomène aléatoire (sexe d'un nouveau-né, température journalière, ...) qu'on appelle observations.
Ces observations sont des copies i.i.d. d'une loi **inconnue** qu'on cherche à retrouver.

Exemple 1 : On observe les sexes de n nouveaux-nés dans une maternité.
Le sexe, fille (F) ou garçon (G), d'un nouveau-né est modélisé par une loi P_θ sur $\{F, G\}$ de paramètre inconnu $\theta \in [0, 1]$, avec $P_\theta(\{G\}) = \theta$ et $P_\theta(\{F\}) = 1 - \theta$.
On souhaite connaître la proportion de garçons à la naissance. Ceci revient à chercher à estimer θ.

Morale :
En probabilités : on se donne une loi P et on cherche les propriétés d'une suite i.i.d. de loi P.
En statistique : on observe une suite i.i.d. et on cherche la loi.

Modèle statistique

Définition

Un modèle statistique est un triplet $(\mathcal{X}, \mathcal{A}, \boldsymbol{\mathcal{P}})$ *où* $\mathcal{X}$ *est l'espace fondamental (l'ensemble des valeurs possibles des observations),* $\mathcal{A}$ *est une tribu sur* $\mathcal{X}$*, et* $\boldsymbol{\mathcal{P}}$ *est une* **famille** *de probabilités sur* $(\mathcal{X}, \mathcal{A})$*.*
Un modèle statistique est dit paramétrique s'il existe un entier p et un ensemble $\Theta \subset \mathbb{R}^p$ *tels que* $\boldsymbol{\mathcal{P}}$ *s'écrit sous la forme :*

$$\boldsymbol{\mathcal{P}} = \{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\},$$

où, pour tout $\boldsymbol{\theta} \in \Theta$*,* $P_{\boldsymbol{\theta}}$ *est une probabilité sur* $(\mathcal{X}, \mathcal{A})$*.*

Exemple 1 : Le modèle statistique est $(\{F, G\}, \mathcal{P}(\{F, G\}), \boldsymbol{\mathcal{P}})$ où :

$$\boldsymbol{\mathcal{P}} = \{P_\theta, \theta \in [0, 1]\},$$

et $P_\theta(\{F\}) = 1 - \theta$ et $P_\theta(\{G\}) = \theta$. Ici $\Theta = [0, 1]$.

Observations

Définition

Un n-échantillon est un vecteur aléatoire $\boldsymbol{X} = (X_i)_{i=1,\dots,n}$ de n v.a. i.i.d. de même loi qui appartient au modèle statistique.

On observe un n-échantillon $\boldsymbol{X}$ de loi $P_{\boldsymbol{\theta}}$ ($\boldsymbol{\theta}$ inconnu).
On cherche une quantité $\boldsymbol{f}(\boldsymbol{X})$ qui s'approche du paramètre inconnu $\boldsymbol{\theta}$.
Une telle quantité est appelée estimateur de $\boldsymbol{\theta}$ et est souvent notée $\hat{\boldsymbol{\theta}}_n$.
Un estimateur = une fonction déterministe $\boldsymbol{f} : \mathcal{X}^n \to \Theta$.
Exemple 1 : On a $X_i = G$ si la i^{eme} naissance donne un garçon, et $X_i = F$ si c'est une fille. $(X_1, \dots, X_n)$ est un n-échantillon de la loi P_θ.
Un estimateur de θ est la proportion empirique de garçons :

$$\hat{\theta}_n = \frac{\mathrm{Card}(i = 1, \dots, n,\, X_i = G)}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_G(X_i)\,.$$

On a $\hat{\theta}_n = f(X_1, \dots, X_n)$, avec $f(x_1, \dots, x_n) = (1/n) \sum_{i=1}^{n} \mathbf{1}_G(x_i)$.

Exemple 2 : L'autonomie d'une batterie de téléphone portable est modélisée par une loi exponentielle de paramètre $\lambda \in]0, +\infty[$. Pour estimer λ, on laisse marcher n téléphones jusqu'à ce qu'ils s'éteignent, on observe donc n durées, à partir desquelles on essaye d'estimer λ.

• Le modèle statistique est $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \boldsymbol{\mathcal{P}})$ où

$$\boldsymbol{\mathcal{P}} = \{P_\lambda,\, \lambda > 0\},$$

et P_λ est la probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ de densité $p_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{]0,+\infty[}(x)$.

