Tests Statistiques

Master 1: 2020-2021

Prof. Abdelhakim Necir

Département de Mathématiques (Univ. Biskra)

Samedi 19 Décembre 2020

- Estimation paramétrique poncuelle
- Estimation paramétrique par intervalle (**intervalle de confiance**)

Estimation paramétrique ponctuelle

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, \ldots$

Exemple

1. $X \rightsquigarrow U(a, b) : \theta = (a, b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, ...$

Exemple

1. $X \rightsquigarrow U(a, b) : \theta = (a, b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$
2. $X \rightsquigarrow \exp(\theta)$, $\theta \in \Theta = \mathbb{R}_+^*$

Estimation paramétrique ponctuelle

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, ...$

Exemple

1. $X \rightsquigarrow U(a, b) : \theta = (a, b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$
2. $X \rightsquigarrow \exp(\theta)$, $\theta \in \Theta = \mathbb{R}_+^*$
3. $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*$

Estimation paramétrique ponctuelle

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, \ldots$

Exemple

1. $X \rightsquigarrow U(a,b) : \theta = (a,b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$
2. $X \rightsquigarrow \exp(\theta)$, $\theta \in \Theta = \mathbb{R}_+^*$
3. $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*$
4. $X \rightsquigarrow \text{Bernoulli}(p)$, $\theta = p \in \Theta \in [0,1]$

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, \ldots$

Exemple

1. $X \rightsquigarrow U(a,b) : \theta = (a,b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$
2. $X \rightsquigarrow \exp(\theta)$, $\theta \in \Theta = \mathbb{R}_+^*$
3. $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*$
4. $X \rightsquigarrow \text{Bernoulli}(p)$, $\theta = p \in \Theta \in [0,1]$
5. $X \rightsquigarrow \text{Binomial}(n,p)$, $\theta = (n,p) \in \Theta \in \mathbb{N}^* \times [0,1]$

Estimation paramétrique ponctuelle

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, ...$

Exemple

1. $X \rightsquigarrow U(a,b) : \theta = (a,b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$
2. $X \rightsquigarrow \exp(\theta)$, $\theta \in \Theta = \mathbb{R}_+^*$
3. $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*$
4. $X \rightsquigarrow$Bernoulli(p), $\theta = p \in \Theta \in [0,1]$
5. $X \rightsquigarrow$Binomial(n,p), $\theta = (n,p) \in \Theta \in \mathbb{N}^* \times [0,1]$
6. $X \rightsquigarrow$Pois(λ), $\theta = \lambda \in \Theta \in \mathbb{R}_+^*$

Soit X une va ayant une densité de probabilité $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, $m = 1, 2, 3, ...$

Exemple

1. $X \rightsquigarrow U(a, b) : \theta = (a, b) \in \Theta = \mathbb{R} \times \mathbb{R}^*$
2. $X \rightsquigarrow \exp(\theta)$, $\theta \in \Theta = \mathbb{R}_+^*$
3. $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*$
4. $X \rightsquigarrow$Bernoulli(p), $\theta = p \in \Theta \in [0, 1]$
5. $X \rightsquigarrow$Binomial(n, p), $\theta = (n, p) \in \Theta \in \mathbb{N}^* \times [0, 1]$
6. $X \rightsquigarrow$Pois(λ), $\theta = \lambda \in \Theta \in \mathbb{R}_+^*$
7. $X \rightsquigarrow \chi^2(r)$, $\theta = r \in \Theta \in \mathbb{N}^*$

Estimation paramétrique ponctuelle

Notations: $\mathbf{E}(X,Y) := (\mathbf{E}[X], \mathbf{E}[Y])$ et $\widehat{\theta}_n := \widehat{\theta}$.

Problème: construire un estimateur $\widehat{\theta} = \widehat{\theta}(X_1, ..., X_n)$ pour le paramètre θ.

- Estimateur sans biais: $\mathbf{E}\left[\widehat{\theta}\right] = \theta$.
- Estimateur asymptotiquement san bias: $\mathbf{E}\left[\widehat{\theta}\right] \to \theta$ quand $n \to \infty$.
- Estimateur biaisé: $\mathbf{E}\left[\widehat{\theta}\right] = \theta + b_n(\theta)$, avec $b_n(\theta) := \mathbf{E}\left[\widehat{\theta} - \theta\right]$ est le biais de l'estimateur.
- Estimateur consistent: $\widehat{\theta} \xrightarrow{P} \theta$.
- Estimateur asymptotiquement normal: $\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(\theta)\right)$.

