

- 3.1 On considère trois modèles de régression emboîtés afin de modéliser le nombre d'accidents de voiture selon la région (region). La variable catégorielle niveaux de risque (risque) a 3 niveaux et le nombre d'années d'expérience (expe) au volant est un facteur à 4 niveaux.

Modèle	variables	$p+1$	$\ell(\hat{\beta})$	AIC	BIC
M ₁	risque	3	-244.566	495.132	510.362
M ₂	risque + region	★	-151.620	★	★
M ₃	risque + region + expe	10	-139.734	299.468	350.235

Table 1: Mesures d'adéquation de trois modèles emboîtés avec les covariables, le nombre de coefficients du modèle ($p+1$), la valeur de la log-vraisemblance évaluée au maximum de vraisemblance ($\ell(\hat{\beta})$) et les critères d'information.

Quelle est la différence entre le AIC et le BIC du modèle M₂ (en valeur absolue)?

Solution

La différence est approximativement de 35.54. Les seules informations requises ici sont le nombre de paramètres du modèle M₂ et la taille de l'échantillon. On peut utiliser la première ligne pour calculer cette dernière,

$$\text{BIC} = -2\ell(M_1) + 3\ln(n)$$

et on trouve en arrondissant $n = 1184$. Le nombre de paramètres dépend du nombre de niveaux de la variable région. On trouve $\text{DF}(M_3) - \text{DF}(M_1) - 3 = 7$; il y a $K = 4$ catégories pour les années d'expérience, mais seulement trois paramètres additionnels dans le modèle. La différence $\text{BIC} - \text{AIC} = 7 \cdot \{\ln(1184) - 2\}$.

- 3.2 Une variable aléatoire X suit une loi géométrique de paramètre p si sa fonction de masse est

$$P(X = x) = (1 - p)^{x-1} p, \quad x = 1, 2, \dots$$

- Écrivez la vraisemblance et la log-vraisemblance d'un échantillon aléatoire de taille n si les observations sont indépendantes.
- Dérivez l'estimateur du maximum de vraisemblance pour le paramètre p .
- Calculez l'information observée.
- Supposons qu'on a un échantillon de 15 observations, $\{5, 6, 3, 7, 1, 2, 11, 8, 7, 34, 1, 7, 10, 1, 0\}$, dont la somme est 103. Calculez l'estimé du maximum de vraisemblance et son erreur-type approximative.
- Calculez la statistique du rapport de vraisemblance et la statistique de Wald pour un test à niveau 5% de l'hypothèse $\mathcal{H}_0 : p_0 = 0.1$ contre l'alternative bilatérale $\mathcal{H}_a : p_0 \neq 0.1$.

Solution

(a)

$$L(p; \mathbf{x}) = \prod_{i=1}^n (1 - p)^{x_i - 1} p = (1 - p)^{\sum_{i=1}^n (x_i - 1)} p^n$$

$$\ell(p; \mathbf{x}) = \ln(1 - p) \sum_{i=1}^n (x_i - 1) + n \ln(p)$$

(b)

$$\frac{d}{dp} \ell(p; \mathbf{x}) = -\frac{1}{(1 - p)} \sum_{i=1}^n (x_i - 1) + \frac{n}{p}$$

Si on fixe la fonction de score à zéro et qu'on réarrange l'expression, on obtient

$$\frac{1}{p} = \frac{1}{n} \sum_{i=1}^n x_i$$

La dérivée deuxième est négative, donc l'estimateur du maximum de vraisemblance est la réciproque de la moyenne, $\hat{p} = \bar{X}^{-1}$.

- (c) La fonction d'information observée vaut -1 fois la hessienne de la log-vraisemblance,

$$-\frac{d^2}{dp^2} \ell(p; \mathbf{x}) = \frac{n(\bar{x} - 1)}{(1 - p)^2} + \frac{n}{p^2}.$$

Si on évalue à l'estimé du maximum de vraisemblance, on obtient $j(\hat{p}) = n\bar{x}^3/(\bar{x} - 1)$.

