



## 1. Übungsblatt

### 1. Schätzung im Poisson-Modell.

- (a) In einem Krankenhaus gab es in den vergangenen  $n$  Tagen  $N_1, \dots, N_n$  Geburten. Begründen Sie, weshalb  $N_1, \dots, N_n$  näherungsweise als unabhängig und  $\text{Pois}(\lambda)$ -verteilt mit unbekanntem Parameter  $\lambda > 0$  angesehen werden können. Geben Sie das entsprechende statistische Modell formal an.
- (b) Belesen Sie sich zur Momentenmethode 1.14 im Buch und konstruieren Sie im Modell aus (a) Momentenschätzer  $\hat{\lambda}_1, \hat{\lambda}_2$  von  $\lambda$  basierend auf dem Erwartungswert  $\mathbb{E}_\lambda[N_i] = \lambda$  sowie auf der Varianz  $\text{Var}_\lambda(N_i) = \lambda$ . Bestimmen Sie außerdem die Likelihoodfunktion im Modell und den MLE  $\hat{\lambda}_3$ .
- (c) Simulieren Sie in 1000 Monte-Carlo-Iterationen die Beobachtungen  $N_1, \dots, N_n$  für  $n = 30$  und  $\lambda = 1$  und stellen Sie die Ergebnisse für die Schätzer aus (b) jeweils in einem *Boxplot* dar. Können Sie die unterschiedliche Qualität der Schätzer mathematisch begründen?

### 2. Für eine $N(\mu, \sigma^2)$ -verteilte mathematische Stichprobe $X_1, \dots, X_n$ betrachte den Schätzer $\hat{\sigma}_\alpha^2 := \frac{\alpha}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ von $\sigma^2$ , wobei $\alpha > 0$ frei gewählt wird.

- (a) Zeigen Sie für den mittleren quadratischen Fehler (MSE)

$$\mathbb{E}_{\mu, \sigma^2}[(\hat{\sigma}_\alpha^2 - \sigma^2)^2] = (\alpha - 1)^2 \sigma^4 + \alpha^2 \text{Var}_{\mu, \sigma^2}(\hat{\sigma}_1^2).$$

- (b) Begründen Sie, weshalb  $\text{Var}_{\mu, \sigma^2}(\hat{\sigma}_1^2) = \sigma^4 V_n$  mit einer Konstanten  $V_n$  unabhängig von  $\mu$  und  $\sigma^2$  gilt. Schließen Sie, dass der MSE minimal ist für  $\alpha = (1 + V_n)^{-1}$ .
- (c) Bestimmen Sie  $V_n$ . Sie dürfen  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) - \bar{X}^2$  sowie  $\text{Var}(X^2) = 2$  für  $X \sim N(0, 1)$  verwenden.

### 3. ChatGPT behauptet manchmal, dass der MLE im Binomialmodell minimax ist (s. Rückseite) und manchmal nicht. Finden Sie heraus, was stimmt, und erklären Sie dies im Detail.

### Is the MLE in the Binomial model minimax?

To determine whether the Maximum Likelihood Estimator (MLE) in the Binomial model is minimax, we need to analyze it in the context of decision theory, specifically minimax estimation under some loss function—typically squared error loss.

#### 1. Setup

Let  $X \sim \text{Binomial}(n, \theta)$ , where  $\theta \in [0, 1]$  is the unknown parameter.

- The MLE of  $\theta$  is:

$$\hat{\theta}_{\text{MLE}} = \frac{X}{n}$$

#### 2. What Does "Minimax" Mean Here?

An estimator  $\hat{\theta}(X)$  is minimax under a loss function  $L(\theta, \hat{\theta})$  if it minimizes the maximum risk over all  $\theta \in [0, 1]$ :

$$\sup_{\theta \in [0,1]} \mathbb{E}_{\theta}[L(\theta, \hat{\theta}(X))] \leq \sup_{\theta \in [0,1]} \mathbb{E}_{\theta}[L(\theta, \tilde{\theta}(X))]$$

for any other estimator  $\tilde{\theta}$ .

For squared error loss:

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$$

#### 3. Risk of the MLE

The risk of the MLE under squared error loss is:

$$R(\theta, \hat{\theta}_{\text{MLE}}) = \text{Var}_{\theta}\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot \text{Var}(X) = \frac{1}{n^2} \cdot n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}$$

So the maximum risk over  $\theta \in [0, 1]$  is:

$$\sup_{\theta \in [0,1]} R(\theta, \hat{\theta}_{\text{MLE}}) = \frac{1}{4n}$$

(because  $\theta(1-\theta)$  is maximized at  $\theta = 1/2$ ).

#### 4. Is the MLE Minimax?

Yes — in this case, the MLE is minimax under squared error loss.

- It achieves the constant risk bound of  $\frac{1}{4n}$ , which is the minimax risk in the Binomial model.
- In fact, this risk is uniformly bounded above and matches the Bayes risk under the least favorable prior (Beta(1/2, 1/2) in some formulations).

Furthermore, for the Binomial( $n, \theta$ ) model, the minimax estimator under squared error loss is known to be the MLE ( $X/n$ ). This has been proved in classical statistical decision theory literature (see references like Lehmann & Casella).

#### Conclusion

Yes, the MLE  $\hat{\theta} = X/n$  in the Binomial( $n, \theta$ ) model is minimax under squared error loss.

Let me know if you want to explore minimaxity under other loss functions (like absolute error or 0-1 loss), or in a Bayesian context.

Abgabe zu Beginn der Vorlesung am Montag, 27.10.25.  
Klausurtermin: Freitag 13.2.26, 9-11 Uhr.

Vorlesung *Methoden der Statistik*  
Wintersemester 2025/26  
Humboldt-Universität zu Berlin  
Prof. Dr. Markus Reiß  
MSc. Shanshan Meng



## 2. Übungsblatt

1. Unter  $N$  Wahlberechtigten gibt es  $N_{AfD}$  AfD-Wähler sowie jeweils  $N/2$  Männer und Frauen.
  - (a) Für gerades  $n \in \mathbb{N}$  ziehe jeweils zufällig (mit Zurücklegen)  $n/2$  Männer und  $n/2$  Frauen und bestimme jeweils die Anzahl  $X_m$  bzw.  $X_w$  von AfD-Wählern/innen in den Stichproben. Formalisieren Sie das statistische Modell und zeigen Sie, dass  $\hat{p}_{m/w} := \frac{X_m + X_w}{n}$  ein erwartungstreuer Schätzer von  $p = N_{AfD}/N$  ist, dessen MSE nicht größer ist als der von  $\hat{p} = \frac{X}{n}$ , wenn sich  $X$  AfD-Wähler unter  $n$  geschlechtsneutral gezogenen Wahlberechtigten befinden.
  - (b) Angenommen die Anzahl der AfD-Wähler unter den Männern ist etwa doppelt so groß wie der unter den Frauen. Wie können Sie durch geeignete Stichprobenzahlen einen erwartungstreuen Schätzer von  $p$  konstruieren, der noch kleineren MSE als  $\hat{p}_{m/w}$  besitzt? (Stichwort: *stratified sampling*).
2. Es sei  $Y$  eine reellwertige Zufallsvariable mit Dichte  $f^Y$  bezüglich einem  $\sigma$ -endlichen Maß  $\nu$  auf  $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ . Zeigen Sie:
  - (a)  $m := \inf\{q \in \mathbb{R} \mid \int_{-\infty}^q f^Y(y)\nu(dy) \geq 1/2\}$  minimiert  $\mathbb{E}[|Y - r|]$  über  $r \in \mathbb{R}$ .
  - (b)  $M := \operatorname{argmax}_{y \in \mathbb{R}} f^Y(y)\nu(\{y\})$  (dies möge existieren) minimiert  $\mathbb{E}[\mathbf{1}(Y \neq r)]$  über  $r \in \mathbb{R}$ .

