Inférence statistique

Cours 4 : Méthode de Moments

Michal W. Urdanivia[*]

[*]UGA, Faculté d'Économie, GAEL,
e-mail : michal.wong-urdanivia@univ-grenoble-alpes.fr

19 septembre 2022

Plan

Plan

1. Introduction

Théorème d'approximation de Weierstrass

- **Théorème**
  - Soit f une fonction continue sur l'intervalle $[a, b]$, alors, pour tout $\epsilon > 0$, il existe $a_0, a_1, \ldots, a_d \in \mathbb{R}$ tels que,

$$\max_{x \in [a,b]} \left| f(x) - \sum_{k=1}^{d} a_k x^k \right| < \epsilon.$$

  - Autrement dit : **les fonctions continues peuvent être approchées arbitrairement par des polynômes**.

Application statistique

- Soit un échantillon de v.a. i.i.d., $X_1, X_2, \ldots, X_n$ associé à un modèle statistique qu'on suppose identifié, $(\mathcal{E}, \mathcal{F}, (\mathrm{P}_\theta)_{\theta\in\Theta})$.
- On note θ^* le vrai paramètre.
- Supposons que pour tout θ, la loi P_θ a la fonction de densité f_θ.
- Si nous trouvons un θ tel que,

$$\int h(x) f_{\theta^*}(x)\mathrm{d}x = \int h(x) f_\theta(x)\mathrm{d}x$$

  pour toutes les fonctions (continues et bornées) h, alors $\theta = \theta^*$.
- En remplaçant les espérances par des moyennes : il s'agit de trouver un estimateur $\hat{\theta}$ tel que,

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) = \int h(x) f_{\hat{\theta}}(x)\mathrm{d}x$$

  pout toutes les fonctions continues et bornées h.
- **Problème :** il y a une **infinité** de fonctions de la sort(infaisable).

Application statistique

- Par application du TAW, il suffit de considérer des polynômes :

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{d} a_k X_i^k = \int \sum_{k=0}^{d} a_k x^k f_{\hat{\theta}}(x)\mathrm{d}x, \ \forall a_0, a_1, \ldots, a_d \in \mathbb{R}.$$

  Soit encore une infinité d'équations :

- Pour sa part, il suffit de considérer,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k = \int x^k f_{\hat{\theta}(x)}\mathrm{d}x, \forall k = 1, 2, \ldots, d,$$

  (seulement $d+1$ équations).

- La quantité $m_k(\theta) := \int x^k f_\theta(x)\mathrm{d}x$ est le moment d'ordre k(ou k-ième moment) de P_θ.
- On peut aussi l'écrire $m_k(\theta) = \mathrm{E}_\theta(X^k)$.

Quadrature de Gauss

- L'approximation de Weierstrass présente un certain nombre de limitations :
  i) ne s'applique qu'à des fonctions continues(ceci peut être réglé),
  ii) ne s'applique que sur des intervalles $[a, b]$,
  iii) ne nous dit pas ce que d doit être(i.e., $\#$ de moments).
- Qu'en est-il pour $\mathcal{E}$ discret(cas où à P_θ on associe une fonction de masse et non de densité) ?

Quadrature de Gauss

- Supposons que $\mathcal{E} = \{x_1, x_2, \ldots, x_r\}$ est fini avec r valeurs possibles.
- La fonction de masse est alors caractérisée par $r - 1$ paramètres,

$$\mathrm{p}(x_1), \mathrm{p}(x_2), \ldots, \mathrm{p}(x_{r-1}),$$

car le dernier est donné par les premiers $r - 1$ paramètres,

$$\mathrm{p}_r = 1 - \sum_{j=1}^{r-1} \mathrm{p}(x_j).$$

- Avec un peu de chance, il ne faudra pas plus que $d = r - 1$ paramètres pour obtenir la fonction de masse $\mathrm{p}(\cdot)$.

Quadrature de Gauss

- Notons que pour tout $k = 1, 2, \ldots, r_1$,

$$m_k := \mathrm{E}(X^k) = \sum_{j=1}^{r} \mathrm{p}(x_j) x_j^k,$$

et,

$$\sum_{j=1}^{r} \mathrm{p}(x_j) = 1$$

- Ceci est un système d'équations linéaires avec les inconnues, $\mathrm{p}(x_1), \mathrm{p}(x_2), \ldots, \mathrm{p}(x_r)$.
- Il est peut être écrit sous une forme compacte(matricielle) :

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1^1 & x_2^1 & \ldots & x_r^1 \\ x_1^2 & x_2^2 & \ldots & x_r^2 \\ \vdots & & \ddots & \vdots \\ x_1^{r-1} & x_2^{r-1} & \ldots & x_r^{r-1} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{p}(x_1) \\ \mathrm{p}(x_2) \\ \vdots \\ \mathrm{p}(x_{r-1}) \\ \mathrm{p}(x_r) \end{pmatrix} = \begin{pmatrix} 1 \\ m_1 \\ m_2 \\ \vdots \\ m_{r-1} \end{pmatrix}$$

Quadrature de Gauss

- Il faut vérifier que l'inversibilité : **déterminant de Vandermonde**

$$\det\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^1 & x_2^1 & \dots & x_r^1 \\ x_1^2 & x_2^2 & \dots & x_r^2 \\ \vdots & & \ddots & \vdots \\ x_1^{r-1} & x_2^{r-1} & \dots & x_r^{r-1} \end{pmatrix} = \prod_{1<j<k<r} (x_j - x_k) \neq 0$$

