Fachbereich Mathematik
Prof. Dr. Michael Kohler
Dipl.-Math. Andreas Fromkorth
Dipl.-Inf. Jens Mehnert

TECHNISCHE UNIVERSITÄT DARMSTADT

SS 2009
17.07.09

Klausur

„Anal. u. Stoch. f. LaGM (PO 98), Teil 2: Stochastik, Stochastik f. LaGM (PO 05) Praktische Mathematik, Aufbaumodul Stoch. f. BSc. MCS, Einf. i. d. Stoch. f. BSc. Mathe, WiMathe, MCS, Econ.“

Name: .... | Vorname: ....

Matrikel-Nr.: .... | Studiengang: ....

Aufgabe	1	2	3	4	max	Note
Bearbeitet (ankreuzen)	☐	☐	☐	☐	3 (2) /4 Aufgaben	
Punktzahl	10	10	10	10	30 (20)	
erreichte Punktzahl						

Hinweise

1. Es sind **keine Hilfsmittel** zugelassen mit Ausnahme eines Fremdsprachenwörterbuchs.
2. Bei der Prüfung „Analysis und Stochastik für LaGM (PO 98)“ werden **zwei der folgenden vier Aufgaben** verlangt und gewertet. Für alle anderen im Titel genannten Prüfungen werden **drei der folgenden vier Aufgaben** verlangt und gewertet. Kreuzen Sie in obiger Tabelle an, welche Aufgaben gewertet werden sollen.
3. Bei der Prüfung „Analysis und Stochastik für LaGM (PO 98)“ beträgt die Prüfungsdauer 60 Minuten. Bei allen anderen im Titel genannten Prüfungen beträgt die Prüfungsdauer 90 Minuten.
4. Lösungsschritte und Teilergebnisse sind ausreichend zu begründen. Eine Angabe des Endergebnisses allein genügt nicht.
5. Es ist eigenes Papier zu verwenden. Verwenden Sie für jede Aufgabe einen neuen Papierbogen. Jedes Blatt ist **vor der Verwendung** deutlich mit Name, Matrikelnummer und Aufgabennummer zu versehen. Beachten Sie, dass nicht beschriftete Blätter nicht gewertet werden müssen.

6. Füllen Sie bevor Sie mit der eigentlichen Bearbeitung der Klausur beginnen die obigen Felder sorgfältig aus. **Aus dem Eintrag im Feld Studiengang muss hervorgehen, an welcher Prüfung Sie teilnehmen.**

7. Legen Sie Studien- und Lichbildausweis zur Kontrolle bereit.

1. Aufgabe (10 Punkte)

Gegeben sei eine zweidimensionale Messreihe

$$(x_1, y_1), \dots, (x_n, y_n)$$

vom Umfang n. Anstelle einer Geraden (wie bei der linearen Regression) könnte man analog auch eine Parabel

$$y = a + b \cdot x + c \cdot x^2$$

durch Minimierung von

$$F(a, b, c) := \sum_{i=1}^{n} \left(y_i - (a + b \cdot x_i + c \cdot x_i^2) \right)^2$$

an die Daten anpassen. Zeigen Sie (durch Nullsetzen geeigneter Ableitungen), dass die Werte a, b, c, für die $F(a, b, c)$ minimal wird, Lösungen des linearen Gleichungssystems

$$\begin{aligned} a + b \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + c \cdot \frac{1}{n} \sum_{i=1}^{n} x_i^2 &= \frac{1}{n} \sum_{i=1}^{n} y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + b \cdot \frac{1}{n} \sum_{i=1}^{n} x_i^2 + c \cdot \frac{1}{n} \sum_{i=1}^{n} x_i^3 &= \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^{n} x_i^2 + b \cdot \frac{1}{n} \sum_{i=1}^{n} x_i^3 + c \cdot \frac{1}{n} \sum_{i=1}^{n} x_i^4 &= \frac{1}{n} \sum_{i=1}^{n} x_i^2 \cdot y_i \end{aligned}$$

sind.

Hinweis: Sie müssen nicht zeigen, dass es sich bei der Lösung des Gleichungssystems wirklich um ein Minimum handelt.