Folgern Sie die Aussage zu Bayesschätzern unter absolutem und 0-1-Verlust aus der Vorlesung.

3. Bayesklassifizierer. Es sei  $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \{0,1\}})$  ein binäres statistisches Modell sowie  $\pi$  eine a-priori-Verteilung auf  $\mathcal{P}(\{0,1\})$ .  $L(\vartheta, x)$  bezeichne die Likelihoodfunktion bezüglich einem dominierenden Maß  $\mu$ .
- (a) Bestimmen Sie die a-posteriori-Verteilung von  $\vartheta$ .
  - (b) Charakterisieren Sie die Bayes-optimalen Schätzer bezüglich 0-1-Verlust  $\ell(\vartheta, r) = \mathbf{1}(\vartheta \neq r)$  und berechnen Sie das zugehörige Bayesrisiko. Welche einfachen Entscheidungsregeln (zwischen 0 und 1) ergeben sich für  $\pi(\{0\}) = \pi(\{1\})$ ?
4. Es sei  $X_1, \dots, X_n$  eine  $N(\mu, \Sigma)$ -verteilte mathematische Stichprobe mit  $\mu \in \mathbb{R}^d$  unbekannt und  $\Sigma \in \mathbb{R}^{d \times d}$  bekannt. Betrachte die a-priori-Verteilung  $\pi = N(a, B)$  für  $\mu$  mit  $a \in \mathbb{R}^d$ ,  $B \in \mathbb{R}^{d \times d}$ .
- (a) Bestimmen Sie die a-posteriori-Verteilung von  $\mu$ .
  - (b) Zeigen Sie, dass der Bayesschätzer  $\hat{\mu}_{a,B}$  bezüglich quadratischem Risiko als Konvexkombination von  $\bar{X}$  und  $a$  dargestellt werden kann. Diskutieren Sie den Einfluss von  $n$ ,  $\Sigma$  und  $B$  auf die Gewichtung.
  - (c) Wenn ein Maß  $\pi$ , das kein Wahrscheinlichkeitsmaß ist, formal in die Berechnung des Bayesschätzers eingesetzt wird, spricht man von einem *improper Bayes estimator*. Zu welchem Maß  $\pi$  ergibt sich der MLE  $\bar{X}$  als improper Bayes estimator?

---

Abgabe **zu Beginn** der Vorlesung am Montag, 3.11.25.

Vorlesung *Methoden der Statistik*  
Wintersemester 2025/26  
Humboldt-Universität zu Berlin  
Prof. Dr. Markus Reiß  
MSc. Shanshan Meng



### 3. Übungsblatt

1. Es werde eine  $\mathbb{P}_\vartheta$ -verteilte mathematische Stichprobe  $X_1, \dots, X_n$  beobachtet mit  $\vartheta \in \{\vartheta_0, \vartheta_1\}$ ,  $0 < \vartheta_0 < \vartheta_1$ , unbekannt. Bestimmen Sie die Form der Neyman-Pearson-Tests für  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta = \vartheta_1$  in folgenden Modellen:
  - (a)  $\mathbb{P}_\vartheta = N(\vartheta, 1)$ ,
  - (b)  $\mathbb{P}_\vartheta = \text{Exp}(\vartheta)$ ,
  - (c)  $\mathbb{P}_\vartheta = U([0, \vartheta])$ .
2. Einseitiger Gaußtest. Es sei  $X_1, \dots, X_n$  eine  $N(\vartheta, 1)$ -verteilte mathematische Stichprobe und  $\vartheta_0 \in \mathbb{R}$ . Zeigen Sie, dass der einseitige Gaußtest  $\varphi_\alpha = \mathbf{1}(\bar{X} \geq c_\alpha)$  mit geeignetem kritischen Wert  $c_\alpha \in \mathbb{R}$  gleichmäßig bester Test (UMP-Test) vom Niveau  $\alpha$  für  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$  ist. Verwenden Sie dazu Aufgabe 1(a), um zunächst die UMP-Eigenschaft für  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$  zu erhalten. Schließen Sie dann mit der Monotonie der Gütefunktion auf die UMP-Eigenschaft auch allgemeiner für  $H_0 : \vartheta \leq \vartheta_0$ .
3. Recherchieren Sie den Begriff des *p-Wertes* eines Tests (z.B. im Buch). Bestimmen Sie für die folgenden Datensätze den *p*-Wert für einen zweiseitigen Test unter der Nullhypothese, dass die Wahrscheinlichkeit für *männlich* und *weiblich* gleich groß ist. Die Binomialverteilung darf durch die Normalverteilung approximiert werden. Klären Sie jeweils, ob die Nullhypothese zu den Niveaus  $\alpha = 0,05$  bzw.  $\alpha = 0,01$  akzeptiert oder abgelehnt wird.
  - (a) 2015 wurden in Berlin 19 614 Jungen und 18 416 Mädchen geboren.
  - (b) In der 4. Coronawelle gab es in Berlin 51 949 Neuinfektionen, darunter 50,2% weibliche und 48,4% männliche Personen (beim Rest ist das Geschlecht unbekannt oder divers).

4. Es werde eine  $N(\mu, \sigma^2)$ -verteilte mathematische Stichprobe  $X_1, \dots, X_n$  beobachtet mit  $\sigma > 0$  bekannt und  $\mu \in \mathbb{R}$  unbekannt. Für welche  $a, b > 0$  ist  $C = [\bar{X} - a, \bar{X} + b]$  mit Stichprobenmittel  $\bar{X}$  ein Konfidenzintervall für  $\mu$  vom Niveau  $1 - \alpha$ ,  $\alpha \in (0, 1)$ ? Für welche dieser Werte von  $a, b$  besitzt  $C$  minimale Länge?

---

Abgabe **zu Beginn** der Vorlesung am Montag, 10.11.25.