• On cherche un estimateur de λ.

• Estimateur naturel : on sait que

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

est proche de l'espérance $\mathbb{E}_\lambda(X_1) = 1/\lambda$ d'après la Loi Forte des Grands Nombres. On peut donc proposer comme estimateur de λ :

$$\hat{\lambda}_n = f(\boldsymbol{X}) \text{ avec } f(\boldsymbol{x}) = \Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big)^{-1}$$

• Cependant, on pourrait s'y prendre autrement. En notant

$$\check{\lambda}_n = \check{f}(\boldsymbol{X}) \text{ avec } \check{f}(\boldsymbol{x}) = \Big(\frac{1}{2n}\sum_{i=1}^{n} x_i^2\Big)^{-1/2},$$

on sait aussi que $\check{\lambda}_n$ est proche de $\big(\frac{1}{2}\mathbb{E}_\lambda(X_1^2)\big)^{-1/2} = \lambda$ d'après la LFGN.

• Ceci montre que $\hat{\lambda}_n$ et $\check{\lambda}_n$ sont tous les deux susceptibles de nous donner une estimation raisonnable de λ (lorsque n est grand).

Qualités d'un estimateur : convergence

On examine les qualités d'une suite d'estimateurs $(\hat{\boldsymbol{\theta}}_n)_{n\in\mathbb{N}^*}$ définie par une suite de fonctions $\boldsymbol{f}_n : \mathcal{X}^n \to \Theta$.

Exemple 1 : On examine la suite $\hat{\theta}_n = f_n(X_1, \ldots, X_n)$ définie par :

$$f_n : \left| \begin{array}{l} \{F, G\}^n \to \mathbb{R}\,, \\ (x_1, \ldots, x_n) \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_G(x_i)\,. \end{array} \right.$$

Sous $\mathbb{P}_{\boldsymbol{\theta}}$, les X_i sont i.i.d. de loi $P_{\boldsymbol{\theta}}$.

Définition (Convergence)

On dit que $\hat{\boldsymbol{\theta}}_n$ est convergent si, pour tout $\boldsymbol{\theta} \in \Theta$:

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_n \xrightarrow{n\to+\infty} \boldsymbol{\theta}\right) = 1\,.$$

Exemple 1 : Sous $\mathbb{P}_\theta$, les v.a. $\mathbf{1}_G(X_i)$ sont i.i.d. de loi de Bernoulli de paramètre θ. Par la loi forte des grands nombres :

$$\mathbb{P}_\theta\left(\lim_{n\to+\infty} \hat{\theta}_n = \theta\right) = \mathbb{P}_\theta\left(\lim_{n\to+\infty} \frac{1}{n}\sum_{j=1}^n \mathbf{1}_G(X_j) = \theta\right) = 1\,.$$

Qualités d'un estimateur : biais

Définition (Biais)

Un estimateur est non-biaisé si $\mathbb{E}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\theta}$ *pour tout* $\boldsymbol{\theta} \in \Theta$ *et pour tout* $n \in \mathbb{N}^*$.

Exemple 1 :

$$\mathbb{E}_\theta(\hat{\theta}_n) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}_\theta(\mathbf{1}_G(X_j)) = \mathbb{E}_\theta(\mathbf{1}_G(X_1)) = \theta\,.$$

L'estimateur $\hat{\theta}_n$ est donc non-biaisé.

Qualités d'un estimateur : risque quadratique moyen

Définition (Risque quadratique moyen)

Si $g : \Theta \to \mathbb{R}$, *alors le risque quadratique moyen d'un estimateur* $\hat{g}_n$ *de* $g(\boldsymbol{\theta})$ *est la moyenne quadratique des écarts à la valeur à estimer :*

$$\mathrm{RQM}_{\boldsymbol{\theta}}(\hat{g}_n) = \mathbb{E}_{\boldsymbol{\theta}}\left[(\hat{g}_n - g(\boldsymbol{\theta}))^2\right].$$

Le risque quadratique moyen se décompose en une partie "variance" et une partie "biais" :

$$\mathrm{RQM}_{\boldsymbol{\theta}}(\hat{g}_n) = \mathrm{Var}_{\boldsymbol{\theta}}(\hat{g}_n) + (\mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - g(\boldsymbol{\theta}))^2 .$$

Un "bon" estimateur a un RQM "petit".

