Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

1. Übungsblatt

1. Es seien X und Y reellwertige Zufallsvariablen, $\mathbb{E}\left[|X|^2\right] < \infty$. Die gemeinsame Verteilung von X und Y möge die Dichte $f^{X,Y}$ bezüglich des Lebesgue-Maßes besitzen.
   Man bezeichnet
   $$f^{X|Y=y}(x) := \frac{f^{X,Y}(x,y)}{\int f(z,y)\,\mathrm{d}z}$$
   als *bedingte Dichte von* X*, gegeben* $Y = y$.

   a) Betrachte die messbare, reellwertige Funktion
   $$g(y) := \int x f^{X|Y=y}(x)\,\mathrm{d}x.$$
   Zeige, dass $g(Y)$ die definierenden Eigenschaften der bedingten Erwartung von X, gegeben Y, erfüllt. Es gilt also $g(Y) = \mathbb{E}[X|Y]$.

   b) X und Y mögen gemeinsam normalverteilt sein mit Parametern Σ und μ. Gib die bedingte Erwartung von X, gegeben $Y = y$, explizit an.

2. Es seien X und Y reellwertige Zufallsvariablen, $\mathbb{E}[|X|] < \infty$. X und Y mögen gemeinsam verteilt sein gemäß einer Dichte $f^{X,Y}$. $m(y)$ heißt ein *bedingter Median von* X*, gegeben* $Y = y$, falls $m(y)$ ein Median der bedingten Verteilung von X, gegeben $Y = y$ ist, falls also gilt:
   $$\int_{m(y)}^{\infty} f^{X|Y=y}(x)\,\mathrm{d}x = 1/2 \quad \text{und} \quad \int_{-\infty}^{m(y)} f^{X|Y=y}(x)\,\mathrm{d}x = 1/2.$$
   Zeige, dass die Zufallsvariable $m(Y)$ die Minimalitätseigenschaft
   $$\mathbb{E}\left[|X - m(Y)|\right] = \inf_h \mathbb{E}\left[|X - h(Y)|\right]$$
   besitzt, wobei das Infimum über allen reellwertigen, messbaren Funktionen h betrachtet wird.

3. Ist $A \in \mathbb{R}^{n\times k}$, dann nennt man $B \in \mathbb{R}^{k\times n}$ eine *Moore-Penrose-Inverse* von A, wenn gilt:

   - $ABA = A$ und $BAB = B$,
   - AB und BA sind symmetrisch.

(a) Sei $k \leqslant n$ und $A \in \mathbb{R}^{n \times k}$ eine Matrix mit vollem Rang k. Zeige, dass $A^\top A$ invertierbar ist und dass $(A^\top A)^{-1} A^\top$ eine Moore-Penrose-Inverse von A ist.

(b) Sei $A \in \mathbb{R}^{n \times k}$ und $b \in \mathbb{R}^n$. Sei A^+ eine Moore-Penrose-Inverse von A. Zeige: Wenn das Gleichungssystem $Ax = b$ lösbar ist, dann ist $A^+ b$ eine Lösung und hat unter allen Lösungen die kleinste euklidische Norm.

4. Bei acht Absolventen werden anhand einer Befragung die Studiendauer und das Einstiegsgehalt (in 1000€) ermittelt:

Studiendauer x_i	10	9	11	9	11	12	10	11
Einstiegsgehalt Y_i	35	35	34	36	41	39	40	38

(a) Modelliere dies als ein lineares Modell und bestimme die Regressionsgerade. Zeichne Messwerte und Regressionsgerade in ein geeignetes Koordinatensystem ein.

(b) Es stellt sich heraus, dass die ersten Vier ein anderes Fach studiert haben als die anderen Vier. Bestimme die Regressionsgraden für beide Studienfächer getrennt und zeichne sie ein.

(c) Wie erklären Sie die unterschiedlichen Ergebnisse in (a) und (b)?

Abgabe vor der Vorlesung am Dienstag, den 26.04.11.

Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

2. Übungsblatt

1. Formuliere und beweise den Satz von Gauß-Markov für das lineare Modell mit allgemeiner Kovarianzmatrix $\Sigma > 0$.

2. Beweise für Entscheidungsregeln ρ basierend auf einem statistischen Experiment $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ mit Verlustfunktion l:

   (a) Ist ρ minimax und eindeutig in dem Sinn, dass jede andere Minimax-Regel die gleiche Risikofunktion besitzt, so ist ρ zulässig.

   (b) Ist ρ zulässig mit konstanter Risikofunktion, so ist ρ minimax.

   (c) Ist ρ eine Bayesregel (bzgl. π) und eindeutig in dem Sinn, dass jede andere Bayesregel (bzgl. π) die gleiche Risikofunktion besitzt, so ist ρ zulässig.

   (d) Die Parametermenge Θ bilde einen metrischen Raum mit Borel-σ-Algebra $\mathcal{F}_\Theta$. Ist ρ eine Bayesregel (bzgl. π), so ist ρ zulsïg, falls (i) $R_\pi(\rho) < \infty$; (ii) für jede nichtleere offene Menge U in Θ gilt $\pi(U) > 0$; (iii) für jede Regel ρ' mit $R_\pi(\rho') \leq R_\pi(\rho)$ ist $\theta \mapsto R(\theta, \rho')$ stetig.

3. Eine Krankheit kommt bei ca. $0,1\%$ der Bevölkerung vor. Ein Test zur Erkennung der Krankheit führt bei 97% der Kranken, aber auch bei 2% der Gesunden zu einer Reaktion. Auf Grund des Tests wird eine Person als krank bzw. gesund klassifiziert. Mit $\ell_0 \geqslant 0$ (bzw. $\ell_1 \geqslant 0$) werde der Verlust bei der Klassifizierung *krank* (bzw. *gesund*) eines gesunden (bzw. kranken) Patienten bewertet. Formuliere dies als Bayessches Entscheidungsproblem und gib eine Bayes-optimale Entscheidungsregel in Abhängigkeit von ℓ_0, ℓ_1 an.