Lösung: Wie in der Vorlesung müssen wir die partiellen Ableitungen nullsetzen. Dies ergibt

$$\frac{\partial}{\partial a} F(a, b, c) = \frac{\partial}{\partial b} F(a, b, c) = \frac{\partial}{\partial c} F(a, b, c) = 0.$$

Wir berechnen also die partiellen Ableitungen

$$\begin{aligned} 0 = \frac{\partial}{\partial a} F(a, b, c) &= \frac{\partial}{\partial a} \left(\sum_{i=1}^{n} \left((y_i - (a + bx_i + cx_i^2) \right)^2 \right) = \sum_{i=1}^{n} \frac{\partial}{\partial a} \left((y_i - (a + bx_i + cx_i^2) \right)^2 \right) \\ &= \sum_{i=1}^{n} 2 \left(y_i - (a + bx_i + cx_i^2) \right) \frac{\partial}{\partial a} \left(y_i - (a + bx_i + cx_i^2) \right) \\ &= -2 \sum_{i=1}^{n} \left(y_i - (a + bx_i + cx_i^2) \right) \\ &= -2 \left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i - \sum_{i=1}^{n} cx_i^2 \right) \\ \Rightarrow \qquad & \frac{1}{n} \sum_{i=1}^{n} y_i = a + b \frac{1}{n} \sum_{i=1}^{n} x_i + c \frac{1}{n} \sum_{i=1}^{n} x_i^2. \end{aligned}$$

$$
\begin{aligned}
0 = \frac{\partial}{\partial b} F(a,b,c) &= \frac{\partial}{\partial b}\left(\sum_{i=1}^{n}\left((y_i - (a + bx_i + cx_i^2)\right)^2\right) = \sum_{i=1}^{n}\frac{\partial}{\partial b}\left(\left(y_i - (a + bx_i + cx_i^2)\right)^2\right) \\
&= \sum_{i=1}^{n} 2\left(y_i - (a + bx_i + cx_i^2)\right)\frac{\partial}{\partial b}\left(y_i - (a + bx_i + cx_i^2)\right) \\
&= -2\sum_{i=1}^{n}\left(y_i - (a + bx_i + cx_i^2)\right)\cdot x_i \\
&= -2\left(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} a x_i - \sum_{i=1}^{n} b x_i^2 - \sum_{i=1}^{n} c x_i^3\right) \\
\Rightarrow \quad & \frac{1}{n}\sum_{i=1}^{n} x_i y_i = a\frac{1}{n}\sum_{i=1}^{n} x_i + b\frac{1}{n}\sum_{i=1}^{n} x_i^2 + c\frac{1}{n}\sum_{i=1}^{n} x_i^3.
\end{aligned}
$$

und

$$
\begin{aligned}
0 = \frac{\partial}{\partial c} F(a,b,c) &= \frac{\partial}{\partial c}\left(\sum_{i=1}^{n}\left((y_i - (a + bx_i + cx_i^2)\right)^2\right) = \sum_{i=1}^{n}\frac{\partial}{\partial c}\left(\left(y_i - (a + bx_i + cx_i^2)\right)^2\right) \\
&= \sum_{i=1}^{n} 2\left(y_i - (a + bx_i + cx_i^2)\right)\frac{\partial}{\partial c}\left(y_i - (a + bx_i + cx_i^2)\right) \\
&= -2\sum_{i=1}^{n}\left(y_i - (a + bx_i + cx_i^2)\right)\cdot x_i^2 \\
&= -2\left(\sum_{i=1}^{n} x_i^2 y_i - \sum_{i=1}^{n} a x_i^2 - \sum_{i=1}^{n} b x_i^3 - \sum_{i=1}^{n} c x_i^4\right) \\
\Rightarrow \quad & \frac{1}{n}\sum_{i=1}^{n} x_i^2 y_i = a\frac{1}{n}\sum_{i=1}^{n} x_i^2 + b\frac{1}{n}\sum_{i=1}^{n} x_i^3 + c\frac{1}{n}\sum_{i=1}^{n} x_i^4.
\end{aligned}
$$

Dies sind die gewünschten Gleichungen.
Anmerkung für Studenten ab dem 3. Semester: Um zu zeigen, dass es sich in der Tat um ein Minimum handelt, muss man jetzt noch nachrechnen, dass die zugehörige Hessematrix positiv definit ist.

2. Aufgabe (10 Punkte)

Ein Glücksrad bleibt nach dem Drehen rein zufällig auf einem von insgesamt 50 Feldern stehen. Bleibt es auf einem der 10 blau gefärbten Felder stehen, so wird ein Gewinn von 5 Euro ausgezahlt. Bleibt es auf einem der 5 grün gefärbten Felder stehen, so wird ein Gewinn von 10 Euro ausgezahlt. Und bleibt es auf dem *einzigen* roten Feld stehen, so wird ein Gewinn von 100 Euro ausgezahlt. Auf den übrigen 34 weiß gefärbten Feldern wird kein Gewinn ausgezahlt.