Exemple 1 :

$$\mathrm{RQM}_\theta(\hat{\theta}_n) = \mathrm{Var}_\theta\left(\hat{\theta}_n\right) = \frac{1}{n}\mathrm{Var}_\theta\left(\mathbf{1}_G(X_1)\right) = \frac{\theta(1-\theta)}{n}.$$

Qualités d'un estimateur : normalité asymptotique

Définition (Normalité asymptotique)

Soit $g : \Theta \to \mathbb{R}$. Un estimateur $\hat{g}_n$ de $g(\boldsymbol{\theta})$ est dit asymptotiquement normal s'il existe deux fonctions $m_n(\boldsymbol{\theta})$ et $\sigma_n(\boldsymbol{\theta})$ telles que, sous $\mathbb{P}_{\boldsymbol{\theta}}$, la suite de variables aléatoires $(\hat{g}_n - m_n(\boldsymbol{\theta}))/\sigma_n(\boldsymbol{\theta})$ converge en loi quand $n \to +\infty$ vers une loi gaussienne centrée réduite.

Très souvent, mais pas toujours, $m_n(\theta) = g(\boldsymbol{\theta})$ et $\sigma_n(\boldsymbol{\theta}) = \sigma(\boldsymbol{\theta})/\sqrt{n}$ pour une fonction $\sigma^2(\boldsymbol{\theta})$ qu'on appelle variance asymptotique.
Exemple 1 : Par le théorème de la limite centrale, sous $\mathbb{P}_\theta$:

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_G(X_j) - \theta\right) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \theta(1-\theta)\right),$$

en loi. La variance asymptotique est ici $\sigma^2(\theta) = \theta(1-\theta)$.

Estimateurs empiriques

On considère un modèle statistique de la forme $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \boldsymbol{\mathcal{P}})$.
Il arrive souvent que le modèle statistique soit paramétré par la moyenne et/ou la variance de la loi inconnue.
Exemple : modèle gaussien :

$$\boldsymbol{\mathcal{P}} = \left\{\mathcal{N}(\mu, \sigma^2),\ \mu \in \mathbb{R}, \sigma^2 \in]0, +\infty[\right\}.$$

Proposition (Estimateur de la moyenne)

Soit $(X_i)_{i=1,\ldots,n}$ un n-échantillon d'un modèle statistique $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \boldsymbol{\mathcal{P}})$ dont les lois sont intégrables.
La moyenne (l'espérance) d'une v.a. de loi $P_{\boldsymbol{\theta}}$ est notée $\mu_{\boldsymbol{\theta}}$.
La moyenne empirique $\overline{X}_n$ définie par :

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j\,,$$

est un estimateur de la moyenne.

Proposition (Estimateur de la moyenne - suite)

La moyenne empirique $\overline{X}_n$ définie par :

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j\,,$$

est un estimateur de la moyenne tel que :
(i) $\overline{X}_n$ est non-biaisé.
(ii) $\overline{X}_n$ est convergent.
(iii) Si de plus les lois sont de carré intégrable, alors $\mathrm{RQM}_{\boldsymbol{\theta}}(\overline{X}_n) = \sigma^2_{\boldsymbol{\theta}}/n$, avec $\sigma^2_{\boldsymbol{\theta}}$ la variance de la loi $P_{\boldsymbol{\theta}}$ et $\overline{X}_n$ est asymptotiquement normal, avec

$$\sqrt{n}\big(\overline{X}_n - \mu_{\boldsymbol{\theta}}\big) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \sigma^2_{\boldsymbol{\theta}}) \text{ en loi}$$

Proposition (Estimateur de la variance)

Soit $(X_i)_{i=1,\dots,n}$ *un n-échantillon d'un modèle statistique* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$ *dont les lois sont de carré intégrable.*
La variance d'une v.a. de loi $P_{\boldsymbol{\theta}}$ *est notée* $\sigma^2_{\boldsymbol{\theta}}$.
La variance empirique de l'échantillon est définie par :

$$\overline{V}_n = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2 \,.$$

Cet estimateur de la variance vérifie les propriétés suivantes.
(i) L'estimateur est biaisé :

$$\mathbb{E}_{\boldsymbol{\theta}}(\overline{V}_n) = \frac{n-1}{n}\sigma^2_{\boldsymbol{\theta}}$$