## 4. Übungsblatt

1. Lesen Sie Kapitel 1 *Grundlagen der Statistik* im Buch und finden Sie mindestens drei Fehler. Die Angabe von Fehlern kann auch gerne durch konkrete Verbesserungsvorschläge zur Darstellung oder Erklärung ersetzt werden.
2. Betrachten Sie das lineare Modell  $Y = X\beta + \varepsilon$  mit normalverteilten Fehlern  $\varepsilon \sim N(0, \Sigma)$  mit bekannter positiv-definiter Matrix  $\Sigma \in \mathbb{R}^{n \times n}$ . Weisen Sie nach, dass der Kleinste-Quadrate-Schätzer von  $\beta$  Maximum-Likelihood-Schätzer ist. Diskutieren Sie auch, wann die Schätzer existieren und wann sie eindeutig sind.
3. Betrachten Sie das gewöhnliche lineare Modell  $Y = X\beta + \varepsilon$  mit  $\varepsilon \sim N(0, \sigma^2 E_n)$ . Bestimmen Sie den Likelihood-Quotiententest für die Nullhypothese  $H_0 : \beta = \beta_0$  gegen  $H_1 : \beta \neq \beta_0$  zum Niveau  $\alpha \in (0, 1)$ . Hierbei seien  $X \in \mathbb{R}^{n \times p}$ ,  $\beta_0 \in \mathbb{R}^p$ ,  $\sigma > 0$  bekannt.  
*Tipp:* Stellen Sie den Test über die Teststatistik  $|\Pi_X Y - X\beta_0|^2$  dar.

4. (*Polynomregression*) Wir beobachten

$$Y_i = \beta_0 + \beta_1 x_i + \cdots + \beta_{p-1} x_i^{p-1} + \varepsilon_i, \quad i = 1, \dots, n,$$

wobei  $x_1, \dots, x_n \in \mathbb{R}$  bekannte Designpunkte sind und  $\varepsilon_i \sim N(0, \sigma^2)$  i.i.d. gilt.

- (a) Schreiben Sie die Beobachtungen als ein lineares Modell  $Y = X\beta + \varepsilon$  mit einer geeigneten Designmatrix  $X \in \mathbb{R}^{n \times p}$  und weisen Sie nach, dass für  $p \leq n$  und paarweise verschiedene  $x_i$  die Designmatrix vollen Rang  $p$  besitzt (Tipp: *Vandermonde-Determinante*).
- (b) Für  $n = 100$ ,  $x_i = (i - 1)/(n - 1)$ ,  $\varepsilon_i \sim N(0; 0, 1)$  simuliere Beobachtungen  $Y_i = \sin(\pi x_i) + \varepsilon_i$ ,  $i = 1, \dots, n$ . Für Dimensionen  $p = 1, 2, 3, 5, 10, 50$  bestimme jeweils das Regressionspolynom  $\hat{f}_p(x) = \sum_{i=0}^{p-1} \hat{\beta}_i x^i$  mit Kleinste-Quadrate-Schätzer  $\hat{\beta}$ . Zeichnen Sie  $f$ , alle  $\hat{f}_p$  und die Beobachtungen  $Y_i$  in ein gemeinsames Koordinatensystem. Diskutieren Sie das Ergebnis.



## 5. Übungsblatt

1. Betrachten Sie das lineare Modell  $Y = X\beta + \varepsilon$ , wobei die  $\varepsilon_i$  unabhängig und gemäß der Laplaceverteilungsdichte  $f(x) = \frac{\lambda}{2}e^{-\lambda|x|}$ ,  $x \in \mathbb{R}$ , für ein  $\lambda > 0$  verteilt seien.

- (a) Weisen Sie nach, dass der Maximum-Likelihood-Schätzer für  $\beta$  Lösung des folgenden Minimierungsproblems ist:

$$\hat{\beta}^{ML} = \min_{b \in \mathbb{R}^p} \sum_{i=1}^n |Y_i - (Xb)_i|.$$

- (b) Schließen Sie, dass im Fall  $p = 1$  und  $X = (1, \dots, 1)^\top \in \mathbb{R}^{n \times p}$  jeder Stichprobenmedian von  $Y_1, \dots, Y_n$  Maximum-Likelihood-Schätzer ist.
- (c) Betrachten Sie den Fall in (b). Zeigen Sie  $\text{Var}(\hat{\beta}_{KQ}) = \frac{2}{\lambda^2 n}$  für den Kleinste-Quadrate-Schätzer  $\hat{\beta}_{KQ}$ . Vergleichen Sie den MSE von  $\hat{\beta}_{KQ}$  und  $\hat{\beta}_{ML}$  in Simulationen für  $n = 99$  und  $\beta = 0$ ,  $\lambda = 1$ .
- (d\*\*) Beweisen Sie  $\sqrt{n}(\hat{\beta}_{ML} - \beta) \xrightarrow{d} N(0, \lambda^{-2})$  für  $n \rightarrow \infty$  in der Situation von (b).
2. Zeigen Sie für die stochastische Ordnung:  $X_n = \mathcal{O}_P(a_n)$ ,  $Y_n = \mathcal{O}_P(b_n)$  impliziert  $X_n + Y_n = \mathcal{O}_P(a_n + b_n)$ ,  $X_n Y_n = \mathcal{O}_P(a_n b_n)$ .
3. Zeigen Sie für den MSE des Ridge-Schätzers  $\hat{\beta}_\lambda = (X^\top X + \lambda E_p)^{-1} X^\top Y$  im gewöhnlichen linearen Modell:

$$\mathbb{E}[|\hat{\beta}_\lambda - \beta|^2] = |(\frac{n}{\lambda} \Sigma_n + E_p)^{-1} \beta|^2 + \frac{\sigma^2}{n} \text{trace}((\Sigma_n + \frac{\lambda}{n} E_p)^{-2} \Sigma_n)$$

mit  $\Sigma_n = \frac{1}{n} X X^\top$ . Diskutieren Sie das Verhalten für  $\lambda \downarrow 0$  und  $\lambda \uparrow \infty$ . Betrachten Sie das sogenannte *spiked covariance model*  $\Sigma_n = \text{diag}(1, \dots, 1, \varepsilon, \dots, \varepsilon)$  mit  $k$  Eigenwerten 1 und  $p - k$  Eigenwerten  $\varepsilon \in [0, 1]$ . Von welcher Größenordnung sollte man  $\lambda = \lambda(\varepsilon, n)$  mindestens wählen für  $n \rightarrow \infty$  und  $\varepsilon \downarrow 0$ , damit der Fehler beschränkt bleibt?

4. Betrachten Sie das statistische Modell

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

mit  $n$  verschiedenen Punkten  $x_i \in \mathbb{R}$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}$  beliebig und unbekannt sowie  $\mathbb{E}[\varepsilon] = 0$ ,  $\text{Cov}(\varepsilon) = \sigma^2 E_n$ . Unter der polynomiellen Regressionsannahme  $f = q_a^d$  mit

$$q_a^d(x) = a_0 + a_1 x + \dots + a_d x^d$$

sei  $\hat{a} = (\hat{a}_0, \dots, \hat{a}_d)$  der Kleinste-Quadrate-Schätzer sowie  $q_{\hat{a}}^d(x)$  die Vorhersage des Funktionswerts an der Stelle  $x \in \mathbb{R}$ . Zeigen Sie unter der Verwendung der empirischen Norm  $\|g\|_n^2 := \frac{1}{n} \sum_{i=1}^n g(x_i)^2$ :

(a) Unter der Annahme  $f = q_a^d$  gilt  $\mathbb{E}_{q_a^d}[\|q_{\hat{a}}^d - q_a^d\|_n^2] = \frac{\sigma^2}{n}(d+1)$ .