(a) Wie groß ist der Gewinn "im Mittel", und wie groß ist die "mittlere quadratische Abweichung" zwischen dem zufälligen Gewinn und dem Gewinn "im Mittel" ?

(b) Für einmaliges Drehen verlangt der Besitzer des Glücksrads einen Einsatz von 5 Euro. Damit beträgt sein Verdienst bei einmaligen Drehen $Y = 5 - X$, wobei X der ausgezahlte Gewinn ist. Wie groß ist sein Verdienst "im Mittel", und wie groß ist die "mittlere quadratische Abweichung" zwischen dem zufälligen Verdienst und dem Verdienst "im Mittel"?

(c) Der Besitzer betreibt sein Glücksrad einen Monat lang auf einem Jahrmarkt. In dieser Zeit drehen dabei $n = 6000$ Personen (unbeeinflusst voneinander) am Glücksrad. Bestimmen Sie mit Hilfe des Zentralen Grenzwertsatzes eine untere Schranke für den Verdienst des Besitzers in diesem Monat, die (ungefähr) mit Wahrscheinlichkeit 0.95 überschritten wird.

Hinweise: Seien $Y_1, \ldots, Y_n$ unabhängige Zufallsvariablen mit $\mathbf{P}_{Y_i} = \mathbf{P}_Y$ $(i = 1, \ldots, n)$, wobei Y die in b) eingeführte ZV ist. Gesucht ist dann $x \in \mathbb{R}$ mit $\mathbf{P}[\sum_{i=1}^n Y_i \geq x] \approx 0.95$.
Ist Z eine standardnormalverteilte Zufallsvariable, so gilt

$$\mathbf{P}[Z \geq -1.65] \approx 0.95.$$

Eine numerische Berechnung des Ergebnisses wird nicht verlangt. Es sollte aber in einer Form angegeben werden, die eine sofortige Lösung mittels Taschenrechner ermöglicht.

Lösung: Siehe Übung.

3. Aufgabe (10 Punkte)

Sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und seien $X, Y : \Omega \to \mathbb{R}$ reelle Zufallsvariablen, für die $\mathbf{E}X, \mathbf{E}Y$ und $\mathbf{E}(X \cdot Y)$ existieren. Dann heißt

$$Cov(X, Y) = \mathbf{E}\left((X - \mathbf{E}X) \cdot (Y - \mathbf{E}Y)\right)$$

die *Kovarianz* von X und Y.

(a) Zeigen Sie: Für $a_1, a_2, b_1, b_2 \in \mathbb{R}$ gilt:

$$Cov(a_1 X + b_1, a_2 Y + b_2) = a_1 \cdot a_2 \cdot Cov(X, Y).$$

(b) Zeigen Sie:

$$Cov(X, Y) = \mathbf{E}(X \cdot Y) - \mathbf{E}X \cdot \mathbf{E}Y.$$

(c) Seien (X, Y), (X_1, Y_1), $(X_2, Y_2), \ldots$ unabhängig identisch verteilt. Ist der Schätzer

$$\hat{C}(X, Y) = \frac{1}{n} \sum_{i=1}^n X_i \cdot Y_i - \bar{X} \cdot \bar{Y} \quad \text{mit } \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \text{ und } \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$$

ein konsistenter bzw. ein erwartungstreuer Schätzer für $Cov(X, Y)$? Begründen Sie ihre Antwort.

Hinweise: Für $i \neq j$ sind X_i und Y_j unabhängig und daher gilt in diesem Fall

$$\mathbf{E}(X_i \cdot Y_j) = \mathbf{E}X_i \cdot \mathbf{E}Y_i.$$

(d) Geben Sie (ohne Begründung) einen konsistenten **und** erwartungstreuen Schätzer für $Cov(X,Y)$ an.

Lösung:

(a) Mit der Linearität des Erwartungswerts folgt:

$$\begin{aligned} Cov(a_1X+b_1, a_2Y+b_2) &= \mathbf{E}((a_1X+b_1-\mathbf{E}(a_1X+b_2))(a_2Y+b_2-\mathbf{E}(a_2Y+b_2))) \\ &= \mathbf{E}((a_1(X-\mathbf{E}(X))a_2(Y-\mathbf{E}(Y)))) \\ &= a_1a_2\mathbf{E}((X-\mathbf{E}(X))(Y-\mathbf{E}(Y))) \\ &= Cov(X,Y). \end{aligned}$$

