

Skriptum (Vers. 2024-03) zur LV
Angewandte (Mathematische) Statistik

LV-Nr. 405.170

Univ.-Prof. Dr. Wolfgang Trutschnig

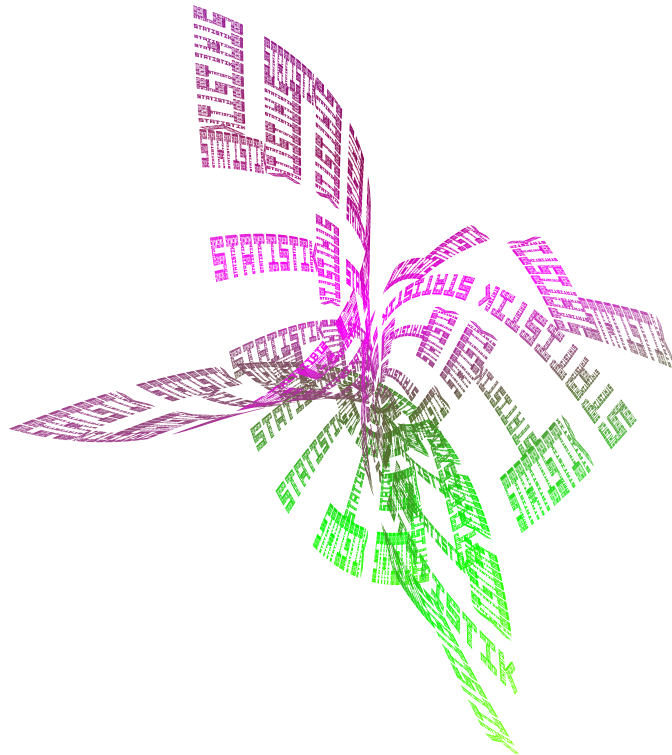
www.trutschnig.net

Fachbereich AIHI

Paris Lodron Universität Salzburg

Hellbrunner Strasse 34

A-5020 Salzburg



Inhaltsverzeichnis

1	Konstruktion von Schätzern	3
1.1	Maximum Likelihood Schätzer	3
1.2	Momentenschätzer	7
2	Ergänzungen zur mathematischen Statistik	10
3	Hilfsresultate über Verteilungen	13
4	Konfidenzbereiche	19
4.1	Konfidenzbereiche für endlichdimensionale Parameter	19
4.2	Konfidenzbänder	25
5	Hypothesentests	29
6	A glimpse on Regression	35
6.1	Lineare Regression	36
6.2	Logistische Regression	38
6.3	Overfitting, Cross-Validation, Feature Selection	40
7	Übungsblätter	42

Kapitel 1

Konstruktion von Schätzern

Wir betrachten im Folgenden die zwei gängigsten Methoden zur Konstruktion von Schätzern - die Maximum Likelihood Methode und die Momentenmethode.

1.1 Maximum Likelihood Schätzer

Sei $X \sim (P_\theta)_{\theta \in \Theta}$, (X_1, \dots, X_n) eine Zufallsstichprobe von X und (x_1, \dots, x_n) eine konkrete Realisierung. Wie gewohnt setzen wir voraus, dass $(P_\theta)_{\theta \in \Theta}$ identifizierbar ist. Zusätzlich verwenden wir dieselbe Notation für den diskreten und den absolut stetigen Fall:

- Falls X diskret mit (vom Parameter θ unabhängigen) Wertebereich $W = \{\alpha_1, \alpha_2, \dots\}$ ist, schreiben wir

$$f(x; \theta) := P_\theta(\{x\})$$

für jedes $x \in W$.

- Falls X absolut stetig ist, bezeichnet $f(x; \theta)$ wie gewohnt die Dichte von $P^X = P_\theta$ (bezüglich λ).

Definition 1.1. Sei $X \sim (P_\theta)_{\theta \in \Theta}$, (X_1, \dots, X_n) eine Zufallsstichprobe von X und (x_1, \dots, x_n) eine konkrete Realisierung. Dann heißt die Funktion

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \tag{1.1}$$

Likelihood Funktion (kurz: Likelihood).

Im diskreten Fall entspricht $L(x_1, \dots, x_n; \theta)$ also genau der Wahrscheinlichkeit

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n),$$

im absolut stetigen Fall genau dem Wert der Wahrscheinlichkeitsdichte des Vektors (X_1, \dots, X_n) an der Stelle (x_1, \dots, x_n) .

Die Grundidee der Maximum Likelihood Methode besteht nun einfach darin, als Schätzer $\hat{\theta}_n$ für θ jenen Wert aus Θ zu verwenden, der $L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$ maximiert (falls existent). Im diskreten Fall bedeutet dies also, dass man genau jenen Parameter sucht, für den die beobachtete Realisierung maximale Wahrscheinlichkeit hat.

Bevor wir Maximum Likelihood Schätzer (kurz: MLEs) allgemein definieren, betrachten wir die folgenden zwei Beispiele, die die Nützlichkeit und mögliche Güte von MLEs gut illustrieren.

Beispiel 1.2. Sei $X \sim (\text{Bin}(1, \theta))_{\theta \in [0,1]}$. Dann erhalten wir für die Likelihood Funktion

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

Wir unterscheiden drei Fälle:

1. Für $\sum_{i=1}^n x_i = 0$ folgt $L(x_1, \dots, x_n; \theta) = (1 - \theta)^n$, die Likelihood Funktion ist also streng monoton fallend in θ und das Maximum wird in $\theta = 0 = \frac{1}{n} \sum_{i=1}^n x_i$ angenommen.
2. Für $\sum_{i=1}^n x_i = n$ folgt $L(x_1, \dots, x_n; \theta) = \theta^n$, die Likelihood Funktion ist also streng monoton wachsend in θ und das Maximum wird in $\theta = 1 = \frac{1}{n} \sum_{i=1}^n x_i$ angenommen.
3. $\sum_{i=1}^n x_i \in \{1, 2, \dots, n - 1\}$. Nachdem $L(x_1, \dots, x_n; \theta) \geq 0$ für jedes $\theta \in [0, 1]$ und $L(x_1, \dots, x_n; 0) = 0 = L(x_1, \dots, x_n; 1)$ reicht es, $\theta \in (0, 1)$ zu betrachten und, zur Vereinfachung der Rechenarbeit, lokale Extrema von $\ln L(x_1, \dots, x_n; \theta)$ zu berechnen: Wir setzen $k := \sum_{i=1}^n x_i$, erhalten

$$\ln L(x_1, \dots, x_n; \theta) = k \ln \theta + (n - k) \ln(1 - \theta)$$

und damit

$$\begin{aligned} \frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} &= \frac{k}{\theta} - \frac{n - k}{1 - \theta} = \frac{k - \theta n}{\theta(1 - \theta)} \\ \frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial \theta^2} &= -\frac{k}{\theta^2} - \frac{n - k}{(1 - \theta)^2} < 0. \end{aligned}$$

Damit ist $\ln L(x_1, \dots, x_n; \theta)$ genau dann maximal wenn $\theta = \frac{k}{n} = \frac{1}{n} \sum_{i=1}^n x_i$.

Alle drei Fälle liefern also dasselbe Resultat

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n,$$

i.e. der durch Maximierung des Likelihood erhaltenen Schätzer ist genau \bar{X}_n . Nachdem \bar{X}_n in dieser Situation effizient für θ ist, liefert also die Maximierung des Likelihood einen effizienten und noch dazu stark konsistenten Schätzer.

Beispiel 1.3. Sei $X \sim (\text{Pois}(\theta))_{\theta \in [0, \infty)}$. Dann ist die Likelihoodfunktion L gegeben durch

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \frac{1}{\prod_{i=1}^n x_i!} \theta^{\sum_{i=1}^n x_i} e^{-n\theta}.$$

Wir setzen $k := \sum_{i=1}^n x_i$, $c := \frac{1}{\prod_{i=1}^n x_i!}$ und betrachten die folgenden beiden Fälle:

- (i) Für $k = 0$ gilt $L(x_1, \dots, x_n; \theta) = e^{-n\theta}$, der MLE ist also gegeben durch $\hat{\theta}_n = 0 = \bar{X}_n$.
- (ii) Für $k \neq 0$ erhalten wir

$$\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} (\ln c + k \ln \theta - n\theta) = \frac{k}{\theta} - n.$$

Die zweite Ableitung ist offensichtlich negativ, $L(x_1, \dots, x_n; \theta)$ ist daher genau für $\theta = \frac{k}{n}$ maximal und der MLE ist auch in diesem Fall gegeben durch $\hat{\theta}_n = \bar{X}_n$. Der erhaltene MLE hat also sehr gute Eigenschaften, er ist stark konsistent und effizient.