*Tipp:* Betrachten Sie  $\|X\hat{\beta} - X\beta\|^2$  allgemein im linearen Modell.

(b) Für  $f$  beliebig im Modell ergibt sich die Bias-Varianz-Zerlegung

$$\mathbb{E}_f[\|q_{\hat{a}}^d - f\|_n^2] = \min_{q_a^d} \|f - q_a^d\|_n^2 + \frac{\sigma^2}{n}(d+1),$$

wobei sich der Bias durch den kleinsten Abstand von  $f$  zum Raum der Polynome vom maximalen Grad  $d$  in empirischer Norm ergibt.

(c) Gilt für jedes quadratische Polynom  $f$ , dass der MSE  $\mathbb{E}_f[\|q_{\hat{a}}^d - f\|_n^2]$  für  $d = 1$  größer als für  $d = 2$  ist?

---

Abgabe **zu Beginn** der Vorlesung am Montag, 24.11.25.



## 6. Übungsblatt

1. Es seien  $X_1, \dots, X_n$  unabhängige  $N(0, \Sigma_X)$ -verteilte Zufallsvektoren im  $\mathbb{R}^p$  und  $\Sigma_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$ .

- (a) Zeigen Sie für jeden deterministischen Vektor  $v \in \mathbb{R}^p$ , dass

$$\begin{aligned} \mathbb{E}[|(\Sigma_n - \Sigma_X)v|^2] &= \frac{1}{n} \mathbb{E}[|(\Sigma_1 - \Sigma_X)v|^2] \\ &= \frac{1}{n} (\langle \Sigma_X v, v \rangle \text{trace}(\Sigma_X) + |\Sigma_X v|^2) \end{aligned}$$

gilt. Schließen Sie  $|(\Sigma_n - \Sigma_X)v| = \mathcal{O}_P(\sqrt{p/n})$ , sofern  $\|\Sigma_X\|$  und  $|v|$  beschränkt bleiben.

*Hinweis:* Sind  $Z_1, Z_2$  gemeinsam normalverteilt mit Erwartungswert Null, so gilt  $\text{Var}(Z_1 Z_2) = \mathbb{E}[Z_1^2] \mathbb{E}[Z_2^2] + \mathbb{E}[Z_1 Z_2]^2$ .

- (b) Die *Frobeniusnorm* (auch *Hilbert-Schmidt-Norm*)  $\|M\|_2$  einer Matrix  $M \in \mathbb{R}^{p \times p}$  ist gegeben durch

$$\|M\|_2^2 = \sum_{k,l=1}^p M_{kl}^2 = \sum_{k=1}^p |M e_k|^2$$

mit kanonischen Einheitsvektoren  $e_k$ . Weisen Sie mit Teil (a)

$$\mathbb{E}[\|\Sigma_n - \Sigma_X\|_2^2] = \frac{1}{n} (\text{trace}(\Sigma_X)^2 + \|\Sigma_X\|_2^2)$$

nach. Welche stochastische Ordnung hat also  $\|\Sigma_n - \Sigma_X\|_2$  in  $n$  und  $p$  für  $\|\Sigma_X\|$  beschränkt?

2. Es gelte  $\|\Sigma_X^{-1/2} \Sigma_n \Sigma_X^{-1/2} - E_p\| = \mathcal{O}_P(\sqrt{p/n})$ . Beweisen Sie im Detail, dass dann für  $p/n \rightarrow 0$  die Wahrscheinlichkeit, dass  $\Sigma_n$  invertierbar ist, gegen eins konvergiert und  $\|\Sigma_n^{-1/2} \Sigma_X \Sigma_n^{-1/2} - E_p\| = \mathcal{O}_P(\sqrt{p/n})$  gilt.

*Hinweis:* Arbeiten Sie dazu die Begründungen nach Satz 2.31 in der neuen Buchfassung aus.

3. Stellen Sie die Quantile  $q_{t(k), 1-\alpha/2}$  für  $k = 1, \dots, 100$  und  $\alpha \in \{0,01; 0,05; 0,10\}$  graphisch dar (vergleiche Aufgabe 4). Gegen die Quantile welcher Verteilung konvergiert  $q_{t(k), 1-\alpha/2}$  für  $k \rightarrow \infty$ ?

4. Im linearen Modell  $Y = X\beta + \varepsilon$  mit  $\varepsilon \sim N(0, \sigma^2 E_n)$  sei  $\rho = \langle v, \beta \rangle$  für ein  $v \in \mathbb{R}^p \setminus \{0\}$  ein abgeleiteter Parameter. Zeigen Sie für das Testproblem  $H_0 : \rho = \rho_0, \sigma > 0$  gegen  $H_1 : \rho \neq \rho_0, \sigma > 0$ , dass

$$\varphi_\alpha := \mathbf{1}\left(|\hat{\rho} - \rho_0| > \hat{\sigma} \sqrt{\langle (X^\top X)^{-1} v, v \rangle} q_{t(n-p), 1-\alpha/2}\right)$$

mit  $\hat{\rho} = \langle v, \hat{\beta} \rangle$  Likelihood-Quotienten-Test zum Niveau  $\alpha \in (0, 1)$  ist.

---

Abgabe **zu Beginn** der Vorlesung am Montag, 1.12.25.



## 7. Übungsblatt

1. Lesen Sie Kapitel 2 *Das lineare Modell* im Buch und finden Sie mindestens drei Fehler. Die Angabe von Fehlern kann auch gerne durch konkrete Verbesserungsvorschläge zur Darstellung oder Erklärung ersetzt werden.
2. Gegeben sei das lineare Modell  $Y = X\beta + \varepsilon$  mit  $\varepsilon \sim N(0, \sigma^2 E_n)$ . Wir testen auf signifikanten Einfluss der ersten Kovariablen auf die Responsevariable:  $H_0 : \beta_1 = 0, \sigma > 0$  gegen  $H_1 : \beta_1 \neq 0, \sigma > 0$  für  $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$  und  $1 < p < n$ .
  - (a) Stellen Sie diese Hypothese in der Form  $H_0 : K\beta = c$  mit geeigneter Kontrastmatrix  $K$  und mit einem Vektor  $c$  dar.
  - (b) Arbeiten Sie den Beweis von Satz 2.42 durch und vereinfachen Sie ihn im vorliegenden Fall (z.B. durch Verwendung konkreter Einträge von Vektoren und Matrizen).
  - (c) Konstruieren Sie die zugehörige F-Statistik und bestimmen Sie ihre Verteilung.
3. Betrachten Sie die lineare Hypothese  $H_0 : K\beta = 0$  und den auf  $H_0$  eingeschränkten Kleinste-Quadrate-Schätzer  $\hat{\beta}_{H_0}$ . Zeigen Sie im Fall von orthogonalem Design  $X^\top X = nE_p$  mit der expliziten Formel für  $\hat{\beta}_{H_0}$ :
  - (a)  $\hat{\beta}_{H_0} = (E_p - \Pi_{K^\top})\hat{\beta}$ ;
  - (b)  $\hat{\beta}_{H_0}$  ist die Orthogonalprojektion von  $\hat{\beta}$  auf den Kern  $\ker(K) = \{\beta \in \mathbb{R}^p \mid K\beta = 0\}$  von  $K$ .
  - (c) Interpretieren Sie im Fall orthogonalen Designs  $RSS_{H_0} - RSS = |X(\hat{\beta} - \hat{\beta}_{H_0})|^2$  geometrisch.
4. Analysieren Sie die Klimadaten im Januar analog zu den Julitemperaturen von Beispiel 2.45 im Buch, indem Sie auch dort Polynomregression bis zum Grad 4 durchführen und geeignete Hypothesen testen.