(b) Hier nutzt man ebenfalls die Linearität des Erwartungswerts:

$$\begin{aligned} Cov(X,Y) &= \mathbf{E}((X-\mathbf{E}X)(Y-\mathbf{E}Y)) \\ &= \mathbf{E}(XY-Y\mathbf{E}X-X\mathbf{E}Y+(\mathbf{E}X)(\mathbf{E}Y)) \\ &= \mathbf{E}(XY)-\mathbf{E}(Y\mathbf{E}X)-\mathbf{E}(X\mathbf{E}Y)+(\mathbf{E}X)(\mathbf{E}Y) \\ &= \mathbf{E}(XY)-2(\mathbf{E}X)(\mathbf{E}Y)+(\mathbf{E}X)(\mathbf{E}Y) \\ &= \mathbf{E}(XY)-(\mathbf{E}X)(\mathbf{E}Y) \end{aligned}$$

(c) Nach dem starken Gesetz der großen Zahlen gitl

$$\bar{X} \to X \qquad f.s.(n \to \infty),$$

$$\bar{Y} \to Y \qquad f.s.(n \to \infty)$$

und

$$\frac{1}{n}\sum_{i=1}^{n} X_iY_i \to XY \qquad f.s.(n \to \infty).$$

Dabei ging ein, dass aufgrund von $(X,Y),(X_1,Y_1),\ldots$ u.i.v. auch $X,X_1,\ldots$ u.i.v., $Y,Y_1,\ldots$ u.i.v sowie $XY,X_1Y_1,\ldots$ u.i.v.. Insgesamt gilt dann

$$\hat{C}(X,Y) \to C(X,Y) \qquad f.s.(n \to \infty).$$

Somit ist der Schätzer konsistent.

Der Schätzer ist aber nicht erwartungstreu:

$$\begin{aligned}
\mathbf{E}(\hat{C}(X,Y)) &= \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_iY_i - \bar{X}\bar{Y}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}(X_iY_i) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}(X_iX_j) \\
&= \mathbf{E}(XY) - \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\mathbf{E}(X_iY_j) + \sum_{i=1}^{n}\mathbf{E}(X_iY_i)\right) \\
&= \mathbf{E}(XY) - \frac{1}{n^2}\left(n(n-1)\mathbf{E}(X)\mathbf{E}(Y) + n\mathbf{E}(XY)\right) \\
&= \frac{n-1}{n}\left(\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)\right) = \frac{n-1}{n}Cov(X,Y).
\end{aligned}$$

Dabei wurde zunächst die Lineariatät des Erwartungswerts ausgenutzt und danach noch verwendet, dass $(X,Y),(X_1,Y_1),\ldots$ u.i.v..

(d) die Rechnugn im vorigen Teil zeigt, dass

$$\bar{C}(X,Y) = \frac{n}{n-1}\hat{C}(X,Y)$$

ein konsistenter und Erwartungstreuer Schätzer für die Covarianz ist.

4. Aufgabe (10 Punkte)

Bei einer Lotterie gibt es drei Sorten von Losen. Zieht man ein Los der Sorte 1 bzw. 2 bzw. 3, so bekommt man keinen Preis bzw. einen Trostpreis bzw. einen Hauptpreis. Der Anteil der Lose der Sorte 1 bzw. 2 bzw. 3 sei p_1 bzw. p_2 bzw. p_3, wobei $p_1, p_2, p_3 \in [0,1]$ mit $p_1 + p_2 + p_3 = 1$. Es kann davon ausgegangen werden, dass sich der Anteil der Lose durch Ziehen eines Loses nicht bzw. nur vernachlässigbar ändert.

(a) Wie groß darf der Anteil p_1 der Lose der Sorte 1 höchstens sein, dass man bei Kauf von zehn Losen mit Wahrscheinlichkeit größer oder gleich 0.99 mindestens einen Preis (Trost- oder Hauptpreis) bekommt ?

Eine numerische Berechnung des Ergebnisses wird nicht verlangt. Es sollte aber in einer Form angegeben werden, die eine sofortige Lösung mittels Taschenrechner ermöglicht.

(b) Um die unbekannten Anteile p_1, p_2 und p_3 der verschiedenen Losarten zu schätzen, kauft Dozent K. $n = 1000$ Lose und beobachtet, dass darunter $n_1 = 917$ bzw. $n_2 = 74$ bzw. $n_3 = 9$ Lose vom Typ 1 bzw. 2 bzw. 3 sind. Bestimmen Sie die zu diesen Daten gehörende Maximum-Likelihood-Schätzung von (p_1, p_2, p_3).

