

1 Estimation non-paramétrique

Moments à savoir

$$\begin{aligned}\mu'_k &= E[X^k] \\ \mu_k &= E[(X - \mu)^k] \\ CV &= \frac{\sigma}{E[X]} \\ \gamma_1 &= \frac{\mu_3}{\sigma^3} \\ \gamma_2 &= \frac{\mu_4}{\sigma^4}\end{aligned}$$

Fonction empirique

$$\begin{aligned}F_n(x) &= \frac{1}{n} \sum_{j=1}^n I_{\{x_j \leq x\}} \\ f_n(x) &= \frac{1}{n} \sum_{j=1}^n I_{\{x_j = x\}} \\ nF_n(x) &\sim \text{bin}(n, F(x)) \\ E[F_n(x)] &= \frac{nF_n(x)}{n} = F_n(x) \\ \widehat{Var}[F_n(x)] &= \frac{nF_n(x)(1 - F_n(x))}{n^2} \\ &= \frac{F_n(x)(1 - F_n(x))}{n} \\ \widehat{Var}[S_n(x)] &= \frac{S_n(x)(1 - S_n(x))}{n} \\ F_n(x) &= \begin{cases} 0, & x < y_1 \\ 1 - \frac{r_j}{n}, & y_{j-1} \leq x < y_j, j = 2, \dots, k \\ 1, & x > y_k \end{cases}\end{aligned}$$

Estimateur de Nelson-Aalen

Estimateur de Kaplan-Meier

2 Frequentist estimation

Méthode des moments

On résoud p équations à p inconnus, telles que

$$\hat{\mu}'_k = \mu'_k$$

Méthode des percentiles

On résoud p équations à p inconnus (paramètres) telles que

$$F_n(\hat{\pi}_{g_i}) = g_i \quad i = 1, \dots, p$$

où $\hat{\pi}_{g_i}$ est le g_i^e quantile de la fonction empirique.

Smoothed empirical estimate

Parfois, le quantile recherché tombe entre 2 marches de la fonction empirique. On utilise l'approximation linéaire suivante avec les statistiques d'ordre $X_{(j)}$:

$$\hat{\pi}_g = (1 - h)X_{(j)} + hX_{(j+1)}$$

avec $j = \lfloor (n+1)g \rfloor$ et $h = (n+1)g - j$.

Méthode du maximum de vraisemblance (MLE)

Données complètes

On définit la fonction de vraisemblance $L(\theta)$ telle que

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

Et la fonction de log-vraisemblance $\ell(\theta)$

$$\ell(\theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

L'estimateur du maximum de vraisemblance θ maximiser $L(\theta)$ ou $\ell(\theta)$, i.e.

$$\left. \frac{\partial}{\partial \theta} \ell(\theta) \right|_{\theta = \hat{\theta}_{MLE}} = 0$$

Données groupées

Si les données sont groupées, alors on utilise une forme plus générale de la fonction de vraisemblance :

$$L(\theta) = \prod_{j=1}^k \left(F_X(c_j; \theta) - F_X(c_{j-1}; \theta) \right)^{n_j}$$

où F_X est la fonction de répartition théorique de la distribution qu'on suppose la distribution de notre estimateur MLE . Si les données sont censurées à la classe c_{j-1} , alors on utilise $(1 - F_X(c_{j-1}; \theta))$.

Variance des estimateurs et intervalle de confiance

Estimation de la variance de $\hat{\theta}$

L'information de Fisher $I(\theta)$ est définie par

$$I(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = E \left[\left(\frac{\partial}{\partial \theta} \ell(\theta) \right)^2 \right]$$

Si l'information n'est pas connue, on peut l'estimer avec l'information observée :

$$\hat{I}(\hat{\theta}) = \sum_{i=1}^n \left(\left. \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right|_{\theta = \hat{\theta}} \right)^2 = - \sum_{i=1}^n \left. \frac{\partial^2}{\partial \theta^2} \ln f(x_i; \theta) \right|_{\theta = \hat{\theta}}$$

Ainsi, on peut calculer la variance de l'estimateur $\hat{\theta}_{MLE}$ telle que

$$\text{Var}(\hat{\theta}) = I(\theta)^{-1}$$

Intervalle de confiance pour $\hat{\theta}$

Lorsque $n \rightarrow \infty$, $\hat{\theta} \sim N(\theta, \text{Var}(\hat{\theta}))$. Alors, on peut trouver un IC pour l'estimateur au seuil $1 - \alpha$:

$$\theta \in \left[\hat{\theta} \pm z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta})} \right]$$

Méthode delta pour estimer la variance d'une transformation de $\hat{\theta}$

Lorsqu'on veut calculer la variance d'une autre quantité que le paramètre $\hat{\theta}$ lui-même, on peut utiliser la méthode Delta :

$$\text{Var}(h(\hat{\theta})) = \left(\frac{\partial}{\partial \theta} h(\theta) \right)^2 \text{Var}(\hat{\theta})$$

Dans un contexte multivarié, où $\hat{\theta}$ est un vecteur d'estimateurs, alors on a

$$\text{Var}(h(\hat{\theta})) = h^\top I(\theta)^{-1} h$$

où h est le vecteur des dérivées partielles de $h(\theta)$:

$$h = \begin{bmatrix} \frac{\partial}{\partial \theta_1} h(\theta) \\ \frac{\partial}{\partial \theta_2} h(\theta) \\ \dots \\ \frac{\partial}{\partial \theta_k} h(\theta) \end{bmatrix}$$

Test du rapport de vraisemblance (LRT)

