Chapitre 4: Estimation par régions de confiance

Rabah Messaci

Département de Probabilités-Statistique
USTHB

Novembre 2011

Proprètés asymptôtiques des EMV

Les estimateurs du maximum de vraisemblance n'ont pas de propriétés connus pour des échantillons de taille finie (sauf la propriété d'invariance fonctionnelle).
Cependant ils ont de trés bonnes propriètés asymptotiques.

Théorème

Soit $\widehat{\theta}_n$ l'EMV de θ basé sur un n-échantillon et $I_n(\theta)$ l'information de Fischer ramenée par les observations sur θ. (On suppose las conditions de régularité de Fischer vérifiées). On a

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\dfrac{1}{I_n(\theta)}}} \underset{n \longrightarrow +\infty}{\xrightarrow{Loi}} X$$

Proprètés asymptôtiques des EMV

Théorème

Soit g une fonction dérivable et $g(\widehat{\theta})$ l'EMV de $g(\theta)$ On a sous les hypothèses du théorème précédent

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\dfrac{g'(\theta)^2}{I_n(\theta)}}} \underset{n \longrightarrow +\infty}{\xrightarrow{Loi}} X$$

Les estimateurs du maximum de vraisemblance sont donc asymptotiquement gaussiens, asymptotiquement sans biais et asymptotiquement efficaces.

Proprètés asymptôtiques des EMV

Théorème

Soit g une fonction dérivable et $g(\widehat{\theta})$ l'EMV de $g(\theta)$ On a sous les hypothèses du théorème précédent

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\frac{g'(\theta)^2}{I_n(\theta)}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

Les estimateurs du maximum de vraisemblance sont donc asymptotiquement gaussiens, asymptotiquement sans biais et asymptotiquement efficaces. On a si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\theta)^2}{I_n(\theta)}\right)$$

Proprètés asymptôtiques des EMV

Théorème

Soit g une fonction dérivable et $g(\widehat{\theta})$ l'EMV de $g(\theta)$ On a sous les hypothèses du théorème précédent

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\dfrac{g'(\theta)^2}{I_n(\theta)}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

Les estimateurs du maximum de vraisemblance sont donc asymptotiquement gaussiens, asymptotiquement sans biais et asymptotiquement efficaces. On a si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\theta)^2}{I_n(\theta)}\right)$$

Proprètés asymptôtiques des EMV

Pour les applications le fait que la variance soit inconnue (dépend de θ) est génant. On a cependant les théorèmes suivants obtenues en remplaçant la variance par la variance estimée.

Théorème

On a sous certaines hypothèses de régularité

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\dfrac{1}{I_n(\widehat{\theta}_n)}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

Théorème

On a sous certaines hypothèses de régularité

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\dfrac{g'(\widehat{\theta}_n)^2}{I_n(\widehat{\theta}_n)}}} \xrightarrow[n \longrightarrow +\infty]{Loi} X$$

On peut alors écrire si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\widehat{\theta}_n)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\widehat{\theta}_n)^2}{I_n(\widehat{\theta}_n)}\right)$$

Introduction

Estimer poctuellement un paramètre θ, c'est lui associer une valeur $\widehat{\theta}(X_1, ..X_n)$ qui est censée l'approcher correctement. En statistique cette valeur est fonction des observations. Cette estimation se fait toujours avec une erreur. En physique, les mesures effectuées le sont toujours avec erreurs, chaque mesure m est accompagnée d'un intervalle de la forme $[m - \Delta m, m + \Delta m]$, Δm étant l'erreur absolue de mesure. En statistique on fait accompagner chaque estimation d'une région $C_\alpha(X_1, ..X_n)$ (dépendant des observations) ayant une forte probabilité de contenir la vraie valeur inconnue du paramètre.

