Estimation d'un paramètre

Objectif : Soient X un caractère, θ un paramètre inconnu de X et $x_1, \ldots, x_n$ des données (observations de X sur n individus d'une population). L'objectif est d'évaluer θ à l'aide des données $x_1, \ldots, x_n$.

Modélisation : On modélise X par une *var* X et on considère n *var* $X_1, \ldots, X_n$ *iid* suivant la même loi que celle de X. On appelle n-échantillon de X le vecteur $(X_1, \ldots, X_n)$.

Estimateur : On appelle estimateur (aléatoire) $\widehat{\theta}_n$ de θ toute fonction de $X_1, \ldots, X_n$. Un estimateur ponctuelle θ^* de θ s'obtient en remplaçant $X_1, \ldots, X_n$ par $x_1, \ldots, x_n$ dans la définition de $\widehat{\theta}_n$.

Biais : On appelle biais de $\widehat{\theta}_n$ le réel

$$B(\widehat{\theta}_n, \theta) = \mathbb{E}(\widehat{\theta}_n) - \theta.$$

On dit que $\widehat{\theta}_n$ est sans biais (de θ) si et seulement si $B(\widehat{\theta}_n, \theta) = 0$ (donc $\mathbb{E}(\widehat{\theta}_n) = \theta$).

Risque quadratique : On appelle risque quadratique de $\widehat{\theta}_n$ le réel

$$R(\widehat{\theta}_n, \theta) = \mathbb{E}((\widehat{\theta}_n - \theta)^2).$$

Plus $R(\widehat{\theta}_n, \theta)$ est petit, plus $\widehat{\theta}_n$ estime bien θ. Si $\widehat{\theta}_n$ est sans biais, alors on a $R(\widehat{\theta}_n, \theta) = \mathbb{V}(\widehat{\theta}_n)$.

Estimateur consistant : On dit que $\widehat{\theta}_n$ est consistant si et seulement si $(\widehat{\theta}_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers θ.

Estimateur asymptotiquement normal : On dit que $\widehat{\theta}_n$ est asymptotiquement normal si, pour n très grand, la loi de $\widehat{\theta}_n$ est assimilable à une loi normale. C'est une propriété utile pour construire des intervalles de confiance ou des tests statistiques pour θ.

Méthode des moments (cas simple) : Un estimateur des moments $\widehat{\theta}_n$ de θ s'obtient de la manière suivante. On détermine une fonction g telle que $\mathbb{E}(X) = g(\theta)$ et on exprime θ en fonction de $\mathbb{E}(X)$: $\theta = g^{-1}(\mathbb{E}(X))$. Un estimateur $\widehat{\theta}_n$ de θ est donné par

$$\widehat{\theta}_n = g^{-1}(\overline{X}_n), \qquad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Un estimateur ponctuel θ^* de θ est donné par $\theta^* = g^{-1}(\overline{x})$ avec $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Un autre estimateur des moments peut être obtenu avec les *var* à une puissance m au choix.

Avantage de la méthode des moments : Sous certaines hypothèses, on peut montrer que l'estimateur des moments est consistant et asymptotiquement normal.

Fonction de vraisemblance : On appelle fonction de vraisemblance pour $(x_1, \ldots, x_n)$ la fonction de θ :

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

avec $f(x;\theta) = \begin{cases} \mathbb{P}_\theta(X = x), & \text{si } X \text{ est une } var \text{ discrète,} \\ f_\theta(x), & \text{si } X \text{ est une } var \text{ à densité de densité } f_\theta. \end{cases}$

Estimateurs du maximum de vraisemblance : On appelle estimateur du maximum de vraisemblance *(emv)* de θ pour $(x_1, \ldots, x_n)$ un réel θ^* qui maximise la fonction de vraisemblance $L_n(x_1, \ldots, x_n; \theta)$ en θ, *i.e.* pour tout θ,

$$L_n(x_1, \ldots, x_n; \theta) \le L_n(x_1, \ldots, x_n; \theta^*).$$

Une expression alternative est : $\theta^* \in \operatorname{argmax}_\theta L_n(x_1, \ldots, x_n; \theta)$ où argmax désigne l'argument du maximum qui est l'ensemble des points en lesquels une expression atteint sa valeur maximale. Ainsi, θ^* est une estimation ponctuelle de θ.

Fonction de Log-vraisemblance : On appelle fonction de log-vraisemblance pour $(x_1, \ldots, x_n)$ la fonction de θ définie par :

$$\ell_n(x_1, \ldots, x_n; \theta) = \ln(L_n(x_1, \ldots, x_n; \theta)).$$

La fonction logarithme népérien étant croissante, l'*emv* θ^* de θ pour $(x_1, \ldots, x_n)$ vérifie : $\theta^* \in \operatorname{argmax}_\theta L_n(x_1, \ldots, x_n; \theta) = \operatorname{argmax}_\theta \ell_n(x_1, \ldots, x_n; \theta)$.

Expression analytique de l'*emv* : Pour envisager d'avoir une expression analytique de l'*emv* θ^* de θ pour $(x_1, \ldots, x_n)$, une idée est d'exprimer $L_n(x_1, \ldots, x_n; \theta)$ en fonction de produits de termes exponentiels/puissances, puis de considérer la fonction de log-vraisemblance $\ell_n(x_1, \ldots, x_n; \theta)$. Si cette dernière est dérivable en θ, une condition nécessaire que doit vérifier θ^* est d'être solution de l'équation de vraisemblance donnée par :

$$\frac{\partial}{\partial \theta} \ell_n(x_1, \ldots, x_n; \theta) = 0.$$

Il faut ensuite vérifier que θ^* est bien un maximum pour $\ell_n(x_1, \ldots, x_n; \theta)$, soit en étudiant les variations de $\ell_n(x_1, \ldots, x_n; \theta)$, soit en montrant que $\dfrac{\partial^2}{\partial \theta^2} \ell_n(x_1, \ldots, x_n; \theta^*) < 0$.

Avantage de la méthode du maximum de vraisemblance : Un estimateur $\widehat{\theta}_n$ de θ par la méthode du maximum de vraisemblance s'obtient en remplaçant $x_1, \ldots, x_n$ par $X_1, \ldots, X_n$ dans l'expression de θ^*. Sous certaines hypothèses, on peut montrer que cet estimateur est consistant et asymptotiquement normal. De plus, il est asymptotiquement efficace : quand n est très grand, il a le plus petit risque quadratique de tous les estimateurs de θ.

Pratique de l'*emv* : Bien souvent, la résolution de l'équation de vraisemblance n'amène pas une expression analytique pour θ^* de θ pour $(x_1, \ldots, x_n)$. En tant que problème d'optimisation, on peut quand même approcher la valeur de θ^* avec précision à l'aide d'algorithmes itératifs efficaces comme l'algorithme de Newton-Raphson ou l'algorithme de Gauss-Newton.