4. a) Es sei Y gemäß dem Modell

$$Y = \mu + \varepsilon$$

verteilt mit $\mu \in \mathbb{R}^n$ und unabhängigen, $\mathcal{N}(0, \sigma^2)$-verteilten Fehlern.
Man betrachtet den Kleinste-Quadrate-Schätzer $\hat{\beta}$ im misspezifizierten linearen Modell

$$Y = X\beta + \varepsilon$$

mit $X \in \mathbb{R}^{n\times p}, \beta \in \mathbb{R}^p$.
Zeige für den Vorhersagefehler die Darstellung

$$\mathbb{E}\left[|X\hat{\beta} - \mu|^2\right] = |(E_n - \Pi_X)\mu|^2 + p\sigma^2.$$

b) Es sei Y verteilt gemäß

$$Y_i = a_0 + a_1 x_i + a_2 x_i^2 + \varepsilon_i,\ i = 1, \cdots, n.$$

Man betrachtet den Kleinste-Quadrate-Schätzer im Modell

$$Y_i = a_0 + a_1 x_i + \varepsilon_i,\ i = 1, \cdots, n.$$

Bestimme den Vorhersagefehler.

Abgabe vor der Vorlesung am Dienstag, den 03.05.11.

Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

3. Übungsblatt

1. Die Beta-Verteilung $B(a,b)$ auf $[0,1]$ ist gegeben durch die Dichte

$$f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad x \in (0,1),$$

wobei $a,b > 0$ und Γ die Gamma-Funktion bezeichnet. $B(a,b)$ hat Erwartungswert $\mu_{a,b} = \frac{a}{a+b}$ und Varianz $\sigma^2_{a,b} = \frac{ab}{(a+b)^2(a+b+1)}$.

   (a) Skizziere $f_{a,b}$ für $(a,b) \in \{0.5; 1; 10\}^2$ (Computereinsatz gestattet).

   (b) Es sei eine $\mathrm{Bin}(n,p)$-verteilte math. Stichprobe X gegeben, wobei $n \geqslant 1$ bekannt ist sowie p gemäß $B(a,b)$ a priori verteilt ist. Zeige, daß die bedingte Dichte von p gegeben $X = x$ zur Beta-Verteilung $B(a+x, b+n-x)$ gehört.

   (c) Schließe, dass der Bayesschätzer unter quadratischem Risiko gegeben ist durch $\hat{p}_{a,b} = \frac{a+X}{a+b+n}$. Bestimme sein quadratisches Risiko als Funktion von p und sein zugehöriges Bayesrisiko.

2. Gegeben sei das gewöhnliche lineare Modell $Y = X\beta + \epsilon$ mit der Kovarianzmatrix $\Sigma = \sigma^2 E_n$. In der *ridge regression* verwendet man den Schätzer $\hat{\beta}_a = (X^\top X + a^2 E_k)^{-1} X^\top Y$. Die a-priori-Verteilung π von β sei eine zentrierte Normalverteilung mit Varianz $\eta^2 E_k$. Zeige: Für quadratisches Risiko ist der Bayes-optimale Schätzer $\hat{\beta}_\pi$ gleich dem ridge-regression-Schätzer $\hat{\beta}_{\frac{\sigma}{\eta}}$.

3. Wenn man in die Bayesformel statt einer Dichte $f_T(\theta)$ eine nichtnegative, messbare Funktion $f_T(\theta)$ einsetzt und $f_{T|X=x}(\theta)$ weiterhin wohldefiniert ist, so ergibt sich aus der a-posteriori-Verteilung ein *verallgemeinerter Bayesschätzer*. Es sei nun $X_1, \ldots, X_n$ eine $N(\mu, E_d)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}^d$ unbekannt.

   (a) Zeige: $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ ist ein verallgemeinerter Bayesschätzer von μ zum quadratischen Risiko bzgl. des Lebesguemaßes als verallgemeinerter a-priori-Verteilung.

   (b) Berechne den verallgemeinerten Bayesschätzer $\hat{\mu}_{a,b}$ zum quadratischen Risiko für $d = 1$ und $f_T(\theta) = \mathbf{1}_{(a,b)}(\theta)$ mit $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$. Zeichne $\hat{\mu}_{0,1}$ für $n = 1$ als Funktion von $\bar{X}$.

4. Gegeben sei $X \sim N(\mu, \sigma^2 E_d)$ mit $\sigma > 0$ bekannt und $\mu \in \mathbb{R}^d$ unbekannt.

   (a) Zeige: Soll in einem statistischen Experiment $g(\theta) \in \mathbb{R}^d$ durch $\hat{g}$ geschätzt werden, so gilt die *Bias-Varianz-Zerlegung*:

   $$\mathbb{E}_\theta[|\hat{g} - g(\theta)|^2] = |\,\mathbb{E}_\theta[\hat{g}] - g(\theta)|^2 + \mathbb{E}_\theta[|\hat{g} - \mathbb{E}_\theta[\hat{g}]|^2]$$

   (b) Berechne die Bias-Varianz-Zerlegung für $\hat{\mu}_\alpha = \alpha X$, $\alpha \in \mathbb{R}$, und zeige, dass $\alpha_{\text{Orakel}} := 1 - \frac{\sigma^2 d}{|\mu|^2 + \sigma^2 d}$ das quadratische Risiko minimiert, falls μ der wahre Parameter ist.

