Feuille 2 : Estimateur du maximum de vraisemblance

Allan Merino

Master 1 Mathématiques, UE Statistiques, Février 2016

Quelques rappels ne font jamais de mal ...

Le but de cette feuille est de familiariser avec la méthode dite du maximum de vraisemblance. Rappelons-en les grandes lignes. Pour cette méthode, on doit être dans un modèle statistique dominé.

Soit X une variable aléatoire réelle (discrète ou continue) dont on veut estimer un paramètre θ. On appelle vraisemblance de θ au vu des observations $(x_1, \ldots, x_n)$ d'un n-échantillon indépendamment distribué le nombre :

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta),$$

avec $f(x, \theta) = \begin{cases} f_\theta(x) & \text{si } X \text{ est une variable aléatoire continue} \\ \mathbb{P}_\theta(X = x) & \text{si } X \text{ est une variable aléatoire discrète} \end{cases}$.

On va alors chercher à maximiser la vraisemblance. Le théorème suivant nous sera très utile : soit g une fonction bijective de Θ sur Λ un ouvert de $\mathbb{R}^p$. Alors, $\hat{\theta}$ est un estimateur du maximum de vraisemblance de θ si et seulement si $g \circ \theta$ est un estimateur du maximum de vraisemblance de $\lambda = g(\theta)$. Ce théorème sera souvent utilisé pour une fonction particulière : le logarithme népérien.

Exercice 1

On considère le modèle normal n-échantillonné $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathcal{N}(\mu, \sigma^2)^{\otimes n}, (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+)$.

1. Estimation de $\mu = g(\mu, \sigma^2)$.
   (a) Trouver l'estimateur de μ par la méthode des moindres carrés (on le notera $\tilde{\mu}$).
   (b) Trouver l'estimateur de μ par la méthode des moments (on le notera $\mu^\star$).
   (c) Trouver l'estimateur de μ par la méthode du maximum de vraisemblance (on le notera $\hat{\mu}$).
2. Estimation de σ^2 ou σ (dans le cas où μ est connu).
   (a) Trouver l'estimateur de σ^2 par la méthode des moments.
   (b) Trouver l'estimateur de σ par la méthode des moments.
   (c) Trouver l'estimateur de σ^2 par la méthode du maximum de vraisemblance.
   (d) Trouver l'estimateur de σ par la méthode du maximum de vraisemblance.

3. Trouver l'estimateur de (μ, σ^2) par la méthode du maximum de vraisemblance.

Exercice 2 - Loi de Pareto

La loi aide Pareto a été introduite pour modéliser la distribution des revenus supérieurs à un seuil donné, puis s'est avérée utile pour d'autres phénomènes (par exemple la distribution de la taille de grains de sables passés au travers d'un tamis). Elle a deux paramètres strictement positifs : le paramètre de seuil a et un paramètre de forme θ. La fonction de répartition de cette loi est donnée par :

$$F(x) = \begin{cases} 1 - \left(\frac{a}{x}\right)^\theta & \text{si } x \geqslant a \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer la densité d'une variable aléatoire X suivant une loi de Pareto.
2. Soit X une variable aléatoire suivant une loi de Pareto.
   (a) Montrer que $E(X) = \frac{a\theta}{\theta-1}$,
   (b) Montrer que $V(X) = \frac{a^2\theta}{(\theta-1)^2(\theta-2)}$,
   (c) Commenter ces résultats.
3. Montrer que si X suit une loi de Pareto de paramètre θ avec seuil a connu, alors, $log(\frac{X}{a})$ suit une loi exponentielle $\Gamma(1, \theta)$.
4. On se place maintenant dans le modèle de loi de Pareto de paramètre θ avec seuil a connu n-échantillonné.
   (a) Ecrire formellement ce modèle.
   (b) Trouver l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ dans ce modèle.
   (c) Calculer $E(\hat{\theta})$ puis le biais de $\hat{\theta}$. Que conclure ?
   (d) En déduire un estimateur sans biais de θ (on le notera $\tilde{\theta}$).
   (e) Montrer que $\tilde{\theta}$ (et également $\hat{\theta}$) est asymptotiquement efficace.

