Aufgabe 1

6 Punkte

Let a family of distributions be given by their pdfs (probability density functions) defined for $\alpha > 0, \gamma > 0$ as

$$f(x;\alpha,\gamma) = \sqrt{\frac{\gamma}{2\pi x^2}}\,\exp\left(-\frac{\gamma(x-\alpha)^2}{2\alpha^2 x}\right), \quad x > 0. \qquad (1)$$

For fixed (known) $\gamma > 0$, $f_\gamma(x;\alpha) = f(x;\alpha,\gamma)$ defines a subfamily of the exponential dispersion family (EDF) of distributions with

$$c(x,\phi) = \frac{1}{2}\left(\ln(\gamma) - \ln(2\pi x^2) - \frac{\gamma}{x}\right).$$

Find the missing numerical values. **For all numerical results the exact values have to be provided without any rounding.**

6 Punkte

1 Punkt

(a) Let $\gamma = 2$ and $X \sim f_2(\,\cdot\,;\alpha)$.

(i) Determine the values of the natural parameter θ and the dispersion $a(\phi)$, when $\alpha = 1$.

$$c(x, \phi) = \frac{1}{2}\left(\ln(\gamma) - \ln(2\pi x^2) - \frac{\gamma}{x}\right).$$

Find the missing numerical values. **For all numerical results the exact values have to be provided without any rounding.** 6 Punkte

1 Punkt

(a) Let $\gamma = 2$ and $X \sim f_2(\,\cdot\,; \alpha)$.

(i) Determine the values of the natural parameter θ and the dispersion $a(\phi)$, when $\alpha = 1$.

$\theta =$ -0.5

1 Punkt

$a(\phi) =$ 0.5

1 Punkt

(ii) Calculate the expectation $\mathrm{E}(X)$, when $\alpha = 1$.

$\mathrm{E}(X) =$ -1

1 Punkt

1 Punkt

(iii) Calculate the variance $\mathbf{Var}(X)$, when $\alpha = 1$.

$\mathbf{Var}(X) =$

0.5

Zahl

1 Punkt

(b) Further assume that Y is a binary response variable and let

$$\pi(x) = P(Y = 1 \mid X = x).$$

Suppose that

$$(X|Y = j) \sim f(\cdot\,|\alpha_j, \gamma_j),$$

that is, conditionally on $Y = j$ the explanatory variable X has pdf (1) with parameters $\alpha_j, \gamma_j > 0$, $j \in \{0, 1\}$, and consider the model

$$\mathrm{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x^{-1} + \beta_2 x.$$

Assume that $\alpha_0 = \gamma_0 = 2$ and $\alpha_1 = \gamma_1 = 1$. Calculate the values of β_1 and β_2.

Hint: Use Bayes' Theorem applied to probability distributions: For random variables X_1, X_2 with pdfs or pmfs f^{X_1}, f^{X_2}, respectively, it holds

$$f^{X_1|X_2=x_2}(x_1) = \frac{f^{X_2|X_1=x_1}(x_2) f^{X_1}(x_1)}{f^{X_2}(x_2)} I_{\mathrm{supp}(X_2)}(x_2), \quad x_1 \in \mathbb{R},$$

1 Punkt

(b) Further assume that Y is a binary response variable and let

$$\pi(x) = P(Y = 1 \mid X = x).$$

Suppose that

$$(X|Y = j) \sim f(\cdot\,|\alpha_j, \gamma_j),$$

that is, conditionally on $Y = j$ the explanatory variable X has pdf (1) with parameters $\alpha_j, \gamma_j > 0$, $j \in \{0, 1\}$, and consider the model

$$\operatorname{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x^{-1} + \beta_2 x.$$

Assume that $\alpha_0 = \gamma_0 = 2$ and $\alpha_1 = \gamma_1 = 1$. Calculate the values of β_1 and β_2.