## 8. Übungsblatt

1. Ein Smartphone wurde in drei europäischen Ländern zu folgenden Euro-Preisen verkauft:

Land A: 200, 210, 220, 230, 250

Land B: 180, 190, 200, 210, 220

Land C: 210, 220, 230, 240, 250

Überprüfen Sie die Hypothese, dass die vom Hersteller kalkulierten Preise nicht vom Land abhängen, indem Sie eine Varianzanalyse durchführen. Geben Sie die zugehörige ANOVA-Tabelle an.

2. Was ist der Unterschied zwischen  $t$ -Test und  $F$ -Test in der Varianzanalyse? Fragen Sie eine *künstliche Intelligenz* Ihrer Wahl und überprüfen Sie die Antwort auf Korrektheit.
3. Bestimmen Sie für  $\lambda_0, \lambda_1 > 0$  die Kullback-Leibler-Divergenz  $\text{KL}(\text{Poiss}(\lambda_1) \mid \text{Poiss}(\lambda_0))$ . Wie schnell muss die Konvergenz  $\lambda_n \rightarrow \lambda_0$  für  $n \rightarrow \infty$  sein, damit  $\text{KL}(\text{Poiss}(\lambda_n)^{\otimes n} \mid \text{Poiss}(\lambda_0)^{\otimes n})$  beschränkt bleibt?
4. Der Totalvariationsabstand zweier Wahrscheinlichkeitsmaße  $\mathbb{P}, \mathbb{Q}$  auf  $(\mathcal{X}, \mathcal{F})$  ist gegeben durch

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} := \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

- (a) Zeigen Sie, dass  $\|\mathbb{P}_n - \mathbb{P}\|_{TV} \rightarrow 0$  die schwache Konvergenz  $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$  impliziert.
- (b) Besitzen  $\mathbb{P}$  und  $\mathbb{Q}$  Dichten  $p$  und  $q$  bezüglich einem Maß  $\mu$  (z.B.  $\mu = \mathbb{P} + \mathbb{Q}$ ), so gilt

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} = \frac{1}{2} \int |p - q| d\mu.$$

Beweisen Sie dies, indem Sie  $A_0 = \{x \in \mathcal{X} \mid q(x) \geq p(x)\}$  betrachten und  $\|\mathbb{P} - \mathbb{Q}\|_{TV} = \mathbb{Q}(A_0) - \mathbb{P}(A_0)$  nachweisen.

(c) Beweisen Sie die Pinsker-Ungleichung

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} \leq \sqrt{\text{KL}(\mathbb{P} \mid \mathbb{Q})/2}.$$

*Anleitung:* für  $\psi(y) = y \log(y) - y + 1$ ,  $y > 0$ , gilt  $|y - 1|^2 \leq (\frac{4}{3} + \frac{2}{3}y)\psi(y)$  und schätze damit  $\|\mathbb{P} - \mathbb{Q}\|_{TV} = \frac{1}{2} \int_{q>0} |p/q - 1| q d\mu$  gegen die Wurzel von  $\text{KL}(\mathbb{P} \mid \mathbb{Q}) = \int_{q>0} \psi(p/q) q d\mu$  ab.

---

Anfertigung für die **Übung** am Montag, 15.12.25.



## 9. Übungsblatt

Entscheiden Sie bei den folgenden Aussagen, ob sie wahr oder falsch sind.

1. Für eine  $N(\mu, \sigma^2)$ -verteilte mathematische Stichprobe  $X_1, \dots, X_n$  ist  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  erwartungstreuer Schätzer von  $\sigma^2$ .
2. Für eine  $\text{Exp}(\lambda)$ -verteilte mathematische Stichprobe  $X_1, \dots, X_n$  ist  $\hat{\lambda} = (\frac{1}{n} \sum_{i=1}^n X_i)^{-1}$  Momentenschätzer für  $\lambda > 0$ .
3. Der Schätzer  $\hat{\lambda}$  in 2 ist auch MLE.
4. Der Schätzer  $\hat{\lambda}$  in 2 ist erwartungstreu.
5. Die Eigenschaft des MLE hängt von der Wahl des dominierenden Maßes ab.
6. Jeder Bayesschätzer ist erwartungstreu.
7. Die Menge der (randomisierten) Tests vom Niveau  $\alpha$  ist konvex.
8. Es gibt stets einen randomisierten Test vom Niveau  $\alpha$ , dessen Wahrscheinlichkeit für einen Fehler 2. Art  $1 - \alpha$  beträgt; nämlich den trivialen Test  $\varphi(x) = \alpha$ .
9. Der zweiseitige Binomialtest ist stets UMP.
10. Der Beweis des Neyman-Pearson-Lemmas zeigt, dass ein Test  $\tilde{\varphi}$  mit  $\mathbb{E}_0[\tilde{\varphi}] \leq \mathbb{E}_0[\varphi]$  und  $\mathbb{E}_1[\tilde{\varphi}] = \mathbb{E}_1[\varphi]$  für einen Neyman-Pearson-Test  $\varphi$  sogar  $\tilde{\varphi} = \varphi$  fast sicher unter  $\mathbb{P}_0$  und  $\mathbb{P}_1$  erfüllt.
11. Eine Konfidenzmenge  $C$  vom Niveau  $1 - \alpha$  gibt an, dass der wahre Parameter mit Wahrscheinlichkeit  $1 - \alpha$  in  $C$  liegt.
12. Für  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  i.i.d. und  $\alpha \in (0, 1)$  ist  $C = [\bar{X} + q_\alpha \sigma n^{-1/2}, \infty)$  ein einseitiges Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$  (mit  $\alpha$ -Quantil  $q_\alpha$  von  $N(0, 1)$ ).
13. Im linearen Modell existiert stets ein Kleinste-Quadrate-Schätzer.
14. Im linearen Modell ist der Kleinste-Quadrate-Schätzer stets eindeutig.