Régions de confiance

Définition : Région de confiance

Soit $X_1, ..X_n$ un n-échantillon d'une v.a X de loi $P_\theta, \theta \in \Theta$.
On appelle région de confiance de niveau $1-\alpha$, $(0 \leq \alpha \leq 1)$ pour θ une région $C_\alpha(X_1, ..X_n) \subset \Theta$ telle que

$$P_\theta(\theta \in C_\alpha(X_1, ..X_n)) = 1-\alpha \qquad \forall \theta \in \Theta$$

Définition : Région de confiance

Soit $X_1, ..X_n$ un n-échantillon d'une v.a X de loi $P_\theta, \theta \in \Theta$.
On appelle région de confiance de niveau $1-\alpha$, $(0 \leq \alpha \leq 1)$ pour θ une région $C_\alpha(X_1, ..X_n) \subset \Theta$ telle que

$$P_\theta(\theta \in C_\alpha(X_1, ..X_n)) = 1 - \alpha \qquad \forall \theta \in \Theta$$

Notation : Pour simplifier l'écriture on notera plus simplement la région C_α au lieu de $C_\alpha(X_1, ..X_n)$

Régions de confiance

Définition : Région de confiance

Soit $X_1, ..X_n$ un n-échantillon d'une v.a X de loi $P_\theta, \theta \in \Theta$.
On appelle région de confiance de niveau $1-\alpha$, $(0 \leq \alpha \leq 1)$ pour θ une région $C_\alpha(X_1, ..X_n) \subset \Theta$ telle que

$$P_\theta(\theta \in C_\alpha(X_1, ..X_n)) = 1 - \alpha \qquad \forall \theta \in \Theta$$

Notation : Pour simplifier l'écriture on notera plus simplement la région C_α au lieu de $C_\alpha(X_1, ..X_n)$

Définition : Région de confiance asymptotique

On appelle région de confiance de niveau asymptotique $1-\alpha$, $(0 \leq \alpha \leq 1)$ une région $C_{n,\alpha} \subset \Theta$ telle que

$$\lim_{n \longrightarrow +\infty} P_\theta(\theta \in C_{n,\alpha}) = 1 - \alpha \qquad \forall \theta \in \Theta$$

Interprétation :
La région de confiance C_α est telle qu'elle contient la vraie valeur du paramètre avec une probabilité $1-\alpha$. C'est le niveau de confiance qu'on accorde à l'intervalle. Il n'y adonc qu'une probabilité α qu'elle ne la contienne pas.

Interprétation :
La région de confiance C_α est telle qu'elle contient la vraie valeur du paramètre avec une probabilité $1-\alpha$. C'est le niveau de confiance qu'on accorde à l'intervalle. Il n'y adonc qu'une probabilité α qu'elle ne la contienne pas. C'est l'erreur. Généralement les valeurs de α sont prises égales à 0.05, 0.02, 0.01 ect...Les régions de confiance sont alors généralement de niveau 95%, 98%, 99% ect..

Interprétation :
La région de confiance C_α est telle qu'elle contient la vraie valeur du paramètre avec une probabilité $1-\alpha$. C'est le niveau de confiance qu'on accorde à l'intervalle. Il n'y adonc qu'une probabilité α qu'elle ne la contienne pas. C'est l'erreur. Généralement les valeurs de α sont prises égales à 0.05, 0.02, 0.01 ect...Les régions de confiance sont alors généralement de niveau 95%, 98%, 99% ect.. Lorsque $\Theta \subset R$ on choisit souvent des intervalles : on parle alors d'intervalles de confiance I_α et ils sont tels que :

$$P_\theta(\theta \in [a(X_1,..X_n), b(X_1,..X_n)] = 1-\alpha \qquad \forall\theta \in \Theta$$

Régions de confiance

Une technique très utile pour la détermination de régions de confiance se base sur les fonctions pivotales.

Une technique très utile pour la détermination de régions de confiance se base sur les fonctions pivotales.

