E-test 3 01.07.22

Freitag, 1. Juli 2022 10:25

Aufgabe 1

13 Punkte

For $k \in \mathbb{N}$ and $\alpha \in (1, \infty)$, let Y be a discrete random variable with values in $M = \frac{1}{k}\mathbb{N}_0 = \{\frac{j}{k} \mid j \in \mathbb{N}_0\}$ and corresponding probability mass function $p(\bullet; k, \alpha) : M \to [0,1]$, given by

$$(*) \quad p(y; k, \alpha) = \binom{ky + k - 1}{k - 1} \frac{(\alpha - 1)^{ky}}{\alpha^{ky + k}}, \; y \in M.$$

It can be shown that the given probability mass function is a member of an Exponential Dispersion Family, i.e. it can be expressed in the following form:

$$(**) \quad p(y; k, \alpha) = \exp\left(\frac{\theta y - b(\theta)}{a(\phi)} + c(y, \phi)\right), \; y \in M,$$

with $\phi = k$, $a(\phi) = a(k) := \frac{1}{k}$, $c(y, \phi) = c(y, k) := \ln\left(\binom{ky+k-1}{k-1}\right)$ for $y \in M$, an appropriately chosen natural parameter θ depending on α and an appropriately chosen function b of θ.

(a) For the following tasks, you need to derive a representation of the form $(**)$ for the probability mass function given by $(*)$.

(i) For $\alpha = 4$, calculate the value of the corresponding natural parameter θ. — 3 Punkte

Give the value of θ (rounded to three decimal places) — 3 Punkte

$-0{,}288$

(ii) For $\alpha = 4$ and $k = 2$, calculate the expectation $E(Y)$. — 2 Punkte

Give the value of $E(Y)$. — 2 Punkte

Zahl 3

(iii) For $\alpha = 4$ and $k = 2$, calculate the variance $\mathrm{Var}(Y)$. — 2 Punkte

Give the value of $\mathrm{Var}(Y)$. — 2 Punkte

Zahl 6

(b) Consider three stochastically independent discrete random variables Y_1, Y_2, Y_3, each having a probability mass function given by $(*)$ for $k = 1$ and some parameters $\alpha_1, \alpha_2, \alpha_3 \in (1, \infty)$, respectively, which are chosen such that it holds:

$$\mathrm{Var}(Y_1) = 6 \;, \quad \mathrm{Var}(Y_2) = 12 \;, \quad \mathrm{Var}(Y_3) = 2 \,.$$

Further, let $Y = (Y_1, Y_2, Y_3)'$ fulfill a GLM given by $g(E(Y)) = g(\mu) = X\beta$ with **canonical** link function g, parameter vector $\beta = (\beta_1, \beta_2)' \in \mathbb{R}^2$ and design matrix

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

1 2 3 4

In this framework and for the variances of Y_1, Y_2, Y_3 given above, determine the expected Fisher information matrix

$$\mathcal{I}_F(\beta) = \begin{pmatrix} i_{11} & i_{12} \\ i_{21} & i_{22} \end{pmatrix}$$

for the corresponding parameter vector β.

Hint: Part (b) can be solved independently of part (a), without deriving a representation of the form $(**)$ for the given probability mass function.

Calculate the entries of $\mathcal{I}_F(\beta)$. — 6 Punkte

Give the value of i_{11}. — 2 Punkte

Zahl

Give the value of i_{12}. — 2 Punkte

1 2 3 4

a) i)

$$f(y; k, \alpha) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)$$

$$= \exp\left(\frac{y\theta - b(\theta)}{\frac{1}{k}} + \ln\left(\binom{ky+k-1}{k-1}\right)\right)$$

$$= \exp\left((y\theta - b(\theta))k\right) \cdot \binom{ky+k-1}{k-1}$$

$\rightsquigarrow$ coefficient comparison:

$$\exp(y\theta - b(\theta))^k = \left(\frac{(\alpha-1)^y}{\alpha^y \cdot \alpha}\right)^k = \left(\frac{(\alpha-1)^y}{\alpha^y} \cdot \frac{1}{\alpha}\right)^k$$

$$\Leftrightarrow \frac{\exp(\theta)^y}{\exp(b(\theta))} = \frac{\exp(\theta)^y}{\exp(b(\theta))} = \left(\frac{\alpha-1}{\alpha}\right)^y \cdot \frac{1}{\alpha}$$