Preuve : $\overline{V}_n = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \Big(\frac{1}{n}\sum_{j=1}^{n} X_j\Big)^2 = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \frac{1}{n^2}\sum_{i,j=1}^{n} X_i X_j.$

$$\mathbb{E}(\overline{V}_n) = \mathbb{E}(X^2) - \frac{1}{n}\mathbb{E}(X^2) - \frac{n-1}{n}\mathbb{E}(X)^2 = \frac{n-1}{n}\left(\mathbb{E}(X^2) - \mathbb{E}(X)^2\right)$$

Proposition (Estimateur de la variance - suite)

(ii) L'estimateur est convergent.
(iii) Si les lois sont de quatrième moment fini, alors :

$$\mathrm{RQM}_{\boldsymbol{\theta}}(\overline{V}_n) = \frac{\mu_{\boldsymbol{\theta}}^{(4)} - \sigma_{\boldsymbol{\theta}}^4}{n} + O\big(\frac{1}{n^2}\big)\,,$$

où $\mu_{\boldsymbol{\theta}}^{(4)} = \mathbb{E}_{\boldsymbol{\theta}}\big((X_1 - \mathbb{E}_{\boldsymbol{\theta}}(X_1))^4\big)$ *et* $\overline{V}_n$ *est asymptotiquement normal :*

$$\sqrt{n}(\overline{V}_n - \sigma_{\boldsymbol{\theta}}^2) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}(0, \mu_{\boldsymbol{\theta}}^{(4)} - \sigma_{\boldsymbol{\theta}}^4) \text{ en loi}$$

Proposition (Variance empirique non-biaisée)

La variance empirique non-biaisée V_n est définie par :

$$V_n = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2 \,.$$

Cet estimateur de la variance vérifie les propriétés suivantes.
(i) V_n est non-biaisé : pour tout $n \geq 2$, $\mathbb{E}_{\boldsymbol{\theta}}(V_n) = \sigma^2_{\boldsymbol{\theta}}$.
(ii) V_n est convergent.
(iii) Si les lois sont de quatrième moment fini, alors :

$$\mathrm{RQM}_{\boldsymbol{\theta}}(V_n) = \frac{\mu^{(4)}_{\boldsymbol{\theta}} - \sigma^4_{\boldsymbol{\theta}}}{n} + O(\frac{1}{n^2}) \,,$$

et V_n est asymptotiquement normal :

$$\sqrt{n}(V_n - \sigma^2_{\boldsymbol{\theta}}) \overset{n \to +\infty}{\longrightarrow} \mathcal{N}(0, \mu^{(4)}_{\boldsymbol{\theta}} - \sigma^4_{\boldsymbol{\theta}}) \text{ en loi}$$

Méthode des moments

Soit $(X_i)_{i=1,\dots,n}$ un n-échantillon d'un modèle $\mathcal{P} = \{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$. On cherche un estimateur de $\boldsymbol{\theta}$.
Supposons : On a trouvé une fonction $\boldsymbol{g} : \Theta \to \Theta' \subset \mathbb{R}^d$ inversible et de fonction réciproque continue et une fonction $\boldsymbol{\psi} : \mathcal{X} \to \mathbb{R}^d$ telle que $\mathbb{E}_{\boldsymbol{\theta}}(|\boldsymbol{\psi}(X_1)|) < \infty$ pour tout $\boldsymbol{\theta} \in \Theta$ et

$$\boldsymbol{g}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\big(\boldsymbol{\psi}(X_1)\big).$$

Résultat : L'estimateur des moments

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{g}^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\psi}(X_i)\Big).$$

est convergent.
En effet, $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\psi}(X_i)$ est un estimateur convergent de $\boldsymbol{g}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{\psi}(X_1))$ d'après la Loi Forte des Grands Nombres, et $\boldsymbol{g}^{-1}$ est continue.

Exemple : On considère le modèle uniforme $\mathcal{P} = \{P_\theta, \theta \in]0, +\infty[\}$, où P_θ est la loi uniforme sur $[0, \theta]$. Si X_1 a pour loi P_θ, alors on a :

$$\mathbb{E}_\theta(X_1) = \frac{\theta}{2}\,.$$

Par conséquent, si $(X_i)_{i=1,\dots,n}$ est un n-échantillon de loi P_θ, alors

$$\hat{\theta}_n = \frac{2}{n}\sum_{i=1}^{n} X_i$$

est un estimateur de θ.