Hinweise: Hier ist p_3 gegeben durch $1 - p_1 - p_2$. Seien $x_1, \ldots, x_n \in \{1,2,3\}$ die Typen der n gekauften Lose. Berechnen Sie die Likelihood-Funktion als Funktion

$$L((p_1,p_2)) := L((p_1,p_2); x_1, \ldots, x_n).$$

Lösung: a)

$$
\begin{aligned}
& P[\text{Bei Kauf von 10 Losen mindestens 1 Preis}] \\
=& 1 - P[\text{Bei Kauf von 10 losen kein Preis}] \\
=& 1 - P[\text{1. Los Sorte 1,...,10. Los Sorte 1}] \\
\overset{\text{unabh.}}{=}& 1 - P[\text{1. Los Sorte 1}] \cdot \ldots \cdot [\text{10. Los Sorte 1}] \\
=& 1 - p_1^{10}
\end{aligned}
$$

Damit diese Wk. mindestens 0.99 ist, muss gelten

$$
\begin{aligned}
1 - p_1^{10} \geq& \ 0.99 \\
\Leftrightarrow \quad p_1^{10} \leq& \ 0.01 \\
\Leftrightarrow \quad p_1 \leq& \ 0.01^{1/10}
\end{aligned}
$$

b) Ist $x \in \{1, 2, 3\}$, so gilt

$$
\begin{aligned}
& P[\text{Kauf eines Loses ergibt Los vom Typ x}] \\
=& \begin{cases} p_1 & \text{falls } x = 1 \\ p_2 & \text{falls } x = 2 \\ p_3 & \text{falls } x = 3 \end{cases} \\
=& p_1^{1_{[x=1]}} \cdot p_2^{1_{[x=2]}} \cdot p_3^{1_{[x=3]}} \\
=& p_1^{1_{[x=1]}} \cdot p_2^{1_{[x=2]}} \cdot (1 - p_1 - p_2)^{1_{[x=3]}}
\end{aligned}
$$

Damit gilt für die Likelihood-Funktion:

$$
\begin{aligned}
L((p_1, p_2)) &= \prod_{i=1}^{n} p_1^{1_{[x_i=1]}} \cdot p_2^{1_{[x_i=2]}} \cdot (1 - p_1 - p_2)^{1_{[x_i=3]}} \\
&= p_1^{\sum_{i=1}^n 1_{[x_i=1]}} \cdot p_2^{\sum_{i=1}^n 1_{[x_i=2]}} \cdot (1 - p_1 - p_2)^{\sum_{i=1}^n 1_{[x_i=3]}} \\
&= p_1^{n_1} \cdot p_2^{n_2} \cdot (1 - p_1 - p_2)^{n_3}
\end{aligned}
$$

Maximum- Likelihood-Schätzer von (p_1, p_2) ist der Wert, der diese Likelihood-Funktion maximiert. Also muss gelte

$$
\begin{aligned}
0 &= \frac{\partial}{\partial p_1} L((p_1, p_3)) \\
&= p_1^{n_1 - 1} \cdot p_2^{n_2} \cdot (n_1 \cdot (1 - p_1 - p_2) - n_3 \cdot p_1) \\
&\rightsquigarrow n_1 = (n_1 + n_3) p_1 + n_1 p_2 (*_1)
\end{aligned}
$$

und analoge Rechnung (in einer Klausur ausrechnen!) liefert

$$\rightsquigarrow n_2 = n_2 p_1 + (n_2 + n_3) p_2 (*_2)$$

$*_1(-n_2) + *_2(n_1 + n_3)$ ergibt:

$$\begin{aligned} & -n_1n_2 + n_2(n_1 + n_3) \\ = \quad & (-n_1n_2 + (n_2 + n_3)(n_1 + n_3))| \cdot p_2 \\ \Rightarrow \quad & \hat{p_2} = \tfrac{n_1n_2+n_2n_3-n_1n_2}{n_1n_3+n_2n_3+n_3n_3} = \tfrac{n_2}{n_1+n_2+n_3} = \tfrac{n_2}{n} \end{aligned}$$

In $*_2$ eingesetzt ergibt sich

$$\hat{p_1} = \text{Zwischenschritte ausrechnen!} = \frac{n_1}{n}$$

Damit

$$\hat{p_3} = 1 - \hat{p_1} - \hat{p_2}$$

Das Ergebnis ist damit $\hat{p_1} = \frac{917}{1000}$, $\hat{p_2} = \frac{74}{1000}$ und $\hat{p_3} = \frac{9}{1000}$.