   (c) Wähle $R > 0$. Weise nach, dass $|X|^2$ ein erwartungstreuer Schätzer von $|\mu|^2 + \sigma^2 d$ ist und setze $\hat{\alpha} := 1 - \frac{\sigma^2 d}{|X|^2}$. Schließe durch Berechnen von $\text{Var}(|X|^2)$, dass $\forall \epsilon > 0 \; \exists K > 0 :$

   $$\mathbb{P}_\mu\left(\left|\frac{|X|^2}{\sigma^2 d} - \frac{|\mu|^2 + \sigma^2 d}{\sigma^2 d}\right| \geqslant \frac{K}{\sqrt{d}}\right) \leqslant \epsilon, \quad \forall d \geqslant 1 \; \forall \mu \in \mathbb{R}^d \text{ mit } |\mu| \leqslant R.$$

   Weise nach, dass $\forall \epsilon > 0 \; \exists K' > 0 :$

   $$\mathbb{P}_\mu(|\hat{\alpha} - \alpha_{\text{Orakel}}| > K'\, d^{-1/2}) \leqslant \epsilon, \quad \forall d \geqslant 1 \; \forall \mu \in \mathbb{R}^d \text{ mit } |\mu| \leqslant R.$$

   Folgere, dass insbesondere für $|\mu| \leqslant R$ die Normen $|\hat{\alpha} - \alpha_{\text{Orakel}}|$ für $d \to \infty$ stochastisch gegen 0 konvergieren.

Abgabe vor der Vorlesung am Dienstag, den 10.05.11

Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

4. Übungsblatt

1. a) Es sei $g : \Theta \to A \subseteq \mathbb{R}$ und $\ell(\theta, \rho) := (g(\theta) - \rho)^2$ der quadratische Verlust. Zeige: Eine Entscheidungsregel $\hat{g} : \mathcal{X} \to \mathcal{A}$ mit $\mathbb{E}_\theta[\hat{g}^2] < \infty$ und $\mathbb{E}_\theta[\hat{g}] \in g(\Theta)$ für alle $\theta \in \Theta$ ist genau dann unverzerrt, wenn sie erwartungstreu ist.

   b) Es sei $\Theta = \Theta_0 \dot\cup \Theta_1$ und $A = [0, 1]$.
   Zeige: Für den Verlust $\ell(\theta, a) = l_0 a 1_{\Theta_0}(\theta) + l_1(1 - a)1_{\Theta_1}(\theta)$ ist eine Entscheidungsregel ρ (ein randomisierter Test von $H_0 : \theta \in \Theta_0$ gegen $H_1 : \theta \in \Theta_1$) genau dann unverzerrt, wenn sie zum Niveau $\alpha := \frac{l_1}{l_0+l_1}$ unverfälscht ist, d.h.

$$\forall \theta \in \Theta_0 : \mathbb{E}_\theta[\rho] \le \alpha, \ \forall \theta \in \Theta_1 : \mathbb{E}_\theta[\rho] \ge \alpha.$$

2. Es sei $X_1, \ldots, X_n$ eine $N(\mu, E_d)$-verteilte mathematische Stichprobe. Der James-Stein-Schätzer mit positivem Gewicht ist definiert als $\hat{\mu}_{JS+} = \big(1 - \frac{d-2}{n|\overline{X}|^2}\big)_+ \overline{X}$. Beweise für alle $d \geqslant 3$ und $\mu \in \mathbb{R}^d$ schrittweise folgenden Risikovergleich mit dem klassischen James-Stein-Schätzer:

$$\mathbb{E}_\mu[|\hat{\mu}_{JS+} - \mu|^2] < \mathbb{E}_\mu[|\hat{\mu}_{JS} - \mu|^2].$$

   (a) Die Abschätzung ist korrekt für $\mu = 0$.

   (b) Die Abschätzung folgt aus der Ungleichung $\mathbb{E}_\mu[\mu_i \overline{X}_i |G| \mathbf{1}_{\{G \leqslant 0\}}] > 0$ für $G = 1 - \frac{d-2}{n|\overline{X}|^2}$ und alle $i = 1, \ldots, d$ mit $\mu_i \neq 0$.

   (c) Für $a > 0$ und $\mu_i \neq 0$ gilt $\mathbb{E}_\mu[\mu_i \overline{X}_i \,|\, (\overline{X}_i)^2 = a^2] = a\mu_i \tanh(na\mu_i) > 0$. Dies ergibt die Ungleichung in (b) durch Einfügen einer auf $((\overline{X}_1)^2, \ldots, (\overline{X}_d)^2)$ bedingten Erwartung.

3. Es sei $X_1, \ldots, X_n$ eine $N(\mu, 1)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}$ unbekannt.

   (a) Gib das zugehörige statistische Experiment auf $\mathcal{X} = \mathbb{R}^n$ an und zeige, dass es vom Produktmaß $N(0, 1)^{\otimes n}$ dominiert wird.

   (b) Bestimme die Likelihoodfunktion für das dominierende Maß in (a). Welcher Wert $\mu \in \mathbb{R}$ maximiert die Likelihoodfunktion zu gegebenem $x \in \mathbb{R}^n$ (dies ist der Maximum-Likelihood-Schätzer bei Beobachtung $X = x$)?

4. Beweise oder widerlege die Aussage, dass folgende Verteilungen Exponentialfamilien bilden. Bestimme gegebenenfalls den natürlichen Parameterraum.

(a) Multinomialverteilung $(M(p_0,\ldots,p_s;n))_{0<p_i<1,\sum p_i=1}$;

(b) Poissonverteilung $(\mathrm{Poiss}(\lambda))_{\lambda>0}$;

(c) Gleichmäßige Verteilung $(U([0,\theta]))_{\theta>0}$;

(d) Gammaverteilung $(\Gamma(a,b))_{a,b>0}$.