Beispiel 1.4. Wir betrachten den Fall $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$. Dann ergibt sich (Übungsaufgabe) für den MLE $\hat{\theta}_n$

$$\hat{\theta}_n = ((\hat{\theta}_1)_n, (\hat{\theta}_2)_n) = \left(\bar{X}_n, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right).$$

Aus den vorhergehenden Kapiteln wissen wir, dass sowohl $(\hat{\theta}_1)_n$ als auch $(\hat{\theta}_2)_n$ stark konsistente Schätzer von $\theta_1 = \mu$ bzw. $\theta_2 = \sigma^2$ sind. Weiters ist $(\hat{\theta}_1)_n$ erwartungstreu, $(\hat{\theta}_2)_n$ asymptotisch erwartungstreu.

Definition 1.5. Sei $X \sim (P_\theta)_{\theta \in \Theta}$ und (X_1, \dots, X_n) eine Zufallsstichprobe von X . Dann heißt ein Schätzer $\hat{\theta}_n \in \Theta$ von θ Maximum Likelihood Schätzer (MLE) genau dann wenn

$$L(X_1, \dots, X_n; \hat{\theta}_n) = \sup_{\theta \in \Theta} L(X_1, \dots, X_n; \theta),$$

i.e. wenn er die Likelihoodfunktion maximiert.

Bemerkung 1.6. MLEs müssen (im Allgemeinen) weder existieren noch eindeutig sein. MLEs sind spezielle M-Schätzer, 'M' für Maximierung.

Wie vorhin schon erwähnt entspricht der MLE im diskreten Fall jenem Parameter θ für den die beobachtete Realisierung maximale Wahrscheinlichkeit hat. Obwohl diese Interpretation offensichtlich nicht direkt auf den absolut stetigen Fall übertragen werden kann (jede Realisierung hat Wahrscheinlichkeit Null) kann das 'Funktionieren' der MLE Methode auch in diesem Fall zumindest *heuristisch* begründet werden: Wir starten mit dem folgenden einfachen Lemma:

Lemma 1.7. Gelte $X \sim (P_\theta)_{\theta \in \Theta}$, jedes P_θ sei absolut stetig mit Dichte $f(\cdot; \theta)$ und $\theta_0 \in \Theta$. Weiter sei die Funktion $m : \Theta \rightarrow [-\infty, \infty]$ definiert durch

$$m(\theta) = \mathbb{E}_{\theta_0} \ln \left(\frac{f(X; \theta)}{f(X; \theta_0)} \right).$$

Dann nimmt die Funktion m ihr Maximum genau in θ_0 an.

Beweis: (i) Offensichtlich gilt $m(\theta_0) = \mathbb{E}_{\theta_0} \ln(1) = 0$. Es genügt daher zu zeigen, dass $m(\theta) < 0$ für alle $\theta \neq \theta_0$. (ii) Nachdem der Logarithmus konkav auf $(0, \infty)$ ist und $t \mapsto t - 1$ genau die Tangente im Punkt 1 ist, gilt offensichtlich $\ln(t) \leq t - 1$ auf $(0, \infty)$. Für $t = \sqrt{x}$ folgt daher die Gültigkeit der Ungleichung $\ln(x) \leq 2(\sqrt{x} - 1)$ auf $(0, \infty)$ und wir erhalten

$$\begin{aligned} m(\theta) &= \mathbb{E}_{\theta_0} \ln \left(\frac{f(X; \theta)}{f(X; \theta_0)} \right) \leq \mathbb{E}_{\theta_0} \left(2\sqrt{\frac{f(X; \theta)}{f(X; \theta_0)}} - 2 \right) \\ &= 2 \int \sqrt{\frac{f(x; \theta)}{f(x; \theta_0)}} f(x; \theta_0) d\lambda(x) - 2 = 2 \int \sqrt{f(x; \theta) f(x; \theta_0)} d\lambda(x) - 2 \\ &= - \int \left(\sqrt{f(x; \theta)} - \sqrt{f(x; \theta_0)} \right)^2 d\lambda(x) < 0 \end{aligned}$$

wobei die letzte Ungleichung genau dann gilt, wenn $\theta \neq \theta_0$. ■

Falls nun die Zufallsvariable $\ln(f(X; \theta))$ integrierbar ist, dann liefert das SLLN sofort

$$\frac{1}{n} \ln L(X_1, \dots, X_n; \theta) = \frac{1}{n} \sum_{i=1}^n \ln(f(X_i; \theta)) \xrightarrow{[P]} \mathbb{E}_{\theta_0}(\ln(f(X; \theta)))$$

Unter Verwendung von Lemma 1.7 wird die rechte Seite genau für $\theta = \theta_0$ maximal - es macht also Sinn, die linke Seite - die Likelihoodfunktion - nach θ zu maximieren.

Allgemein haben Maximum-Likelihood Schätzer (allgemeiner, M-Schätzer) für den Fall, dass $f(x; \theta)$ gewisse, nicht allzu restriktive Regularitätsbedingungen erfüllt (im Wesentlichen Glattheit von $f(x; \theta)$, Kompaktheit des Parameterraums Θ und Eindeutigkeit des MLE) gute Eigenschaften. Nachdem wir an dieser Stelle nicht näher auf diese Bedingungen eingehen können, zählen wir die wesentlichsten 'guten' asymptotischen Eigenschaften nur ohne Beweis und genaue Formulierung der Voraussetzungen im folgenden Satz auf.

Satz 1.8. *Sei $X \sim (P_\theta)_{\theta \in \Theta}$ mit $\Theta \subseteq \mathbb{R}$ und $\hat{\theta}_n$ Maximum Likelihood Schätzer von θ . Unter gewissen Regularitätsvoraussetzung hat $\hat{\theta}_n$ die folgenden Eigenschaften:*

1. $\hat{\theta}_n$ ist konsistent für θ , i.e. für jedes feste $\theta_0 \in \Theta$ gilt: falls $X \sim P_{\theta_0}$ dann folgt $\hat{\theta}_n \xrightarrow{P} \theta_0$.
2. $\hat{\theta}_n$ ist invariant, i.e. für eine messbare Funktion g ist $g \circ \hat{\theta}_n$ Maximum Likelihood Schätzer von $g(\theta)$.
3. $\hat{\theta}_n$ ist asymptotisch normal, i.e. für jedes feste $\theta_0 \in \Theta$ gilt: falls $X \sim P_{\theta_0}$ dann folgt

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{w} Z$$

wobei $Z \sim \mathcal{N}(0, \frac{1}{I(\theta_0)})$.

Beispiel 1.9. Wir kehren nochmals zurück zur Alternativverteilung. Für $X \sim \text{Bin}(1, \theta)$ rechnet man leicht nach, dass $\hat{\theta}_n = \bar{X}_n$ der MLE für θ ist. In diesem Fall ist der MLE also sogar stark konsistent und effizient, es gilt also etwas mehr als Aussage eins in Satz 1.8. Falls $X \sim \text{Bin}(1, \theta_0)$, dann impliziert das CLT, dass

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \theta_0}{\sqrt{\theta_0(1 - \theta_0)}} \xrightarrow{w} Z$$

wobei $Z \sim \mathcal{N}(0, 1)$. Wegen $I(\theta_0) = \frac{1}{\theta_0(1 - \theta_0)}$ folgt daraus sofort die dritte Aussage in Satz 1.8.

1.2 Momentenschätzer

Wir beginnen mit einem einfachen Beispiel, das die, der Momentenmethode zugrundeliegende Idee veranschaulicht:

Beispiel 1.10. Sei $X \sim (\mathcal{E}(\theta))_{\theta \in (0, \infty)}$. Für Zufallsstichproben X_1, X_2, \dots, X_n von X impliziert das SLLN

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{[P]} \mathbb{E}_\theta(X) = \frac{1}{\theta} =: g(\theta).$$

Nachdem $g : (0, \infty) \rightarrow (0, \infty)$ bijektiv und stetig ist, ist

$$\hat{\theta}_n := g^{-1}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{\bar{X}_n}$$

ein stark konsistenter Schätzer von θ (vergleiche mit dem MLE für θ in den Übungen).

