Économétrie non-linéaire

Chapitre 1: Introduction - Maximum Likelihood Estimation Method

Gilles de Truchis, Elena Dumitrescu

Master 2 BMM - EIPMC - GDA

Septembre 2016

Plan

Méthode du Maximum de Vraisemblance

- Partons d'un exemple : soit un échantillon $X_t = X_1, \cdots, X_n \sim P(\theta)$
  - $P(\theta)$ dénote la distribution de Poisson dont la fonction de masse est

$$\Pr(X_i = x) = \frac{\exp(-\theta)\theta^x}{x!}, \quad \theta > 0, \quad \forall x \in \mathbb{N}$$

- Soit une réalisation de l'échantillon $x_t = x_1, \cdots, x_n$

- La probabilité d'observer cette réalisation est

$$\Pr\left((X_1 = x_1) \cap, \cdots, \cap (X_n = x_n)\right)$$

- L'indépendance des tirages donne l'équivalence avec le produit des probabilités marginales

$$\Pr\left((X_1 = x_1) \cap, \cdots, \cap (X_n = x_n)\right) = \prod_{i=1}^{n} \Pr(X_i = x_i)$$

L'estimateur du Maximum de Vraisemblance

- En remplaçant par la fonction de masse de la loi de Poisson on obtient

$$\Pr\left((X_1 = x_1)\cap, \cdots, \cap(X_n = x_n)\right) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = e^{-n\theta}\frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

  - Il s'agit donc d'une fonction dépendant de $x_1, \cdots, x_n$ et de θ
  - θ est un paramètre inconnu mais on observe $x_1, \cdots, x_n$
  - Par la suite on notera :

$$L_n(\theta; x_1, \cdots, x_n) = \Pr\left((X_1 = x_1)\cap, \cdots, \cap(X_n = x_n)\right)$$

- Le principe du maximum de vraisemblance est le suivant :
  - Trouver le θ qui maximise la probabilité d'apparition de $x_1, \cdots, x_n$

- L'estimateur du maximum de vraisemblance est donc :

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}^+} L_n(\theta; x_1, \cdots, x_n)$$

L'estimateur du Maximum de la log-Vraisemblance

- Dans le cas de l'exemple reposant sur la loi de Poisson on a

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}^+} e^{-n\theta} \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

  - La formule est complexe et la présence d'un produit n'arrange rien

- Simplifions le programme de maximisation en considérant la log-vraisemblance

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}^+} \ln L_n(\theta; x_1, \cdots, x_n)$$

- Dans le cadre de notre exemple la log-vraisemblance est

$$\ln L_n(\theta; x_1, \cdots, x_n) = -n\theta + \ln(\theta) \sum_{i=1}^n x_i - \ln\left(\prod_{i=1}^n x_i!\right)$$

Conditions nécessaire et suffisante

- La condition nécessaire répond à la question
  - Le problème admet-il une solution ?
  - ⇒ Pour répondre on annule la dérivée première par rapport à θ

$$\left.\frac{\partial \ln L_n(\theta; x_1, \cdots, x_n)}{\partial \theta}\right|_{\hat{\theta}} = -n + \hat{\theta}^{-1} \sum_{i=1}^{n} x_i = 0 \Longleftrightarrow \hat{\theta} = n^{-1} \sum_{i=1}^{n} x_i$$

  - Ici, la log vraisemblance est maximisée par la moyenne empirique

- La condition suffisante répond à la question
  - Cette solution est-elle un maximum ?
  - ⇒ Pour répondre on regarde le signe de la dérivée seconde par rapport à θ

$$\left.\frac{\partial^2 \ln L_n(\theta; x_1, \cdots, x_n)}{\partial \theta^2}\right|_{\hat{\theta}} = -\hat{\theta}^{-2} \sum_{i=1}^{n} x_i < 0$$