Abgabe vor der Vorlesung am Dienstag, dem 17.05.11

Vorlesung *Mathematische Statistik*
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl.-Math. Johanna Kappus

5. Übungsblatt

1. Ein Physiker untersucht die Radioaktivität bei zwei verschiedenen Präparaten. Die unabhängig gemessene Zahl der Zerfälle in einer Zeiteinheit bei Präparat 1 sei $X_1, \ldots, X_{m_1}$ (m_1 Messungen), bei Präparat 2 $Y_1, \ldots, Y_{m_2}$ (m_2 Messungen). Gib eine vernünftige Regel an, um zu entscheiden, welches Präparat stärker radioaktiv ist. Begründe dazu, weshalb die Annahme einer Poissonverteilung gerechtfertigt ist, und gib ein Suffizienzargument.

2. Beweise: Es sei $(\mathbb{P}_\theta)_{\theta \in \mathcal{Z}}$ eine Exponentialfamilie mit natürlichem Parameterraum $\mathcal{Z} \subseteq \mathbb{R}^k$ und Darstellung
$$\frac{d\,\mathbb{P}_\theta}{d\mu}(x) = C(\theta)h(x)\exp(\langle \theta, T(x)\rangle) = h(x)\exp(\langle \theta, T(x)\rangle - A(\theta)),$$
wobei $A(\theta) = \log\big(\int h(x)\exp(\langle \theta, T(x)\rangle)\mu(dx)\big)$. Ist $\bar\theta$ ein innerer Punkt von $\mathcal{Z}$, so ist die *erzeugende Funktion* von T, $\psi_{\bar\theta}(s) = \mathbb{E}_{\bar\theta}[e^{\langle T,s\rangle}]$, $s \in \mathbb{R}^k$, in einer Umgebung der Null wohldefiniert und beliebig oft differenzierbar. Es gilt $\psi_{\bar\theta}(s) = \exp(A(\bar\theta + s) - A(\bar\theta))$ für alle s mit $\bar\theta + s \in \mathcal{Z}$. Für $i, j = 1, \ldots, k$ folgt $\mathbb{E}_{\bar\theta}[T_i] = \frac{dA}{d\theta_i}(\bar\theta)$ und $\operatorname{Cov}_{\bar\theta}(T_i, T_j) = \frac{d^2 A}{d\theta_i d\theta_j}(\bar\theta)$.

3. Eine suffiziente Statistik T^* heißt *minimalsuffizient*, wenn es zu jeder suffizienten Statistik T eine messbare Funktion h gibt, so dass $T^* = h(T)$ $\mathbb{P}_\theta$-f.s. für alle $\theta \in \Theta$ gilt. Beweise, dass jede $\mathbb{R}^d$-wertige, suffiziente und vollständige Statistik minimalsuffizient ist, sofern eine minimalsuffiziente Statistik überhaupt existiert. Gilt die Umkehrung für $\mathbb{R}^d$-wertige Statistiken?
Hinweis: Man kann zeigen, dass minimalsuffiziente Statistiken für dominierte Experimente auf separablen Messräumen (wie $(\mathbb{R}^d, \mathscr{B}_{\mathbb{R}^d})$) stets existieren.

4. Es sei $(B_t, t \geqslant 0)$ eine Brownsche Bewegung. Es wird $X_t := \sigma B_t + at$ mit $\sigma > 0$ unbekannt und $a \in \mathbb{R}$ unbekannt zu den n Zeitpunkten $h, 2h, \ldots, T := nh$ mit $h > 0$ beobachtet.

   (a) Bestimme die gemeinsame Verteilung der $\Delta X_k := X_{kh} - X_{(k-1)h}$, $k \in \{1, \ldots, n\}$.

   (b) $\mathbb{P}_{a,\sigma^2}$ bezeichne die Verteilung von $(\Delta X_1, \Delta X_2, \ldots, \Delta X_n)$ mit $X_t := \sigma B_t + at$. Bestimme die Likelihoodfunktion bezüglich $\mathbb{P}_{0,1}$ und weise nach, dass $(X_T, \sum_{k=1}^n (\Delta X_k)^2)$ eine suffiziente Statistik ist.

   (c) Berechne das quadratische Risiko von $\hat a = X_T/T$ und $\hat\sigma^2 = \sum_{k=1}^n (\Delta X_k)^2/T$ und diskutiere jeweils das Verhalten für $T \to \infty$ bei festem h und für $h \to 0$ bei festem T.

(*d) Simuliere 1000 Realisierungen von $X_t = B_t$ sowie $X_t = 0.5B_t + 4t$ auf dem Intervall $[0,1]$ und bestimme $\hat{\sigma}^2$ jeweils für $h \in \{0.1,\, 0.01,\, 10^{-4}\}$ anhand der Beobachtungen $X_h,\, X_{2h}, \dots, X_1$. Stelle in jedem der sechs Fälle die Verteilung des Schätzfehlers $\hat{\sigma} - \sigma$ in einem Histogramm dar. Äußere eine Vermutung gegen welche Verteilung $\hat{\sigma} - \sigma$ bei richtiger Skalierung für $h \to 0$ konvergiert. (+4P)

Hinweis: Eine *Brownsche Bewegung* $(B_t, t \geqslant 0)$ ist durch folgende Eigenschaften charakterisiert:

(i) es gilt $B_0 = 0$ und $B_t \sim N(0,t)$, $t > 0$;

(ii) die Inkremente sind stationär und unabhängig: für $0 \leqslant t_0 < t_1 < \cdots < t_m$ gilt $(B_{t_1} - B_{t_0}, \dots, B_{t_m} - B_{t_{m-1}}) \sim N(0, \mathrm{diag}(t_1 - t_0, \dots, t_m - t_{m-1}))$;

(iii) B hat stetige Pfade.