Für $X \sim (P_\theta)_{\theta \in \Theta}$ mit $\Theta \subseteq \mathbb{R}^d$ kann analog vorgegangen werden: Angenommen X^d ist integrierbar und es existieren Funktionen g_1, g_2, \dots, g_d sodass $\mathbb{E}_\theta(X^k) = g_k(\theta_1, \theta_2, \dots, \theta_d) = g_k(\theta)$ für jedes $k \in \{1, \dots, d\}$. Dann impliziert das SLLN

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i &\xrightarrow{[P]} g_1(\theta_1, \dots, \theta_d) \\ \frac{1}{n} \sum_{i=1}^n X_i^2 &\xrightarrow{[P]} g_2(\theta_1, \dots, \theta_d) \\ &\vdots \\ \frac{1}{n} \sum_{i=1}^n X_i^d &\xrightarrow{[P]} g_d(\theta_1, \dots, \theta_d) \end{aligned}$$

und wir erhalten einen Schätzer $\hat{\theta}_n = ((\hat{\theta}_1)_n, (\hat{\theta}_2)_n, \dots, (\hat{\theta}_d)_n)$ für θ falls das d -dimensionale Gleichungssystem

$$\left. \begin{aligned} m_1 &:= \frac{1}{n} \sum_{i=1}^n X_i = g_1(\theta_1, \dots, \theta_d) \\ m_2 &:= \frac{1}{n} \sum_{i=1}^n X_i^2 = g_2(\theta_1, \dots, \theta_d) \\ &\vdots \\ m_d &:= \frac{1}{n} \sum_{i=1}^n X_i^d = g_d(\theta_1, \dots, \theta_d) \end{aligned} \right\} =: \mathbf{g}(\theta) \quad (1.2)$$

eindeutig nach θ auflösbar ist. Der resultierende Schätzer $\hat{\theta}_n = \mathbf{g}^{-1}(m_1, \dots, m_d)$ heißt *Momentenschätzer von θ* . $\hat{\theta}_n = \mathbf{g}^{-1}(m_1, \dots, m_d)$ ist automatisch stark konsistent wenn \mathbf{g}^{-1} stetig ist.

Beispiel 1.11. Sei $X \sim \mathcal{N}(\theta_1, \theta_2)_{(\theta_1, \theta_2) \in \mathbb{R} \times (0, \infty)}$. Wir berechnen den Momentenschätzer von $\theta = (\theta_1, \theta_2)$: Wegen $\mathbb{E}_\theta(X) = \theta_1$ und $\mathbb{E}_\theta(X^2) = \mathbb{V}_\theta(X) + \mathbb{E}_\theta(X)^2 = \theta_2 + \theta_1^2$ nimmt das Gleichungssystem (1.2) die folgende einfache Form an:

$$\left. \begin{aligned} m_1 &:= \frac{1}{n} \sum_{i=1}^n X_i = \theta_1 = g_1(\theta_1, \theta_2) \\ m_2 &:= \frac{1}{n} \sum_{i=1}^n X_i^2 = \theta_2 + \theta_1^2 = g_2(\theta_1, \theta_2) \end{aligned} \right\} =: \mathbf{g}(\theta) \quad (1.3)$$

und wir erhalten als Momentenschätzer sofort

$$\hat{\theta}_n = \left(\bar{X}_n, \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 \right) = \left(\bar{X}_n, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right),$$

also genau den im vorigen Abschnitt erhaltenen MLE (der asymptotisch erwartungstreu und stark konsistent ist).

Genauso wie Maximum Likelihood Schätzer haben auch Momentenschätzer unter nicht allzu restriktiven Regularitätsbedingungen gute asymptotische Eigenschaften - wie im vorigen Abschnitt formulieren wir nur das eindimensionale Resultat ohne genaue Angabe hinreichender Bedingungen; analoge Eigenschaften gelten auch im allgemeinen multivariaten Fall $\Theta \subseteq \mathbb{R}^d$:

Satz 1.12. Sei $X \sim (P_\theta)_{\theta \in \Theta}$ mit $\Theta \subseteq \mathbb{R}$ und $\hat{\theta}_n$ der Momentenschätzer von θ . Weiters gelte $g(\theta) = \mathbb{E}_\theta(X)$ für alle $\theta \in \Theta$. Unter gewissen Regularitätsvoraussetzung hat $\hat{\theta}_n$ die folgenden Eigenschaften:

1. Die Wahrscheinlichkeit, dass $\hat{\theta}_n$ existiert, geht für $n \rightarrow \infty$ gegen 1.
2. $\hat{\theta}_n$ ist konsistent für θ , i.e. für jedes feste $\theta_0 \in \Theta$ gilt: falls $X \sim P_{\theta_0}$ dann folgt $\hat{\theta}_n \xrightarrow{P} \theta_0$.
3. $\hat{\theta}_n$ ist asymptotisch normal, i.e. für jedes feste $\theta_0 \in \Theta$ gilt: falls $X \sim P_{\theta_0}$ dann folgt

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{w} Z$$

$$\text{wobei } Z \sim \mathcal{N}\left(0, \frac{\mathbb{V}_{\theta_0}(X)}{(g'(\theta_0))^2}\right).$$

Beispiel 1.13. Wir betrachten wieder die in Beispiel 1.10 behandelte Situation $X \sim (\mathcal{E}(\theta))_{\theta \in (0, \infty)}$. Die Existenz des Schätzers $\hat{\theta}_n = 1/\bar{X}_n$ ist klar, starke Konsistenz ebenfalls. Auch die im vorhergehenden Satz behauptete asymptotische Normalität lässt sich mit Hilfe des CLT und der sogenannten δ -Methode unschwer überprüfen.

Beispiel 1.14. Für den Fall $X \sim \text{Bin}(1, \theta)_{\theta \in [0, 1]}$ ist der Momentenschätzer offensichtlich gegeben durch $\hat{\theta}_n = \bar{X}$, er stimmt also mit dem MLE überein und ist stark konsistent. Wir haben schon in Beispiel 1.9 gezeigt, dass für $X \sim \text{Bin}(1, \theta_0)$

$$Z_n = \sqrt{n} \frac{\hat{\theta}_n - \theta_0}{\sqrt{\theta_0(1 - \theta_0)}} \xrightarrow{w} Z$$

gilt, wobei $Z \sim \mathcal{N}(0, 1)$. Nachdem in diesem Fall $g' = 1$ und $\mathbb{V}_{\theta_0}(X) = \theta_0(1 - \theta_0)$ erhalten wir also genau die dritte, im vorhergehenden Satz behauptete Eigenschaft.

Abschließend geben wir noch ein Resultat, das zeigt, wie sich die asymptotische Normalität der zugrundeliegenden Zufallsvariablen auf eine Funktion dieser Zufallsvariablen vererbt, falls die Funktion gewisse Eigenschaften erfüllt.

Satz 1.15 (Spezialfall δ -Methode). $U \subseteq \mathbb{R}$ sei offen, die Abbildung $\phi : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ sei differenzierbar im Punkt $\theta \in U$ und erfülle $\phi'(\theta) \neq 0$. Die Zufallsvariable T, T_1, T_2, \dots seien U -wertig. Dann impliziert $\sqrt{n} \cdot (T_n - \theta) \xrightarrow{w} T \sim \mathcal{N}(0, \sigma^2)$ für $n \rightarrow \infty$ die Gültigkeit von

$$\sqrt{n} (\phi(T_n) - \phi(\theta)) \xrightarrow{w} \mathcal{N}(0, \sigma^2 \cdot (\Phi'(\theta)^2))$$

für $n \rightarrow \infty$.

Beweis: Übungsaufgabe.

Als Anwendung der δ -Methode betrachten wir wiederum Beispiel 1.10 und zeigen die asymptotische Normalität des Schätzers.

Beispiel 1.16. Sei $X \sim (\mathcal{E}(\theta))_{\theta \in (0, \infty)}$. Aus dem CLT wissen wir $Z_n \xrightarrow{w} Z \sim \mathcal{N}(0, 1)$ wobei Z_n für jedes n gegeben ist durch

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \frac{1}{\theta}}{\frac{1}{\theta}}.$$

Insbesondere gilt also

$$\sqrt{n} \left(\bar{X}_n - \frac{1}{\theta} \right) \xrightarrow{w} Y \sim \mathcal{N}\left(0, \frac{1}{\theta^2}\right).$$

Wir setzen $\mu = \frac{1}{\theta}$ sowie $\phi(\theta) = \frac{1}{\theta}$ und erhalten als direkte Folgerung von Satz 1.15 sofort

$$\sqrt{n} \left(\frac{1}{\bar{X}_n} - \theta \right) \xrightarrow{w} Z \sim \mathcal{N}\left(0, \frac{1}{\mu^4} \mu^2\right) = \mathcal{N}(0, \theta^2).$$

$\hat{\theta}_n$ ist asymptotisch normalverteilt.

Kapitel 2

Ergänzungen zur mathematischen Statistik

Ziel dieses Kapitels ist es, zwei Verschärfungen des Satzes von Glivenko-Cantelli und eine direkte Folgerung des CLT kennenzulernen. Sei $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge von Zufallsvariablen mit Verteilungsfunktion F , \hat{F}_n bezeichne die empirische Verteilungsfunktion, i.e.