$$\rightsquigarrow \exp(\theta) = \frac{\alpha - 1}{\alpha} = \frac{3}{4} \;, \text{ for } \alpha = 4$$

$$\theta = \ln\left(\frac{3}{4}\right) \approx -0.288$$

ii) II.2.4

$$E(Y) = b'(\theta) = \frac{d}{d\theta}\left(\ln(\alpha)\right)$$

$$\exp(\theta) = 1 - \frac{1}{\alpha}$$

$$1 - e^\theta = \frac{1}{\alpha} \Rightarrow \alpha = \frac{1}{1 - e^\theta}$$

$$\rightsquigarrow E(Y) = \frac{d}{d\theta}\left(\ln\left(\frac{1}{1 - e^\theta}\right)\right)$$

$$= \frac{d}{d\theta}\left(\underbrace{\ln(1)}_{=0} - \ln(1 - e^\theta)\right)$$

$$= -\frac{d}{d\theta}\left(-\ln(1 - e^\theta)\right)$$

$$= -\left(\frac{-e^\theta}{1 - e^\theta}\right) =$$

$$\rightsquigarrow -\left(\frac{-3/4}{1/4}\right) = 3$$

$$\frac{1}{1 - e^\theta} = 4$$

$$e^\theta = 1 - \frac{1}{4} = \frac{3}{4}$$

iii)

$$\mathrm{Var}(Y) = b''(\theta) \cdot a(\phi) \;, \quad a(\phi) = \frac{1}{4}$$

$$= \frac{d}{d\theta}\left(\frac{e^\theta}{1 - e^\theta}\right)\frac{1}{2} = \frac{1}{2}\left(\frac{e^\theta}{1 - e^\theta} + \frac{e^{2\theta}}{(1 - e^\theta)^2}\right)$$

$$= \frac{e^\theta(1 - e^\theta) + e^{2\theta}}{(1 - e^\theta)^2} \cdot \frac{1}{2}$$

$$= \frac{e^\theta \cancel{- e^{2\theta} + e^{2\theta}}}{(e^\theta - 1)^2} \cdot \frac{1}{2}$$

with $e^\theta = 3/4$

$$\Rightarrow \frac{\frac{3}{4}}{\left(\frac{3}{4} - \frac{4}{4}\right)^2} \cdot \frac{1}{2} = \frac{3/4}{1/8}$$

$$= 3 \cdot 2 = 6$$

b)

$$\mathrm{Var}(Y_i) = a \cdot b''(\theta_i)$$

$$= \frac{1}{k} \cdot b''(\theta_i) \quad \text{with } k = 1$$

$$\rightsquigarrow \mathrm{Var}(Y_i) = b''(\theta_i)$$

$$W = \begin{pmatrix} b''(\theta_1) & 0 & 0 \\ 0 & b''(\theta_2) & 0 \\ 0 & 0 & b''(\theta_3) \end{pmatrix}$$

$$\Rightarrow W = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\rightsquigarrow \mathcal{I}_F = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 6 & & \\ & 12 & \\ & & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 18 & -12 \\ -12 & 14 \end{pmatrix}$$

II.2.22 Theorem (large sample Normal distribution of MLE $\hat{\beta}$)

The asymptotic covariance matrix of $\hat{\beta}$ is $\mathrm{Cov}(\hat{\beta}) = \mathcal{I}_F^{-1}$, where $\mathcal{I}_F$ is the expected Fisher information matrix

$$\mathcal{I}_F = \ldots = X'WX,$$

II.2.19 Likelihood equations for GLMs with canonical link

$b''(\theta_i)$

Proof

II.2.23 Definition (observed information matrix)

II.2.24 Information matrix for GLMs with canonical link

We are all work together

Hint: Part (b) can be solved independently of part (a), without deriving a representation of the form (**) for the given probability mass function.

Calculate the entries of $\mathcal{I}_F(\beta)$. 6 Punkte

Give the value of i_{11}. 2 Punkte

18 Zahl

Give the value of i_{12}. 2 Punkte

-42

Give the value of i_{22}. 2 Punkte

14

Aufgabe 2

7 Punkte

Let $\mu_1, \mu_2 > 0$ and $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ be two sequences of stochastically independent random variables with $X_i \sim \mathcal{P}(\mu_1)$ and $Y_i \sim \mathcal{P}(\mu_2)$ for $i \in \mathbb{N}$, where $\mathcal{P}(\mu_j)$ denotes the Poisson distribution with parameter μ_j for $j \in \{1,2\}$.