Abgabe vor der Vorlesung am Dienstag, dem 17.05.2011

Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

6. Übungsblatt

1. Es sei (X_n) eine zeitlich homogene Markovkette mit Werten in $S = \{1, \cdots, m\}$, deterministischem Anfangswert $X_0 = x_0$ und Übergangswahrscheinlichkeiten $p_{kl} = \mathbb{P}(X_{n+1} = l | X_n = k)$. Man beobachtet $X = (X_0, \cdots, X_n)$ mit $(p_{kl})_{k,l \in S}$ unbekannt.
   Zeige, dass $N_{kl} = \operatorname{card}\{n < N : X_n = k, X_{n+1} = l\}$, $k, l \in S$ suffizient für die unbekannten Übergangswahrscheinlichkeiten ist. Ist sie auch vollständig?

2. Bestimme die Fisher-Informationsmatrix für eine $N(\mu, \sigma^2)$-verteilte mathematische Stichprobe $X_1, \ldots X_m$ mit unbekannten Werten $\mu \in \mathbb{R}$ und $\sigma > 0$ sowie für $X \sim \operatorname{Bin}(n, p)$ mit $p \in (0, 1)$ unbekannt und n bekannt. Finde jeweils einen erwartungstreuen Schätzer für μ und für p, der die Cramér-Rao-Schranke erreicht. Finde einen erwartungstreuen Schätzer für σ^2 bei $m \geqslant 2$ Beobachtungen der zumindest asymptotisch für $m \to \infty$ die Cramér-Rao-Schranke erreicht (bei Reskalierung mit m).

3. Betrachte eine mathematische Stichprobe $X_1, \ldots, X_n$ mit Dichte $f_{\mu,\sigma}(x) = \sigma^{-1} f((x - \mu)/\sigma)$, $f \in C^1(\mathbb{R})$, $\mu \in \mathbb{R}$, $\sigma > 0$ (*Lokations-Skalen-Familie*). Bestimme die Fisher-Information für die Fälle, dass (a) f bekannt und μ, σ unbekannt sowie (b) f, σ bekannt und μ unbekannt sind. Unter welchen Bedingungen an f ist die Fisher-Information für μ unabhängig von der Kenntnis von σ?
   Hinweis: Zeige für eine symmetrische, positiv-definite Matrix $A \in \mathbb{R}^{k \times k}$, dass $(A_{ii})^{-1} \leqslant (A^{-1})_{ii}$, $1 \leqslant i \leqslant k$, gilt mit Gleichheit im Fall einer Diagonalmatrix.

4. Es seien $\mathbb{P}$ und $\mathbb{Q}$ Wahrscheinlichkeitsmaße mit Dichten p und q bezüglich eines dominierenden Maßes μ. Der *Hellinger-Abstand* ist definiert durch $\mathrm{H}(\mathbb{P}, \mathbb{Q}) := \left(\int (\sqrt{p} - \sqrt{q})^2 \, \mathrm{d}\mu\right)^{1/2}$. Der *Totalvariationsabstand* ist $\| \mathbb{P} - \mathbb{Q} \|_{\mathrm{TV}} := \sup_{A \in \mathbb{A}} | \mathbb{P}(A) - \mathbb{Q}(A)|$. Die Kullback-Leibler- Divergenz ist

$$\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) := \begin{cases} \int \left(\log \frac{d\mathbb{P}}{d\mathbb{Q}}\right) \mathrm{d}\,\mathbb{P}, & \text{falls } \mathbb{P} \ll Q \\ \infty, & \text{sonst} \end{cases}$$

   a) Es sei $\mathbb{P} \sim \mathbb{Q}$ und $\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) < \infty$. Zeige, dass

$$\| \mathbb{P} - \mathbb{Q} \|_{\mathrm{TV}}^2 \leq 1 - \exp(-\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q})).$$

b) Zeige für Wahrscheinlichkeitsmaße $\mathbb{P}_i, \mathbb{Q}_i,\ i = 1, \cdots, n$:

$$1 - \exp\left(- \sum_{i=1}^{n} \frac{1}{2} \mathrm{H}^2(\mathbb{P}_i, \mathbb{Q}_i) \right) \leq \frac{1}{2} \mathrm{H}^2(\otimes_{i=1}^{n} \mathbb{P}, \otimes_{i=1}^{n} \mathbb{Q}_i) \leq \sum_{i=1}^{n} \frac{1}{2} \mathrm{H}^2(\mathbb{P}_i, \mathbb{Q}_i).$$

Abgabe vor der Vorlesung am Dienstag, dem 31.05.2011

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

8. Übungsblatt

1. (a) Zeige, dass $f : (0,\infty) \to \mathbb{R}, x \mapsto x \log(x)$ konvex ist, und schließe (benutze $\mathrm{d}\,\mathbb{P} = \frac{\mathrm{d}\,\mathbb{P}}{\mathrm{d}\,\mathbb{Q}}\,\mathrm{d}\,\mathbb{Q}$)

$$\mathrm{KL}(\mathbb{P} \mid \mathbb{Q}) \geqslant 0 \quad \text{und} \quad \mathrm{KL}(\mathbb{P} \mid \mathbb{Q}) = 0 \Longleftrightarrow \mathbb{P} = \mathbb{Q}\,.$$

   Finde zwei äquivalente Wahrscheinlichkeitsmaße $\mathbb{P}$ und $\mathbb{Q}$ mit

$$\mathrm{KL}(\mathbb{P} \mid \mathbb{Q}) \neq \mathrm{KL}(\mathbb{Q} \mid \mathbb{P}).$$

