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`	⋆	−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 ($\ell(\widehat{\boldsymbol{\beta}})$) et les critères d'information.

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

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 $\mathscr{H}_a: p_0\neq 0.1$.

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 $\widehat{a}_b$.

Compilé le 04/10/2020 à 22:45