Estimation paramétrique ponctuelle: critère de comparaison

Un bon critère de comparaison est l'erreur quadratique:
Soit $\widehat{\theta}$ un estimateur de θ. L'**erreur quadratique moyenne (EQM), en anglais mean squared error (MSE),** est définie par:

$$EQM\left(\widehat{\theta},\theta\right)=\mathbf{E}\left[\widehat{\theta}-\theta\right]^2.$$

Soit $\widehat{\theta}_1$ et $\widehat{\theta}_2$ deux estimateurs de θ. On dit que $\widehat{\theta}_1$ est meilleur que $\widehat{\theta}_2$ (au sense de **EQM**), si

$$\mathbf{EQM}\left(\widehat{\theta}_1,\theta\right)<\mathbf{EQM}\left(\widehat{\theta}_2,\theta\right).$$

L'**EQM** se décompose en deux termes, le carré du biais et la variance:

$$\begin{aligned}\mathbf{EQM}\left(\widehat{\theta},\theta\right) &= \left(\mathbf{E}\left[\widehat{\theta}-\theta\right]\right)^2+\mathbf{Var}\left[\widehat{\theta}\right]\\ &= b_\theta^2+\mathbf{Var}\left[\widehat{\theta}\right].\end{aligned}$$

Supposons que $\widehat{\theta}_1$ et $\widehat{\theta}_2$ deux estimateurs **sans biais** de θ. On dit que $\widehat{\theta}_1$ est plus éfficace que $\widehat{\theta}_2$ si

$$\mathbf{Var}\left[\widehat{\theta}_1\right] < \mathbf{Var}\left[\widehat{\theta}_2\right].$$

Construction d'un estimateur: méthode du maximum de vraisemblance

La fonction de vraisemblance est la densité de probabilité de l'echantillons $(X_1, ..., X_n)$:

$$\mathbf{f}_\theta (X_1, ..., X_n) = \prod_{i=1}^{n} f_\theta (X_i) .$$

La fonction de log-vraisemblance est définie par

$$\begin{aligned} L_\theta (X_1, ..., X_n) &= \log \mathbf{f}_\theta (X_1, ..., X_n) \\ &= \sum_{i=1}^{n} \ell_\theta (X_i) . \end{aligned}$$

où $\ell_\theta (x) := \log f_\theta (x)$.

Construction d'un estimateur: méthode du maximum de vraisemblance

L'estimateur de maximum de vraisemblance (**EMV**) de $\theta \in \Theta \subset \mathbb{R}^m$, la statistique définie par:

$$\widehat{\theta}_{EMV} := \arg\max_{\theta \in \Theta} \mathbf{f}_\theta \left(X_1, ..., X_n \right).$$

En d'autres termes, on cherche $\widehat{\theta}_{EMV}$ est le point tel que

$$\mathbf{f}_{\widehat{\theta}_{EMV}} \left(X_1, ..., X_n \right) \text{ est maximale.}$$

Pour tout $\theta \in \Theta$

$$\mathbf{f}_\theta \left(X_1, ..., X_n \right) \leq \mathbf{f}_{\widehat{\theta}_{EMV}} \left(X_1, ..., X_n \right).$$

Ou de manière équivalente

$$\widehat{\theta}_{EMV} := \arg\max_{\theta \in \Theta} L_\theta \left(X_1, ..., X_n \right).$$

Construction d'un estimateur: méthode du maximum de vraisemblance

Supposons que: $\theta \longrightarrow \ell_\theta$ admet une derivée continue $\ell'_\theta := d\ell_\theta / d\theta$, alors $\widehat{\theta}_{EMV}$ est une solution de l'équation

$$L'_\theta\left(X_1, ..., X_n\right) = \sum_{i=1}^{n} \ell'_\theta\left(X_i\right) = 0.$$

S'il existe une solution $\theta_n = \theta_n\left(X_1, ..., X_n\right)$ de cette equation telle que

$$L''_\theta\left(X_1, ..., X_n\right) \leq 0, \text{ pour tout } \theta \in \Theta,$$

alors $\theta_n \equiv \widehat{\theta}_{EMV}$.

L'**existence** et l'**unicité** de l'EMV seront discutés par la suite.