- (d) L'estimé du maximum de vraisemblance est 0.1456 et son erreur-type est 0.0347.
 (e) La statistique de Wald et la statistique du rapport de vraisemblance valent

$$W = (\hat{p} - p_0)^2 / \text{Va}(\hat{p}) = (0.1456 - 0.1)^2 / 0.0347 = 1.72$$

$$R = 2\{\ell(\hat{p}) - \ell(p_0)\} = 2\{-45.11 + 45.39\} = 0.558.$$

Les deux peuvent être comparées à une loi χ_1^2 : les valeurs- p sont 0.19 et 0.45 (obtenues à l'aide d'un logiciel), donc on ne rejette pas $\mathcal{H}_0 : p = 0.1$.

- 3.3 On considère le temps avant défaillance de machines sur la base de leur niveau de corrosion w . Spécifiquement, le temps avant défaillance, T , est modélisé à l'aide d'une loi exponentielle de densité $f(t) = \lambda \exp(-\lambda t)$, mais d'intensité $\lambda = aw^b$; si $b = 0$, le temps de défaillance moyen est constant et vaut $E(T_i) = a^{-1}$. On suppose que les n observations sont indépendantes et que les niveaux de corrosion w_i sont supposés connus (donc fixes). [Coles (2001)]

- (a) Écrivez la log-vraisemblance du modèle
 (b) Dérivez les matrices d'informations observées et de Fisher.
 (c) Montrez que la log-vraisemblance profilée pour b est

$$\ell_p(b) = n \ln(\hat{a}_b) + b \sum_{i=1}^n \ln(w_i) - \hat{a}_b \sum_{i=1}^n w_i^b t_i,$$

et dérivez une formule explicite pour l'estimateur du maximum de vraisemblance partiel \hat{a}_b .

Solution

- (a) La log-vraisemblance est

$$\ell(a, b; \mathbf{w}, \mathbf{t}) = n \ln(a) + b \sum_{i=1}^n \ln(w_i) - a \sum_{i=1}^n w_i^b t_i$$

- (b) L'information de Fisher est l'espérance du négatif de la hessienne. L'information observée est

$$j(a, b) = - \begin{pmatrix} \partial^2 \ell / \partial a^2 & \partial^2 \ell / \partial a \partial b \\ \partial^2 \ell / \partial b \partial a & \partial^2 \ell / \partial b^2 \end{pmatrix} = \begin{pmatrix} na^{-2} & \sum_{i=1}^n w_i^b t_i \ln(w_i) \\ \sum_{i=1}^n w_i^b t_i \ln(w_i) & a \sum_{i=1}^n w_i^b t_i \ln^2(w_i) \end{pmatrix}.$$

Pour obtenir l'information de Fisher, on calcule l'espérance de chaque entrée de la matrice. Seuls les fonctions des T_i sont aléatoires et chaque entrée n'a que des facteurs linéaires (T_i fois une constante). Puisque $E(T_i) =$

$a^{-1}w_i^{-b}$, l'information de Fisher est

$$I(a, b) = \begin{pmatrix} na^{-2} & a^{-1} \sum_{i=1}^n \ln(w_i) \\ a^{-1} \sum_{i=1}^n \ln(w_i) & \sum_{i=1}^n \ln^2(w_i) \end{pmatrix}$$

- (c) On obtient l'estimateur du maximum de vraisemblance partiel \hat{a}_b en dérivant la log-vraisemblance par rapport à a tout en traitant b comme fixe,

$$\frac{\partial \ell}{\partial a} = \frac{n}{a} - \sum_{i=1}^n w_i^b t_i = 0$$

et on voit que la valeur de a qui maximise cette expression est $\hat{a}_b = n / \sum_{i=1}^n w_i^b t_i$, puisque la dérivée deuxième calculée précédemment est négative.