   (b) Beweise für Produktmaße:

$$\mathrm{KL}(\mathbb{P}_1 \otimes \mathbb{P}_2 \mid \mathbb{Q}_1 \otimes \mathbb{Q}_2) = \mathrm{KL}(\mathbb{P}_1 \mid \mathbb{Q}_1) + \mathrm{KL}(\mathbb{P}_2 \mid \mathbb{Q}_2).$$

2. (a) Zeige: Bildet $(\mathbb{P}_\theta)_{\theta\in\Theta}$ eine natürliche Exponentialfamilie und ist θ_0 innerer Punkt von Θ, so gilt $\mathrm{KL}(\mathbb{P}_{\theta_0} \mid \mathbb{P}_\theta) = A(\theta) - A(\theta_0) + \langle \dot{A}(\theta_0), \theta_0 - \theta\rangle$. Folgere

$$\ddot{\mathrm{KL}}(\mathbb{P}_{\theta_0} \mid \mathbb{P}_\theta)|_{\theta=\theta_0} = I(\theta_0).$$

   (b) Finde allgemeine Voraussetzungen, so dass folgende Gleichungen gelten:

$$\dot{\mathrm{KL}}(\mathbb{P}_{\theta_0} \mid \mathbb{P}_\theta)\Big|_{\theta=\theta_0} = 0, \quad \ddot{\mathrm{KL}}(\mathbb{P}_{\theta_0} \mid \mathbb{P}_\theta)\Big|_{\theta=\theta_0} = -\int \ddot{\ell}(\theta_0)\,\mathrm{d}\,\mathbb{P}_{\theta_0}\,.$$

3. Sei $X_1, \ldots, X_n$ eine mathematische Stichprobe bezüglich der Lebesguedichte

$$f_\theta(x) = \frac{1-\theta}{\varphi(\theta)}\left(1 - \frac{|x-\theta|}{\varphi(\theta)}\right)^+ + \frac{\theta}{2}\mathbf{1}_{[-1,1]}(x),$$

   wobei $\theta \in [0,1]$ und $\varphi : [0,1] \to [0,1]$ eine stetige, fallende Funktion mit $\varphi(0) = 1$ und $0 < \varphi(\theta) \leqslant 1-\theta$ für $\theta \in (0,1)$ ist. Ziel ist es, für geeignetes φ zu sehen, dass für alle $\theta \in [0,1]$ jeder MLE fast sicher gegen Eins konvergiert und insbesondere inkonsistent ist. Zeige:

   (a) Es existiert ein Maximum-Likelihood-Schätzer $\hat{\theta}_n$.

   (b) Für $\theta < 1$ ist $f_\theta(x) < 1/\varphi(\theta) + 1/2$ und daraus folgt, dass für die Loglikelihoodfunktion ℓ_n bei n Beobachtungen und für jedes $\alpha < 1$

$$\max_{0\leqslant\theta\leqslant\alpha} \frac{\ell_n(\theta)}{n} \leqslant \log\left(\frac{1}{\varphi(\alpha)} + \frac{1}{2}\right) < \infty$$

   gilt. Um zu beweisen, dass $\lim_{n\to\infty}\hat{\theta}_n = 1$ f.s. für alle $\theta \in [0,1]$, reicht es $\max_{0\leqslant\theta\leqslant 1} \ell_n(\theta)/n \to \infty$ f.s. zu zeigen.

(c) Mit $X_{(n)} = \max\{X_1, \ldots, X_n\}$ gilt

$$\max_{0 \leqslant \theta \leqslant 1} \frac{\ell_n(\theta)}{n} \geqslant \frac{n-1}{n} \log\left(\frac{X_{(n)}}{2}\right) + \frac{1}{n} \log\left(\frac{1 - X_{(n)}}{\varphi(X_{(n)})}\right).$$

(d) Aus dem Lemma von Borel-Cantelli folgt $n^{1/4}(1 - X_{(n)}) \to 0$ f.s. für $\theta = 0$ und auch für alle $\theta \in [0,1]$. Mit $\varphi(\theta) := (1-\theta)\exp(-(1-\theta)^{-4} + 1)$ folgt $\liminf_{n\to\infty}(1/n)\log((1 - X_{(n)})/\varphi(X_{(n)})) = \infty$ f.s. und damit die gewünschte Aussage.

Abgabe vor der Vorlesung am Dienstag, dem 14.06.2011

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

9. Übungsblatt

1. Im nichtlinearen Regressionsmodell der Beobachtungen

$$Y_i = g_\theta(i/n) + \varepsilon_i, \quad i = 1, \dots, n,\ g_\theta \in C([0,1]),\ (\varepsilon_i)_{1 \leqslant i \leqslant n} \text{ iid},$$

mit $\mathbb{E}[\varepsilon_i] = 0$, $\operatorname{Var}(\varepsilon_i) = \sigma^2$, $\mathbb{E}[\varepsilon_i^4] < \infty$, $\sigma > 0$ betrachte den Kleinste-Quadrate-Schätzer $\hat\theta_n = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^n (Y_i - g_\theta(i/n))^2$. Gib Voraussetzungen für die Parametrisierung $\theta \mapsto g_\theta$ an, um auf die asymptotische Normalität von $\hat\theta_n$ für $n \to \infty$ zu schließen und bestimme die asymptotische Varianz.

2. Im linearen Regressionsmodell der Beobachtungen

$$Y_i = g_\theta(i/n) + \varepsilon_i, \quad i = 1, \dots, n,\ (\varepsilon_i)_{1 \leqslant i \leqslant n} \text{ iid},$$

mit $g_\theta(x) = \sum_{l=1}^k \theta_l\, g_l(x)$, $g_l \in C([0,1])$, $k < n$, $\mathbb{E}[\varepsilon_i] = 0$, $\operatorname{Var}(\varepsilon_i) = \sigma^2$, $\sigma > 0$ wird $\theta \in \Theta = \mathbb{R}^k$ geschätzt.

