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	$p+1$	$\ell(\widehat{\boldsymbol{\beta}})$	AIC	BIC
M_1	`risque`	3	-244.566	495.132	510.362
M_2	`risque + region`	$\star$	-151.620	$\star$	$\star$
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 ($\ell(\widehat{\boldsymbol{\beta}})$) et les critères d'information.

Quelle est la différence entre le AIC et le BIC du modèle M_2 (en valeur absolue)?

Solution

La différence est approximativement de 35.54. Les seules informations requises ici sont le nombre de paramètres du modèle M_2 et la taille de l'échantillon. On peut utiliser la première ligne pour calculer cette dernière,

$$\mathsf{BIC} = -2\ell(\mathrm{M}_1) + 3\ln(n)$$

et on trouve en arrondissant $n = 1184$. Le nombre de paramètres dépend du nombre de niveaux de la variable région. On trouve $\mathrm{DF}(\mathrm{M}_3) - \mathrm{DF}(\mathrm{M}_1) - 3 = 7$; il y $K = 4$ catégories pour les années d'expérience, mais seulement trois paramètres additionnels dans le modèle. La différence $\mathsf{BIC} - \mathsf{AIC} = 7 \cdot \{\ln(1184) - 2\}$.

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

$$\mathsf{P}\,(X = x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots$$

(a) Écrivez la vraisemblance et la log-vraisemblance d'un échantillon aléatoire de taille n si les observations sont indépendantes.
(b) Dérivez l'estimateur du maximum de vraisemblance pour le paramètre p.
(c) Calculez l'information observée.
(d) 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.
(e) Calculez la statistique du rapport de vraisemblance et la statistique de Wald pour un test à niveau 5% de l'hypothèse $\mathscr{H}_0 : p_0 = 0.1$ contre l'alternative bilatérale $\mathscr{H}_a : p_0 \neq 0.1$.

Solution

(a)

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

(b)

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

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, $\widehat{p} = \overline{X}^{-1}$.

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

$$-\frac{\mathrm{d}^2}{\mathrm{d}p^2}\ell(p;\boldsymbol{x}) = \frac{n(\overline{x}-1)}{(1-p)^2} + \frac{n}{p^2}.$$

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

(d) L'estimé du maximum de vraisemblance est 0.1456 et son erreur-type est 0.0347.

(e) La statistique de Wald vaut 1.72 et la statistique du rapport de vraisemblance 0.558. Les deux peuvent être comparées à une loi χ^2_1: les valeurs-p sont 0.19 et 0.45, donc on ne rejette pas $\mathscr{H}_0 : p = 0.1$.

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 $\mathsf{E}(T_i) = a^{-1}$. 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

$$\ell_{\mathsf{p}}(b) = n\ln(\widehat{a}_b) + b\sum_{i=1}^{n}\ln(w_i) - \widehat{a}_b\sum_{i=1}^{n} w_i^b t_i,$$

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

Solution

(a) La log-vraisemblance est

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

(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 $\mathsf{E}(T_i) = a^{-1}w_i^{-b}$, 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 $\widehat{a}_b$ en dérivant la log-vraisemblance par rapport à a tout en traitant b comme fixe,

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

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