  - Négatif donc bien un maximum

Log-Vraisemblance Gaussienne

- Dans l'exemple, il s'agissait de variables aléatoires discrètes
- Dans le cas de variables aléatoires continues, l'intuition est la même
  - Néanmoins, on raisonnera sur la densité de la loi jointe des variables

$$L_n(\theta; x_1, \cdots, x_n) = f_{X_1,\cdots,X_n}(x_1, \cdots, x_n; \theta)$$

- Soit une séquence $X_n \sim$ i. i. d. (μ, σ^2) selon une loi normale
- La densité de la loi normale implique 2 paramètres $\theta = (\mu, \sigma^2)'$

$$L_n(\theta; x_1, \cdots, x_n) = \prod_{i=1}^{n} (\sigma\sqrt{2\pi})^{-1} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\ln L_n(\theta; x_1, \cdots, x_n) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

L'estimateur du Maximum de Vraisemblance

- Estimateur du maximum de vraisemblance

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}^+} \ln L_n(\theta; x_1, \cdots, x_n)$$

- Hypothèses
  - $\theta = (\mu, \sigma^2)'$ est identifiable : $\forall \theta^*, \theta$ avec $\theta^* \neq \theta$, les lois jointes de $x_1, \cdots, x_n$ sont différentes
  - Condition nécessaire du gradient :

$$g_n(\hat{\theta}; x_1, \cdots, x_n) = \left. \frac{\partial \ln L_n(\theta; x_1, \cdots, x_n)}{\partial \theta} \right|_{\hat{\theta}} = 0$$

  - Condition suffisante de la hessienne :

$$H_n(\hat{\theta}; x_1, \cdots, x_n) = \left. \frac{\partial^2 \ln L_n(\theta; x_1, \cdots, x_n)}{\partial \theta^2} \right|_{\hat{\theta}} < 0$$

Condition nécessaire du MLE gaussien

- Notons $\ln L_n(\theta; x_1, \cdots, x_n) = \ell_n(\theta; x)$ et commençons par le gradient :

Condition nécessaire du MLE gaussien

- Notons $\ln L_n(\theta; x_1, \cdots, x_n) = \ell_n(\theta; x)$ et commençons par le gradient :

$$\frac{\partial \ell_n(\theta; x)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell_n(\theta;x)}{\partial \mu} \\ \\ \frac{\partial \ell_n(\theta;x)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{pmatrix}$$

$$\Rightarrow \left. \frac{\partial \ell_n(\theta; x)}{\partial \theta} \right|_{\hat{\theta}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \hat{\theta} = \begin{pmatrix} \hat{\mu} = n^{-1} \sum_{i=1}^n x_i = \bar{x} \\ \\ \hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix}$$

- Le programme de maximisation a donc une solution
  - Les réalisations du ML sont $\hat{\mu} = n^{-1} \sum_{i=1}^n x_i = \bar{x}$ et $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ (variance empirique non-corrigée)
  - Les estimateurs du ML sont $\hat{\mu} = n^{-1} \sum_{i=1}^n X_i = \bar{X}$ et $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Condition suffisante du MLE gaussien

- La solution est-elle bien un maximum ?

Condition suffisante du MLE gaussien

- La solution est-elle bien un maximum ?

$$\frac{\partial^2 \ell_n(\theta; x)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu^2} & \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu\partial\sigma^2} \\ \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu\partial\sigma^2} & \frac{\partial^2 \ell_n(\theta;x)}{\partial\sigma^4} \end{pmatrix}$$

- On obtient alors

Condition suffisante du MLE gaussien

- La solution est-elle bien un maximum ?

$$\frac{\partial^2 \ell_n(\theta;x)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu^2} & \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu\partial\sigma^2} \\ \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu\partial\sigma^2} & \frac{\partial^2 \ell_n(\theta;x)}{\partial\sigma^4} \end{pmatrix}$$

- On obtient alors

$$\left.\frac{\partial^2 \ell_n(\theta;x)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i - \hat{\mu}) \\ -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i - \hat{\mu}) & \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\sum_{i=1}^n (x_i - \hat{\mu})^2 \end{pmatrix}$$

- D'après l'étude du gradient, on sait que $n \times \hat{\mu} = \sum_{i=1}^n x_i$ et donc

$$-\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i - \hat{\mu}) = -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n x_i + \frac{1}{\hat{\sigma}^4} n \times \hat{\mu} = \frac{1}{\hat{\sigma}^4}\sum_{i=1}^n x_i - \frac{1}{\hat{\sigma}^4}\sum_{i=1}^n x_i = 0$$