   (a) Schreibe dies unter einer Rangbedingung als ein gewöhnliches lineares Modell und bezeichne mit $\hat\theta$ den Kleinste-Quadrate-Schätzer.

   (b) Unter dem zugrundeliegenden Wahrscheinlichkeitsmaß $\mathbb{P}$ gilt allerdings $Y_i = g(i/n) + \varepsilon_i$ mit einer Funktion $g \in C([0,1])$. Bestimme $g_{\hat\theta}(i/n)$ und das quadratische Risiko $\mathbb{E}_{\mathbb{P}}[\|g_{\hat\theta} - g\|_n^2]$.

3. Es sei (Y, Z) gemäß der Dichte $f(y, z, \theta)$, $\theta \in \Theta$, bezüglich $\mu \otimes \nu$ verteilt, wobei μ und ν σ-endliche Maße seien. Nur Y wird beobachtet. Der EM-Algorithmus zur Berechnung eines MLE besteht aus der Wahl eines Startwertes θ_0 mit $L(\theta_0) = f_Y(y, \theta_0) > 0$ und aus der Wiederholung für $j = 0, 1, \dots$ der Schritte (1) und (2):

   (1) Berechne
$$J(\theta, \theta_j) = \mathbb{E}_{\theta_j}\left[\log\left(\frac{f(Y, Z, \theta)}{f(Y, Z, \theta_j)}\right) \middle| Y = y\right].$$

   (2) Setze $\theta_{j+1} = \operatorname{argmax}_\theta J(\theta, \theta_j)$.

Zeige die Gleichung

$$J(\theta_{j+1}, \theta_j) = \log\left(\frac{f_Y(y, \theta_{j+1})}{f_Y(y, \theta_j)}\right) + \int \log\left(\frac{f_{Z|Y=y}(z, \theta_{j+1})}{f_{Z|Y=y}(z, \theta_j)}\right) f_{Z|Y=y}(z, \theta_j)\nu(\,\mathrm{d}z)$$

und folgere, dass im EM-Algorithmus $L(\theta_{j+1}) \geqslant L(\theta_j)$ gilt.

4. Betrachte eine mathematische Stichprobe $Y_1, \dots, Y_n$, die gemäß einer Mischung zweier Normalverteilungen verteilt ist: Gegeben $Z_i = 0$ ist $Y_i \sim N(a,1)$ und gegeben $Z_i = 1$ ist $Y_i \sim N(b,1)$, wobei $\mathbb{P}_{a,b}(Z_i = 0) = \mathbb{P}_{a,b}(Z_i = 1) = 1/2$ und $(Y_1, Z_1), \dots, (Y_n, Z_n)$ unabhängig. μ sei das LebesguemaÃŸ auf $\mathbb{R}^n$ und ν sei gegeben durch $\nu(\{z\}) = 1/2^n$ für alle $z \in \{0,1\}^n$.

   (a) Bestimme $f_Y(y,a,b)$ für $n = 1$ und zeige für beliebige n

   $$f(y,z,a,b) = \prod_{i=1}^{n} \varphi(y_i - a)^{1-z_i} \varphi(y_i - b)^{z_i}, \quad \text{mit } \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

   (b) Zeige: Es gilt im EM-Algorithmus

   $$a_{j+1} = \frac{\sum_{i=1}^n (1-\tau_i) y_i}{\sum_{i=1}^n (1-\tau_i)} \quad \text{und} \quad b_{j+1} = \frac{\sum_{i=1}^n \tau_i y_i}{\sum_{i=1}^n \tau_i},$$

   wobei $\tau_i := \varphi(y_i - b_j)/(\varphi(y_i - a_j) + \varphi(y_i - b_j))$.

   (c*) Simuliere einen numerischen MLE und den EM-Algorithmus für $a = 1$, $b = 2$, $n = 100$ und für verschiedene Werte von j . Konvergiert θ_j fÃ¼r $j \to \infty$ gegen den numerischen MLE?

Abgabe vor der Vorlesung am Dienstag, dem 21.06.2011

Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

10. Übungsblatt

1. Im Gaußschen linearen Regressionsmodell der Beobachtungen

$$Y_i = g_\theta(i/n) + \varepsilon_i, \qquad i = 1, \ldots, n, \quad \varepsilon \sim N(0, \sigma^2 E_n),$$

mit $g_\theta(x) = \sum_{l=1}^k \theta_l \, g_l(x)$, $g_l \in C([0,1])$, $n > k$, $\sigma > 0$ wird $\theta \in \Theta = \mathbb{R}^k$ geschätzt. Die (g_l) seien orthonormal bezüglich $\langle f, f' \rangle_n := (1/n) \sum_{i=1}^n f(i/n) f'(i/n)$.

(a) Zeige, dass im falsch spezifizierten Modell für den Maximum-Likelihood-Schätzer $\hat\theta$ die Darstellung $\|g_{\hat\theta} - g\|_n^2 = \inf_{\theta \in \Theta} \|g_\theta - g\|_n^2 + \frac{Z}{n}$ mit $Z := n \sum_{l=1}^k \langle \varepsilon, g_l \rangle_n^2$ gilt und dass $(\langle \varepsilon, g_l \rangle_n)_{l=1,\ldots,k}$ unabhängige $N(0, \sigma^2/n)$-verteilten Zufallsvariablen sind.

(b) Schließe, dass $U := (1/\sigma^2) Z$ eine $\chi^2(k)$-verteilte Zufallsvariable ist und berechne $\mathbb{E}[e^{\alpha U}]$ für $\alpha < 1/2$.