For $n \in \mathbb{N}$, the corresponding arithmetic means of $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, respectively, are denoted by

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i .$$

In each of the following two tasks, determine the asymptotic variance, i.e. the variances of the corresponding limit distributions for the sequences of random variables considered in these parts.

(a) For $\mu_1 = 4$, calculate the asymptotic variance σ_1^2 of the sequence $\left(\sqrt{n}(\overline{X}_n - \mu_1)\right)_{n\in\mathbb{N}}$. 3 Punkte

Give the value of σ_1^2. 3 Punkte

4 Zahl

Central limit theorem:

$\lim_{n\to\infty} (\sqrt{n}(\overline{X}_n - \mu)/\sigma) \xrightarrow{d} N(0,1)$

$\sigma = 2$

$\rightsquigarrow \sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \underbrace{\sigma N(0,1)}_{N(0,\sigma^2\cdot 1)}$

since for Poisson-Distribution $Var = \mu$

$\rightsquigarrow Var = \sigma^2 = 4$

b) delta-method

$\mu_2 = 1$

$\lim_{n\to\infty} \sqrt{n}(\sqrt{\overline{Y}_n} - \sqrt{\mu_2}) \xrightarrow{d} N(0,1)$

define $g = \sqrt{\mu_2}$ and $g' = \frac{1}{2}\frac{1}{\sqrt{\mu_2}} = \frac{1}{2}$

$\rightsquigarrow \lim_{n\to\infty} g' \sqrt{n}(\overline{Y}_n - \mu_2) \sim \underbrace{g' N(0,1)}_{N(0,g'^2) = N(0,\frac{1}{4})}$

$\rightarrow \sigma_2^2 = \frac{1}{4}$

(b) For $\mu_2 = 1$, calculate the asymptotic variance σ_2^2 of the sequence $\left(\sqrt{n}\left(\sqrt{\overline{Y}_n} - \sqrt{\mu_2}\right)\right)_{n\in\mathbb{N}}$. 4 Punkte

Give the value of σ_2^2. 4 Punkte

1/4 Zahl

← Vorherige Aufgabe

Nächste Aufgabe →

Alle Antworten wurden gespeichert!

2 3 4

Aufgabe 3

10 Punkte

Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.

Consider the following **ungrouped** data file called `Beetles` which contains data from one of the first studies using a binary regression model in 1935. The study divided a sample of $n_{tot} = 481$ adult flour beetles to 8 groups of size m_j, $j = 1, \dots, 8$, with $\sum_{j=1}^{8} m_j = n_{tot}$. The beetles were exposed to gaseous carbon disulfide at 8 distinct dosages (in mg/liter), one for every group. The study observed for the beetles of all groups whether they are alive or dead after a 5 hours exposure, which gives us the corresponding response variable Y, taking values $y_i \in \{0, 1\}$, where the value 0 denotes the survival of beetle i, $i = 1, \dots, n_{tot}$. The explanatory variable is the dosage in log-scale. Hence, if $dose_i$ is the dose to which beetle i is exposed to, then $x_i = \log_{10}(dose_i)$, $i = 1, \dots, n_{tot}$.

Beetles

(a) Give the number m_1 of beetles exposed to the dose for which $x_i = \log_{10}(dose_i) = 1.691$. Further, give the proportion of deaths among the beetles exposed to this particular dose. 1 Punkt

m_1 **(requested precision: whole numbers)** 0.5 Punkte

59 Zahl

proportion of deaths **(requested precision: 2 digits)** 0.5 Punkte

0.10 Zahl

(b) Fit a generalized linear model using the canonical link function for the response Y and treating the explanatory variable as a continuous one. Calculate the sum of squared errors (SSE). 2 Punkte

SSE **(requested precision: 2 digits)** 2 Punkte

59.14 Zahl

(c) Provide the corresponding AIC and BIC values for the model in (b). 2 Punkte

(c) Provide the corresponding AIC and BIC values for the model in (b). 2 Punkte

AIC **(requested precision: 2 digits)** 1 Punkt

376.25 Zahl

BIC **(requested precision: 2 digits)** 1 Punkt

384.7 Zahl

(d) For the model fitted in (b), give the 95% asymptotic (Wald) confidence interval for the parameter corresponding to the explanatory variable. 2 Punkte

lower bound of CI **(requested precision: 3 digits)** 1 Punkt

28.867 Zahl

(d) For the model fitted in (b), give the 95% asymptotic (Wald) confidence interval for the parameter corresponding to the explanatory variable. — 2 Punkte

lower bound of CI **(requested precision: 3 digits)** — 1 Punkt

28.867

upper bound of CI **(requested precision: 3 digits)** — 1 Punkt

40.318

(e) What is the proportion of correctly classified observations from the model fitted in (b) using 0.5 as threshold probability? — 1 Punkt

proportion of correctly classified observations **(requested precision: 2 digits)** — 1 Punkt