Définition : Fonctions pivotales

Une variable aléatoire $g(\theta, T(\widetilde{X}))$ est dite fonction pivotale pour θ si sa loi est indépendante de θ.

Régions de confiance

Une technique très utile pour la détermination de régions de confiance se base sur les fonctions pivotales.

Définition : Fonctions pivotales

Une variable aléatoire $g(\theta, T(\widetilde{X}))$ est dite fonction pivotale pour θ si sa loi est indépendante de θ.

Application à la détermination des intervalles de confiance
On veut déterminer a et b telles que $P_\theta(a \leq \theta \leq b) = 1 - \alpha$
$P_\theta(a' \leq g(\theta, T(\widetilde{X}) \leq b') = 1 - \alpha$
$P_\theta(h_1(a', T(\widetilde{X})) \leq \theta \leq h_2(b', T(\widetilde{X}))) = 1 - \alpha$
Point important : a' et b' ne dépendent pas de θ, car g fonction pivotale

Exemples : Fonctions pivotales

- $X \sim N(m, \sigma^2)$: $\dfrac{X - m}{\sigma} \sim N(0, 1)$

Régions de confiance

Exemples : Fonctions pivotales

- $X \sim N(m, \sigma^2)$: $\dfrac{X - m}{\sigma} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

Régions de confiance

Exemples : Fonctions pivotales

- $X \sim N(m, \sigma^2)$: $\dfrac{X - m}{\sigma} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \sim St_{n-1}$

Régions de confiance

Exemples : Fonctions pivotales

- $X \sim N(m, \sigma^2)$: $\dfrac{X - m}{\sigma} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \sim St_{n-1}$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\sum_{i=1}^{n} \dfrac{(X_i - m)^2}{\sigma^2} \sim \chi^2_n$

Exemples : Fonctions pivotales

- $X \sim N(m, \sigma^2)$: $\dfrac{X - m}{\sigma} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\dfrac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \sim St_{n-1}$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$: $\sum_{i=1}^{n} \dfrac{(X_i - m)^2}{\sigma^2} \sim \chi^2_n$
- $(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m, \sigma^2)$:
  $\sum_{i=1}^{n} \dfrac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$

Exemples : Fonctions pivotales

$(X_1, X_2, ..., X_n)$ échantillon de $X \sim N(m_1, \sigma_1^2)$, $(Y_1, Y_2, ..., Y_p)$ échantillon de $Y \sim N(m_2, \sigma_2^2)$

X et Y indépendantes

$$\frac{p}{n}\frac{\displaystyle\sum_{i=1}^{n}\frac{(X_i-m_1)^2}{\sigma_1^2}}{\displaystyle\sum_{i=1}^{p}\frac{(Y_i-m_2)^2}{\sigma_2^2}} \sim F_{(n,p)}$$

$$\frac{p-1}{n-1}\frac{\displaystyle\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma_1^2}}{\displaystyle\sum_{i=1}^{p}\frac{(Y_i-\overline{Y})^2}{\sigma_2^2}} \sim F_{(n-1,p-1)}$$

Tests d'hypothèses bilatères

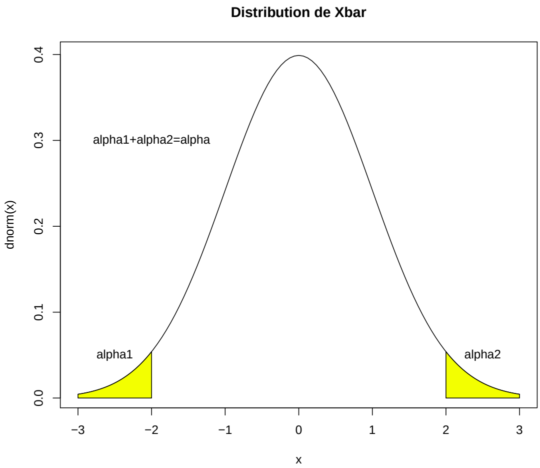

FIGURE: Région critique 3