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(X_i).$$

Definieren wir die Zufallsvariable D_n durch

$$D_n(w) := \sup_{x \in \mathbb{R}} |\hat{F}_n(x)(\omega) - F(x)|, \quad (2.1)$$

dann gilt nach dem Satz von Glivenko-Cantelli

$$\lim_{n \rightarrow \infty} D_n = 0 \quad [\mathbb{P}].$$

Die folgenden beiden nichttrivialen Resultate geben Aufschluss über die asymptotische Verteilung von D_n - beachten Sie, dass die konkrete Form von F nicht in die Ausdrücke auf der rechten Seite eingeht.

Satz 2.1. *Sei $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge von Zufallsvariablen mit stetiger Verteilungsfunktion F . Dann gilt:*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n} D_n \leq x) = F_{KS}(x) = \begin{cases} 0 & , x \leq 0 \\ 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2} & , x > 0 \end{cases}$$

Die Funktion F_{KS} ist die sogenannte Kolmogorov-Smirnov Verteilungsfunktion.

Für allgemeines, nicht notwendigerweise stetiges F gilt die folgende Dvoretzky, Kiefer, and Wolfowitz Ungleichung (bestmöglich im Sinne, dass Zufallsvariable existieren, für die die Ungleichung scharf ist).

Satz 2.2 ([3]). *Sei $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge von Zufallsvariablen mit Verteilungsfunktion F . Dann gilt für jedes $x > 0$ und jedes $n \in \mathbb{N}$*

$$\mathbb{P}(\sqrt{n} D_n > x) \leq 2e^{-2x^2}$$

Wir überlegen uns nun, dass im stetigen Fall nicht nur die Grenzverteilung von $\sqrt{n}D_n$ sondern sogar die Verteilung von D_n nicht von F abhängt, Schlüssel zu dem Resultat ist die PIT. Für

Satz 2.3. Sei F eine stetige Verteilungsfunktion und $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge von (nicht notwendigerweise integrierbaren) Zufallsvariablen mit Verteilungsfunktion F . Weiters bezeichne $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}$ die Ordnungsstatistik einer Zufallsstichprobe U_1, \dots, U_n von $U \sim \mathcal{U}(0, 1)$.ⁱ Definieren wir die Zufallsvariable d_n durch

$$d_n(\omega) = \max_{i=1, \dots, n} \left\{ \max \left\{ \left| U_{(i)} - \frac{i}{n} \right|, \left| U_{(i)} - \frac{i-1}{n} \right| \right\} \right\}, \quad (2.2)$$

dann haben D_n und d_n die selbe Verteilung.

Beweis: Sei $n \in \mathbb{N}$ fest (aber beliebig). Da F stetig ist, sind X_1, \dots, X_n mit Wahrscheinlichkeit 1 paarweise verschieden, i.e., mit Wahrscheinlichkeit 1 gilt $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. Aufgrund der Stetigkeit von F wird $\sup_{x \in \mathbb{R}} |\hat{F}_n(x)(\omega) - F(x)|$ in einer der Sprungstellen der empirischen Verteilungsfunktion \hat{F}_n , also in (zumindest) einem der Werte $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ angenommen (Skizze!) und es gilt

$$D_n(\omega) = \max_{i=1, \dots, n} \left\{ \max \left\{ \left| F \circ X_{(i)}(\omega) - \frac{i}{n} \right|, \left| F \circ X_{(i)}(\omega) - \frac{i-1}{n} \right| \right\} \right\}.$$

Setzen wir $U_i = F \circ X_i \sim \mathcal{U}(0, 1)$ für jedes $i \in \{1, \dots, n\}$, dann sind U_1, \dots, U_n unabhängig und stetig gleichverteilt auf $[0, 1]$ (PIT); zusätzlich gelten mit Wahrscheinlichkeit 1 die folgenden zwei Aussagen (Übungsaufgabe)

1. $U_{(1)} < U_{(2)} < \dots < U_{(n)}$
2. $F \circ X_{(i)} = U_{(i)}$ für jedes $i \in \{1, \dots, n\}$.

Die behauptete Eigenschaft ist damit bewiesen. ■

Bemerkung 2.4. Beachten Sie, dass d_n als Funktion der Stichprobe (U_1, \dots, U_n) von $U \sim \mathcal{U}(0, 1)$ nicht von F abhängt. Die Verteilungsfunktion F_{d_n} von d_n (die wir später zur Konstruktion von exakten (nicht asymptotischen) Konfidenzbändern verwenden werden) lässt sich zwar nicht schön analytisch darstellen, Sie kann aber unter Verwendung des Satzes von Glivenko Cantelli beliebig genau mittels Simulationen in R approximiert werden.

Im Sinne der nächsten Übungsaufgaben skizzieren wir weiters vorab eine direkte Anwendung des CLTs zur Konstruktion von asymptotischen Konfidenzintervallen, die Sie schon in der VO Wahrscheinlichkeitsrechnung kennengelernt haben:

Beispiel 2.5. Wir berechnen ein asymptotisches Konfidenzintervall (CI) für den Parameter θ der Exponentialverteilung. Es sei $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge von Zufallsvariablen mit $X_1 \sim \mathcal{E}(\theta)$. Bekanntlich gilt dann $\mathbb{E}(X_1) = \frac{1}{\theta}$ und $\mathbb{V}(X_1) = \frac{1}{\theta^2}$. Nach CLT konvergiert Z_n , definiert durch

$$Z_n = \sqrt{n} \cdot \frac{\bar{X}_n - \frac{1}{\theta}}{\frac{1}{\theta}}$$

ⁱ $U_{(1)}$ bezeichne den kleinsten Wert, $U_{(2)}$ den zweitkleinsten Wert, etc.

schwach gegen $Z \sim \mathcal{N}(0, 1)$. Für $\alpha \in (0, 1]$ folgt damit sofort

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(-z_{1-\frac{\alpha}{2}} \leq Z_n \leq z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha,$$

wobei, wie gewohnt, $z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$. Offensichtlich gilt $-z_{1-\frac{\alpha}{2}} \leq Z_n(\omega) \leq z_{1-\frac{\alpha}{2}}$ genau dann, wenn

$$\frac{1}{\bar{X}_n(\omega)} \left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right) \leq \theta \leq \frac{1}{\bar{X}_n(\omega)} \left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right).$$

Definieren wir $C_n^{1-\alpha}$ durch

$$C_n^{1-\alpha} := \frac{1}{\bar{X}_n} \left[1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, 1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right] \quad (2.3)$$

dann gilt also $\lim_{n \rightarrow \infty} \mathbb{P}_\theta(\theta \in C_n^{1-\alpha}) = 1 - \alpha$, das zufällige Intervall $C_n^{1-\alpha}$ enthält den wahren Parameter also asymptotisch mit Wahrscheinlichkeit $1 - \alpha$.

Kapitel 3

Hilfsresultate über Verteilungen

Für die in den kommenden Kapiteln studierten Resultate aus der klassischen Statistik benötigen wir einige Hilfsresultate über Verteilungen, die in diesem Abschnitt zusammengefasst werden. Wir starten mit einer Erweiterung des univariaten Transformationssatzes für Dichten. Gegeben sei ein Zufallsvektor $X = (X_1, \dots, X_d)$ gegeben sowie eine Funktion $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Im Allgemeinen muss der Zufallsvektor $Y = T \circ X$ nicht absolut stetig sein - wenn T jedoch ausreichend regulär ist, ist Y absolut stetig und die Dichte von Y lässt sich explizit angeben:

Satz 3.1 (Transformationssatz für Wahrscheinlichkeitsdichten; einfacher Fall).

Sei $\mathcal{X} = (X_1, \dots, X_d)$ absolut stetig mit Dichte $f : \mathbb{R}^d \rightarrow [0, \infty]$, $U \subseteq \mathbb{R}^d$ offen und $\mathbb{P}(\mathcal{X} \in U) = 1$. Weiters sei $T : U \rightarrow \mathbb{R}^d$ injektiv und stetig differenzierbar, die Funktionaldeterminante J_T erfülle $J_T(x) \neq 0$ für jedes $x \in U$. Dann gilt:

1. $T(U)$ ist offen
2. $\mathcal{Y} = T \circ \mathcal{X}$ ist absolut stetig mit Dichte

$$g(y) = \frac{f \circ T^{-1}(y)}{|J_T(T^{-1}(y))|} \mathbb{1}_{T(U)}(y)$$

Beweis: Mit Analysis 3 einfach zu lösende Übungsaufgabe (Transformationsformel für das Lebesgue Integral).