Cet estimateur est non-biaisé, convergent, son RQM est

$$\mathrm{RQM}_\theta(\hat{\theta}_\mathrm{n}) = \mathrm{Var}_\theta(\hat{\theta}_n) = \frac{4}{n}\mathrm{Var}_\theta(X_1) = \frac{\theta^2}{3n}\,,$$

et il est asymptotiquement normal :

$$\sqrt{n}\big(\hat{\theta}_n - \theta\big) \overset{n\to+\infty}{\longrightarrow} \mathcal{N}(0, \theta^2/3) \text{ en loi.}$$

Estimation par maximum de vraisemblance

On considère un modèle statistique $\boldsymbol{\mathcal{P}} = \{P_{\boldsymbol{\theta}}, \, \boldsymbol{\theta} \in \Theta\}$ tel que :
- ou bien χ est fini ou dénombrable et pour tout $\boldsymbol{\theta} \in \Theta$, la loi $P_{\boldsymbol{\theta}}$ est discrète et à valeurs dans $\mathcal{X}$. On pose alors $p(x, \boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(\{x\})$ pour tout $x \in \mathcal{X}$. On pose pour tout $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$:

$$p_n(\boldsymbol{x}, \boldsymbol{\theta}) = p(x_1, \boldsymbol{\theta}) \cdots p(x_n, \boldsymbol{\theta}),$$

qu'on appelle vraisemblance.
- ou bien $\chi = \mathbb{R}$ ou $\mathbb{R}^d$ et pour tout $\boldsymbol{\theta} \in \Theta$, la loi $P_{\boldsymbol{\theta}}$ est à densité. On note alors $p(x, \boldsymbol{\theta})$ la densité de la loi $P_{\boldsymbol{\theta}}$ et $p_n(\boldsymbol{x}, \boldsymbol{\theta})$ la densité jointe d'un n-échantillon, qu'on appelle vraisemblance.

Définition

On suppose que pour toute réalisation $\boldsymbol{x} = (x_1, \ldots, x_n)$ *d'un n-échantillon* $\boldsymbol{X} = (X_1, \ldots, X_n)$, *il existe une unique valeur* $\boldsymbol{\theta}_n(\boldsymbol{x}) \in \Theta$ *qui maximise la vraisemblance (vue comme fonction de* $\boldsymbol{\theta}$*) :*

$$\boldsymbol{\theta}_n(\boldsymbol{x}) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \; p_n(\boldsymbol{x}, \boldsymbol{\theta}).$$

Alors l'estimateur $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n(\boldsymbol{X})$ *est appelé Estimateur du Maximum de Vraisemblance (EMV) de* $\boldsymbol{\theta}$.

- Un tel estimateur n'existe pas toujours et il peut être multiple.

Cependant dans la plupart des cas, cet estimateur existe, est unique, facile à calculer, et possède de bonnes propriétés.

- Parfois il est plus facile de chercher $\boldsymbol{\theta}_n(\boldsymbol{x}) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \; \ln(p_n(\boldsymbol{x}, \boldsymbol{\theta}))$
- En pratique, on cherche souvent l'EMV, puis on étudie ses propriétés.

Interprétation bayésienne : Supposons que $\mathcal{X}$ et Θ sont finis.
Avant de recueillir des données, on ne sait rien sur le paramètre $\boldsymbol{\theta}$, à part qu'il est dans Θ, donc on peut considérer que c'est une réalisation d'une variable aléatoire discrète $\boldsymbol{T}$ de loi uniforme sur Θ.
La loi jointe des données et du paramètre est donc :

$$P(\boldsymbol{x},\boldsymbol{\theta}) := \mathbb{P}(\boldsymbol{X}=\boldsymbol{x},\boldsymbol{T}=\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}=\boldsymbol{x})\mathbb{P}(\boldsymbol{T}=\boldsymbol{\theta}) = p_n(\boldsymbol{x},\boldsymbol{\theta})\frac{1}{\mathrm{card}(\Theta)}.$$