0.82

Zahl

(f) Fit a generalized linear model using a probit link. Calculate the sum of squared errors (SSE). Based on the value of SSE, would you prefer the model using a probit link, the model of (b) or both? — 2 Punkte

SSE **(requested precision: 2 digits)** — 1 Punkt

59.21

preference based on SSE — 1 Punkt

Bitte auswählen ▾

← Vorherige Aufgabe — Nächste Aufgabe →

Aufgabe 4 — 10 Punkte

Before II.2.3

The answer of an item of this task is correct (and the corresponding point (points) is (are) granted) only if the correct set of right statements is specified **exactly**.

(a) Choose the statement (statements) that is (are) **true**. — 1 Punkt

- [ ] Logistic regression assumes a linear relationship between the response variable Y and the explanatory variables.
- [x] Logistic regression assumes a linear relationship between the logarithm of the odds of the response and the explanatory variables.
- [x] If we have a binary response variable, we always have to use logistic regression.
- [ ] The link function used for obtaining the logistic regression model is the identity link.
- [ ] The link function used for obtaining the logistic regression model is the log link.

(b) Choose the assumption (assumptions) that is (are) **not** an assumption in the GLM framework where Y is the response variable and X the explanatory variable. — 2 Punkte

- [x] The response is binary.
- [ ] The conditional probability density (or mass) function (pdf or pmf) of Y given $X = x$ belongs to the exponential dispersion family.
- [x] The conditional probability density (or mass) function (pdf or pmf) of X given $Y = y$ belongs to the exponential dispersion family.
- [ ] For a random sample of size n, the responses $Y_i, i = 1, \dots, n$, are independent and identically distributed.
- [ ] The link function links the expectation of the response with the linear predictor.

- [ ] The link function links the expectation of the response with the linear predictor.

(c) Choose the statement (statements) that is (are) **true**. — 2 Punkte

- [ ] Generalized linear models allow the linear predictor to be non-linear in the parameters β.
- [x] Generalized linear models are more sensitive to outliers than linear models.
- [ ] Generalized linear models can fit complex relationships between the response and the explanatory variables.
- [ ] Generalized linear models can handle both continuous and categorical data while linear models can just handle one type of them.
- [ ] In a generalized linear model, the distribution of the error term has to be a normal distribution.

(d) Choose the statement (statements) that is (are) **true** for a GLM. — 2 Punkte

- [ ] For a poisson distributed random response variable, the canonical link is the logit link.
- [x] The link function links the expected value of the random response variable to the linear predictor.
- [ ] The link function transforms the expected value of the random response variable to the natural parameter θ of the exponential dispersion family corresponding to the random response variable.
- [ ] The link function is used to transform the values of the response variable.

(e) Choose the statement (statements) that is (are) **true** for a GLM. — 1 Punkt

- [ ] The degrees of freedom of a saturated model are always equal to 0.

(f) Consider a simple logistic regression model with parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ and model matrix $\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. Choose the statement (statements) that is (are) **true**.

2 Punkte

- [ ] Increasing the explanatory variable by one unit, the odds of success for the response variable will be multiplied by $\exp(\beta_2)$.
- [ ] Increasing the explanatory variable by one unit, the odds of success for the response variable will increase additively by $\exp(\beta_2)$.
- [ ] If $\beta_2 = 0$ the success probability of the response variable is equal to zero.

model matrix $\boldsymbol{X} = \begin{pmatrix} \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. Choose the statement (statements) that is (are) **true**.

2 Punkte

- [x] Increasing the explanatory variable by one unit, the odds of success for the response variable will be multiplied by $\exp(\beta_2)$.
- [ ] Increasing the explanatory variable by one unit, the odds of success for the response variable will increase additively by $\exp(\beta_2)$.
- [ ] If $\beta_2 = 0$ the success probability of the response variable is equal to zero.
- [x] If $\beta_2 = 0$ the success probability of the response variable is a constant function of the explanatory variable.
- [ ] The median effective level is the point where the success probability of the response variable is maximized.