Hint: Use Bayes' Theorem applied to probability distributions: For random variables X_1, X_2 with pdfs or pmfs f^{X_1}, f^{X_2}, respectively, it holds

$$f^{X_1|X_2=x_2}(x_1) = \frac{f^{X_2|X_1=x_1}(x_2) f^{X_1}(x_1)}{f^{X_2}(x_2)} I_{\operatorname{supp}(X_2)}(x_2), \quad x_1 \in \mathbb{R},$$

where $f^{X_i|X_j}, i \neq j, i, j \in \{1, 2\}$ is the conditional probability density or mass function of X_i given X_j and I_A is the indicator function on a set A.

$\beta_1 =$

Zahl

$$\operatorname{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x + \beta_2 x.$$

Assume that $\alpha_0 = \gamma_0 = 2$ and $\alpha_1 = \gamma_1 = 1$. Calculate the values of β_1 and β_2.

Hint: Use Bayes' Theorem applied to probability distributions: For random variables X_1, X_2 with pdfs or pmfs f^{X_1}, f^{X_2}, respectively, it holds

$$f^{X_1|X_2=x_2}(x_1) = \frac{f^{X_2|X_1=x_1}(x_2) f^{X_1}(x_1)}{f^{X_2}(x_2)} I_{\text{supp}(X_2)}(x_2), \quad x_1 \in \mathbb{R},$$

where $f^{X_i|X_j}, i \neq j, i, j \in \{1, 2\}$ is the conditional probability density or mass function of X_i given X_j and I_A is the indicator function on a set A.

$\beta_1 =$

Zahl

1 Punkt

$\beta_2 =$

Zahl

Nächste Aufgabe →

Alle Antworten wurden gespeichert!

Aufgabe 2

3 Punkte

Consider a GLM $\mathbf{E}(\boldsymbol{Y}) = g(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta}$, where g is the **canonical** link, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and

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

Further assume that the responses Y_1, Y_2, Y_3 are jointly independent with $Y_i \sim \mathcal{P}(\mu_i)$, where $\mathcal{P}(\mu_i)$ denotes the Poisson distribution with parameters $\mu_i = 2i^2$, $i \in \{1, 2, 3\}$.

Determine the expected Fisher information matrix

$$\mathcal{I}_F = \begin{pmatrix} i_1 & i \\ i & i_2 \end{pmatrix}$$

with respect to the parameter vector $\boldsymbol{\beta}$.

Calculate the missing numerical values. **For all numerical results the exact values have to be provided without any rounding.**

3 Punkte

1 Punkt

$i_1 =$

Zahl

with respect to the parameter vector β.

Calculate the missing numerical values. **For all numerical results the exact values have to be provided without any rounding.** 3 Punkte

$i_1 =$ 20 1 Punkt
Zahl

$i_2 =$ 26 1 Punkt
Zahl

$i =$ -18 1 Punkt
Zahl

Vorherige Aufgabe

Nächste Aufgabe

1 2 3 4 5 6

Aufgabe 3

4 Punkte

Let $(X_i)_i \overset{iid}{\sim} \mathcal{P}(\mu_1)$, $(Y_i)_i \overset{iid}{\sim} \mathcal{P}(\mu_2)$ be two jointly independent sequences of Poisson random counts with parameter μ_i, $i \in \{1, 2\}$, respectively. Further, define

$$\hat{\mu}_{1(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\mu}_{2(n)} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad n \in \mathbb{N}.$$

Derive the asymptotic variance $\sigma^2 > 0$, say, of

$$Z_n = \sqrt{n}(\hat{\mu}_{1(n)} + \hat{\mu}^2_{2(n)})$$

as $n \to \infty$ for $\mu_1 = 1$ and $\mu_2 = 2$.

Calculate the missing numerical value. **The numerical result has to be given as an exact values without any rounding.**

4 Punkte

4 Punkte

$\sigma^2 =$

Zahl

$\frac{1}{n}\sum_{i=1}^{n}$ $\frac{1}{n}\sum_{i=1}^{n}$

Derive the asymptotic variance $\sigma^2 > 0$, say, of

$$Z_n = \sqrt{n}(\hat{\mu}_{1(n)} + \hat{\mu}^2_{2(n)})$$

as $n \to \infty$ for $\mu_1 = 1$ and $\mu_2 = 2$.