Exercice 3

1. On se place dans un modèle de Poisson n-échantillonné. Trouver l'estimateur du maximum de vraisemblance de λ, noté $\hat{\lambda}_{MV}$.
2. Dans le même genre d'idée, déterminer les estimateurs de θ par la méthode du maximum de vraisemblance dans les cas suivants :
   (a) $(\{0,1\}^n, \mathscr{P}(\{0,1\}^n), \mathscr{B}(p)^{\otimes n}, p \in]0,1[)$ avec $p = -ln(\theta)$,
   (b) $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathscr{U}([0,\theta])^{\otimes n}, \theta \in \mathbb{R}_+^*)$,
   (c) $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathscr{U}([\theta,\theta+1])^{\otimes n}, \theta \in \mathbb{R})$,
   (d) $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathscr{E}(\theta)^{\otimes n}, \theta \in \mathbb{R}_+^*)$.

Exercice 4

Soit X une variable aléatoire réelle de densité

$$f(x) = \begin{cases} \frac{1}{\theta} x^{-\frac{\theta+1}{\theta}} & \text{si } x \geqslant 1 \\ 0 & \text{sinon} \end{cases}$$

où $\theta > 0$ est un paramètre réel inconnu. Pour $n \in \mathbb{N}^*$, on notera $(X_1, \ldots, X_n)$ un échantillon indépendants et identiquement distribués de X.

1. (a) Calculer l'espérance $E(X)$ de X lorsqu'elle existe.
   (b) On note $Y = log(X)$. Démontrer que $E(Y) = \theta$ et $E(Y^2) = 2\theta^2$.
2. (a) Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ.
   (b) Cet estimateur est-il sans biais ? Est-il convergent ?
   (c) Justifier que, pour n assez grand, on peut approcher la loi de probabilité de $\sqrt{n}\frac{\hat{\theta}-\theta}{\theta}$ par une loi normale centrée réduite $\mathcal{N}(0, 1)$. On supposera que cette approximation est satisfaisante dès que $n \geqslant 30$.
3. (a) On souhaite estimer θ par un estimateur T_n combinaison linéaire des $Y_i = log(X_i)$:

$$T_n = \sum_{i=1}^{n} \alpha_i log(X_i).$$

   (b) Quelle condition doivent vérifier les coefficients α_i pour que T_n soit non biaisé ?
   (c) On note $\mathrm{T}_\mathrm{n}^* = \frac{1}{n} \sum_{i=1}^{n} log(X_i)$ obtenu avec le choix particulier $\alpha_1 = \ldots = \alpha_n = \frac{1}{n}$. Démontrer que parmi les estimateurs T_n non biaisés, l'estimateur T_n^* est à variance minimale.
   (d) Comparer T_n^* à l'estimateur du maximum de vraisemblance $\hat{\theta}$.

Exercice 5

Soit X une variable aléatoire suivant une loi exponentielle symétrique décalée de paramètre λ, de fonction de densité de probabilité $f(x) = \frac{\lambda}{2} e^{-\lambda|x-a|}$, $\lambda > 0$, $x \in \mathbb{R}$.

1. Déterminer $E(X)$.
2. On prélève un échantillon $X_1, \ldots, X_n$ de variables aléatoires mutuellement indépendantes et suivant toutes la loi exponentielle symétrique décalée. Donner l'estimateur classique du paramètre a.
3. En utilisant l'estimation du paramètre a définie à la question précédente, déterminer l'estimateur du paramètre λ par la méthode du principe du maximum de vraisemblance.

Exercice 6

La hauteur maximale H de la crue annuelle d'un fleuve est observée avec attention, car une crue supérieure à 6 mètres est catastrophique. On a modélisé la distribution de la variable aléatoire H par une loi de Rayleigh, dont la densité de probabilité est donnée par :

$$f(x) = \frac{x}{a} e^{-\frac{x^2}{a}} \text{ si } x > 0 \qquad f(x) = 0 \text{ si } x \leqslant 0.$$

où a est un paramètre inconnu. Durant une période de 8 ans, on a observé les hauteurs (supposées être indépendantes) de crues suivantes en mètres :

2,1	2,8	1,7	0,9	1,8	2,5	2,2	2,9

1. Donner l'estimateur $\hat{\theta}$ du principe du maximum de vraisemblance du paramètre a.
2. Déterminer la fonction de répartition de la variable H.
3. Calculer la probabilité d'avoir une catastrophe (c'est-à-dire une crue de hauteur supérieure à 6 mètres) durant une année et calculer la probabilité de n'avoir aucune catastrophe en cent ans, en supposant l'indépendance entre les années.

Exercice 7

Soit la famille de densités :

$$f(x, \rho) = \rho(\rho + 1)x(1 - x)^{\rho-1} \qquad x \in [0, 1],\ \rho > 0.$$

Donner l'estimateur du maximum de vraisemblance pour ρ.