15. Im gewöhnlichen linearen Modell ist der Kleinste-Quadrate-Schätzer unabhängig von der Kenntnis von  $\sigma^2$ .
16. Die Spur einer Orthogonalprojektion auf einen  $p$ -dimensionalen Unterraum beträgt  $p$ , so dass  $\text{tr}(\Pi_{X_\Sigma}) = p$  gilt.
17. Im gewöhnlichen linearen Modell  $Y = X\beta + \varepsilon$  mit  $p = n = 1$  ist  $\hat{\sigma}^2 = Y^2$  stets erwartungstreuer Schätzer von  $\sigma^2$ .
18. Die Ridge-Regression verringert durch die Einführung des  $l_2$ -Strafterms den Bias des Schätzers, erhöht aber seine Varianz.
19. Für  $X \sim t(n)$  gilt  $\mathbb{E}[|X|^p] < \infty$ ,  $p > 0$ , genau dann, wenn  $p < n$ .
20. Im gewöhnlichen linearen Modell unter Normalverteilungsannahme der Fehler und unbekanntem  $\sigma^2$  folgt die t-Statistik zur Prüfung des linearen Parameters  $\rho = \langle v, \beta \rangle$  unter der Nullhypothese einer Standardnormalverteilung.
21. Für  $X, Y \sim N(0, 1)$  unabhängig ist  $X/Y$   $t(1)$ -verteilt, also Cauchy-verteilt.
22. Die Beobachtungen  $Y_i = \mu_1 + \varepsilon_i$ ,  $i = 1, \dots, n_1$ , und  $Y_i = \mu_2 + \varepsilon_i$ ,  $i = n_1 + 1, \dots, n_1 + n_2$ , mit  $\varepsilon \sim N(0, \sigma^2 E_n)$ ,  $n = n_1 + n_2$ , lassen sich als lineares Modell mit  $p = 2$ ,  $\beta = (\mu_1, \mu_2)^\top$  und Designmatrixeinträgen  $X_{ij} = \mathbf{1}(i \leq n_1, j = 1) + \mathbf{1}(i > n_1, j = 2)$  schreiben. Das  $(1 - \alpha)$ -Konfidenzellipsoid für  $\beta$  ist stets ein Kreis.
23. Das Quadrat einer  $t(n)$ -verteilten Zufallsvariable ist  $F(1, n)$ -verteilt.
24. Die Differenz zweier unabhängiger  $F(m, n)$ -verteilten Zufallsvariablen ist  $F(2m, n)$ -verteilt.
25. Aus  $\sup_n \mathbb{E}[|X_n|^p] < \infty$  für ein  $p > 0$  folgt  $X_n = O_P(1)$ .

---

Abgabe **zu Beginn** der Vorlesung am Montag, 05.01.2026.



## 10. Übungsblatt

1. In der Vorlesung wurde für die penalisierte Modellwahl die Wahrscheinlichkeitsabschätzung

$$\mathbb{P}(Z_d \geq (1 + \kappa + \sqrt{\tau/d})^2 d) \leq \exp(- (d\kappa^2 + \tau)/2), \quad \kappa, \tau > 0,$$

für eine  $\chi^2(d)$ -verteilte Zufallsvariable  $Z_d$  verwendet. Arbeiten Sie den Beweis davon in Satz 3.19 (Buch) aus, insbesondere die Berechnung des Erwartungswerts.

2. Betrachten Sie für Klassifizier  $C : \mathcal{X} \rightarrow \{-1, +1\}$  und  $\alpha \in (0, 1)$  den gewichteten Klassifikationsfehler

$$R_\alpha(C) = \alpha \mathbb{P}(C(X) = -1, Y = +1) + (1 - \alpha) \mathbb{P}(C(X) = +1, Y = -1).$$

Bestimmen Sie unter Benutzung von  $\eta(x) = \mathbb{P}(Y = +1 | X = x)$  den zugehörigen Bayes-Klassifizierer  $C_\alpha^*$ , der  $R_\alpha(C_\alpha^*) = \min_C R_\alpha(C)$  erfüllt. Wie groß ist  $R_\alpha(C_\alpha^*)$ ?

3. Für einen Klassifizier  $C$  und eine mathematische Stichprobe (*Trainingsdaten*)  $(X_i, Y_i)_{1 \leq i \leq n}$  bezeichnet

$$R_n(C) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_i \neq C(X_i))$$

das *empirische Risiko*.  $\hat{C}$  heißt *ERM-Klassifizierer* (*empirical risk minimizer*) in einer Klasse  $\mathcal{C}$  von Klassifizierern, falls  $R_n(\hat{C}) = \min_{C \in \mathcal{C}} R_n(C)$  gilt. Zeigen Sie für das *Risiko* (den Klassifizierungsfehler)  $R$  die Fundamentalungleichung

$$R(\hat{C}) \leq \inf_{C \in \mathcal{C}} R(C) + 2 \sup_{C \in \mathcal{C}} |R_n(C) - R(C)|.$$

4. Betrachten Sie eine endliche Familie  $\mathcal{C} = \{C_1, \dots, C_M\}$  von Klassifizierern und den zugehörigen ERM-Klassifizierer  $\hat{C}$ . Verwenden Sie Aufgabe 3 und die Hoeffding-Ungleichung aus der Literatur, um für alle  $\tau > 0$  zu zeigen, dass mit Wahrscheinlichkeit mindestens  $1 - e^{-\tau}$

$$R(\hat{C}) \leq \min_{1 \leq m \leq M} R(C_m) + \frac{\sqrt{8(\log(2M) + \tau)}}{\sqrt{n}}.$$

Vorlesung *Methoden der Statistik*  
Wintersemester 2025/26  
Humboldt-Universität zu Berlin  
Prof. Dr. Markus Reiß  
MSc. Shanshan Meng



## 11. Übungsblatt

1. Lesen Sie den Abschnitt zum allgemeinen Klassifikationsproblem im Buch und finden Sie mindestens zwei Fehler oder machen Sie mindestens zwei Verbesserungsvorschläge.
2. Beweisen Sie die Aussage aus der Vorlesung, dass das Bayesrisiko im LDA-Modell gegeben ist durch

$$R^* = \pi_{-1} \Phi\left(\left(\log\left(\frac{\pi_{+1}}{\pi_{-1}}\right) - \Delta^2/2\right)/\Delta\right) + \pi_{+1} \left(1 - \Phi\left(\left(\log\left(\frac{\pi_{+1}}{\pi_{-1}}\right) + \Delta^2/2\right)/\Delta\right)\right)$$

mit der  $N(0, 1)$ -Verteilungsfunktion  $\Phi$  und dem *Mahalanobisabstand*  $\Delta$ . Diskutieren Sie jeweils die Asymptotik für  $\pi_{+1} \rightarrow 1$ ,  $\Delta \rightarrow \infty$  sowie  $\Delta \rightarrow 0$ .

3. Betrachten Sie für  $K \geq 2$  Klassen mit Klassenwahrscheinlichkeiten  $\pi_k \in (0, 1)$ ,  $\sum_{k=1}^K \pi_k = 1$ , Mittelwerten  $\mu_k \in \mathbb{R}^p$  und einer invertierbaren Kovarianzmatrix  $\Sigma_X \in \mathbb{R}^{p \times p}$  die Normalverteilungsmischungs-dichte von  $X$

$$p(x) = \sum_{k=1}^K \pi_k p_k(x) \text{ mit } p_k(x) = \varphi_{\mu_k, \Sigma_X}(x), \quad x \in \mathbb{R}^p.$$

Dabei bezeichnet  $p_k$  die Dichte von  $X$  in Klasse  $k$ , also die Dichte von  $\mathbb{P}(X \in \bullet | Y = k)$ .