- De plus, $n \times \hat{\sigma}^2 = \sum_{i=1}^n (x_i - \hat{\mu})^2$, ce qui donne

$$\left.\frac{\partial^2 \ell_n(\theta;x)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} - \frac{n\hat{\sigma}^2}{\hat{\sigma}^6} \end{pmatrix} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

Condition suffisante du MLE gaussien

- Pour conclure, il faut montrer que la hessienne est définie négative

$$\left.\frac{\partial^2 \ell_n(\theta;x)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

- Pour cela on s'intéresse aux mineurs principaux, Δ_1 et Δ_2. Le premier mineur est

Condition suffisante du MLE gaussien

- Pour conclure, il faut montrer que la hessienne est définie négative

$$\left.\frac{\partial^2 \ell_n(\theta; x)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

- Pour cela on s'intéresse aux mineurs principaux, Δ_1 et Δ_2. Le premier mineur est

$$\Delta_1 = -\frac{n}{\hat{\sigma}^2} < 0$$

- Le second mineur est

Condition suffisante du MLE gaussien

- Pour conclure, il faut montrer que la hessienne est définie négative

$$\left.\frac{\partial^2 \ell_n(\theta; x)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

- Pour cela on s'intéresse aux mineurs principaux, Δ_1 et Δ_2. Le premier mineur est

$$\Delta_1 = -\frac{n}{\hat{\sigma}^2} < 0$$

- Le second mineur est

$$\Delta_2 = \det \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix} = -\frac{n}{\hat{\sigma}^2} \times -\frac{n}{2\hat{\sigma}^4} - 0 > 0$$

- Les mineurs principaux étant de signes opposés, la hessienne est bien définie négative et la solution du programme est bien un maximum

Le score

- Le score ressemble au gradient mais en diffère pour la raison suivante :
  - Le gradient est déterministe car basé sur les réalisations :

$$\frac{\partial \ell_n(\theta; x_1, \cdots, x_n)}{\partial \theta}$$

  - Le score est une version stochastique du gradient car basé sur les variables aléatoires :

$$S_n(\theta; X) = \frac{\partial \ell_n(\theta; X_1, \cdots, X_n)}{\partial \theta}$$

- Le score étant une variable aléatoire, il convient de s'intéresser à ces moments et notamment son espérance
  - L'espérance nous intéresse afin de calculer la variance
  - La variance nous intéresse car elle permet de calculer la matrice d'information de Fisher

La hessienne stochastique

- De même que pour le gradient, on peut considérer une version stochastique de la hessienne
  - La hessienne déterministe est basée sur les réalisations :

$$H_n(\theta, x) = \frac{\partial^2 \ell_n(\theta; x_1, \cdots, x_n)}{\partial\theta\partial\theta'}$$

  - La hessienne stochastique est basés sur les variables aléatoires :

$$H_n(\theta, X) = \frac{\partial^2 \ell_n(\theta; X_1, \cdots, X_n)}{\partial\theta\partial\theta'}$$

- La hessienne stochastique étant une variable aléatoire elle a des moments :
  - l'espérance de la hessienne nous permet de calculer la matrice d'information de Fisher

L'information de Fisher

- La matrice d'information de Fisher peut se calculer de plusieurs manières

Remark

La quantité d'information de Fisher associée à l'échantillon est une constante définie par la variance du score ou l'espérance de l'opposée de la hessienne stochastique :

$$I_n(\theta) = \mathbb{V}(S_n(\theta; X)) = \mathbb{E}(S_n^2(\theta; X)) - \mathbb{E}(S_n(\theta; X))^2$$

ou

$$I_n(\theta) = \mathbb{E}(-H_n(\theta, X))$$

L'information de Fisher et MLE Gaussien

- Repartons du MLE Gaussien et calculons l'information de Fisher :

L'information de Fisher et MLE Gaussien

- Repartons du MLE Gaussien et calculons l'information de Fisher :

$$S_n(\theta; X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \hat{\theta} = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) \\ \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (X_i - \mu)^2 \end{pmatrix}$$

$$H_n(\theta; X) = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum_{i=1}^n (X_i - \mu) \\ -\frac{1}{\sigma^4} \sum_{i=1}^n (X_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (X_i - \mu)^2 \end{pmatrix}$$

