

dynexite.rwth-aachen.de

 Learner | Orbit | Dynexite
 01:26:59

Aufgabe 1 2 Punkte

Consider $X_1, \dots, X_n, n \in \mathbb{N}$, independently and identically distributed random variables with probability density and cumulative distribution function, respectively, given by

$$f(x; \alpha, \gamma) = \begin{cases} \alpha \gamma^\alpha x^{-(\alpha+1)}, & \gamma \leq x, \\ 0, & \text{else,} \end{cases}, \quad F(x; \alpha, \beta) = \begin{cases} 1 - \gamma^\alpha x^{-\alpha}, & \gamma \leq x \\ 0, & \text{else} \end{cases} \quad (1)$$

with real parameters $\alpha > 1, \gamma > 0$; that is, the (full) likelihood of the parameters for a sample $\mathbf{x} = (x_1, \dots, x_n)$ is given by

$$L(\alpha, \gamma; \mathbf{x}) = \begin{cases} \alpha^n \gamma^{n\alpha} \prod_{i=1}^n x_i^{-(\alpha+1)}, & \gamma \leq x_{(1)}, \\ 0, & \text{else,} \end{cases}$$

where $x_{(1)} = \min\{x_1, \dots, x_n\}$ denotes the minimum of the observed sample values.

Consider the profile likelihood approach, where γ is the nuisance parameter, while α is the parameter of interest. Denote by $\hat{\gamma}(\alpha)$ the nuisance parameter estimate of γ for fixed α .

Suppose we have observed the sample

$$x_1 = 6, x_2 = 4, x_3 = 3$$

of size $n = 3$.

Find the missing numerical values with a precision of two decimals. 2 Punkte

(a) Calculate $\hat{\gamma}(\alpha)$ and give the value of $\hat{\gamma}(\alpha)$ for $\alpha = 2$. 1 Punkt

(b) Calculate the profile maximum likelihood estimate for the parameter α based on the observed sample given above (with a precision of two decimals). 1 Punkt

$\hat{\alpha}^{profile} =$

1 2 3 4 5

A screenshot of a web browser window titled "dynexite.rwth-aachen.de". The URL bar shows "Kurs: Applied Data Analysis (VO) [21ss-11.40010]". The main content area is a Dynexite assignment page. At the top, there's a formula: $D(\alpha, \gamma, \omega) = \begin{cases} 0, & \text{else,} \end{cases}$. Below it, text states: "where $x_{(1)} = \min\{x_1, \dots, x_n\}$ denotes the minimum of the observed sample values." It continues: "Consider the profile likelihood approach, where γ is the nuisance parameter, while α is the parameter of interest. Denote by $\hat{\gamma}(\alpha)$ the nuisance parameter estimate of γ for fixed α ". A note says: "Suppose we have observed the sample". Below this, a sample is given: $x_1 = 6, x_2 = 4, x_3 = 3$. The text "of size $n = 3$ " follows. There are three questions listed as tasks:

- Task 1:** Find the missing numerical values with a precision of two decimals. (2 Punkte) - This task is marked as completed with a green checkmark.
- Task 2:** Calculate $\hat{\gamma}(\alpha)$ and give the value of $\hat{\gamma}(2)$. (1 Punkt) - The input field contains "3".
- Task 3:** Calculate the profile maximum likelihood estimate for the parameter α based on the observed sample given above (with a precision of two decimals). (1 Punkt) - The input field contains "3,06".

 At the bottom right of the assignment area, there's a button labeled "Nächste Aufgabe →". Below the assignment area, a message says "Alle Antworten wurden gespeichert!". At the very bottom, there are navigation links: "ÜBERSICHT" on the left and "ABGABE" on the right.

□ < > 🔍 dynexite.rwth-aachen.de 🌐 ↻

◀ Kurs: Applied Data Analysis (VO) [21ss-11.40010] ▶ Learner | Orbit | Dynexite 01:26:53 ⓘ

DYNEXITE

← Vorherige Aufgabe → Nächste Aufgabe

Aufgabe 2 **3 Punkte**

Let $Y \mid X = k \sim \mathcal{P}(k)$ be a Poisson distributed random variable conditional on the value of X . Further suppose that

$$P(X = k) = (1 - p)^k p, \quad k \in \mathbb{N}_0,$$

for $p \in (0, 1)$ with $E(X) = \frac{1-p}{p}$ and $\text{Var}(X) = \frac{1-p}{p^2}$.