Beispiel 3.2. Seien $X, Y \sim \mathcal{N}(0, 1)$ und unabhängig. Wir berechnen die Dichte von $Z := \frac{X}{Y}$. Nachdem $(x, y) \mapsto \frac{x}{y}$ nicht injektiv ist, gehen wir wie folgt vor: Wir setzen $\mathcal{X} = (X, Y)$, $U := \mathbb{R}^2 \setminus (\mathbb{R} \times \{0\})$ und betrachten die Transformation $T : U \rightarrow U$, definiert durch $T(x, y) := (\frac{x}{y}, y)$. Dann ist U offen, \mathcal{X} ist absolut stetig mit Dichte

$$f_{\mathcal{X}}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}},$$

und es gilt $\mathbb{P}(\mathcal{X} \in U) = 1 - \mathbb{P}(X \in \mathbb{R}, Y = 0) = 1$. Weiters ist $T : U \rightarrow U$ injektiv und für $J_T(x, y)$ gilt

$$|J_T((x, y))| = \begin{vmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{y} \neq 0,$$

alle Voraussetzungen von Satz 3.1 sind also erfüllt. T^{-1} ist offensichtlich gegeben durch $T^{-1}(v, w) = (vw, w)$, der Transformationssatz liefert also, dass $\mathcal{Y} = T \circ \mathcal{X}$ absolut stetig mit Dichte

$$g(v, w) = \frac{1}{2\pi} e^{-\frac{v^2 w^2 + w^2}{2}} |w| \mathbb{1}_U(v, w)$$

ist. Die gewünschte Dichte von $\frac{X}{Y}$ erhalten wir schliesslich als erste Randdichte von g , i.e.,

$$\begin{aligned} g_1(v) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{w^2(v^2+1)}{2}} |w| dw = \frac{2}{2\pi} \int_0^\infty e^{-\frac{w^2(v^2+1)}{2}} w dw \\ &= \frac{1}{2\pi} \int_0^\infty e^{-z \frac{v^2+1}{2}} dz = \dots = \frac{1}{\pi(1+v^2)}. \end{aligned}$$

Mit anderen Worten: $\frac{X}{Y}$ ist Cauchy-verteilt.

Für den Spezialfall, dass T linear ist liefert Satz 3.1 sofort folgende Konsequenz:

Folgerung 3.3. $\mathcal{X} = (X_1, \dots, X_d)$ habe Dichte f , weiters sei $A \in \mathbb{R}^{d \times d}$ regulär (invertierbar). Dann hat $\mathcal{Y} = A \cdot \mathcal{X}$ Dichte g mit

$$g(y) = \frac{f(A^{-1}(y))}{|\det(A)|}.$$

Satz 3.1 lässt sich unschwer auf die folgende, etwas allgemeinere Form erweitern:

Satz 3.4 (Transformationssatz für W-Dichten, Allgemeiner Fall).

Sei $\mathcal{X} = (X_1, \dots, X_d)$ absolut stetig mit Dichte f , U_1, \dots, U_n seien paarweise disjunkte offene Mengen mit $\mathbb{P}^{\mathcal{X}}(\bigcup_{i=1}^n U_i) = 1$. Weiters sei $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ eine Transformation mit der Eigenschaft, dass $T_i := T|_{U_i}$ injektiv und stetig differenzierbar ist und zusätzlich $J_{T_i}(x) \neq 0$ für jedes $i = 1 \in \{1, \dots, n\}$ erfüllt. Dann gilt:

1. $T_i(U_i)$ ist offen für jedes $i \in \{1, \dots, n\}$.
2. $\mathcal{Y} = T \circ \mathcal{X}$ ist absolut stetig mit Dichte

$$g(y) = \sum_{i=1}^n \frac{f \circ T^{-1}(y)}{|J_{T_i}(T_i^{-1}(y))|} \mathbb{1}_{T_i(U_i)}(y)$$

Definition 3.5. Es sei $\mu = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ symmetrisch und positiv definit. $\mathcal{X} = (X_1, \dots, X_d)$ heißt (multivariat) normalverteilt mit Erwartungswert μ und Kovarianzmatrix Σ [wir schreiben $X \sim \mathcal{N}_d(\mu, \Sigma)$] genau dann wenn \mathcal{X} absolut stetig mit Dichte

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-\frac{1}{2} \langle (x-\mu), \Sigma^{-1}(x-\mu) \rangle}$$

Der Nachweis, dass f tatsächlich eine Wahrscheinlichkeitsdichte definiert, erfolgt in den Übungen.

Im folgenden werden wir häufig mit der sog. t -Verteilung zu tun haben, die sich auf natürliche Art und Weise wie folgt aus der Normalverteilung ergibt: Seien X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. Wir möchten ein Konfidenzintervall für den Parameter μ berechnen.

1. Einfacher (und unrealistischer) Fall: $\sigma > 0$ ist bekannt. Aus der VO Mathematische Statistik wissen wir, dass $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ gilt und wir erhalten

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Daraus folgt sofort

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

und damit

$$\mathbb{P}\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

$C_{1-\alpha}^n$, definiert durch

$$C_{1-\alpha}^n = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right],$$

ist daher ein exaktes Konfidenzintervall für μ mit Überdeckungswahrscheinlichkeit $1 - \alpha$.

2. Allgemeiner Fall: $\sigma > 0$ ist unbekannt. Wenn wir analog zum vorigen Fall vorgehen und den Parameter σ^2 wie gewohnt durch $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ schätzen landen wir bei

$$T_n := \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}}$$

Die Verteilung von T_n sowie die gemeinsame Verteilung von (\bar{X}_n, S_n^2) leiten wir in mehreren einfach Schritten her:

Lemma 3.6. Sei $X \sim \mathcal{N}(0, 1)$, $Y \sim \chi_n^2$ und X, Y unabhängig. Dann ist die Zufallsvariable

$$\frac{X}{\sqrt{\frac{Y}{n}}}$$

absolut stetig mit Dichte

$$g_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \frac{1}{\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}. \quad (3.1)$$

Definition 3.7. Eine Zufallsvariable mit Dichte (3.1) heißt *t-verteilt* mit n Freiheitsgraden.

Lemma 3.8. Seien X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2.$$

Beweis: Übungsaufgabe.

Bemerkung 3.9. Wenn zwei Zufallsvariable X, Y unabhängig sind schreiben wir im Folgenden auch oft $X \perp Y$.

Satz 3.10. Seien X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und $n \geq 2$. Dann gelten folgende Aussagen:

1. $\bar{X}_n \perp S_n^2$
2. $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
3. $\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$

Beweis: Die zweite Aussage ist schon aus der VO Mathematische Statistik bekannt.

Zum Beweis der ersten Aussage nehmen wir o.B.d.A. $\mu = 0$ an. Dann gilt offensichtlich und $\mathcal{X} = (X_1, \dots, X_n) \sim \mathcal{N}_n(0, \sigma^2 E_n)$. Die sogenannte Helmert Matrix $A \in \mathbb{R}^{n \times n}$, definiert durch

$$A := \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2 \cdot 1}} & -\sqrt{\frac{1}{2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{3 \cdot 2}} & \frac{1}{\sqrt{3 \cdot 2}} & -\sqrt{\frac{2}{3}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{4 \cdot 3}} & \frac{1}{\sqrt{4 \cdot 3}} & \frac{1}{\sqrt{4 \cdot 3}} & -\sqrt{\frac{3}{4}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} & -\sqrt{\frac{n-1}{n}} \end{pmatrix}$$

ist offensichtlich orthogonal, es gilt also insbesondere $A^T A = E_n = A A^T$, wobei E_n die n -dimensionale Einheitsmatrix bezeichnet. Für den Zufallsvektor $\mathcal{Y} := A \cdot \mathcal{X}$ folgt (direktes Nachrechnen oder Transformationssatz für Dichten anwenden)

$$\mathcal{Y} \sim \mathcal{N}_n(0, A(\sigma^2 E_n)A^T) = \mathcal{N}_n(0, \sigma^2 E_n) \sim \mathcal{X}.$$

Für $\mathcal{Y} = (Y_1, \dots, Y_n)$ gilt daher Y_1, \dots, Y_n i.i.d. mit $Y_1 \sim \mathcal{N}(0, \sigma^2)$; für Y_1 erhalten wir ausserdem

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sqrt{n} \bar{X}_n.$$

Als orthogonale Matrix ist A eine Isometrie, wir erhalten daher $\|Y\|_2^2 = \|AX\|_2^2 = \|X\|_2^2$, i.e. $\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2$, sowie

$$\begin{aligned} \sum_{i=2}^n Y_i^2 &= \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2 = n \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right) \\ &= n \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = (n-1)S_n^2. \end{aligned}$$

Daraus folgt sofort die behauptete Unabhängigkeit von $\sqrt{n} \bar{X}_n \perp (n-1)S_n^2$ und $\bar{X}_n \perp S_n^2$. Die dritte Behauptung ist eine direkte Konsequenz aus der Tatsache dass Y_1, \dots, Y_n i.i.d. mit $Y_1 \sim \mathcal{N}(0, \sigma^2)$ gilt:

$$\frac{(n-1)S_n^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=2}^n Y_i^2 \sim \chi_{n-1}^2.$$

Folgerung 3.11. Seien X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t_{n-1}$$

Beweis: Der Ausdruck auf der linken Seite lässt sich umschreiben zu

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\frac{S_n}{\sigma}} = \frac{\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}}{\sqrt{\frac{S_n^2}{\sigma^2}}}$$

Der Zähler ist offensichtlich standardnormalverteilt und nach Satz 3.10 wissen wir, dass $\frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2$ gilt, Lemma 3.6 liefert daher sofort die gewünschte Aussagen. ■

Bemerkung 3.12. Im Allgemeinen sind das Stichprobenmittel und die Stichprobenvarianz natürlich nicht unabhängig.