- Les deux méthodes peuvent être utilisée. Par exemple :

$$\begin{aligned} I_n(\theta) &= \mathbb{E}(-H_n(\theta, X)) \\ &= \mathbb{E} \begin{pmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4} \sum_{i=1}^n (X_i - \mu) \\ \frac{1}{\sigma^4} \sum_{i=1}^n (X_i - \mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{i=1}^n (X_i - \mu)^2 \end{pmatrix} \end{aligned}$$

L'information de Fisher et MLE Gaussien

- Les quantités déterministe n'étant pas affectées par l'espérance on obtient

L'information de Fisher et MLE Gaussien

- Les quantités déterministe n'étant pas affectées par l'espérance on obtient

$$I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4}\sum_{i=1}^n \mathbb{E}(X_i-\mu) \\ \frac{1}{\sigma^4}\sum_{i=1}^n \mathbb{E}(X_i-\mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^n \mathbb{E}((X_i-\mu)^2) \end{pmatrix}$$

- Or, $\mathbb{E}(X_i) = \mu$ donc $\mathbb{E}(X_i - \mu) = 0$
- De plus, par définition, $\mathbb{E}\big((X_i-\mu)^2\big) = \sigma^2$ ce qui nous donne

$$I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

- La borne informationnelle de Cramer-Rao définissant l'efficacité du MLE Gaussien est donc :

L'information de Fisher et MLE Gaussien

- Les quantités déterministe n'étant pas affectées par l'espérance on obtient

$$I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4}\sum_{i=1}^n \mathbb{E}(X_i - \mu) \\ \frac{1}{\sigma^4}\sum_{i=1}^n \mathbb{E}(X_i - \mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^n \mathbb{E}((X_i - \mu)^2) \end{pmatrix}$$

- Or, $\mathbb{E}(X_i) = \mu$ donc $\mathbb{E}(X_i - \mu) = 0$
- De plus, par définition, $\mathbb{E}\big((X_i - \mu)^2\big) = \sigma^2$ ce qui nous donne

$$I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

- La borne informationnelle de Cramer-Rao définissant l'efficacité du MLE Gaussien est donc :

$$I_n^{-1}(\theta_0) = \begin{pmatrix} \frac{\sigma_0^2}{n} & 0 \\ 0 & \frac{2\sigma_0^4}{n} \end{pmatrix}$$

Propriétés de maximum de vraisemblance

- Commençons par poser 3 hypothèses dites de régularité
  - Hypothèse 1 : la fonction de densité $f_X(\theta; x_i)$ est trois fois différentiable par rapport à θ et ses dérivées sont continues et finies $\forall x, \theta$
  - Hypothèse 2 : les espérances des dérivées première et seconde de $\ln f_X(\theta; X_i)$ par rapport à θ existent
  - Hypothèse 3 : la vraie valeur de θ, i.e. θ_0, appartient à un ensemble compact Θ
  - Par ensemble compact il faut comprendre un ensemble fermé et petit dont on ne peut pas s'échapper

Propriétés de maximum de vraisemblance

- Sous cet ensemble d'hypothèses il est possible de montrer
  - que le MLE est convergent
  $$\hat{\theta} \xrightarrow{p} \theta_0$$
  - que le MLE est asymptotiquement efficace
  $$\mathbb{V}(\hat{\theta}) = I_n^{-1}(\theta_0)$$
  - que le MLE est asymptotiquement normalement distribué
  $$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I_n^{-1}(\theta_0))$$

Maximum de vraisemblance conditionnelle

- Soit un modèle économétrique du type $Y_t = g(\theta; X_t) + \varepsilon_t$
- Une approche par MLE nécessite de considérer la distribution conditionnelle de Y sachant les réalisations de X

$$f_{Y|X}(y|x;\theta)$$

Remark (Vraisemblance conditionnelle)

