10. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse B
Abgabe: 21.11.2019 23:59

WiSe 19/20
Prof. W. Rhode

Zeit	Raum	Abgabe im Moodle; Mails mit Betreff: [SMD19]
Di. 16–18	CP-03–150	dominik.baack@udo.edu und noah.biederbeck@udo.edu
Do.14–16	CP-03–123	kevin3.schmidt@udo.edu und maximilian.sackel@udo.edu
Fr. 16–18	CP-03–150	felix.geyer@udo.edu und rune.dominik@udo.edu

Aufgabe 25: *γ-Astronomie* **5 P.**

Bei einer typischen Messung in der γ-Astronomie wird das Teleskop auf eine Position (*on*-Position) gerichtet, an der eine γ-Strahlungs-Quelle vermutet wird. In der anschließenden Messung werden N_{on} Ereignisse über einen Zeitraum t_{on} aufgezeichnet. In den gemessenen Ereignissen N_{on} befinden sich sowohl Untergrund- als auch Signalphotonen. Um zu ermitteln, wie viel Untergrund vorhanden ist, wird ebenfalls an einer anderen Position ohne Quelle (*off*-Position) gemessen. Bei dieser Messung werden N_{off} Photonen-Ereignisse in einer Zeit t_{off} gemessen.

Um zu entscheiden, ob sich an der *on*-Position eine Quelle befindet, soll mit einem Likelihood-Quotienten-Test getestet werden, ob ein signifikanter Überschuss an Photonen über der Untergrunderwartung für die *on*-Position gemessen wurde (hier noch nicht, erst in Kapitel *Testen*).

Ziel dieser Aufgabe ist es, vorbereitend die richtige Likelihood-Funktion für den späteren Likelihood-Quotienten-Test aufzustellen.

Nutzen Sie für die Bearbeitung der Aufgabe die Ausdrücke:

- $\alpha = \frac{t_{\text{on}}}{t_{\text{off}}}$: Quotient der unterschiedlichen Messzeiten
- $b = \langle N_{\text{off}} \rangle$: Unbekannter Erwartungswert für die Zahl der Untergrundphotonen während der Messzeit t_{off}
- s: Unbekannter Erwartungswert für die Zahl der Signal Photonen während der Messzeit t_{on} aus der γ-Strahlungs-Quelle. Nicht zu verwechseln mit der gesamten Erwartung für die *on*-Position.

(a) Wie groß ist der Erwartungswert $\langle N_{\text{on}} \rangle$, ausgedrückt durch s, b und α?

(b) Welchen Wahrscheinlichkeitsverteilungen folgen N_{on} und N_{off}?
Tipp: Die gezählten Ereignisse kommen unabhängig voneinander im Detektor an.

(c) Wie sieht die Likelihoodfunktion $\mathcal{L}(b, s)$ für die Parameter b und s aus?

(d) Welche Werte $\hat{b}$ und $\hat{s}$ maximieren $\mathcal{L}$?
Tipp: Nutzen Sie die negative Likelihood-Funktion, dann wird die Rechnung einfacher.

(e) Berechnen Sie die Kovarianzmatrix von $\hat{b}$ und $\hat{s}$. Wie hängt die Kovarianzmatrix mit der Likelihood zusammen? Ist diese Art der Fehlerberechnung exakt?

Aufgabe 26: *Maximum-Likelihood* **5 P.**

Eine Zufallsvariable x soll einer Gleichverteilung

$$f(x) = \begin{cases} 1/b & 0 \leq x \leq b \\ 0 & x < 0 \quad \text{oder} \quad x > b \end{cases}$$

folgen.

(a) Bestimmen Sie einen Schätzer für den Parameter b mit der Maximum Likelihood Methode aus einer Stichprobe $x_1, x_2, \dots, x_n$.

(b) Ist diese Schätzung erwartungstreu? Wenn nein, wie kann das in diesem Fall korrigiert werden?

Aufgabe 27: *Likelihoodkurve* **5 P.**

Aus einer Poisson-Verteilung werden 3 Stichproben, nämlich die Zahlen 13, 8 und 9 entnommen.

(a) Berechnen Sie die negative Log-Likelihood-Funktion als Funktion des einzigen Parameters λ und stellen Sie sie graphisch dar.

(b) Bei welchem Wert von λ liegt das Minimum $-\ln \mathcal{L}_{\text{max}}$?

(c) Für welche Werte von λ nimmt $-\ln \mathcal{L}$ die Werte

$$\begin{aligned} &-\ln \mathcal{L}_{\text{max}} + \frac{1}{2} \quad , \\ &-\ln \mathcal{L}_{\text{max}} + 2 \quad \text{und} \\ &-\ln \mathcal{L}_{\text{max}} + \frac{9}{2} \end{aligned}$$

an und was sagen diese Werte aus?

(d) Vergleichen Sie diese Werte mit der Näherung über eine Taylor-Entwicklung 2. Ordnung, indem Sie die Näherung sowohl zusammen mit der Likelihood graphisch darstellen als auch die Werte aus **(c)** bestimmen. Wofür könnte die Näherung nützlich sein?

Aufgabe 28: *Regularisierte kleinste Quadrate* **5 P.**

Ein Kollege hat für Sie eine Verteilung gemessen. Diese Verteilung wird Teil einer Monte-Carlo-Simulation. Um besser mit der Verteilung arbeiten zu können, suchen Sie nach einer geeigneten Parametrisierung. Sie wissen, dass sich die Verteilung durch ein Polynom sechsten Grades gut beschreiben lässt. Jedoch ist die Messung stark verrauscht und Ihr Kollege war auch nur in der Lage acht Wertepaare (x, y) zu nehmen.

(a) Fitten sie mit der Methode der kleinsten Quadrate ein Polynom sechsten Grades an die Daten der Datei *aufg_a.csv*. Geben Sie die resultierenden Koeffizienten an und zeichnen Sie das gefittete Polynom und die Daten in eine Abbildung ein.

(b) Fitten sie mit der Methode der kleinsten Quadrate ein Polynom sechsten Grades an die Daten der Datei *aufg_a.csv* und nutzen Sie dabei zusätzlich die Regularisierung über die zweite Ableitung ($\Gamma = \sqrt{\lambda} C A$). Für die Regularisierungsstärke nutzen sie $\lambda \in (0.1, 0.3, 0.7, 3, 10)$. Geben Sie die resultierenden Koeffizienten an und zeichnen Sie das gefittete Polynome und die Daten in eine Abbildung.

Ihr Kollege macht sich die Mühe und fertigt 50 neue Messungen des Spektrums an.

(c) Fitten Sie mit der Methode der kleinsten Quadrate ein Polynom sechsten Grades an die Mittelwerte der Daten aus der Datei *aufg_c.csv*. Gewichten Sie die berechneten Mittelwerte mit dem Fehler des Mittelwerts. Nutzen Sie diese Gewichte beim Fitten. Zeichnen Sie das gefittete Polynom und die gemittelten Daten in eine Abbildung ein.