Quadrature de Gauss

- Donc, si nous avons les moments $m_1, m_2, \ldots, m_{r-1}$, il y a une seule fonction de masse avec ces moments. Et elle est donnée par,

$$\begin{pmatrix} \mathrm{p}(x_1) \\ \mathrm{p}(x_2) \\ \vdots \\ \mathrm{p}(x_{r-1}) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1^1 & x_2^1 & \ldots & x_r^1 \\ x_1^2 & x_2^2 & \ldots & x_r^2 \\ \vdots & & \ddots & \vdots \\ x_1^{r-1} & x_2^{r-1} & \ldots & x_r^{r-1} \end{pmatrix}^{-1} . \begin{pmatrix} 1 \\ m_1 \\ m_2 \\ \vdots \\ m_{r-1} \end{pmatrix}$$

Conclusion à partir du TAW et de la quadrature de Gauss

- Les moments d'une v.a., fournissent une information importante sur la fonction de densité ou de masse.
- En estimant avec précision ces moments on devrait être capables de récupérer/retrouver la loi génératrice des données.
- Dans un cadre paramétrique où la loi des observations P_θ est connue aux paramètres θ près, il est fréquent de n'avoir besoin que d'un petit nombre de moments pour obtenir(estimer) θ, ceci variant d'un cas à l'autre.
- Règle basique : quand $\theta \in \Theta \subseteq \mathbb{R}^d$, on a besoin de d moments.

Plan

2. Méthode des Moments

Principe d'estimation de la MM

- Soit $X_1, X_2, \ldots, X_n$ un échantillon i.i.d. pour le modèle statistique $(\mathcal{E}, \mathcal{F}, (\mathrm{P}_\theta)_{\theta\in\Theta})$.
- On suppose $\Theta \subseteq \mathbb{R}^d$, pour un $d \geq 1$.
  - **Moments théorique(ou de Population) :**

$$m_k(\theta) := \mathrm{E}_\theta(X^k), \quad 1 \leq k \leq d.$$

  - **Moments empiriques :**

$$\hat{m}_k := \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad 1 \leq k \leq d.$$

Principe d'estimation de la MM

- Soit,

$$\begin{aligned}\psi : \Theta \subseteq \mathbb{R}^d &\to \mathbb{R}^d\\ \theta &\mapsto (m_1(\theta), m_2(\theta), \ldots, m_d(\theta)).\end{aligned}$$

- Supposons ψ bijective, alors,

$$\theta = \psi^{-1}(m_1(\theta), m_2(\theta), \ldots, m_d(\theta)).$$

- **<u>Définition :</u>** un estimateur des moments de θ est défini par,

$$\hat{\theta}_n^{MM} = \psi^{-1}(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_d),$$

dès lors qu'il existe.

Analyse de $\hat{\theta}_n^{MM}$

- Notons :

$$\begin{aligned} M(\theta) &:= (m_1(\theta), m_2(\theta), \dots, m_d(\theta)), \\ \hat{M} &= (\hat{m}_1, \hat{m}_2, \dots, \hat{m}_d), \\ \mathrm{V}(\theta) &= \underbrace{\mathrm{Var}_\theta\left(X, X^2, \dots, X^d\right)}_{\text{Matrice des variances-covariances de } (X, X^2, \dots, X^d)\text{, quand } X \sim \mathrm{P}_\theta\text{.}} \end{aligned}$$

  .

- Supposons ψ^{-1} continûment dérivable en $M(\theta)$. Notons $\nabla\psi^{-1}_{M(\theta)}$ la matrice $d \times d$ du gradient en ce point.

Analyse de $\hat{\theta}_n^{MM}$

- **LGN** : $\hat{\theta}_n^{MM}$ est faiblement/fortement convergent.
- **TCL** :

$$\sqrt{n}\left(\hat{M} - M(\theta)\right) \xrightarrow{d} \mathcal{N}(0, \mathrm{V}(\theta)) \quad \text{par rapport à } \mathrm{P}_\theta.$$

et en utilisant la méthode du delta(voir diapos suivantes),

Analyse de $\hat{\theta}_n^{MM}$

- **Théorème** :

$$\sqrt{n}\left(\hat{\theta}_n^{MM} - \theta\right) \xrightarrow{d} \mathcal{N}(0, \Gamma(\theta)) \quad \text{par rapport à } \mathrm{P}_\theta,$$

  où

$$\Gamma(\theta)) = \left(\nabla\psi^{-1}_{M(\theta)}\right)^\top \mathrm{V}(\theta) \left(\nabla\psi^{-1}_{M(\theta)}\right).$$

Méthode du Delta dans le cas multivarié

- Soit $(T_n)_{n\geq 1}$ une suite de vecteurs aléatoires dans $\mathbb{R}^p$(pour $p \geq 1$) telle que,

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{d} \mathcal{N}(0, \mathrm{V}),$$

  pour un $\theta \in \mathbb{R}^p$ et une matrice symétrique et semi-définie positive $\mathrm{V} \in \mathbb{R}^p \times p$.
- Soit $g : \mathbb{R}^p \to \mathbb{R}^k$(pour $k \geq 1$) une fonction continûment dérivable en θ.
- Alors,

$$\sqrt{n}\,(g(T_n) - g(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \nabla g(\theta)^\top \mathrm{V}\, \nabla g(\theta)\right),$$

  où,

$$\nabla g(\theta) := \left(\frac{\partial g_j}{\partial \theta_i}\right)_{1\leq i\leq d, 1\leq j\leq k} \in \mathbb{R}^{k\times d}.$$

MV vs MM

- **Risque quadratique** : en général l'estimateur du MV est plus précis.
- **Problème de calcul** : l'estimateur du MV ne peut pas être obtenu en forme analytique :
  - Quand la vraisemblance est concave, on peut utiliser des algorithmes d'optimisation numérique(méthodes du point intérieur, gradient descendant, etc).
  - Si la vraisemblance n'est pas concave : démarche heuristique, maxima locaux.