Weisen Sie nach, dass der beste Klassifizierer (bei Klassifikationsfehler  $R(C) = \mathbb{P}(C(X) \neq Y)$ ) gegeben ist durch  $C^*(x) = \operatorname{argmax}_{k=1, \dots, K} \delta_k(x)$  mit Diskrimanten

$$\delta_k(x) = -\frac{1}{2} \langle \Sigma_X^{-1}(x - \mu_k), x - \mu_k \rangle + \log \pi_k.$$

Skizzieren Sie für  $K = 3$ ,  $\pi_k = 1/3$ ,  $p = 2$  und  $\Sigma_X = E_2$  die Gebiete  $\{x \in \mathbb{R}^2 | C^*(x) = k\}$ .

4. Für unabhängige  $N(\mu, \Sigma_X)$ -verteilte Zufallsvektoren mit  $\mu \in \mathbb{R}^p$ ,  $\Sigma_X \in \mathbb{R}^{p \times p}$  invertierbar betrachte die empirische Kovarianzmatrix

$$\widehat{\Sigma}_X := \frac{1}{n} \sum_{i=1}^n X_i X_i^\top - \left( \frac{1}{n} \sum_{i=1}^n X_i \right) \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^\top.$$

- (a) Schreiben Sie  $Z_i = X_i - \mu$  mit  $Z_i \sim N(0, \Sigma_X)$  und weisen Sie  $\widehat{\Sigma}_X = \widehat{\Sigma}_Z$  nach (mit  $\widehat{\Sigma}_Z$  analog über  $Z_i$  definiert).
- (b) Folgern Sie für  $n \rightarrow \infty$ ,  $p/n \rightarrow 0$  aus  $|\frac{1}{n} \sum_{i=1}^n \Sigma_X^{-1/2} Z_i| = O_{\mathbb{P}}((p/n)^{1/2})$ , dass

$$\left\| \Sigma_X^{-1/2} \widehat{\Sigma}_X \Sigma_X^{-1/2} - \frac{1}{n} \sum_{i=1}^n (\Sigma_X^{-1/2} Z_i) (\Sigma_X^{-1/2} Z_i)^\top \right\| = O_{\mathbb{P}}(p/n)$$

gilt, und weiter unter Verwendung der Resultate für die empirische Kovarianzmatrix im linearen Modell mit zufälligem Design

$$\begin{aligned} \|\Sigma_X^{-1/2} \widehat{\Sigma}_X \Sigma_X^{-1/2} - E_p\| &= O_{\mathbb{P}}((p/n)^{1/2}), \\ \|\widehat{\Sigma}_X^{-1/2} \Sigma_X^{1/2}\| &= 1 + O_{\mathbb{P}}((p/n)^{1/2}). \end{aligned}$$

- (c) Verallgemeinern Sie die Resultate auf die LDA-Situation mit zwei Mittelwerten  $\mu_k$ , Klassenwahrscheinlichkeiten  $\pi_k$  weg von Null beschränkt und  $\widehat{\Sigma}_X = \frac{1}{n-2} \sum_{i=1}^n (X_i - \hat{\mu}_{Y_i})(X_i - \hat{\mu}_{Y_i})^\top$ .

---

Abgabe **zu Beginn** der Vorlesung am Montag, 19.1.26.



## 12. Übungsblatt

1. Lesen Sie den Abschnitt zur linearen Diskriminanzanalyse im Buch und finden Sie mindestens zwei Fehler oder machen Sie mindestens zwei Verbesserungsvorschläge.
2. Nicht-Existenz des logistischen MLE: Betrachten Sie Daten  $(x_1, Y_1), \dots, (x_n, Y_n)$  mit  $x_i \in \mathbb{R}^p$ ,  $Y_i \in \{0, 1\}$  im Modell der logistischen Regression, die sich linear separieren lassen:

$$\exists v_0 \in \mathbb{R}^p \setminus \{0\} \forall i = 1, \dots, n : Y_i = \mathbf{1}(\langle x_i, v_0 \rangle > 0).$$

Zeigen Sie, dass in diesem Fall kein Maximum-Likelihood-Schätzer  $\hat{\beta}$  existiert. Geben Sie aber eine Folge  $\hat{\beta}^{(m)} \in \mathbb{R}^p$  konkret an, so dass  $\lim_{m \rightarrow \infty} L_n(\hat{\beta}^{(m)}) = \sup_{\beta \in \mathbb{R}^p} L_n(\beta)$  gilt.

Sind im eindimensionalen Fall  $p = 1$  Nicht-Existenz des MLE und lineare Separation äquivalent?

3. Es sei  $\hat{\beta}$  der Maximum-Likelihood-Schätzer bei der logistischen Regression mit Absolutglied, das heißt,  $x_{i,1} = 1$  gilt für  $i = 1, \dots, n$ . Zeigen Sie, dass dann die beobachtete Anzahl der Einsen mit der erwarteten Anzahl übereinstimmt:

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n p(\langle x_i, \hat{\beta} \rangle) \text{ mit } p(\eta) := e^\eta (1 + e^\eta)^{-1}.$$

4. Klassifikation in der Praxis: Laden Sie die Cleveland-Daten (<http://www.archive.ics.uci.edu/ml/datasets/heart+disease>, vgl. Beispiel 5.1 im Buch) herunter. Benutzen Sie die ersten  $n = 151$  Patienten mit den Kovariablen Ruheblutdruck und maximale Herzfrequenz und ihren Labels (0=gesund, 1-4=krank) als Trainingsdaten. Die restlichen  $m = 150$  Patienten dienen als *Testdaten*.
- (a) Zeichnen Sie die Kovariablen aller Patienten in ein Koordinatensystem und markieren Sie die Fälle gesund/krank (mit Farben) sowie Training-/Test-Datum (mit dunkel/hell).
  - (b) Führen Sie eine logistische Regression für die Trainingsdaten durch (mit  $p = 3$ , d.h. unter Verwendung eines konstanten Glieds), zeichnen Sie die erhaltene Entscheidungsgrenze in das Koordinatensystem und bestimmen Sie den Klassifikationsfehler jeweils auf den Trainings- und Testdaten (relative Häufigkeit von Fehlklassifikationen).
  - (c) Führen Sie eine lineare Diskriminanzanalyse durch und verfahren Sie wie in (b).

---

Abgabe **zu Beginn** der Vorlesung am Montag, 26.1.26.



## Probeklausur

1. Entscheiden Sie, ob folgende 12 Aussagen wahr oder falsch sind. Jede korrekte Antwort gibt +0,5 Punkte, jede nicht-korrekte -0,5 Punkte. Insgesamt gibt es mindestens 0 und höchstens 5 Punkte.