Find the missing numerical values with a precision of two decimals. **3 Punkte**

(a) Derive the expectation and variance of Y for $p = \frac{1}{3}$. **1 Punkt**

$E(Y) =$

$\text{Var}(Y) =$

(b) Denote by p_{\min} the smallest $p \in (0, 1)$, such that $P(Y = 0) \geq \frac{1}{3}$. Find p_{\min} with a precision of two decimals. **1 Punkt**

Hint: Use the law of total probability.

$p_{\min} =$

1 2 3 4 5

ÜBERSICHT ABGABE

< > 🔍

dynexit.e.rwth-aachen.de

Kurs: Applied Data Analysis (VO) (21ss-11.40010)

Learner | Orbit | Dynexit

01:26:50

DYNEXITE

Let X_1, X_2, \dots, X_n be a i.i.d. binomial distributed random variable conditional on the value of X_1 . Further suppose that

$$P(X = k) = (1-p)^k p, \quad k \in \mathbb{N}_0,$$

for $p \in (0, 1)$ with $E(X) = \frac{1-p}{p}$ and $\text{Var}(X) = \frac{1-p}{p^2}$.

Find the missing numerical values with a precision of two decimals.

3 Punkte

(a) Derive the expectation and variance of Y for $p = \frac{1}{3}$.

$E(Y) =$

1 Punkt

$\text{Var}(Y) =$

1 Punkt

(b) Denote by p_{\min} the smallest $p \in (0, 1)$, such that $P(Y = 0) \geq \frac{1}{3}$. Find p_{\min} with a precision of two decimals.

Hint: Use the law of total probability.

$p_{\min} =$

← Vorherige Aufgabe → Nächste Aufgabe

Alle Antworten wurden gespeichert!

ÜBERSICHT

ABGABE

1 2 3 4 5

Screenshot of a web-based assignment interface for Dynexite.

The top navigation bar includes:

- Back and forward buttons
- Address bar: `dynexite.rwth-aachen.de`
- Search icon
- User icon
- Logout icon
- Course information: `Kurs: Applied Data Analysis (VO) [21ss-11.40010]`
- Page title: `DYNEXITE`
- Time: `01:26:47`

The main content area shows:

Aufgabe 3 (Aufgabe 3) **5 Punkte**

Consider a linear model with design matrix

$$X = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{3}} \\ -1 & \frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

and parameter vector $\beta = (\beta_1, \beta_2)'$. Denote the lasso and ridge regression estimators as solutions to the objective function

$$\min_{\beta \in \mathbb{R}^2} \left\{ \frac{1}{2} \|\mathbf{y} - X\beta\|^2 + \lambda \|\beta\|_q^q \right\}$$

for $q = 1$ or $q = 2$, respectively, as $\hat{\beta}^{lasso}$ and $\hat{\beta}^{ridge}$. Suppose we have observed

$$\mathbf{y} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Hint: The following tasks have to be solved theoretically (see related theoretical exercise). In particular, *do not use R*, since the objective function in `glmnet` - and thus the tuning parameter λ - are scaled differently than above.

Task: Find the missing numerical values with a precision of two decimals. **5 Punkte**

(a) Find the smallest value λ_{\min} , say, for the tuning parameter λ , such that at least one of the lasso estimates $\hat{\beta}_i^{lasso}$, $i \in \{1, 2\}$, equals exactly zero (with a precision of two decimal places). **1 Punkt**

Buttons for grading: 1, 2, **3**, 4, 5

Navigation buttons: Vorherige Aufgabe ←, Nächste Aufgabe →

□ < > 🔍 dynexite.rwth-aachen.de 🌐

Kurs: Applied Data Analysis (VO) [21ss-11.40010] Learner | Orbit | Dynexite 01:26:40

DYNEXITE

Find the missing numerical values with a precision of two decimal places. 5 Punkte