On veut tester si le modèle réduit avec θ_0 , qui est une bonne simplification de θ_1 , le modèle complet. Alors, on teste si la différence dans les log-vraisemblance est significative :

$$T = 2(\ell(\theta) - \ell(\theta_0)) \sim \chi_{dl_1 - dl_0, 1-\alpha}^2$$

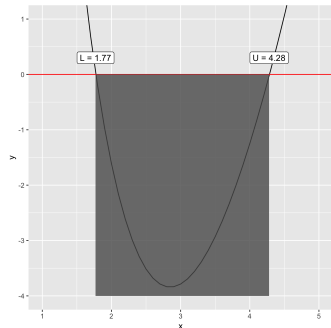
où dl_1 est le nombre de paramètres non-fixés du modèle complet et dl_0 le nombre de paramètres non-fixés du modèle réduit. **On va rejeter H_0 si $T > \chi_{dl_1 - dl_0, 1-\alpha}^2$ (test unilatéral)**, en concluant que le modèle réduit n'est pas une bonne simplification du modèle de l'hypothèse alternative.

Construction d'un intervalle de confiance par inversion du LRT

Si θ_0 est un paramètre adéquat pour le modèle réduit, alors la statistique T du LRT ne dépassera pas le quantile théorique $\chi_{dl_1 - \alpha^2}^2$. Alors, on veut trouver $\hat{\theta}_0$ tel que

$$2(\ell(\theta) - \ell(\theta_0)) \leq \chi_{dl_1 - \alpha^2}^2$$

On trouvera une équation du genre $g(\theta) \leq 0$, où g sera une fonction avec deux racines définies, qui correspondent aux bornes de l'intervalle de confiance pour les valeurs de $\hat{\theta}_0$:



3 Sélection de modèles

Chi-Square Goodness-of-fit

On veut valider l'adéquation du modèle qu'on propose avec ce test. On calcule la quantité X^2 :

$$X^2 = \frac{\sum_{j=1}^k (E_j - O_j)^2}{E_j}$$

où $E_j = n\hat{p}_i$ est le nombre de valeurs qu'on s'attend à avoir dans la i^e classe et $O_j = np_{ni}$ le nombre d'observations dans la i^e classe. On peut prouver que

$$X^2 \sim \chi_{k-p-1}^2$$

On peut aussi faire le test LRT pour valider l'adéquation aussi.

Critères de sélection

Pour choisir entre plusieurs modèles, on peut, entre autres, se baser sur les critères suivants :

1. la plus faible valeur pour le test Kolmogorov-Smirnov;
2. la plus faible valeur pour le test Anderson-Darling;
3. la plus faible valeur pour le test Goodness-of-fit;
4. la plus haute valeur pour la p -value du test Goodness-of-fit;
5. la plus haute valeur pour la fonction de vraisemblance à son maximum.

4 Estimation bayésienne

Distribution *a posteriori*



La distribution *a posteriori* nous permet de savoir avec quelle probabilité non-nulle notre paramètre θ peut prendre une certaine valeur, sachant qu'on a observé certains x , qu'on dénote comme $\pi_{\Theta|X}(\theta|x)$:

$$\pi_{\Theta|X}(\theta|x) = \frac{f_{\Theta, X}(\theta, x)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta)\pi(\theta)}{\int f_{X|\Theta}(x|\theta)\pi(\theta)d\theta} \quad (1)$$

L'idée est de remplacer les différentes distributions dans l'Équation 1, et en déduire une distribution avec une paramétrisation différente^a.

^a. Souvent, la distribution *a posteriori* aura la même distribution que celle *a priori*, mais avec des paramètres différents.

L'estimateur Bayésien L'estimateur Bayésien est défini comme l'espérance du paramètre θ , sachant la distribution de X . En d'autres mots, on veut l'espérance de la distribution *a posteriori* :

$$\hat{\theta}_{BAYES} = E[\Theta|X] \quad (2)$$

Distribution *a priori*



Soit un paramètre θ d'une distribution quelconque. Afin de réaliser une estimation Bayésienne, on connaît *a priori* la distribution que prend le paramètre θ , qu'on dénote par $\pi(\theta)$.

Alors, notre distribution des pertes est conditionnée par rapport à la valeur que θ prend (i.e. $f_{X|\Theta}$).

5 Rappel de probabilité

Certaines lois à savoir

Loi	$\Pr(X = x)$ ou $f_X(x)$	$E[X]$	$Var(X)$	$M_X(t)$
$Bin(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	$((1-p) + p^t)^n$
$Pois(\lambda)$	$\frac{e^{-\lambda} \lambda^x}{x!}$	λ	λ	$e^{\lambda(t-1)}$
$Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^\alpha$
$Normale(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	$e^{\mu t + \frac{\sigma^2 t^2}{2}}$

Rappels d'algèbre linéaire

5.0.1 Matrice transposée

la matrice transposée est définie par A^\top , telle que

$$A^\top = \begin{bmatrix} a & -c \\ -b & d \end{bmatrix}$$

5.0.2 Déterminant d'une matrice

On peut calculer le déterminant $\det(A)$ de la matrice A tel que

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

5.0.3 Inverse d'une matrice

L'équivalent de l'opération $\frac{1}{A}$ en algèbre linéaire est de calculer la matrice inverse de A^{-1} , telle que

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a & -c \\ -b & d \end{bmatrix}$$

où on multiplie par la matrice adjointe de A . Il faut normalement calculer les cofacteurs, mais le cas à 2 dimensions est un cas simplifié.