La loi du paramètre sachant les données est :

$$P(\boldsymbol{\theta}|\boldsymbol{x}) := \mathbb{P}(\boldsymbol{T}=\boldsymbol{\theta}|\boldsymbol{X}=\boldsymbol{x}) = \frac{P(\boldsymbol{x},\boldsymbol{\theta})}{P(\boldsymbol{x})}, \text{ avec } P(\boldsymbol{x}) = \sum_{\boldsymbol{\theta}'\in\Theta} P(\boldsymbol{x},\boldsymbol{\theta}'),$$

ce qui s'écrit donc :

$$P(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p_n(\boldsymbol{x},\boldsymbol{\theta})}{\mathrm{card}(\Theta)P(\boldsymbol{x})}.$$

Le mode de cette loi, i.e. le paramètre $\boldsymbol{\theta}$ le plus probable dans Θ sachant les données $\boldsymbol{x}$, est l'élément de Θ qui maximise $P(\boldsymbol{\theta}|\boldsymbol{x})$, qui est aussi celui qui maximise $p_n(\boldsymbol{x},\boldsymbol{\theta})$: c'est donc l'EMV.

Exemple : Dans le modèle de Bernoulli, $\mathcal{X} = \{0, 1\}$,

$$\boldsymbol{\mathcal{P}} = \big\{\mathcal{B}(\theta), \theta \in [0,1]\big\},$$

la vraisemblance est :

$$\begin{aligned} p_n(\boldsymbol{x}, \theta) &= \prod_{i=1}^{n} \theta^{\mathbf{1}_1(x_i)}(1-\theta)^{\mathbf{1}_0(x_i)} \\ &= \left[\theta^{\overline{x}_n}(1-\theta)^{1-\overline{x}_n}\right]^n, \end{aligned}$$

avec $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$.
A $\boldsymbol{x}$ fixé, en tant que fonction de θ, la vraisemblance est maximale lorsque la fonction $\theta \mapsto \theta^{\overline{x}_n}(1-\theta)^{1-\overline{x}_n}$ est maximale. Le maximum sur $[0,1]$ est unique et est atteint au point où la dérivée s'annule, en $\theta = \overline{x}_n$.
L'EMV est ici l'estimateur empirique.

EMV pour les lois gaussiennes

Etant donnés des paramètres μ, σ^2, le vecteur aléatoire $(X_1, \ldots, X_n)$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ a une loi de densité

$$p_n\big((x_1, \ldots, x_n), (\mu, \sigma^2)\big) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x_i - \mu)^2}{2\sigma^2}\Big)$$

L'EMV est défini par :

$$(\hat{\mu}_n, \hat{\sigma}_n^2) = \operatorname*{argmax}_{\mu, \sigma^2} \; p_n\big((X_1, \ldots, X_n), (\mu, \sigma^2)\big)$$

Dans ce cas gaussien, on trouve $\hat{\mu}_n = \overline{X}_n$ et $\hat{\sigma}_n^2 = \overline{V}_n$ (variance empirique biaisée). C'est l'unique point (μ, σ^2) où les dérivées de $\ln p_n$ s'annulent.

Exemple : On considère le modèle uniforme $\mathcal{P} = \{P_\theta, \theta \in]0, +\infty[\}$, où P_θ est la loi uniforme sur $[0, \theta]$. La vraisemblance d'un n-échantillon est :

$$p_n(\boldsymbol{x}, \theta) = \theta^{-n} \prod_{i=1}^{n} \mathbf{1}_{[0,\theta]}(x_i) = \theta^{-n} \mathbf{1}_{[0,+\infty[}\big(\min_{i=1,\dots,n}(x_i)\big) \mathbf{1}_{[0,\theta]}\big(\max_{i=1,\dots,n}(x_i)\big).$$

• La vraisemblance est maximale pour $\theta_n(\boldsymbol{x}) = \max_{i=1,\dots,n}(x_i)$. L'EMV est donc $\hat{\theta}_n = \max_{i=1,\dots,n}(X_i)$.