Construction d'un estimateur: méthode du maximum de vraisemblance

Exemple 1. $X \rightsquigarrow Exp\left(\theta\right)$: une va qui suit la loi exponentielle de paramétre $\theta > 0$:

$$f_\theta\left(x\right) = \frac{1}{\theta}\exp\left(-\frac{1}{\theta}x\right),\ \ x \geq 0.$$

La fonction $\theta \longrightarrow f_\theta\left(x\right)$ admet une dérivée (par rapport à θ) continue. Nous avons

$$\ell_\theta\left(x\right) = \log f_\theta\left(x\right) = -\log\theta - \frac{1}{\theta}x$$

et

$$\ell'_\theta\left(x\right) = -\frac{1}{\theta} + \frac{1}{\theta^2}x$$

et

$$\ell''_\theta\left(x\right) = \frac{1}{\theta^2} - \frac{2}{\theta^3}x = \frac{1}{\theta^2}\left(1 - \frac{2}{\theta}x\right).$$

Construction d'un estimateur: méthode du maximum de vraisemblance

Donc $\widehat{\theta}_{EMV}$ est une solution de l'équation

$$\sum_{i=1}^{n} \ell'_{\theta}(X_i) = \sum_{i=1}^{n} \left(-\frac{1}{\theta} + \frac{1}{\theta^2} X_i \right) = 0.$$

Cece implique que

$$-n + \frac{1}{\theta} \sum_{i=1}^{n} X_i = 0.$$

Ce qui donne $\theta_n = \overline{X}$.

Construction d'un estimateur: méthode du maximum de vraisemblance

D'autre part:

$$\begin{aligned} L''_{\theta}\left(X_1,...,X_n\right) &= \sum_{i=1}^{n} \ell''_{\theta}\left(X_i\right) = \frac{1}{\theta^2}\sum_{i=1}^{n}\left(1-\frac{2}{\theta}X_i\right) \\ &= \frac{1}{\theta^2}\left(n-\frac{2}{\theta}\sum_{i=1}^{n}X_i\right) = \frac{n}{\theta^2}\left(1-\frac{2\overline{X}_n}{\theta}\right). \end{aligned}$$

Cece implique que

$$L''_{\theta_n}\left(X_1,...,X_n\right) = -\frac{n}{\overline{X}^2} < 0.$$

La solution de cette équation est $\theta_n = \widehat{\theta}_{EMV}$.

Construction d'un estimateur: méthode du maximum de vraisemblance

Exemple 2. $X \rightsquigarrow \mathcal{N}\left(\mu, \sigma^2\right)$: une va qui suit la loi de Gauss de paramètre

$$\theta = \left(\mu, \sigma^2\right) := \left(\theta_1, \theta_2\right) \in \mathbb{R} \times \mathbb{R}_+^* = \Theta.$$

La densité de X est

$$f_\theta\left(x\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \ x \in \mathbb{R}$$

Construction d'un estimateur: méthode du maximum de vraisemblance

Il est claire que $\theta \to f_\theta(x)$ admet $\partial f_\theta(x)/\partial\theta_i$, $i = 1, 2$ continues et

$$\ell_\theta(x) = \log f_\theta(x) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2.$$

Alors

$$\frac{\partial \ell_\theta(x)}{\partial\theta_1} = \frac{\partial \ell_{(\mu,\sigma^2)}(x)}{\partial\mu} = \frac{x-\mu}{\sigma^2}$$

et

$$\frac{\partial \ell_\theta(x)}{\partial\theta_2} = \frac{\partial \ell_{(\mu,\sigma^2)}(x)}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2}(x-\mu)^2\frac{1}{\sigma^2}.$$

Construction d'un estimateur: méthode du maximum de vraisemblance

Donc $\widehat{\theta}_{EMV}$ est une solution de l'équation

$$\sum_{i=1}^{n} \frac{\partial \ell_{(\mu,\sigma^2)}(X_i)}{\partial \mu} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma^2} = 0.$$

$$\sum_{i=1}^{n} \frac{\partial \ell_{(\mu,\sigma^2)}(X_i)}{\partial \sigma^2} = \left(\sum_{i=1}^{n} -\frac{1}{2\sigma^2} + \frac{1}{2}(X_i - \mu)^2 \frac{1}{\sigma^2} \right) = 0.$$

Cece implique que

$$\mu_n = \overline{X} \text{ et } \sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \widetilde{S}_n^2.$$

Construction d'un estimateur: méthode du maximum de vraisemblance

On démontre que

$$\sum_{i=1}^{n} \frac{\partial^2 \ell_{\left(\mu_n, \sigma_2^2\right)} \left(X_i\right)}{\partial \mu^2} \leq 0$$

et

$$\sum_{i=1}^{n} \frac{\partial^2 \ell_{\left(\mu_n, \sigma_n^2\right)} \left(X_i\right)}{\partial \left(\sigma_n^2\right)^2} \leq 0.$$

Donc

$$\widehat{\mu}_{EMV} = \overline{X} \text{ et } \widehat{\sigma}_{EMV}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$