Prof. Dr. E. Cramer, Prof. Dr. M. Kateri RWTH Aachen, SS 2021
T. v. Bentum, M.Sc., L. Kaufmann, M.Sc.
publication: June 25th solutions available online: July 2n

Applied Data Analysis

Exercise Sheet 5

Exercise 19

(a) Consider a GLM with non-canonical link function. Explain why it does not need to be true that $\sum_i \hat{\mu}_i = \sum_i y_i$. Hence, the residuals do not need to have a mean of 0. Further, explain why a GLM with a canonical link function needs an intercept term to ensure that this mean does equal 0.

(b) Illustrate that for a GLM with non-canonical link the observed information matrix may depend on the data and hence differs from the expected information matrix.

(c) Bias-variance trade off: Let $Y_1, \ldots, Y_{100} \overset{\text{iid}}{\sim} \mathscr{B}(1, \pi)$ be given. Consider $\hat{\pi}_1 = \overline{Y}$ and $\hat{\pi}_2 = \frac{\overline{Y}}{2} + \frac{1}{4}$ to be two estimators of π. Which one is unbiased, which has smaller variance and for which values of π has $\hat{\pi}_2$ smaller mean squared error (MSE) than $\hat{\pi}_1$?

Exercise 20

(a) In selecting explanatory variables for a linear model, what is inadequate about the strategy of selecting the model with largest R^2 value?

(b) For a normal linear model $\mathscr{M}_1$ with $p + 1 = q + 2$ parameters $\beta_0, \ldots, \beta_q, \sigma^2$, show that $\text{AIC} = n(\log(2\pi\hat{\sigma}_1^2) + 1) + 2p + 2$, where $\hat{\sigma}_1^2$ denotes the maximum likelihood estimator of σ^2 for $\mathscr{M}_1$ and AIC denotes Akaike's Information Criterion given in Definition II.2.39. Let $\mathscr{M}_2$ be another normal linear model for the same data set with q_2 additional parameters. Show that $\mathscr{M}_2$ has smaller AIC as $\mathscr{M}_1$, if $\frac{\text{SSE}_2}{\text{SSE}_1} < \exp(\frac{-2q_2}{n})$.

(c) Let f be the true probability density or mass function (pdf/pmf) of the random variable Y and let $f_{\mathscr{M}}$ be the model pdf/pmf. Show that for the Kullback-Leibler-divergence it holds that $\text{E}\left(\log\left(\frac{f(Y)}{f_{\mathscr{M}}(Y)}\right)\right) \geq 0$ with "=", if and only if $f = f_{\mathscr{M}}$ P-almost surely.

Hint: You can solve this simultaneously by treating Y as absolutely continuous with respect to a measure μ and denoting the μ-densities of Y and its model by f and $f_{\mathscr{M}}$. Assume that $\text{supp}(f_{\mathscr{M}}) = \text{supp}(f)$ and that μ is restricted to $\text{supp}(f)$.

Exercise 21

(a) Proof the multivariate delta method, that is, for a sequence of p-dimensional random vectors $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$, $p \in \mathbb{N}$, with

$$\sqrt{n}\,\{\boldsymbol{X}_n - \boldsymbol{\mu}\} \xrightarrow{d} \mathscr{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$$

and a function $g\colon \mathbb{R}^p \to \mathbb{R}^q$, $q \in \mathbb{N}$, with continuous partial derivatives we have

$$\sqrt{n}\,\{g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})\} \xrightarrow{d} \mathscr{N}_q(\boldsymbol{0}, (D_g(\boldsymbol{\mu}))'\boldsymbol{\Sigma} D_g(\boldsymbol{\mu})),$$

where

$$D_g(\boldsymbol{x}) = \left(\frac{\partial g_i}{\partial x_j}\right)_{1\le j\le p, 1\le i\le p} \in \mathbb{R}^{p\times q}$$

is the matrix of partial derivatives.

Hint: Taylor expansion and Slutzky's Lemma.

(b) Use the delta method to proof the following:

(i) Corollary II.2.31 of the lecture.

(ii) Let $X_n \sim \mathcal{B}(n, \pi)$, $\pi \in (0, 1)$. Then, even though for each $n \in \mathbb{N}$ the variance of $Y_n := \log(X_n/n)$ does not exist, the asymptotic variance of $(Y_n)_{n\in\mathbb{N}}$ does, and is given by

$$\operatorname{Var}(Y_n) \approx \frac{1-\pi}{n\pi}.$$

Exercise 22

Consider a sample $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

(a) Derive the profile likelihood of the mean μ and the variance σ^2, respectively.

(b) Derive the expectation of the score for σ^2 with respect to its profile likelihood. What do you observe?

Explain possible effects on the consistency of estimators derived from the profile likelihood when the number of samples is small compared to the number of nuisance parameters.

(c) Find the (expected) Fisher information for σ^2 derived from the profile likelihood and compare the result to the (expected) Fisher information for the unrestricted likelihood.

Explain your observation.

Exercise 23

(a) Let a baseline-category logit model with 3 possible outcome categories and a single explanatory variable $x \in \mathbb{R}$ be given, such that the probability to be in response category j conditional on the value of x is given by

$$\pi_j(x) = \frac{\exp(\beta_{j0} + \beta_j x)}{1 + \exp(\beta_{10} + \beta_1 x) + \exp(\beta_{20} + \beta_2 x)}, \qquad j = 1, 2,$$

and $\pi_3(x) = 1 - (\pi_1(x) + \pi_2(x))$. Assume that $\beta_1 \neq 0$, $\beta_2 \neq 0$ and show that

$$\pi_3 \text{ is } \begin{cases} \text{decreasing}, & \text{if } \beta_1 > 0, \beta_2 > 0, \\ \text{increasing}, & \text{if } \beta_1 < 0, \beta_2 < 0, \\ \text{not monotonic}, & \text{else.} \end{cases}$$

(b) Let a baseline-category logit model for grouped data with g observed groups and J outcome categories be given. Denote by n_i the number of responses in group i, $i = 1, \ldots, g$. Further, denote by y_{ij} the sample proportion of responses in category j of group i (the saturated model) and by $\hat{\pi}_{ij}$ the estimate of the probability of responding category j belonging to group i in the model sense, $i = 1, \ldots, g$, $j = 1, \ldots, J$. Show that the deviance

$$G^2 = 2 \sum_{i=1}^{g} \sum_{j=1}^{J} n_i y_{ij} \log\left(\frac{n_i y_{ij}}{n_i \hat{\pi}_{ij}}\right)$$

coincides with the corresponding likelihood-ratio test statistic of the baseline-category logit model against the saturated model.