Aussage	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
wahr												
falsch												
weiß nicht												

- (a) Minimax-optimale Schätzer haben für keinen Parameterwert einen größeren Fehler als Bayesschätzer.
- (b) Der Maximum-Likelihood-Schätzer von  $p$  bei Beobachtung von  $X \sim \text{Bin}(n, p)$  ist minimax bezüglich quadratischem Verlust.
- (c) In der Varianzanalyse wird die Streuung zwischen den Gruppen und die innerhalb der Gruppen mit einem F-Test verglichen.
- (d) Sind im linearen Modell die Fehler  $N(0, \sigma^2 E_n)$ -verteilt, so ist die Stichprobenvarianz  $\hat{\sigma}^2 = |Y - X\hat{\beta}|^2 / (n - p)$   $\chi^2$ -verteilt.
- (e) Der *ridge-regression*-Schätzer hat eine größere Varianz, aber einen kleineren Bias als der Kleinste-Quadrate-Schätzer.
- (f) Für i.i.d.  $(c\Sigma_X, C)$ -subgaußsche Zufallsvektoren  $X_i \in \mathbb{R}^p$  und  $n \rightarrow \infty$  gilt  $\|\Sigma_X^{-1/2} \hat{\Sigma}_n \Sigma_X^{-1/2} - E_p\| = O_{\mathbb{P}}(\sqrt{p/n})$  für  $\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^T$ ,  $\Sigma_X = \mathbb{E}[\hat{\Sigma}_n]$  invertierbar.
- (g) Bei der logistischen Regression modelliert man  $Y_i \sim \text{Bin}(1, p_i)$  unabhängig mit  $p_i = \text{log}((X\beta)_i)$ ,  $i = 1, \dots, n$ .
- (h) Der Median von  $(X_1, \dots, X_n)$  ist Maximum-Likelihood-Schätzer, wenn die  $X_i$  unabhängig sind mit Dichte  $f(x) = \frac{\lambda}{2} e^{-\lambda|x|}$ ,  $x \in \mathbb{R}$ , wobei  $\lambda > 0$  beliebig.
- (i) Die Klassifikationsgrenze beim k-nächste-Nachbar-Klassifizierer ist im Allgemeinen nicht linear.
- (j) Für Folgen  $(P_n)_{n \geq 1}$  und  $(Q_n)_{n \geq 1}$  von Maßen auf einem gemeinsamen messbaren Raum gilt:  $\text{TV}(P_n, Q_n) \rightarrow 0 \Rightarrow \text{KL}(P_n|Q_n) \rightarrow 0$ .

2. Tests und Konfidenzmengen.

- (a) Beweisen Sie: Ist  $(\varphi^{(\vartheta_0)})_{\vartheta_0 \in \Theta}$  eine Familie nicht-randomisierter Tests auf  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$  vom Niveau  $\alpha$ , so ist

$$C_{1-\alpha}(x) = \{\vartheta_0 \in \Theta \mid \varphi^{(\vartheta_0)}(x) = 0\}$$

eine  $(1 - \alpha)$ -Konfidenzmenge. (1P)

- (b) Es seien  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  i.i.d. Beobachtungen mit  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  unbekannt. Weisen Sie für  $\mu_0 \in \mathbb{R}$  nach, dass ein (nicht-randomisierter) Likelihood-Quotienten-Test auf  $H_0 : \mu = \mu_0, \sigma > 0$  gegen  $H_1 : \mu \neq \mu_0, \sigma > 0$  folgende Form mit geeignetem  $c_\alpha \geq 0$  besitzt: (2P)

$$\varphi_\alpha = \mathbf{1}(|T| \geq c_\alpha) \text{ mit } T := \frac{\bar{X} - \mu_0}{\hat{\sigma}}, \hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- (c) Wie ist  $T$  verteilt und wie muss  $c_\alpha$  gewählt werden, damit  $\varphi_\alpha$  Niveau  $\alpha \in (0, 1)$  besitzt? (1P)
- (d) Leiten Sie aus dem Vorangegangenen eine  $(1 - \alpha)$ -Konfidenzmenge für  $\mu$  her. (1P)

3. Betrachten Sie das gewöhnliche lineare Modell  $Y = X\beta + \varepsilon$  mit  $\beta \in \mathbb{R}^p$ ,  $X \in \mathbb{R}^{n \times p}$  mit  $\text{rank}(X) = p \leq n$  und  $\mathbb{E}[\varepsilon] = 0$ ,  $\text{Cov}(\varepsilon) = \sigma^2 E_n$ . Definiere den linearen Schätzer

$$\hat{\beta} = AY \text{ mit } A \in \mathbb{R}^{p \times n} \text{ deterministisch.}$$

- (a) Bestimmen Sie Bias, Varianz und mittleren quadratischen Fehler (MSE) von  $\hat{\beta}$  in Abhängigkeit von  $A$ . (2P)
- (b) Zeigen Sie für Vorhersagefehler und die *residual sum of squares*  $RSS = |X\hat{\beta} - Y|^2$  den allgemeinen Zusammenhang (2P)

$$\mathbb{E}[|X(\hat{\beta} - \beta)|^2] = \mathbb{E}[RSS] + 2\sigma^2 \text{trace}(XA) - \sigma^2 n.$$

- (c) Leiten Sie auf diesem Weg für den Kleinste-Quadrate Schätzer ab, dass das AIC-Kriterium für die Modellwahl die Kullback-Leibler-Devianz erwartungstreu schätzt. (1P)

4. Es seien  $(X, Y)$  ein Zufallsvektor auf  $\mathbb{R}^p \times \{-1, +1\}$  sowie  $(X_1, Y_1), \dots, (X_n, Y_n)$  unabhängige, wie  $(X, Y)$  verteilte Beobachtungen. Für einen Klassifizierer  $C : \mathbb{R}^p \rightarrow \{-1, +1\}$  betrachte den Klassifizierungsfehler  $R(C) := \mathbb{P}(C(X) \neq Y)$ .
- (a) Zeigen Sie, dass der Bayes-Klassifizierer  $C^*(x) = \text{sgn}(2\mathbb{P}(Y = +1 | X = x) - 1)$  den Klassifizierungsfehler minimiert. Wieso ist der Bayes-Klassifizierer für den Statistiker unzugänglich? (2P)
- (b) Auf welche Klassifikationsmethode führt der Ansatz, dass  $X$  gegeben  $\{Y = k\}$   $N(\mu_k, \Sigma)$ -verteilt mit  $\mu_{-1} \neq \mu_{+1}$  und  $\Sigma > 0$  ist? Wie sieht die zugehörige Klassifikationsgrenze (*classification boundary*) im  $\mathbb{R}^p$  geometrisch aus? (1P)
- (c) Setze  $p = 1$  und modelliere  $\mathbb{P}(Y = +1) = \mathbb{P}(Y = -1) = 1/2$  sowie  $f^{X|Y=k}(x) = \lambda_k e^{-\lambda_k x}$ ,  $x \geq 0$ ,  $k \in \{-1, +1\}$  (exponential-verteilt mit  $0 < \lambda_{-1} < \lambda_{+1}$ ). Benutzen Sie die Bayesformel, um den Bayesklassifizierer zu bestimmen. Ersetzen Sie in diesem  $(\lambda_{-1}, \lambda_{+1})$  durch einen Maximum-Likelihood-Schätzer  $(\hat{\lambda}_{-1}, \hat{\lambda}_{+1})$ , basierend auf den Daten  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Wie heißt dieses Konstruktionsprinzip eines Klassifizierers? (2P)

---

Abgabe **vor** der Vorlesung am Montag, 2.2.26.