Frage 3.13. Wie können Satz 3.10 und Folgerung 3.11 mittels Simulationen illustriert werden? Wie kann Unabhängigkeit von \bar{X}_n zu S_n^2 verifiziert werden?

Sei nun X_1, \dots, X_n i.i.d mit $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ und Y_1, \dots, Y_m i.i.d mit $Y_1 \sim \mathcal{N}(\mu_2, \sigma^2)$, wobei $n, m \geq 2$ und σ^2 unbekannt, aber nach Voraussetzung für beide Verteilungen gleich ist. Im Abschnitt über Konfidenzbereiche werden wir insbesondere ein Konfidenzintervall für $\mu_1 - \mu_2$ herleiten - der folgende Satz liefert dafür die Grundlage:

Satz 3.14. Sei X_1, \dots, X_n i.i.d mit $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ und Y_1, \dots, Y_m i.i.d mit $Y_1 \sim \mathcal{N}(\mu_2, \sigma^2)$, $(X_1, \dots, X_n) \perp (Y_1, \dots, Y_m)$ und $n, m \geq 2$. Dann gilt:

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sqrt{S_p^2} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2},$$

wobei $S_p^2 = \frac{(n-1)S_n^2 + (m-1)S_m^2}{n+m-2}$ die sog. gepoolte Stichprobenvarianz bezeichnet.

Beweis: Unter Verwendung des Additionstheorems der Gamma-Verteilung gilt

$$\begin{aligned} S_p^2 &= \frac{1}{n+m-2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{j=1}^m (Y_j - \bar{Y}_m)^2 \right) \\ &= \frac{\sigma^2}{n+m-2} \underbrace{\left(\sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma^2} + \sum_{j=1}^m \frac{(Y_j - \bar{Y}_m)^2}{\sigma^2} \right)}_{\sim \chi_{n+m-2}^2} \end{aligned}$$

In Anbetracht von

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \sqrt{\frac{1}{n} + \frac{1}{m}}}} = \frac{\frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sigma}}{\sqrt{S_p^2 \frac{n+m-2}{\sigma^2} \frac{1}{n+m-2} \sqrt{\frac{1}{n} + \frac{1}{m}}}},$$

reicht es lt. Lemma 3.6 zu zeigen, dass

$$Z := \frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sigma} \sqrt{\frac{nm}{n+m}} \sim \mathcal{N}(0, 1).$$

gilt. Dass Z normalverteilt ist, ist klar, dass $\mathbb{E}(Z) = 0$ gilt ebenso. Für die Varianz von Z erhalten wir

$$\mathbb{V}(Z) = \frac{nm}{(n+m)\sigma^2} \mathbb{V}(\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)) = \frac{nm}{(n+m)\sigma^2} \left(\frac{\sigma^2}{n} + \frac{\sigma^2}{m} \right) = 1. \blacksquare$$

Kapitel 4

Konfidenzbereiche

4.1 Konfidenzbereiche für endlichdimensionale Parameter

Gelte $X \sim (P_\theta)_{\theta \in \Theta \subseteq \mathbb{R}^d}$, weiters sei (X_1, \dots, X_n) eine Zufallsstichprobe von X . Gesucht wird eine Menge $C_n \subseteq \mathbb{R}^d$ mit den folgenden drei Eigenschaften:

- C_n lässt sich aus (X_1, \dots, X_n) berechnen
- C_n enthält den Parameter θ mit mindestens vorgegebener Wahrscheinlichkeit $1 - \alpha$
- C_n ist möglichst klein (genauer: $\lambda_d(C_n)$ ist möglichst klein)

Wir beginnen mit der folgenden grundlegenden Definition:

Definition 4.1. Sei $X \sim (P_\theta)_{\theta \in \Theta \subseteq \mathbb{R}^d}$, X_1, \dots, X_n sei eine Stichprobe von X , und es gelte $\alpha \in (0, 1)$ (typischerweise $\alpha = 0.05$ oder $\alpha = 0.01$). Dann heißt

$$C_n : Rg(X)^n \longrightarrow \mathcal{P}(\mathbb{R}^d)$$

ein Konfidenzbereich für θ mit Überdeckungswahrscheinlichkeit $1 - \alpha$, falls

$$P_\theta(C_n(X_1, \dots, X_n) \ni \theta) \geq 1 - \alpha \tag{4.1}$$

für alle $\theta \in \Theta$ gilt.

C_n heißt exakt genau dann, wenn sogar $P_\theta(\theta \in C_n(X_1, \dots, X_n)) = 1 - \alpha$ für alle $\theta \in \Theta$ gilt.

Bemerkung 4.2. Für $d = 1$ ist C_n oft von der Form eines Intervalls, wir sprechen daher von einem Konfidenzintervall (CI).

Bemerkung 4.3. Beachten Sie, dass in der Definition des Konfidenzbereichs implizit angenommen wird, dass die Wahrscheinlichkeit des entsprechenden Ereignisses berechnet werden kann, i.e., dass für jedes $\theta \in \Theta$

$$\{\omega \in \Omega : C_n(X_1(\omega), \dots, X_n(\omega)) \ni \theta\} \in \mathcal{A}$$

gilt. C_n kann als zufällige Menge interpretiert werden (muss aber kein 'random set' im Sinne der Literatur über random compact sets sein, siehe [4])

Gilt die Überdeckungseigenschaft (4.1) erst asymptotisch so sprechen wir von einem asymptotischen Konfidenzbereich.

Definition 4.4. Sei $X \sim (P_\theta)_{\theta \in \Theta \subset \mathbb{R}^d}$, X, X_1, X_2, \dots sei i.i.d., und $\alpha \in [0, 1]$. Dann heisst $C_n : \text{Rg}(X)^n \rightarrow \mathcal{P}(\mathbb{R}^d)$ (bzw. genauer: die Folge $(C_n)_{n \in \mathbb{N}}$) asymptotischer Konfidenzbereich mit Überdeckungswahrscheinlichkeit $1 - \alpha$ genau dann, wenn für jedes $\theta \in \Theta$ die folgende Aussage gilt:

$$\lim_{n \rightarrow \infty} P_\theta(\theta \in C_n(X_1, \dots, X_n)) \geq 1 - \alpha. \quad (4.2)$$

Falls in (4.2) sogar Gleichheit gilt, dann nennen wir C_n einen asymptotisch exakten Konfidenzbereich.

Asymptotische Konfidenzbereiche werden wir hauptsächlich mit Hilfe des CLT konstruieren, ein erstes konkretes Beispiel haben wir schon in Beispiel 2.5 kennengelernt.

Wir kommen nun zurück zur im letzten Kapitel skizzierten Problemstellung, für $X \sim \mathcal{N}(\mu, \sigma^2)$ und unbekanntem μ, σ^2 ein Konfidenzintervall für μ zu bestimmen.

Satz 4.5. Sei X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und $n \geq 2$. Weiters sei $\alpha \in (0, 1)$. Dann ist

$$C_n(X_1, \dots, X_n) = \left[\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}} \right] \quad (4.3)$$

ein exaktes Konfidenzintervall für den Parameter μ mit Überdeckungswahrscheinlichkeit $1 - \alpha$.

Beweis: Lt. Folgerung 3.11 gilt

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t_{n-1},$$

und damit offensichtlich

$$\mathbb{P}_\mu \left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \in \left[t_{n-1, \frac{\alpha}{2}}, t_{n-1, 1-\frac{\alpha}{2}} \right] \right) = 1 - \alpha,$$

woraus sich durch elementare Umformungen sofort das in (4.3) angegebene Intervall ergibt. ■

Satz 4.6. Sei X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und $n \geq 2$. Weiters sei $\alpha \in (0, 1)$. Dann ist

$$C_n(X_1, \dots, X_n) = \left[\frac{(n-1)S_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right] \quad (4.4)$$

ein exaktes Konfidenzintervall für σ^2 mit Überdeckungswahrscheinlichkeit $1 - \alpha$.

Beweis: Lt. Satz 3.10 gilt unter den gemachten Voraussetzungen $\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$, woraus sich das Intervall in (4.4) sofort ergibt.