Calculate the missing numerical value. **The numerical result has to be given as an exact values without any rounding.** 4 Punkte

4 Punkte

$\sigma^2 =$

Zahl

Vorherige Aufgabe

Nächste Aufgabe

Alle Antworten wurden gespeichert!

Aufgabe 4 **3 Punkte**

The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. **Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.**

Task 24 3 Punkte

0.5 Punkte

Which of the three models of task 24(b) and 24(c) has the best fit in terms of AIC? Type in "1" for the logistic regression model, "2" for the probit and "3" for the cloglog regression model (without quotation marks). Provide the corresponding AIC value for this model **(requested precision: 1 digit)**

best model

3

Zahl

0.5 Punkte

AIC value

128.2

Zahl

1 Punkt

Transform the variable **`sex`** into a factor variable. Fit a cloglog regression model that predicts **`transport`** using **`time, sex`** and their interaction. What is the percentage of correct classified observations for this model? **(requested precision: 4 digits)**

Which of the three models of task 24(b) and 24(c) has the best fit in terms of AIC? Type in "1" for the logistic regression model, "2" for the probit and "3" for the cloglog regression model (without quotation marks). Provide the corresponding AIC value for this model **(requested precision: 1 digit)**

best model

Zahl

0.5 Punkte

AIC value

Zahl

1 Punkt

Transform the variable **`sex`** into a factor variable. Fit a cloglog regression model that predicts **`transport`** using **`time`**, **`sex`** and their interaction. What is the percentage of correct classified observations for this model? **(requested precision: 4 digits)**

0.7479

0.5 Punkte

What are the values for AIC and BIC of the model above? **(requested precision: 1 digit)**

Value AIC

123,8

Value BIC

132,1

Transform the variable **sex** into a factor variable. Fit a cloglog regression model that predicts **transport** using **time, sex** and their interaction. What is the percentage of correct classified observations for this model? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

What are the values for AIC and BIC of the model above? **(requested precision: 1 digit)**

Value AIC

Zahl

0.5 Punkte

Value BIC

Zahl

Vorherige Aufgabe

Nächste Aufgabe

Alle Antworten wurden gespeichert!

Aufgabe 5

5.5 Punkte

The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. **Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.**

Task 25

At the beginning of your code, set the seed to 2021. Load the dataset *credits.csv* from RWTHmoodle into your workspace. **Make sure that your working directory contains an unmodified version of *credits.csv*. In particular, we recommend downloading a fresh version of *credits.csv* from the RWTHmoodle space.**

5.5 Punkte

1 Punkt

In the *credits* dataset, the variable **guaran** stands for wheather the possible credits recipient is "No further guarantor of debtor" (**guaran** $= 1$), "co applicant" (**guaran** $= 2$) or "guarantor" (**guaran** $= 3$). Introduce a new variable **guaran2** with two categories where **guaran2** $= 1$ for **guaran** $= 1$ and **guaran** $= 2$ and **guaran2** $= 2$ otherwise. For how many cases in the data set *credits* it holds that **guaran2=1**? **(requested precision: whole number)**

948

0.5 Punkte

Split randomly the data in training (65 % of the dat) and testing data (the remaining 35 % of the data). Transform **account, behavior, rate, finance, furthered, home, job, pers** into factor variables. Fit the following logistic regression model on the training data:

1 Punkt

In the *credits* dataset, the variable **guaran** stands for wheather the possible credits recipient is "No further guarantor of debtor" (**guaran** $= 1$), "co applicant" (**guaran** $= 2$) or "guarantor" (**guaran** $= 3$). Introduce a new variable **guaran2** with two categories where **guaran2** $= 1$ for **guaran** $= 1$ and **guaran** $= 2$ and **guaran2** $= 2$ otherwise. For how many cases in the data set *credits* it holds that **guaran2=1**? **(requested precision: whole number)**

Zahl

0.5 Punkte

Split randomly the data in training (65 % of the dat) and testing data (the remaining 35 % of the data). Transform **account, behavior, rate, finance, furthered, home, job, pers** into factor variables. Fit the following logistic regression model on the training data:

repayment $\sim$
time+age+account+behavior+savings+rate+guaran2+finance+furthcred+home+job+pers

In the following, we will refer to this model as *model b*.

