Applied Data Analysis

Exercise Sheet 5

Exercise 18

(a) Prove the multivariate delta method, that is:
Let $\boldsymbol{\mu} \in \mathbb{R}^p$ and $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of p-dimensional random vectors with

$$\sqrt{n}\,\big(\boldsymbol{X}_n - \boldsymbol{\mu}\big) \quad \stackrel{d}{\longrightarrow} \quad \mathcal{N}_p(\mathbf{0}, \Sigma) \quad \text{for} \quad n \longrightarrow \infty\ ,$$

where $\Sigma \in \mathbb{R}^{p\times p}$ denotes a positive definite covariance matrix. Further, let $g : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ be a function with continuous partial derivatives.
Then, it holds:

$$\sqrt{n}\,\big(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})\big) \quad \stackrel{d}{\longrightarrow} \quad \mathcal{N}_q\big(\mathbf{0}, (D_g(\boldsymbol{\mu}))'\,\Sigma\, D_g(\boldsymbol{\mu})\big) \quad \text{for} \quad n \longrightarrow \infty\ ,$$

where

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

denotes the matrix of the partial derivatives of the function g evaluated at $x \in \mathbb{R}^p$.

Hint: Taylor expansion and Slutsky's Lemma.

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

(i) The second part of Corollary II.2.31 of the Lecture under the additional assumption that the inverse g^{-1} of the link function g is continuously differentiable.

(ii) Let $X_n \sim \mathcal{B}(n, \pi)$ for $n \in \mathbb{N}$ with some parameter $\pi \in (0, 1)$.
Then, even though for each $n \in \mathbb{N}$ the variance of $Y_n := \ln(X_n/n)$ does not exist, the asymptotic variance of $(Y_n)_{n\in\mathbb{N}}$ does, yielding:

$$\mathrm{Var}\big(\sqrt{n}\,(Y_n - \ln(\pi))\big) \;\approx\; \frac{1-\pi}{\pi} \text{ for sufficiently large } n \in \mathbb{N}\ .$$

Exercise 19

For a multiple linear regression model $\mathscr{M}_1$ (according to I.5.3) with $p+1 = m+2$ parameters $\beta_0, \dots, \beta_m \in \mathbb{R}$ and $\sigma^2 > 0$, show:

$$\mathrm{AIC} \;=\; n\,\big(\ln(2\,\pi\,\widehat{\sigma}_1^2) + 1\big) + 2\,p + 2\ ,$$

where $\widehat{\sigma}_1^2$ denotes the maximum likelihood estimate of the variance σ^2 for model $\mathscr{M}_1$ and AIC denotes Akaike's Information Criterion given in Definition II.2.39.

Let $\mathscr{M}_2$ be another multiple linear regression model for the same data set with q additional parameters and $n \ge p+1+q$. Show that $\mathscr{M}_2$ has a smaller AIC compared to $\mathscr{M}_1$, if

$$\frac{\mathrm{SSE}_2}{\mathrm{SSE}_1} \;<\; \exp\left(-\,\frac{2\,q}{n}\right)$$

with SSE_1 and SSE_2 being defined as in I.5.12 for models $\mathscr{M}_1$ and $\mathscr{M}_2$, respectively.

Exercise 20

(a) Consider a GLM with a non-canonical link function. Explain why it does not need to be true that

$$\sum_{i=1}^{n} \widehat{\mu}_i \;=\; \sum_{i=1}^{n} y_i \;.$$

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 of the residuals does equal 0.

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

Hint: As a counter-example, consider the intercept-GLM for a single random variable $Y \sim \mathcal{B}(n, \pi)$ with $n \in \mathbb{N}$ and $\pi \in (0,1)$ and with the identity link function (which is *not* the canonical one).

(c) Let $Y_1, \ldots, Y_{100}$ be stochastically independent random variables with $X_i \sim \mathcal{B}(1, \pi)$ for $i \in \{1, \ldots, 100\}$ and for some $\pi \in (0,1)$. Consider the following two estimators for the parameter π:

$$\widehat{\pi}_1 \;:=\; \overline{Y} \;=\; \frac{1}{100}\sum_{i=1}^{100} Y_i \quad \text{and} \quad \widehat{\pi}_2 \;:=\; \frac{1}{2}\overline{Y} + \frac{1}{4} \;.$$

(i) Which of the two estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ is unbiased?

(ii) Which of the two estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ has smaller variance?

(iii) For which values of $\pi \in (0,1)$ has $\widehat{\pi}_1$ a smaller mean squared error (MSE) than $\widehat{\pi}_2$?