Les fonctions de vraisemblance et log-vraisemblance conditionnelle d'un échantillon $\{y_t, x_t\}_{t=1}^n$ *sont définies par*

$$L_n(\theta; y|x) = \prod_{t=1}^{n} f_{Y|X}(y_t|x_t;\theta), \quad et \quad \ell_n(\theta; y|x) = \sum_{t=1}^{n} \ln f_{Y|X}(y_t|x_t;\theta)$$

MLE et modèle de régression linéaire

- Dans le cadre simple du modèle $Y_t = \beta X_t + \varepsilon_t \sim$ i.i.d.
- En supposant la normalité des erreurs, i.e. $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, si $X_i = x_i$, on obtient que $Y_i | x_i \sim \mathcal{N}(\beta x_i, \sigma^2)$
- On obtient alors la densité conditionnelle de Y_i suivante

$$f_{Y|X}(y_t|x_t;\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y_i - \beta x_i}{2\sigma}\right)^2, \quad \theta = (\beta, \sigma^2)'$$

- Les fonctions de ML et log-ML conditionnelles sont alors

$$L_n(\theta; y|x) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y_i - \beta x_i}{2\sigma}\right)^2$$

et

$$\ell_n(\theta; y|x) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2$$

Exercice type

- Soit un échantillon $(X_1, \cdots, X_n) \sim$ i.i.d. selon une distribution exponentielle de paramètre θ^{-1}
  - La fonction de densité d'une loi exponentielle est $\theta^{-1} \exp(-\theta^{-1} X)$

- La log-vraisemblance de l'échantillon $(x_1, \cdots, x_n)$ est alors

Exercice type

- Soit un échantillon $(X_1, \cdots, X_n) \sim$ i.i.d. selon une distribution exponentielle de paramètre θ^{-1}
  - La fonction de densité d'une loi exponentielle est $\theta^{-1}\exp(-\theta^{-1}X)$
- La log-vraisemblance de l'échantillon $(x_1, \cdots, x_n)$ est alors

$$\ell_n(\theta; x) = -n\ln(\theta) - \theta^{-1}\sum_{t=1}^{n} x_t$$

- L'estimateur du log-ML est alors

Exercice type

- Soit un échantillon $(X_1, \cdots, X_n) \sim$ i. i. d. selon une distribution exponentielle de paramètre θ^{-1}
  - La fonction de densité d'une loi exponentielle est $\theta^{-1} \exp(-\theta^{-1} X)$

- La log-vraisemblance de l'échantillon $(x_1, \cdots, x_n)$ est alors

$$\ell_n(\theta; x) = -n \ln(\theta) - \theta^{-1} \sum_{t=1}^{n} x_t$$

- L'estimateur du log-ML est alors

$$\frac{\partial \ell_n(\theta; x)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{t=1}^{n} X_t = 0 \Rightarrow \hat{\theta} = n^{-1} \sum_{t=1}^{n} X_t$$

- Sachant que la loi exponentielle de paramètre λ a pour espérance λ^{-1} et pour variance λ^{-2}, $\mathbb{E}(X_t)$ et $\mathbb{V}(X_t)$ sont données par

Exercice type

- Soit un échantillon $(X_1, \cdots, X_n) \sim$ i.i.d. selon une distribution exponentielle de paramètre θ^{-1}
  - La fonction de densité d'une loi exponentielle est $\theta^{-1}\exp(-\theta^{-1}X)$

- La log-vraisemblance de l'échantillon $(x_1, \cdots, x_n)$ est alors

$$\ell_n(\theta; x) = -n\ln(\theta) - \theta^{-1}\sum_{t=1}^{n} x_t$$

- L'estimateur du log-ML est alors

$$\frac{\partial \ell_n(\theta; x)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{t=1}^{n} X_t = 0 \Rightarrow \hat{\theta} = n^{-1}\sum_{t=1}^{n} X_t$$

