Master 1 Informatique
Éléments de statistique inférentielle

Faicel Chamroukhi
Maître de Conférences
UTLN, LSIS UMR CNRS 7296

email: chamroukhi@univ-tln.fr
web: chamroukhi.univ-tln.fr

2014/2015

Plan I

1. Estimation de paramètres

Estimation de paramètres

- Introduction
- Critères de qualité pour les estimateurs
- Méthodes d'estimation

Estimation de paramètres

Pour étudier et caractériser un phénomène physique, naturel ou autre $\Rightarrow$ e.g., adoption d'un modèle probabiliste paramétrique représenté par une fonction de densité de probabilité $f(x;\theta)$ (ou une fonction de masse de probabilité $P(x;\theta)$ dans le cas discret).

$\Rightarrow$ L'explication de ce phénomène nécessite l'estimation de ce modèle probabiliste à partir des données que l'on a observées (l'échantillon que l'on a à notre disposition).

$\Rightarrow$ Ceci consiste donc à estimer le(s) paramètre(s) θ de ce modèle à partir des données observées $(x_1, \ldots, x_n)$ (i.i.d dans le cadre de ce cours.)

Nous considérerons d'abord le cas d'un seul paramètre θ à estimer pour plus de clarté et simplicité et notons par $f(x;\theta)$ la densité ayant θ comme vrai paramètre mais qui est inconnu et que l'on cherche à estimer.

Estimation de paramètres

Définition d'un estimateur

Le problème d'estimation de paramètres est donc celui de déterminer une fonction appropriée des données $(x_1, \ldots, x_n)$, que nous noterons $h(x_1, \ldots, x_n)$ qui donne la “meilleure" estimation de θ au sens de critères d'optimalité que nous verrons.
Nous avons donc

$$\hat{\theta} = h(x_1, \ldots, x_n)$$

et plus généralement, sous forme de variable aléatoire (car en effet pour des nouvelles réalisation des X_i, la valeur de $\hat{\theta}$ change) :

$$\hat{\Theta} = h(X_1, \ldots, X_n).$$

Cette statistique à déterminer s'appelle *un estimateur*.

Critères de qualité pour les estimateurs

- Ce sont des critères selon lesquels la qualité d'une estimation peut être évaluée.
- Ces critères définissent en général des propriétés souhaitables pour un estimateur et fournissent un guide par lequel la qualité d'un estimateur peut être comparée à celle d'un autre.

$\Rightarrow$ Notre objectif est de déterminer un estimateur $\hat{\Theta} = h(X_1, \ldots, X_n)$ de θ.

$\Rightarrow$ Des propriétés comme la moyenne, la variance ou la distribution fournissent une mesure de qualité pour cet estimateur.

Critères de qualité pour les estimateurs

- Ce sont des critères selon lesquels la qualité d'une estimation peut être évaluée.
- Ces critères définissent en général des propriétés souhaitables pour un estimateur et fournissent un guide par lequel la qualité d'un estimateur peut être comparée à celle d'un autre.

$\Rightarrow$ Notre objectif est de déterminer un estimateur $\hat{\Theta} = h(X_1, \ldots, X_n)$ de θ.

$\Rightarrow$ Des propriétés comme la moyenne, la variance ou la distribution fournissent une mesure de qualité pour cet estimateur.

Estimateur vs Estimation

Une fois nous avons observé un échantillon de valeurs $(x_1, \ldots, x_n)$, la valeur de l'estimateur $\hat{\theta} = h(x_1, \ldots, x_n)$ qui est une valeur numérique, est appelé *estimation* du paramètre θ.

Absence de biais

Définition : **Absence de biais**

Un estimateur $\hat{\Theta}$ de θ est dis *sans biais* si

$$\mathbb{E}[\hat{\Theta}] = \theta, \quad (1)$$

$\Rightarrow$ en moyenne, on espère que $\hat{\Theta}$ est égal à la valeur du vrai paramètre θ.

