Statistical Inference 1	Winter Term 2024/2025

Prof. Dr. Thomas Augustin, Prof. Dr. Christian Heumann
Sergio Buttazzo, Jan Anders
Sophie Hopp, Michael Speckbacher

Aufgabe	1	Bonus
Punkte	100	5
erzielt		

Problem Set 2

Hinweise:

1. Dieses Problem Set umfasst (inklusive dieser Seite) 3 Seiten und 1 Aufgabe in deutscher und englischer Fassung. Bei Unstimmigkeiten zwischen der deutschen Angabe und der englischen Version ist die Deutsche bindend.
2. Es können 100 Punkte erreicht werden. Lösungen, die einen besonders sauberen und einheitlichen Gesamteindruck machen, können bis zu fünf zusätzliche Bonuspunkte erhalten.
3. Sie können Ihre Antworten in deutscher oder englischer Sprache formulieren. Achten Sie auf die Nachvollziehbarkeit Ihrer Lösung und die Konsistenz Ihres Lösungswegs über die Teilaufgaben hinweg, eine lesbare Schrift und die Angabe verwendeter Quellen. Wenn Sie über das Vorlesungsmaterial hinausgehende Quellen verwenden, so stellen Sie bitte explizit den Bezug zu den Notationen und Konzepten der Vorlesung her!
4. Die Bearbeitungszeit beträgt drei Wochen. Die Abgabe ist bis zum 09.01.2025 um 23.59 CET auf der Moodle-Kursseite möglich. Bitte geben Sie Ihre Lösungen wenn möglich in einer pdf Datei ab, mit diesem Deckblatt als erster Seite. Im zugehörigen Etherpad auf Moodle können etwaige Fragen zur Bearbeitung gestellt werden.
5. Alle Ihnen zur Verfügung stehenden Hilfsmittel sind zugelassen. Dies schließt insbesondere auch die Zusammenarbeit mit anderen Studierenden mit ein. Beachten Sie, dass Sie jegliche Form fremder Hilfe und alle verwendeten Quellen zitieren müssen. Den Zitationsstil können Sie frei wählen; wir bestehen auf keinem speziellen Zitationsstandard, aber die Nachvollziehbarkeit der Quellenarbeit muss gewährleistet sein. Eine Verletzung dieser Regeln wird als Unterschleif gewertet.
6. Bei Unterschleif gibt es für dieses Problem Set 0 Punkte, und es erfolgt eine Meldung an das Prüfungsamt. Bei einem Verdacht auf Unterschleif sind Sie verpflichtet, bei der Aufklärung dieses Verdachts mitzuwirken. Wir behalten uns insbesondere vor, bei begründetem Verdacht etwaige Zweifel in einem persönlichen Fachgespräch zu klären.
7. Es darf nicht mit Rotstift bzw. roter Farbe geschrieben oder gezeichnet werden.
8. Geben Sie bitte Ihre Lösungen inklusive dem von Ihnen ausgefüllten und unterzeichneten Deckblatt ab und kennzeichnen Sie in jeder Aufgabe, welche Hilfmittel Sie verwendet haben/von und mit welchen Personen Sie Hilfe erhalten haben/zusammengearbeitet haben.

Bitte ausfüllen und unterschreiben!

Name, Vorname: ______________________

Matrikelnummer: ______________________

Studiengang und PO: ______________________

Ich bestätige, dass ich die Hinweise auf dem Deckblatt zur Kenntnis genommen habe und sie befolgt habe.

Unterschrift: ______________________

Aufgabe 1

Sei X eine Zufallsvariable mit der Dichtefunktion

$$f(x;\theta) = c(\alpha) \cdot e^{-(|x-\theta|^{\alpha})},$$

wobei $x, \theta \in \mathbb{R},\ 0 < \alpha \leq 2$ und $c(\alpha)$ eine Normierungskonstante ist, die nicht von x abhängt.

(a) Für welche Werte von α gehört diese Wahrscheinlichkeitsverteilung zur Exponentialfamilie? Begründen Sie Ihre Antwort. Hinweis: Es ist nicht erforderlich, $c(\alpha)$ zu berechnen; nehmen Sie einfach an, dass es eine Konstante für ein gegebenes α ist. 20 P.

Setzen Sie nun $\alpha = 1$ und $c(\alpha) = \frac{1}{2}$. Sei $\boldsymbol{X} = (X_1, \dots, X_n)^T$ eine Stichprobe von n identisch und unabhängig verteilten Zufallsvariablen mit dieser Dichte.

(b) Zeigen Sie, dass für jedes X_i gilt: $\mathbb{E}[X_i] = \theta$ und $\mathrm{Var}(X_i) = 2$. 10 P.

(c) Zeigen Sie, dass die Log-Likelihood der Stichprobe gegeben ist durch: 10 P.

$$\ell(\theta; x) = -n \log 2 - \sum_{i=1}^{n} |x_i - \theta|.$$

(d) Zeigen Sie, dass jeder Stichproben-<u>Median</u> ein Maximum-Likelihood-Schätzer (MLE) für θ ist. 20 P.

(e) Berechnen Sie die Fisher-Information und die Cramér-Rao-Schranke für θ. Schreiben Sie die asymptotische Verteilung des MLE auf. 15 P.

(f) Sei $\bar{x}$ der Stichprobenmittelwert. Ist dies ein unverzerrter Schätzer für θ? Bestimmen Sie die Varianz des Stichprobenmittelwerts. Welcher der beiden Schätzer aus (d) und (f) ist MSE-effizienter und warum? 15 P.

(g) Verwenden Sie das Ergebnis aus (e) und geben Sie ein 95%-Konfidenzintervall für θ an. 10 P.

English translation

Exercise 1

Let X be a random variable with density function

$$f(x;\theta) = c(\alpha) \cdot e^{-(|x-\theta|^{\alpha})}$$

with $x, \theta \in \mathbb{R}$, $0 < \alpha \leq 2$ and $c(\alpha)$ is a normalizing constant that does not depend on x.

(a) For which values of α is this probability distribution a member of the exponential family? Provide a justification for your answer. Note: You are not required to compute $c(\alpha)$, just assume it is some constant for a given α. 20 Pt.

Now, let $\alpha = 1$ and $c(\alpha) = \frac{1}{2}$. Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a sample of n identically and independently distributed random variables with this density.

(b) Show that, for every X_i, $\mathbb{E}[X_i] = \theta$ and $Var(X_i) = 2$. 10 Pt.

(c) Show that the log-likelihood of the sample is: 10 Pt.

$$\ell(\theta; x) = -n \log 2 - \sum_{i=1}^{n} |x_i - \theta|.$$

(d) Show that any sample median is a maximum likelihood estimator (MLE) for θ. 20 Pt.

(e) Calculate the Fisher information and the Cramer-Rao lower bound for θ. Write down the asymptotic distribution of the MLE. 15 Pt.

(f) Let $\bar{x}$ be the sample mean. Is it an unbiased estimator for θ? Evaluate the variance of the sample mean. Which of the two estimators from (d) and (f) is more MSE-efficient and why? 15 Pt.

(g) Using the result in (e), write down a 95% confidence interval for θ. 10 Pt.