- Sachant que la loi exponentielle de paramètre λ a pour espérance λ^{-1} et pour variance λ^{-2}, $\mathbb{E}(X_t)$ et $\mathbb{V}(X_t)$ sont données par

$$\mathbb{E}(X_t) = \theta_0, \quad \mathbb{V}(X_t) = \theta_0^2$$

Exercice type

- Calculez $\mathbb{E}(\hat{\theta})$

Exercice type

- Calculez $\mathbb{E}(\hat{\theta})$

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}\Big(n^{-1}\sum_{t=1}^{n} X_t\Big) = n^{-1}\sum_{t=1}^{n}\mathbb{E}(X_t) = \frac{n \times \theta_0}{n} = \theta_0$$

- Calculez $\mathbb{V}(\hat{\theta})$

Exercice type

- Calculez $\mathbb{E}(\hat{\theta})$

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}\left(n^{-1}\sum_{t=1}^{n} X_t\right) = n^{-1}\sum_{t=1}^{n}\mathbb{E}(X_t) = \frac{n \times \theta_0}{n} = \theta_0$$

- Calculez $\mathbb{V}(\hat{\theta})$

$$\mathbb{V}(\hat{\theta}) = \mathbb{V}\left(n^{-1}\sum_{t=1}^{n} X_t\right) = n^{-2}\sum_{t=1}^{n}\mathbb{V}(X_t) = \frac{n \times \theta_0^2}{n^2} = \frac{\theta_0^2}{n}$$

- Que pouvez-vous conclure ?

Exercice type

- Calculez $\mathbb{E}(\hat{\theta})$

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}\Big(n^{-1}\sum_{t=1}^{n} X_t\Big) = n^{-1}\sum_{t=1}^{n}\mathbb{E}(X_t) = \frac{n \times \theta_0}{n} = \theta_0$$

- Calculez $\mathbb{V}(\hat{\theta})$

$$\mathbb{V}(\hat{\theta}) = \mathbb{V}\Big(n^{-1}\sum_{t=1}^{n} X_t\Big) = n^{-2}\sum_{t=1}^{n}\mathbb{V}(X_t) = \frac{n \times \theta_0^2}{n^2} = \frac{\theta_0^2}{n}$$

- Que pouvez-vous conclure ?
  - L'estimateur est sans biais et asymptotiquement convergent car $\lim_{n\to\infty} \mathbb{V}(\hat{\theta}) = 0$ et donc

$$\hat{\theta} \xrightarrow{p} \theta_0$$

Estimateur de Pseudo maximum de vraisemblance (QMLE)

Remark (Limite du MLE)

En cas d'erreur sur la distribution postulée, l'estimateur MLE n'a pas de fondement.

Proposition (L'idée générale des estimateurs du QMLE)

Cela consiste à démontrer que si l'on commet une erreur sur la distribution conditionnelle des résidus en utilisant à tort une log-vraisemblance fondée sur une loi normale, l'estimateur du MV ainsi obtenu peut tout de même être convergent si la vraie loi des résidus appartient à la même classe de loi que la loi normale (Gourieroux, Montfort, 1989) c'est une hypothèse paramétrique à la base du MV. Si elle est respecté / vraie ok, sinon : si elle est fausse, le MV ne veut rien dire, c'est du n'importe quoi, la log-vraisemblance ne correspond à rien

Propriétés de l'estimateur QMLE

Sous certaines conditions de régularité, il est possible de montrer

- que le QMLE est convergent

$$\hat{\theta} \xrightarrow{p} \theta_0$$

- que le QMLE est asymptotiquement normalement distribué

$$\sqrt{n}(\hat{\theta}_{QMLE} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

- Hypothèse : la 'vraie' distribution n'est pas normale, mais elle fait partie de la famille des lois exponentielles
Estimation :
- on fait une hypothèse à tort d'une distribution normale alors que la distribution n'est pas normale
- on construit la vraisemblance sous l'hypothèse de normalité (c'est une pseudo-vraisemblance car elle n'est pas construite sur la base de la vraie densité $\hat{L}(r_t;\theta)$)
- on maximize cette fonction $\hat{\theta}_{QMLE} = \arg \underset{\theta\in\Theta}{Max} \hat{L}(r_t;\theta)$
- ces estimateurs sont convergents et normalement distribués

$$\hat{\theta}_{QMLE} \xrightarrow{p} \theta_0$$

$\sqrt{T}(\theta_{QMLE} - \theta_0) \underset{T\to\infty}{\xrightarrow{L}} \mathcal{N}(0,V)$, sauf que la matrice de variance-covariance n'est pas $I(\theta_0)^{-1}$ (seule différence par rapport à MLE) - ici matrice robuste