Satz 3.14 liefert das folgende Resultat:

Satz 4.7. Sei X_1, \dots, X_n i.i.d mit $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ und Y_1, \dots, Y_m i.i.d mit $Y_1 \sim \mathcal{N}(\mu_2, \sigma^2)$, wobei $\sigma^2 > 0$ unbekannt aber gleich in beiden Fällen ist. Zusätzlich gelte $(X_1, \dots, X_n) \perp (Y_1, \dots, Y_m)$ und $n, m \geq 2$. Dann ist

$$\left[\bar{X}_n - \bar{Y}_m - t_{n+m-2, 1-\frac{\alpha}{2}} \sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{S_p^2}, \bar{X}_n - \bar{Y}_m + t_{n+m-2, 1-\frac{\alpha}{2}} \sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{S_p^2} \right]$$

ein exaktes Konfidenzintervall mit Überdeckungswahrscheinlichkeit $1 - \alpha$ für $\Delta := \mu_1 - \mu_2$.

Beispiel 4.8. Wir konstruieren einen exakten Konfidenzbereich für den Parameter $\theta = (\mu, \sigma^2)$ der Normalverteilung. Da \bar{X}_n und S_n^2 lt. Satz (3.10) unabhängig sind folgt sofort

$$\mathbb{P}(\bar{X}_n \in B_1, S_n^2 \in B_2) = \mathbb{P}(\bar{X}_n \in B_1) \cdot \mathbb{P}(S_n^2 \in B_2)$$

Unter Verwendung von

$$\begin{aligned} \mathbb{P}_\theta \left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}} \right) &= 1 - \alpha \\ \mathbb{P}_\theta \left(\chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{n-1}{\sigma^2} S_n^2 \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right) &= 1 - \alpha \end{aligned}$$

erhalten wir

$$\begin{aligned} (1 - \alpha)^2 &= \mathbb{P}_\theta \left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}} \right) \cdot \mathbb{P}_\theta \left(\chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{n-1}{\sigma^2} S_n^2 \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right) \\ &= \mathbb{P}_\theta \left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}, \quad \chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{n-1}{\sigma^2} S_n^2 \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right) \\ &= \mathbb{P}_\theta \left((\bar{X}_n - \mu)^2 \leq \frac{\sigma^2}{n} z_{1-\frac{\alpha}{2}}^2, \quad \frac{(n-1)S_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right) \end{aligned}$$

Setzen wir

$$(1 - \beta)^2 = 1 - \alpha \Rightarrow \beta = 1 - \sqrt{1 - \alpha}$$

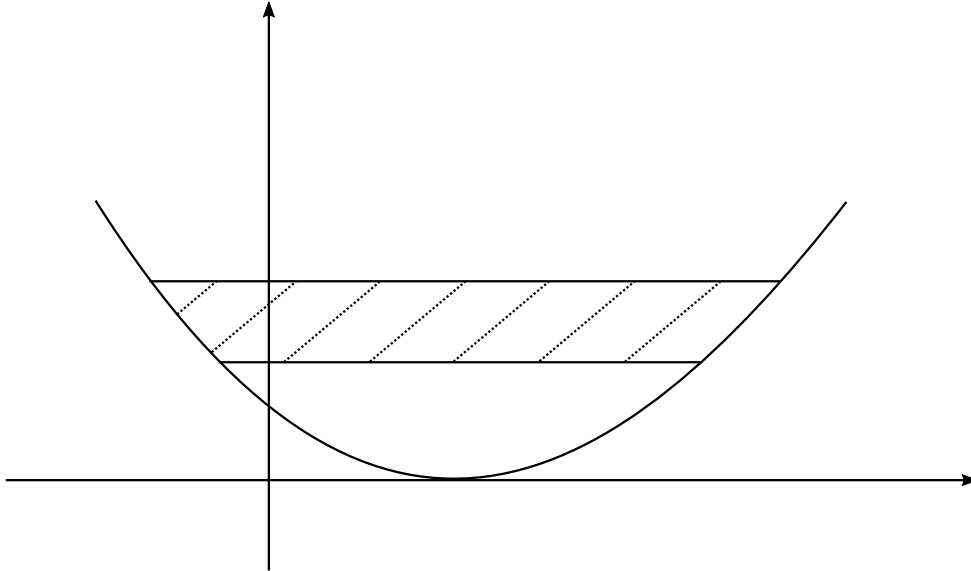
dann erhalten wir insgesamt folgendes Resultat:

Satz 4.9. Sei X_1, \dots, X_n i.i.d mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und $n \geq 2$. Dann ist

$$C_n(X_1, \dots, X_n) = \left\{ (\mu_0, \sigma_0^2) \in \mathbb{R} \times (0, \infty) : (\mu_0 - \bar{X}_n)^2 \leq \frac{\sigma_0^2}{n} z_{\frac{1}{2} + \frac{\sqrt{1-\alpha}}{2}}, \right. \\ \left. \frac{(n-1)S_n^2}{\chi_{n-1, \frac{1}{2} + \frac{\sqrt{1-\alpha}}{2}}^2} \leq \sigma_0^2 \leq \frac{(n-1)S_n^2}{\chi_{n-1, \frac{1}{2} - \frac{\sqrt{1-\alpha}}{2}}^2} \right\}$$

ein exakter Konfidenzbereich mit Überdeckungswahrscheinlichkeit $1 - \alpha$ für $\theta = (\mu, \sigma^2)$.

Allen bisherigen Konfidenzbereichen lag die Idee zugrunde, aus der Stichprobe Zufallsvariable zu konstruieren, deren Verteilung nur mehr von der Sample Size n , nicht aber vom Parameter θ abhängt, und dann Quantile der Verteilung zu betrachten.

Abbildung 4.1: Konfidenzbereich für $\theta = (\mu, \sigma^2)$ lt. Satz 4.9.

Definition 4.10. Sei $X \sim (P_\theta)_{\theta \in \Theta}$, (X_1, \dots, X_n) sei eine Zufallsstichprobe von X . Dann heißt eine messbare Funktion $Q : \Theta \times (\text{Rg}(X))^n \rightarrow \mathbb{R}^d$ Pivotgröße, oder kurz Pivot genau dann, wenn die Verteilung von $Q(\theta; X_1, \dots, X_n)$ nicht vom Parameter θ abhängt.

Bemerkung 4.11. In den bisher abgeleiteten Konfidenzbereichen für $X \sim (\mu, \sigma^2)$ haben wir die folgenden Pivots verwendet:

- i) $Q((\mu, \sigma^2); X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.
- ii) $Q((\mu, \sigma^2); X_1, \dots, X_n) = \frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$
- iii) $Q((\mu, \sigma^2); X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t_{n-1}$

Zahlreiche weitere Beispiele für Pivots und daraus konstruierte Konfidenzbereiche finden Sie auf den Übungsblättern. Einer der wichtigsten Pivots kommt von der PIT - wir illustrieren den PIT-Zugang anhand der stetigen Gleichverteilung auf $[0, \theta]$ mit $\theta > 0$ - beachten Sie, dass die Methode übertragbar ist auf Fälle in denen F_θ stetig (daher lt. PIT $F_\theta \circ X \sim \mathcal{U}(0, 1)$) und der Parameter θ eindimensional ist.

Beispiel 4.12. Sei (X_1, \dots, X_n) eine Zufallsstichprobe von $X \sim \mathcal{U}(0, \theta)$. Dann ist $\frac{X}{\theta} \sim \mathcal{U}(0, 1)$ (Trivialfall der PIT). Nachdem für $Z \sim \mathcal{U}(0, 1)$ und $y \in [0, \infty)$ offensichtlich

$$\mathbb{P}(-\ln Z \leq y) = \mathbb{P}(\ln Z \geq -y) = \mathbb{P}(Z \geq e^{-y}) = 1 - e^{-y}$$

gilt, folgt sofort $-2 \ln(Z) \sim \mathcal{E}(\frac{1}{2}) = \chi_2^2$. Anwendung der Additionstheoreme für die Gamma-Verteilung liefert den Pivot Y_n , definiert durch

$$Y_n = -2 \sum_{i=1}^n \ln \left(\frac{X_i}{\theta} \right) \sim \chi_{2n}^2$$

Unter Verwendung von

$$\begin{aligned} \chi_{2n, \frac{\alpha}{2}}^2 &\leq -2 \left(\sum_{i=1}^n \ln(X_i) - n \ln(\theta) \right) \leq \chi_{2n, 1-\frac{\alpha}{2}}^2 \\ \iff \frac{1}{2n} \chi_{2n, \frac{\alpha}{2}}^2 &\leq \ln(\theta) - \frac{1}{n} \sum_{i=1}^n \ln(X_i) \leq \frac{1}{2n} \chi_{2n, 1-\frac{\alpha}{2}}^2 \\ \iff e^{\frac{1}{2n} \chi_{2n, \frac{\alpha}{2}}^2} &\leq \frac{\theta}{\sqrt[n]{\prod_{i=1}^n X_i}} \leq e^{\frac{1}{2n} \chi_{2n, 1-\frac{\alpha}{2}}^2} \end{aligned}$$

erhalten wir schließlich das gewünschte Konfidenzintervall $C_n(X_1, \dots, X_n)$ für θ zum Niveau $1 - \alpha$:

$$C_n(X_1, \dots, X_n) = \sqrt[n]{\prod_{i=1}^n X_i} \left[e^{\frac{1}{2n} \chi_{2n, \frac{\alpha}{2}}^2}, e^{\frac{1}{2n} \chi_{2n, 1-\frac{\alpha}{2}}^2} \right]$$

Den bisherigen Konfidenzbereichen lagen sehr restriktive Verteilungsannahmen (größtenteils Normalverteilungsannahme) zugrunde. Unter Verwendung des CLT lassen sich asymptotische Konfidenzbereiche auch ohne solche Annahmen einfach konstruieren. Bevor wir die Resultate herleiten ergänzen wir einen dafür wichtigen Satz über schwache Konvergenz.