(a) Find the smallest value λ_{\min} , say, for the tuning parameter λ , such that at least one of the lasso estimates $\hat{\beta}_i^{\text{lasso}}$, $i \in \{1, 2\}$, equals exactly zero (with a precision of two decimal places). 1 Punkt

$\lambda_{\min} =$ 0,5

(b) Calculate the corresponding ridge regression estimates for the value of λ_{\min} derived in (a) with a precision two decimal places. 1 Punkt

(i) The corresponding ridge regression estimate of the first component is
 $\hat{\beta}_1^{\text{ridge}} =$ 0,33

(ii) The corresponding ridge regression estimate of the second component is
 $\hat{\beta}_2^{\text{ridge}} =$ -0,58

(c) Suppose the underlying model is normal, that is
 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

with $\boldsymbol{\epsilon} \sim N_2(\mathbf{0}, 4 \cdot I_2)$, where I_2 denotes the 2-dimensional identity matrix.

Consider the mean squared error (MSE) of the ridge estimator of the first component $\hat{\beta}_1$ of $\boldsymbol{\beta}$,

$$\text{MSE}(\hat{\beta}_1^{\text{ridge}}) = \text{bias}(\hat{\beta}_1^{\text{ridge}})^2 + \text{Var}(\hat{\beta}_1^{\text{ridge}}).$$

Let $\lambda = 1$ and denote by
 $(\hat{\beta}_1^{\text{lower}}, \hat{\beta}_1^{\text{upper}})$

1 2 3 4 5

ÜBERSICHT ABGABE

A screenshot of a web browser window. The address bar shows "dynexite.rwth-aachen.de". The main content area is a Dynexite exercise page for "Applied Data Analysis (VO) [21ss-11.40010]". The exercise is titled "(c) Suppose the underlying model is normal, that is" and is worth "1 Punkt". The problem statement is: "with $\epsilon \sim N_2(\mathbf{0}, 4 \cdot I_2)$, where I_2 denotes the 2-dimensional identity matrix. Consider the mean squared error (MSE) of the ridge estimator of the first component β_1 of β , $MSE(\hat{\beta}_1^{ridge}) = \text{bias}(\hat{\beta}_1^{ridge})^2 + \text{Var}(\hat{\beta}_1^{ridge})$ ". Below this, it says "Let $\lambda = 1$ and denote by $(\beta_1^{lower}, \beta_1^{upper})$ the interval of values for β_1 , for which $MSE(\hat{\beta}_1^{ridge}) < \text{Var}(\hat{\beta}_1^{LS})$, i.e., the MSE of $\hat{\beta}_1^{ridge}$ is smaller than the variance of the least squares estimator $\hat{\beta}_1^{LS}$ of β_1 . Find this interval (with a precision of two decimal places). Hint: Consult Exercise 27(a) and Theorem I.4.25 of the lecture to derive the distribution of the estimator $\hat{\beta}_1^{LS}$. $\beta_1^{lower} =$ " followed by a text input field with value "-3". Below this, another input field for $\beta_1^{upper} =$ has value "3". Navigation buttons at the bottom include "Vorherige Aufgabe" (left arrow), "Nächste Aufgabe" (right arrow), "ÜBERSICHT" (overview), and "ABGABE" (submit).

Screenshot of a web-based assignment interface for a Data Analysis course.

The top navigation bar includes:

- Browser icons (Back, Forward, Stop, Refresh)
- Address bar: `dynexite.rwth-aachen.de`
- Page title: `Kurs: Applied Data Analysis (VO) [21ss-11.40010]`
- User information: `Learner | Orbit | Dynexite`
- Time: `01:26:33`

The main content area shows:

- A navigation bar with `Vorherige Aufgabe` and `Nächste Aufgabe`.
- A section header: **Aufgabe 4** (**8 Punkte**)
- Text instructions: "Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong!"
- Task description: "Load the dataset `Boston` from the package `MASS` into your R workspace. Use `?Boston` to get information about the dataset. Let `n` denote the number of observations included in this dataset."
- Sub-task (a): "Set the seed to 2021. Split the dataset randomly into a training and a test dataset where the training dataset contains $n - 100$ observations. Fit a linear model based on the training data where `medv` is the response variable and `crim, zn, indus, nox, rm, age, dis, tax, ptratio, black, lstat` are the explanatory variables. What is the mean value of the resulting fitted values?"
Score: 1 Punkt
Input: `mean value (requested precision: 4 digits)`
Value: `22,3111`
- Sub-task (b): "Fit the penalized regression model with Lasso with tuning parameter $\lambda = 0.5$. What are the resulting estimates for the intercept and the coefficient of `dis`?"
Score: 2 Punkte
Input: `estimate of intercept (requested precision: 4 digits)`
Value: `10,1197`

At the bottom, there are navigation buttons for `ÜBERSICHT` (Overview) and `ABGABE` (Submission), along with a page footer showing page numbers 1 through 5.

□ < > 🔍 dynexite.rwth-aachen.de 🌐

📅 Kurs: Applied Data Analysis (VO) [21ss-11.40010] Learner | Orbit | Dynexite 01:26:28 ⓘ

DYNEXITE

(b) Fit the penalized regression model with Lasso with tuning parameter $\lambda = 0.5$. What are the resulting estimates for the intercept and the coefficient of `dis`? 2 Punkte

estimate of intercept (requested precision: 4 digits) 1 Punkt
10,1197

estimate of coefficient of `dis` (requested precision: 4 digits) 1 Punkt
-0,0615

(c) Fit the penalized regression model with Lasso using cross validation (CV). What are the values for the minimum of λ (denoted by λ_{min}) and λ_{1-SE} ? 2 Punkte

value for λ_{min} (requested precision: 4 digits) 1 Punkt
0,0227

value for λ_{1-SE} (requested precision: 4 digits) 1 Punkt
1,0293

(d) Fit the penalized regression model with Lasso with tuning parameter λ_{1-SE} from (c). What is the minimum value of the resulting estimated coefficients ? 1 Punkt

minimum value of resulting estimated coefficients (requested precision: 4 digits) 1 Punkt
-0,5996

(e) Fit the penalized regression model with Lasso with tuning parameter λ_{min} from (c). What is the minimum value of the resulting estimated coefficients ? 1 Punkt

ÜBERSICHT ABGABE

1 2 3 4 5

Kurs: Applied Data Analysis (VO) [21ss-11.40010] dynexite.rwth-aachen.de Learner | Orbit | Dynexite 01:26:26

DYNEXITE

(c) Fit the penalized regression model with Lasso using cross validation (CV). What are the values for the minimum of λ (denoted by λ_{min}) and λ_{1-SE} ? 2 Punkte

value for λ_{min} (requested precision: 4 digits) 1 Punkt
0,0227

value for λ_{1-SE} (requested precision: 4 digits) 1 Punkt
1,0293

(d) Fit the penalized regression model with Lasso with tuning parameter λ_{1-SE} from (c). What is the minimum value of the resulting estimated coefficients ? 1 Punkt

minimum value of resulting estimated coefficients (requested precision: 4 digits) 1 Punkt
-0,5996

(e) Fit the penalized regression model with Lasso with tuning parameter λ_{min} from (c). What is the minimum value of the resulting estimated coefficients ? 1 Punkt

minimum value of resulting estimated coefficients (requested precision: 4 digits) 1 Punkt
-10,6274

(f) Test the models fitted in (d) and (e) on the test dataset and compute the values for predicted residual sum of squares (PRESS). Which model has the lowest value of PRESS? Type in "1" for the model fitted in (d) and type in "2" for the model fitted in (e) (without quotation marks). 1 Punkt

model with lowest value of PRESS 1 Punkt

ÜBERSICHT ABGABE

1 2 3 4 5

Kurs: Applied Data Analysis (VO) [21ss-11.40010] | dynexite.rwth-aachen.de | Learner | Orbit | Dynexite

01:26:23

DYNEXITE

(d) Fit the penalized regression model with Lasso with tuning parameter λ_{1-SE} from (c). What is the minimum value of the resulting estimated coefficients ? 1 Punkt

minimum value of resulting estimated coefficients (requested precision: 4 digits)