Propriétés de l'estimateur QMLE

où la matrice de variance covariance asymptotique de l'estimateur QML est

$$V = I_n(\theta)^{-1} J_n(\theta) I_n(\theta)^{-1}$$

avec

$$I_n(\theta) = \mathbb{E}(-H_n(\theta, X)) = \mathbb{E}_0(\frac{\partial^2 \ell_n(\theta; X_1, \cdots, X_n)}{\partial\theta\partial\theta'})$$

et

$$J_n(\theta) = \mathbb{V}(S_n(\theta; X)) = \mathbb{E}_0(\frac{\partial \ell_n(\theta; X_1, \cdots, X_n)}{\partial\theta} \frac{\partial \ell_n(\theta; X_1, \cdots, X_n)}{\partial\theta'})$$

où E_0 désigne l'espérance prise par rapport à la vraie loi.

Propriétés de l'estimateur QMLE

Remark (1)

Dans la pratique les matrice $I_n(\theta)$ *et* $J_n(\theta)$ *sont directement estimées en remplaçant l'espérance* E_0 *par la moyenne empirique et le paramètre inconnu* θ *par son estimateur convergent* $\hat{\theta}_{QMLE}$

Remark (2)

Dans le cas où la vraie loi sous-jacente est normale (Maximum de Vraisemblance), la matrice de variance covariance asymptotique se réduit à

$$V(\hat{\theta}_{QMLE} - \theta_0) = I_n(\theta)^{-1}$$

puisque $I_n(\theta) = J_n(\theta)$

Optimisation numérique de la vraisemblance

On cherche

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}^+} \ell_n(\theta; x_1, \cdots, x_n)$$

La plupart du temps une solution analytique **n'est pas disponible**.

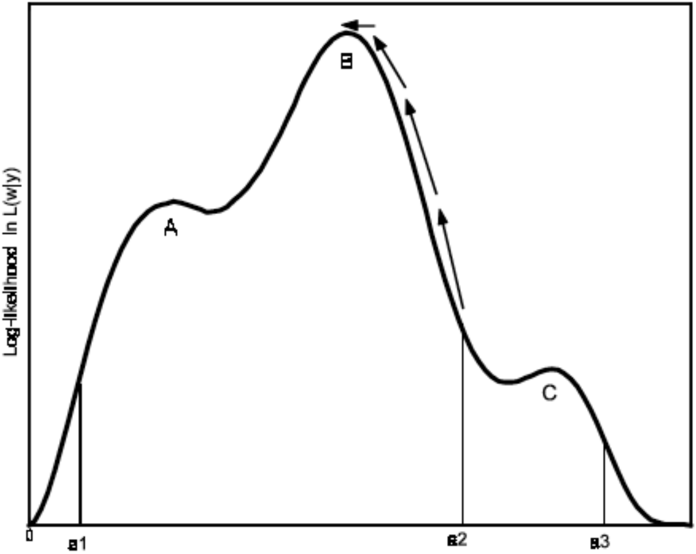
$\Rightarrow$ On utilise des algorithmes d'optimisation numérique (algorithmes itératifs) :

i) condition initiale

ii) règle de passage

iii) règle d'arrêt

Optimisation numérique de la vraisemblance



Conclusions

- Modèle linéaire = modèle de base, fondé sur la normalité
- Vaste classe de modèles non-linéaires, adaptées aux propriétés des séries économiques
- On distingue notamment la non-linéarité en moyenne et en variance
- NB : en prévision il est toujours difficile de faire mieux que le meilleur modèle linéaire !

Bibliographie générale :

Gourieroux, C., Monfort, A., 1989, Statistics and Econometric Models, Volume 2, Cambridge University Press.

Tong, H., 1990. Non-Linear Time Series : A Dynamical Systems Approach, Oxford : Oxford University Press

Teräsvirta, T., Tjøstheim, D., et Granger, C. W. J. (2010), Modelling nonlinear economic time series, Oxford University Press.