• On peut calculer la loi de l'EMV. Sa fonction de répartition est :

$$\mathbb{P}_\theta(\hat{\theta}_n \leq t) = \begin{cases} 1 & \text{si } t \geq \theta, \\ \mathbb{P}_\theta(X_1 \leq t)^n = \left(\dfrac{t}{\theta}\right)^n & \text{si } 0 \leq t < \theta, \\ 0 & \text{si } t < 0. \end{cases}$$

Donc, sous $\mathbb{P}_\theta$, $\hat{\theta}_n$ est une v.a. à densité :

$$p_{\hat{\theta}_n}(t) = n \frac{t^{n-1}}{\theta^n} \mathbf{1}_{[0,\theta]}(t).$$

• L'EMV est biaisé :

$$\mathbb{E}_\theta(\hat{\theta}_n) = \int_0^\theta t p_{\hat{\theta}_n}(t) dt = \frac{n\theta}{n+1} = \theta - \frac{\theta}{n+1}$$

- La variance de l'EMV est

$$\mathrm{Var}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta(\hat{\theta}_n^2) - \mathbb{E}_\theta(\hat{\theta}_n)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 \, ,$$

qui est d'ordre $1/n^2$. Le RQM est donc d'ordre $1/n^2$ lui aussi :

$$\mathrm{RQM}_\theta(\hat{\theta}_n) = \mathrm{Var}_\theta(\hat{\theta}_n) + \left(\mathbb{E}_\theta(\hat{\theta}_n) - \theta\right)^2 = \frac{2}{(n+1)(n+2)}\theta^2$$

$\rightarrow$ plus petit que le RQM de l'estimateur empirique !

-

$$\mathbb{P}_\theta\big(n(\hat{\theta}_n - \theta) \le t\big) = \begin{cases} 1 & \text{si } t \ge 0, \\ \left(1 + \dfrac{t}{n\theta}\right)^n & \text{si } -n\theta \le t < 0, \\ 0 & \text{si } t < -n\theta, \end{cases}$$

et donc, pour tout $t \ge 0$:

$$\mathbb{P}_\theta\big(n(\hat{\theta}_n - \theta) \le -\theta t\big) \stackrel{n \to +\infty}{\longrightarrow} e^{-t},$$

ce qui montre que

$$n\big(\hat{\theta}_n - \theta\big) \stackrel{n \to +\infty}{\longrightarrow} -\theta Z$$

en loi, où Z est une v.a. exponentielle de paramètre 1.

• Les fluctuations de l'EMV sont d'ordre $1/n$ et de loi exponentielle lorsque n est grand. On est donc loin de la normalité asymptotique.
• On peut débiaiser l'EMV en considérant l'estimateur :

$$\widetilde{\theta}_n = \frac{n+1}{n}\hat{\theta}_n = \frac{n+1}{n}\max_{i=1,\ldots,n}(X_i)\,.$$

L'estimateur $\widetilde{\theta}_n$ est non-biaisé, son RQM est :

$$\mathrm{RQM}_\theta(\widetilde{\theta}_n) = \mathrm{Var}_\theta(\widetilde{\theta}_n) = \frac{(n+1)^2}{n^2}\mathrm{Var}_\theta(\hat{\theta}_n) = \frac{1}{n(n+2)}\theta^2\,,$$

qui est plus petit que le RQM de $\hat{\theta}_n$, et il satisfait :

$$n\big(\widetilde{\theta}_n - \theta\big) \overset{n\to+\infty}{\longrightarrow} -\theta(Z-1),$$

en loi, où Z est une v.a. exponentielle de paramètre 1. Notez que $\mathbb{E}((Z-1)^2) = 1 < 2 = \mathbb{E}(Z^2)$, ce qui montre que la variance asymptotique de $\widetilde{\theta}_n$ est deux fois plus petite que celle de $\hat{\theta}_n$.

• On considère un n-échantillon de loi $\mathcal{U}(0, \theta_0)$ avec $n = 10$ et $\theta_0 = 1{,}5$:

$$\boldsymbol{x} = [1{,}222 \;\; 1{,}358 \;\; 0{,}190 \;\; 1{,}370 \;\; 0{,}948 \;\; 0{,}146 \;\; 0{,}417 \;\; 0{,}820 \;\; 1{,}436 \;\; 1{,}447]$$

L'estimateur empirique est $\check{\theta}_{10} = 2\overline{x}_{10} = 1{,}872$.
L'EMV non-biaisé est $\widetilde{\theta}_{10} = (11/10) \max_i x_i = 1{,}592$.

• On considère 1000 échantillons et on trace les histogrammes des 1000 estimations :