-0,5996

(e) Fit the penalized regression model with Lasso with tuning parameter λ_{min} from (c). What is the minimum value of the resulting estimated coefficients ? 1 Punkt

minimum value of resulting estimated coefficients (requested precision: 4 digits)

-10,6274

(f) Test the models fitted in (d) and (e) on the test dataset and compute the values for predicted residual sum of squares (PRESS). Which model has the lowest value of PRESS? Type in "1" for the model fitted in (d) and type in "2" for the model fitted in (e) (without quotation marks). 1 Punkt

model with lowest value of PRESS

2

Vorherige Aufgabe ← Nächste Aufgabe →

Alle Antworten wurden gespeichert!

□ < > 🔍 dynexite.rwth-aachen.de 🌐

🔗 Kurs: Applied Data Analysis (VO) [21ss-11.40010] | Learner | Orbit | Dynexite 01:26:20

Vorherige
← Aufgabe

Aufgabe 5 7 Punkte

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

Load the dataset `UCBAdmissions` from the package `datasets` into your R workspace.

(a) Use `?UCBAdmissions` to get information about the dataset. Create a flat contingency table and add the marginal sums. What are the mean values for the row and the column marginals? 2 Punkte

mean of row marginals (requested precision: 1 digit)
1131,5

mean of column marginals (requested precision: 1 digit)
754,3

(b) Transform the flat contingency table from (a) into a data frame. Fit the saturated model that predicts the count variable `Freq`. What is the value for AIC of the saturated model? 1 Punkt

AIC (requested precision: 1 digit)
207,1

ÜBERSICHT 1 2 3 4 5 ABGABE


Kurs: Applied Data Analysis (VO) [21ss-11.40010]
dynexite.rwth-aachen.de
Learner | Orbit | Dynexite
01:26:16

DYNEXITE

(b) Transform the flat contingency table from (a) into a data frame. Fit the saturated model that predicts the count variable **Freq**. What is the value for AIC of the saturated model? 1 Punkt
AIC (requested precision: 1 digit) 1 Punkt
 207,1

(c) Starting with the saturated model of (b), use a backward selection algorithm to select the best nested model based on AIC. What is the value of the null deviance for this selected model? 1 Punkt
null deviance (requested precision: whole numbers) 1 Punkt
 2650

(d) Fit the model of no three way interactions that predicts the count variable **Freq**. What is the value for λ_i^X for X given by Gender and i chosen to be the category "Female"? 1 Punkt
 λ_i^X (requested precision: 4 digits) 1 Punkt
 0,6819

(e) What is the mean value for the pearsonian residuals for the model of no three way interactions of (d)? 1 Punkt
mean value (requested precision: 3 digits) 1 Punkt
 -0,035

(f) What is the mean value for the deviance residuals for the model of no three way interactions of (d)? 1 Punkt

1 2 3 4 **5**

Screenshot of a web-based assignment interface for Dynexite.

The top navigation bar shows the URL dynexite.rwth-aachen.de, the course name "Kurs: Applied Data Analysis (VO) [21ss-11.40010]", and the Dynexite logo.

The main content area displays a list of questions:

- (d) Fit the model of no three way interactions that predicts the count variable **Freq**. What is the value for λ_i^X for X given by Gender and i chosen to be the category "Female"? **1 Punkt**
Answer: λ_i^X (requested precision: 4 digits)
Value: 0,6819
- (e) What is the mean value for the pearsonian residuals for the model of no three way interactions of (d)? **1 Punkt**
Answer: mean value (requested precision: 3 digits)
Value: -0,035
- (f) What is the mean value for the deviance residuals for the model of no three way interactions of (d)? **1 Punkt**
Answer: mean value (requested precision: 3 digits)
Value: -1,205

At the bottom left, there is a button labeled "Vorherige Aufgabe" with a back arrow icon.

A message at the bottom center states: "Alle Antworten wurden gespeichert!" (All answers were saved!).

The bottom navigation bar includes links for "ÜBERSICHT" (Overview), "ABGABE" (Submission), and page numbers 1, 2, 3, 4, and 5, with the number 5 highlighted in blue.