Absence de biais

Définition : Absence de biais

Un estimateur $\hat{\Theta}$ de θ est dis *sans biais* si

$$\mathbb{E}[\hat{\Theta}] = \theta, \tag{1}$$

$\Rightarrow$ en moyenne, on espère que $\hat{\Theta}$ est égal à la valeur du vrai paramètre θ.

⚠ Remarque : Il est naturel que, si $\hat{\Theta}$ est à qualifier comme un bon estimateur de θ , non seulement sa moyenne doit être très proche du vrai paramètre θ mais aussi il faudrait qu'il y ait une grande probabilité que toute valeur $\hat{\theta}$ soit très proche de θ.

$\Rightarrow$ Cela revient à sélectionner un estimateur de façon à ce que non seulement il soit sans biais mais aussi sa variance soit la plus petite possible.

Variance minimale

Définition **Variance minimale**

Soit $\hat{\Theta}$ un estimateur sans biais de θ. Il est dit à variance minimale pour θ si, pour tout autre estimateur sans biais $\Theta^\star$ de θ, à partir du même échantillon, on a :

$$\mathrm{var}(\hat{\Theta}) \leq \mathrm{var}(\Theta^\star)\cdot \tag{2}$$

Variance minimale

Définition **Variance minimale**

Soit $\hat{\Theta}$ un estimateur sans biais de θ. Il est dit à variance minimale pour θ si, pour tout autre estimateur sans biais $\Theta^\star$ de θ, à partir du même échantillon, on a :

$$\mathrm{var}(\hat{\Theta}) \leq \mathrm{var}(\Theta^\star)\cdot \tag{2}$$

$\Rightarrow$ Étant donné deux estimateurs sans biais pour un paramètre donné, celui ayant la variance plus faible est préférable, car une plus petite variance implique que les estimations ont tendance à être plus proche de sa moyenne qui est la valeur du vrai paramètre.

Variance minimale

Définition **Variance minimale**

Soit $\hat{\Theta}$ un estimateur sans biais de θ. Il est dit à variance minimale pour θ si, pour tout autre estimateur sans biais $\Theta^\star$ de θ, à partir du même échantillon, on a :

$$\mathrm{var}(\hat{\Theta}) \leq \mathrm{var}(\Theta^\star)\cdot \tag{2}$$

$\Rightarrow$ Étant donné deux estimateurs sans biais pour un paramètre donné, celui ayant la variance plus faible est préférable, car une plus petite variance implique que les estimations ont tendance à être plus proche de sa moyenne qui est la valeur du vrai paramètre.

$\Rightarrow$ La question qui se pose donc est, étant donné un échantillon à partir duquel on construit plusieurs estimateurs sans biais, le quel parmi tout ces estimateurs qui a la variance minimale ? $\Rightarrow$ Thèorème : Borne de Cramer-Rao

Variance minimale : Borne de Cramer-Rao

Theorem (Borne de Cramer-Rao (Cramer-Rao Lower Bound (CRLB)))

Soit $(X_1, \dots, X_n)$ *un échantillon de v.a issues d'une population de densité* $f(x;\theta)$ *où* θ *est le paramètre inconnu, et soit* $\hat{\Theta} = h(X_1, \dots, X_n)$ *un estimateur sans biais pour* θ*. La variance de* $\hat{\Theta}$ *satisfait l'inégalité suivante*

$$\mathrm{var}(\hat{\Theta}) \geq \left[n\mathbb{E}\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)^2\right]^{-1} \tag{3}$$

⚠ Remarque : si l'espérance et la dérivée existent. Un résultat analogue avec $p(X;\theta)$ en remplaçant $f(X;\theta)$ est obtenue lorsque X est discrète.

$\Rightarrow$ Cette inéquation fournit donc une borne inférieure de la variance de n'importe quel estimateur sans biais.