(c) Setze $\delta = 1/2 - \alpha$ und zeige für $\delta \in (0, 1/2)$ mit der verallgemeinerten Markov-Ungleichung $\mathbb{P}(Z \geqslant \kappa) \leqslant (2\delta)^{-k/2} \exp(-(1/2 - \delta)\kappa/\sigma^2)$, $\kappa > 0$.

(d*) Bestimme numerisch oder analytisch den Wert von $\mathbb{P}(Z \geqslant \kappa)$ und vergleiche mit der Schranke.

2. Es seien $\mathbb{P}$ und $\mathbb{Q}$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{F})$ mit Dichten p und q bezüglich $(\mathbb{P} + \mathbb{Q})/2$. Setze $\mathbb{Q}^a(A) := \mathbb{Q}(A \cap \{p > 0\})$ und $\mathbb{Q}^\perp(A) := \mathbb{Q}(A \cap \{p = 0\})$. Zeige:

a) Es gilt die *Lebesgue-Zerlegung* $\mathbb{Q} = \mathbb{Q}^a + \mathbb{Q}^\perp$ mit $\mathbb{Q}^a \ll \mathbb{P}$ und $\mathbb{Q}^\perp \perp \mathbb{P}$.

b) Es gilt $\mathbb{Q}^a(A) = \int_A \frac{q}{p} \, \mathrm{d}\,\mathbb{P}$.

3. Zeige: Gilt $\| \mathbb{P}_n - \mathbb{Q}_n \|_{\mathrm{TV}} \to 0$, so sind $(\mathbb{P}_n)$ und $(\mathbb{Q}_n)$ gegenseitig contiguous. Gilt auch die Umkehrung?

4. Beweise für $(\mathbb{Q}_n) \lhd (\mathbb{P}_n)$ mit Dichten (p_n) und (q_n) bezüglich eines dominierenden Maßes und für reellwertige Zufallsvariablen (X_n):
Wenn $(X_n, \frac{q_n}{p_n}) \underset{\mathbb{P}_n}{\xrightarrow{\mathcal{D}}} (X, V)$ gilt, so definiert $R(B) := \mathbb{E}\,[1_B(X) V]$ ein Wahrscheinlichkeitsmaß und es folgt $X_n \underset{\mathbb{Q}_n}{\xrightarrow{\mathcal{D}}} R$.

Abgabe vor der Vorlesung am Dienstag, dem 28.06.2011

Mathematische Statistik
Sommersemester 2011
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

11. Übungsblatt

1. Es seien $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. Es sei $\frac{1}{2\sigma^2}$ gemäß der Gamma-Verteilung $\Gamma(\alpha, \beta)$ mit Dichte
$$\gamma_{\alpha,\beta}(x) := \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x},\ x \geq 0$$
verteilt.
Berechne die a-posteriori-Dichte von $\frac{1}{2\sigma^2}$ und verifiziere die Aussage des Bernstein-von-Mises-Satzes direkt.

2. Vollziehe den Beweis des Neyman-Pearson-Lemmas im Skriptum nach. Beweise, dass jeder beste Test für $H_0 : \theta = 0$ gegen $H_1 : \theta = 1$ fast sicher ein Neyman-Pearson-Test ist.

3. Für den n-maligen Münzwurf möchte man die Hypothese, dass die Münze fair ist, gegen die Alternative, dass die Erfolgswahrscheinlichkeit $0,25$ beträgt, testen. Formuliere dies als Testproblem und gib konkret einen Neyman-Pearson-Test zum Niveau α an.

4. Für $n \geq 1$ sei φ_n ein Neyman-Pearson-Test von $H_0 : \theta = 0$ gegen $H_1 : \theta = 1$, beruhend auf den i.i.d. Beobachtungen $X_1, \cdots, X_n$ mit $\mathbb{E}_0[\varphi_n] = \alpha$. Es ist zu zeigen, dass
$$\lim_{n\to\infty} \frac{1}{n} \log(1 - \mathbb{E}_1[\varphi_n]) = -\mathrm{KL}(\mathbb{P}_0 \,|\, \mathbb{P}_1)$$
gilt.

   a) Es sei $h_n(X) := \frac{1}{n} \sum_{i=1}^{n} \log \frac{p_0}{p_1}(X_i)$. Dann hat φ_n die Gestalt
$$\varphi_n(X) = \begin{cases} 1, & \text{falls } h_n < a_n \\ 0, & \text{falls } h_n > a_n \end{cases}$$
für geeignete a_n.

   b) Zeige die Abschätzung
$$e^{n a_n}\, \mathbb{E}_1[1 - \varphi_n] \leq \mathbb{E}_0[1 - \varphi_n] \leq 1.$$
Folgere aus dem schwachen Gesetz der großen Zahl, dass $a_n > a$ gilt für beliebiges $a < \mathrm{KL}(\mathbb{P}_0 \,|\, \mathbb{P}_1)$ und hinreichend großes n und schließe daraus, dass
$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_1[1 - \varphi_n] \leq -\,\mathrm{KL}(\mathbb{P}_0 \,|\, \mathbb{P}_1)$$
gilt.

c) Zeige für beliebiges $a > \mathrm{KL}(\mathbb{P}_0 \,|\, \mathbb{P}_1)$ und hinreichend großes n die Abschätzung

$$\mathbb{E}_1[1-\varphi_n] \geq e^{-na}\frac{1-\alpha}{2}$$

und schließe daraus:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}_1[1-\varphi_n] \geq -\mathrm{KL}(\mathbb{P}_0 \,|\, \mathbb{P}_1).$$

Damit ist

$$\lim_{n\to\infty} \frac{1}{n} \log(1-\mathbb{E}_1[\varphi_n]) = -\mathrm{KL}(\mathbb{P}_0 \,|\, \mathbb{P}_1)$$

gezeigt.

Abgabe vor der Vorlesung am Dienstag, dem 05.07.2011