What are the AIC and BIC values for this model? **(requested precision: 1 digit)**

AIC value

663.1

0.5 Punkte

BIC value

779.5

1 Punkt

0.5 Punkte

BIC value

Zahl

1 Punkt

What is the AUC (area under the curve) of *model b*? **(requested precision: 3 digits)**

Zahl

1 Punkt

Fit the null model corresponding to *model b*. What is the null deviance for this model? **(requested precision: 4 digits)**

797.5

0.5 Punkte

Select the best model nested in *model b* in terms of AIC using a backward stepwise selection algorithm. In the following, we will refer to this selected model as *model e*. What is the median of the estimated probabilities for on time credit repayment for the training sample, based on this model? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

Select the best model nested in *model b* in terms of BIC using a backward stepwise selection algorithm. In the following, we will refer to this selected model als *model f*. Test *model f* and *model e* on the test data. What are the values for predicted residual sum of squares (PRESS) for both models ? **(requested precision: 4 digits)**

PRESS for *model e*

1 Punkt

Fit the null model corresponding to *model b*. What is the null deviance for this model? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

Select the best model nested in *model b* in terms of AIC using a backward stepwise selection algorithm. In the following, we will refer to this selected model as *model e*. What is the median of the estimated probabilities for on time credit repayment for the training sample, based on this model? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

Select the best model nested in *model b* in terms of BIC using a backward stepwise selection algorithm. In the following, we will refer to this selected model als *model f*. Test *model f* and *model e* on the test data. What are the values for predicted residual sum of squares (PRESS) for both models ? **(requested precision: 4 digits)**

PRESS for *model e*

Zahl

0.5 Punkte

PRESS for *model f*

Zahl

Aufgabe 6

3.5 Punkte

The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. **Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.**

Task 26

3.5 Punkte

0.5 Punkte

Fit a probit regression model that predicts `Good.` using `Dist` as explanatory variable. In the following, we will refer to this model as *model 1*. What is the deviance of this model? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

What is the value of the predicted probability for a good kick when the distance is 47 yards for *model 1*? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

What is the 90 % profile confidence interval for the intercept of *model 1*? **(requested precision: 4 digits)**

lower bound of CI

0.5 Punkte

Fit a probit regression model that predicts `Good.` using `Dist` as explanatory variable. In the following, we will refer to this model as *model 1*. What is the deviance of this model? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

What is the value of the predicted probability for a good kick when the distance is 47 yards for *model 1*? **(requested precision: 4 digits)**

Zahl

0.5 Punkte

What is the 90 % profile confidence interval for the intercept of *model 1*? **(requested precision: 4 digits)**

lower bound of CI

Zahl

0.5 Punkte

upper bound of CI

Zahl

0.5 Punkte

Fit a probit model that predicts `Good.` using the attributes `Dist`, `Blk.`, `Pressure`, `Roof.type`, `Altitude` and `Field`. Select the model with smallest AIC using a backward stepwise selection algorithm. In the following, we will refer to this model as *model 2*. What is the number of parameters included in the selected model beyond the intercept?

upper bound of CI

Zahl

0.5 Punkte

Fit a probit model that predicts `Good.` using the attributes `Dist, Blk., Pressure, Roof.type, Altitude` and `Field`. Select the model with smallest AIC using a backward stepwise selection algorithm. In the following, we will refer to this model as *model 2*. What is the number of parameters included in the selected model beyond the intercept? **(requested precision: whole number)**

Zahl

0.5 Punkte

What are the percentages of correct classified observations for *model 1* and *model 2* when the threshold probability is 0.5? **(requested precision: 4 digits)**

percentage correct classified observations *model 1*

Zahl

0.5 Punkte

percentage correct classified observations *model 2*

Zahl