Variance minimale : Borne de Cramer-Rao

Information de Fisher

La quantité $n\mathbb{E}\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2$ s'appelle l'*information de Fisher* contenue dans un échantillon de taille n et se note $\mathcal{I}_n(\theta)$.

Variance minimale : Borne de Cramer-Rao

Information de Fisher

La quantité $n\mathbb{E}\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2$ s'appelle l'*information de Fisher* contenue dans un échantillon de taille n et se note $\mathcal{I}_n(\theta)$.

Borne de Cramer-Rao et Information de FIsher

La borne inférieure de Cramér-Rao alors se définie aussi par :

$$\text{var}(\hat{\Theta}) \geq \frac{1}{\mathcal{I}_n(\theta)} \tag{4}$$

et énonce donc que l'inverse de l'information de Fisher, $\mathcal{I}_n(\theta)$, d'un paramètre θ, est une borne inférieure de la variance d'un estimateur sans biais de ce paramètre.

⚠ Remarque : En anglais, la borne inférieure de Cramér-Rao s'appelle **Cramér-Rao Lower Bound** abréviée par CRLB.

Variance minimale : Borne de Cramér-Rao

Deuxième forme opérationnelle. Si le modèle est régulier, l'espérance $\left[\mathbb{E}\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right]^{-1}$ dans (3) est équivalente à $-\left[\mathbb{E}\left(\frac{\partial^2 \ln f(X;\theta)}{\partial^2\theta}\right)\right]^{-1}$.

⇒ L'inégalité de Cramér-Rao peut également être alors mise sous la forme :

CRLB : deuxième forme opérationnelle

$$\mathrm{var}(\hat{\Theta}) \geq -\left[n\mathbb{E}\left(\frac{\partial^2 \ln f(X;\theta)}{\partial^2\theta}\right)\right]^{-1}. \quad (5)$$

⇒ Cette expression alternative souvent offre des avantages de point de vue calcul.

⚠ Remarque : : ces résultats concernent le cas d'un seul paramètre θ

⇒ Le résultat peut être facilement étendu au cas de plusieurs paramètres.

Variance minimale : Borne de Cramér-Rao

Cas de plusieurs paramètres : Soit $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$ $(m \leq n)$ le vecteur des paramètres inconnus du modèle (la densité) $f(x; \theta_1, \cdots, \theta_m)$ pour lequel on cherche un estimateur $\hat{\boldsymbol{\Theta}} = (\hat{\Theta}_1, \cdots, \hat{\Theta}_m)^T$.

Variance minimale : Borne de Cramér-Rao

Cas de plusieurs paramètres : Soit $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$ $(m \leq n)$ le vecteur des paramètres inconnus du modèle (la densité) $f(x; \theta_1, \cdots, \theta_m)$ pour lequel on cherche un estimateur $\hat{\boldsymbol{\Theta}} = (\hat{\Theta}_1, \cdots, \hat{\Theta}_m)^T$.

Borne de Cramér-Rao pour un vecteur paramètre

L'inégalité de Cramér-Rao, pour le cas de paramètres multiples, est de la forme

$$\text{cov}(\hat{\boldsymbol{\Theta}}) \geq \frac{\Lambda^{-1}}{n}, \tag{6}$$

ou le terme général de la matrice d'information de Fisher Λ est donné par :

$$\Lambda_{ij} = \Lambda(\theta_i, \theta_j) = \mathbb{E}\left[\left(\frac{\partial \ln f(X; \boldsymbol{\theta})}{\partial \theta_i}\right)\left(\frac{\partial \ln f(X; \boldsymbol{\theta})}{\partial \theta_j}\right)\right], \; i, j = 1, \cdots, m. \tag{7}$$

$\Rightarrow$ On a donc remplacé l'information de Fisher par la matrice Λ qui est la *matrice d'information de Fisher*.

Variance minimale : Borne de Cramér-Rao

Transformée de la CRLB :

La CRLB peut être transformée sous une transformation du paramètre.

Supposons que, au lieu de θ, on s'intéresse à $\phi = g(\theta)$ qui est une transformation un-à-un et différentiable par rapport à θ ; alors

$$\text{CRLB pour var}(\hat{\boldsymbol{\Phi}}) = \left[\frac{\mathrm{d}\, g(\theta)}{\mathrm{d}\,\theta}\right]^2 \times \Big(\text{CRLB pour var}(\hat{\Theta})\Big) \tag{8}$$

où $\hat{\boldsymbol{\Phi}}$ est un estimateur sans biais pour ϕ.

Efficacité d'un estimateur

Définition : **Efficacité d'un estimateur**

Étant donné un estimateur sans biais $\hat{\Theta}$ de θ, le rapport de sa CRLB par sa variance est appelé l'**efficacité** de $\hat{\Theta}$

$$e(\hat{\Theta}) = \frac{\text{CRLB pour var}(\hat{\Theta})}{\text{var}(\hat{\Theta})} \tag{9}$$

$\Rightarrow$ L'efficacité d'un estimateur sans biais est ainsi inférieure ou égale à 1.

Estimateur efficace

Un estimateur sans biais ayant une efficacité égale à 1 est dit **efficace**.

Efficacité d'un estimateur

Définition : Efficacité d'un estimateur

Étant donné un estimateur sans biais $\hat{\Theta}$ de θ, le rapport de sa CRLB par sa variance est appelé l'**efficacité** de $\hat{\Theta}$

$$e(\hat{\Theta}) = \frac{\text{CRLB pour var}(\hat{\Theta})}{\text{var}(\hat{\Theta})} \tag{9}$$

$\Rightarrow$ L'efficacité d'un estimateur sans biais est ainsi inférieure ou égale à 1.

Estimateur efficace

Un estimateur sans biais ayant une efficacité égale à 1 est dit **efficace**.

On souhaite aussi, en augmentant la taille de l'échantillon, pouvoir diminuer l'erreur d'estimation $\Rightarrow$ on parle de convergence

Consistance (ou convergence) d'un estimateur

Définition : Consistance (ou convergence) d'un estimateur

Un estimateur $\hat{\Theta}$ est dit **consistant** (on dit aussi convergent) pour θ si,

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\Theta} - \theta| > \epsilon) = 0 \qquad \forall\, \epsilon > 0. \tag{10}$$

$\Rightarrow$ Convergence en probabilité

$\Rightarrow$ La probabilité de s'éloigner de la vraie valeur du paramètre de plus de ϵ tend vers 0 quand la taille de l'échantillon n augmente.

$\Rightarrow$ L'estimateur donc converge vers la valeur à estimer quand la taille de l'échantillon tends vers l'infini (asymptotiquement).

Consistance (ou convergence) d'un estimateur

Propriété

Un estimateur sans biais et de variance asymptotiquement nulle est convergent.
Soit $\hat{\Theta}$ un estimateur pour θ sur un échantillon de taille n. Alors, si

$$\lim_{n\to\infty} \mathbb{E}[\hat{\Theta}] = \theta, \quad \text{et} \quad \lim_{n\to\infty} \mathrm{var}[\hat{\Theta}] = 0, \tag{11}$$

l'estimateur $\hat{\Theta}$ est dit *consistant* pour θ.

Suffisance d'un estimateur

Statistique suffisante (exhaustive)

Soit X un vecteur $i.i.d.$ de taille n. Soit θ un paramètre de la loi de probabilité des X_i. Une statistique $T(X)$ est dite exhaustive pour le paramètre θ (on dit aussi suffisante) si la probabilité conditionnelle d'observer X sachant $T(X)$ est indépendante de θ :

$$\mathbb{P}(X = x | T(X) = s, \theta) = \mathbb{P}(X = x | T(X) = s), \tag{12}$$

En pratique on se sert peu de cette formule pour montrer qu'une statistique est exhaustive et on préfère utiliser le critère de factorisation suivant (appelé critère de Fisher-Neyman) :

Statistique suffisante (exhaustive) et critère de Fisher-Neyman

Soit $f_\theta(x)$ la densité de probabilité du vecteur aléatoire X. Une statistique S est exhaustive si et seulement s'il existe deux fonctions u et v telles que :

$$f_\theta(x) = u(x)\, v(\theta, T(x))$$

Suffisance d'un estimateur

Définition : **Estimateur suffisant**

Soit $(X_1, X_2, \cdots, X_n)$ un échantillon *i.i.d.* de X de distribution à paramètre θ. Si $Y = T(X_1, X_2, \cdots, X_n)$ est une statistique telle que, pour toute autre statistique $Z = u(X_1, X_2, \cdots, X_n)$, la distribution conditionnelle de Z, étant donné $Y = y$, ne dépend pas de θ, ç.à.d

$$\mathbb{P}(Z = z | Y = y, \theta) = \mathbb{P}(Z = z | Y = y)$$

alors Y est appelée une **statistique exhaustive (suffisante)** pour θ. Si l'on a également $\mathbb{E}[Y] = \theta$, alors Y est dit un **estimateur suffisant** pour θ.

$\Rightarrow$ la définition de la suffisance dit que, si Y est une statistique suffisante pour θ, toute l'information de l'échantillon concernant θ est contenue dans Y.

Suffisance et critère de factorisation de Fisher-Neyman

Si une statistique suffisante pour un paramètre θ existe, le théorème 19 suivant, fournit un moyen de la trouver.

Critère de factorisation de Fisher-Neyman

Soit $Y = T(X_1, \ldots, X_n)$ une statistique basée sur un échantillon i.i.d de taille n. Alors Y est une statistique exhaustive pour θ si et seulement si la densité jointe des X_i peut être factorisée selon la forme :

$$f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = u\left(T(x_1, \ldots, x_n), \theta\right) v(x_1, \ldots, x_n). \quad (13)$$

Dans le cas discret on a :

$$P(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} P(x_i; \theta) = v\left(T(x_1, \ldots, x_n), \theta\right) v(x_1, \ldots, x_n). \quad (14)$$

Suffisance et critère de factorisation de Fisher-Neyman

Le résultat précédent peut être étendu au cas de paramètres multiples.

Soit $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$, $m \leq n$ le vecteur paramètre. Alors $Y_1 = T(X_1, \ldots, X_n), \ldots, Y_r = T(X_1, \ldots, X_n)$, $r \geq m$ est un ensemble de statistique suffisantes pour $\boldsymbol{\theta}$ si et seulement si

$$f(x_1, \ldots, x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) = u\,(T(x_1, \ldots, x_n), \boldsymbol{\theta})\, v(x_1, \ldots, x_n). \quad (15)$$

avec $T = (T_1, \ldots, T_r)^T$.

L'expression dans le cas discret est similaire.

Méthodes d'estimation I

Il existe plusieurs méthodes d'estimation de paramètres, notamment **estimation ponctuelle** comme la méthode des moments, la méthode du maximum de vraisemblance, la méthode du maximum a posteriori ou la méthode d'**estimation par intervalle**.

1. L'estimation d'un paramètre quelconque θ est *ponctuelle* si l'on associe une seule valeur à l'estimateur à partir des données observées sur un échantillon aléatoire.
2. L'estimation *par intervalle* associe quant à elle à un échantillon aléatoire, un intervalle $[\hat{\theta}_a \,,\, \hat{\theta}_b]$ qui recouvre θ avec une certaine probabilité.

Dans ce cours, nous verrons en particulier les méthodes du maximum de vraisemblance (et moindres carées) (estimation ponctuelle) pour leur usage très commun en estimation de modèles probabilistes et l'estimation par intervalle de confiance