Satz 4.13 (Satz von Slutsky). *Seien $(X_n)_{n \in \mathbb{N}}$ und $(Y_n)_{n \in \mathbb{N}}$ Folgen von Zufallsvariablen, X sei eine weitere Zufallsvariable, $c \in \mathbb{R}$ sei eine feste Konstante. Falls $X_n \xrightarrow{w} X$ und $Y_n \xrightarrow{w} c$, dann gelten auch die folgenden drei Eigenschaften:*

1. $X_n + Y_n \xrightarrow{w} X + c$
2. $X_n Y_n \xrightarrow{w} c X$
3. Falls $c \neq 0$ dann folgt $\frac{X_n}{Y_n} \xrightarrow{w} \frac{X}{c}$.

Beweis: Übungsaufgabe

Sei nun X_1, \dots, X_n i.i.d. und $\mathbb{V}(X_1) \in (0, \infty)$. Dann gilt nach dem CLT

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{w} Z \sim \mathcal{N}(0, 1) \quad \text{für } n \rightarrow \infty$$

Nachdem wir einen asymptotischen Konfidenzbereich für μ herleiten möchten, ersetzen wir σ durch die Stichprobenvarianz S_n^2 , verwenden $S_n^2 \xrightarrow{[P]} \sigma^2$ (SLLN), wenden den Satz von Slutsky an und erhalten

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} = \sqrt{n} \frac{\frac{\bar{X}_n - \mu}{\sigma}}{\frac{\sqrt{S_n^2}}{\sigma}} \xrightarrow{w} \mathcal{N}(0, 1).$$

Damit folgt sofort das folgende Resultat:

Satz 4.14. Sei X, X_1, X_2, \dots i.i.d. mit $\mathbb{V}(X) \in (0, \infty)$. Dann ist

$$C_n(X_1, \dots, X_n) = \left[\bar{X}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}} \right] \quad (4.5)$$

ein asymptotisch exaktes Konfidenzintervall mit Überdeckungswahrscheinlichkeit $1 - \alpha$ für den Parameter $\mu = \mathbb{E}(X)$.

Für den Spezialfall der Alternativverteilung lässt sich das Intervall (4.5) ein klein wenig verbessern (i.e., verkleinern):

Satz 4.15. Sei X, X_1, X_2, \dots i.i.d mit $X \sim A(p)$ und $p \in (0, 1)$. Dann ist

$$C_n(X_1, \dots, X_n) = \left[\bar{X}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right]$$

ein asymptotisch exaktes Konfidenzintervall mit Überdeckungswahrscheinlichkeit $1 - \alpha$ für den Parameter p .

Beweis: Wegen

$$\bar{X}_n(1 - \bar{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 = \frac{n-1}{n} S_n^2 < S_n^2$$

ist das Intervall etwas kleiner als jenes in (4.5). Aus dem SLLN folgt

$$\bar{X}_n \xrightarrow{[P_p]} \mathbb{E}(X_1) = p,$$

das Continuous Mapping Theorem liefert

$$\bar{X}_n(1 - \bar{X}_n) \xrightarrow{[P_p]} p(1 - p) = \mathbb{V}(X_1),$$

und der Satz von Slutsky impliziert

$$\frac{\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}}{\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sigma}} \xrightarrow{w} \mathcal{N}(0, 1).$$

Das gewünschte Resultat ergibt sich daraus durch elementare Umformungen. ■

Das folgende Resultat, dessen Beweis eine einfache Übungsaufgabe ist, lässt sich analog herleiten:

Satz 4.16. Seien $X_1, X_2, \dots, Y_1, Y_2, \dots$ unabhängig, X, X_1, X_2, \dots i.i.d. mit Varianz $\sigma_1^2 := \mathbb{V}(X) \in (0, \infty)$ und Y, Y_1, Y_2, \dots i.i.d. mit $\sigma_2^2 := \mathbb{V}(Y) \in (0, \infty)$. Weiters gelte $n = n_1 + n_2$ und $\lim_{n \rightarrow \infty} \frac{n_1}{n} = \kappa \in (0, 1)$ (d.h. beide Stichproben wachsen mit vergleichbarer Geschwindigkeit). Dann ist

$$C_n(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}) = \left[\bar{X}_{n_1} - \bar{Y}_{n_2} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X}_{n_1} - \bar{Y}_{n_2} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

ein asymptotisch exaktes Konfidenzintervall für den Parameter $\theta = \mathbb{E}(X_1) - \mathbb{E}(Y_1)$ mit Überdeckungswahrscheinlichkeit $1 - \alpha$.

4.2 Konfidenzbänder

Wir haben im letzten Abschnitt insbesondere gesehen, wie sich asymptotische Konfidenzbereiche mittels CLT konstruieren lassen. Wenn wir Verteilungsfunktionen als unendlichdimensionalen Parameter interpretieren, dann können wir versuchen, analog vorzugehen, und zumindest asymptotische Konfidenzbereiche für diesen Parameter herzuleiten. Wir definieren zuerst den Begriff des Konfidenzbands, überlegen uns dann, warum es wenig Sinn macht, Konfidenzintervalle für $F(x)$ für jedes feste $x \in \mathbb{R}$ via Satz 4.15 zu betrachten und leiten dann unter Verwendung von Satz 2.1 ein asymptotisches Konfidenzband für den Fall stetiger Verteilungsfunktionen her. Dieses Band vergleichen wir dann mit einem auf Satz 2.3 beruhenden (fast) exakten Konfidenzband.

In der folgenden Definition bezeichnet \mathcal{F} wie in der VO Mathematische Statistik die Familie aller eindimensionalen Verteilungsfunktionen, $2^{\mathcal{F}}$ die Potenzmenge von \mathcal{F} .

Definition 4.17. Sei X, X_1, X_2, \dots i.i.d. mit $X \sim F$, weiters gelte $\alpha \in (0, 1)$. Dann heisst eine Abbildung $\mathcal{C}_n : \mathcal{R}g(X)^n \rightarrow 2^{\mathcal{F}}$ Konfidenzband für F mit Überdeckungswahrscheinlichkeit $1 - \alpha$, genau dann wenn gilt

$$P(F \in \mathcal{C}_n(X_1, \dots, X_n)) \geq 1 - \alpha. \quad (4.6)$$

\mathcal{C}_n heisst exakt, genau dann wenn statt $\geq 1 - \alpha$ sogar $= 1 - \alpha$ gilt.

\mathcal{C}_n heisst asymptotisches Konfidenzband für F mit Überdeckungswahrscheinlichkeit $1 - \alpha$ falls

$$\lim_{n \rightarrow \infty} P(F \in \mathcal{C}_n(X_1, \dots, X_n)) \geq 1 - \alpha. \quad (4.7)$$

und asymptotisch exaktes Konfidenzband für F mit Überdeckungswahrscheinlichkeit $1 - \alpha$ falls

$$\lim_{n \rightarrow \infty} P(F \in \mathcal{C}_n(X_1, \dots, X_n)) = 1 - \alpha. \quad (4.8)$$

Bemerkung 4.18. Beachten Sie, dass in der Definition des Konfidenzbands implizit angenommen wird, dass die Wahrscheinlichkeit des entsprechenden Ereignisses berechnet werden kann, i.e., dass

$$\{\omega \in \Omega : F \in \mathcal{C}_n(X_1(\omega), \dots, X_n(\omega))\} \in \mathcal{A}$$

gilt. Selbiges gilt für die Existenz der Limiten in (4.7) und (4.8).

Seien nun X_1, \dots, X_n i.i.d. mit Verteilungsfunktion F und $x, y \in \mathbb{R}$ mit $x \neq y$. Aus der VO Mathematische Statistik wissen wir, dass die empirische Verteilungsfunktion

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_i)$$

an der Stelle x ein stark konsistenter, effizienter Schätzer für $F(x)$ ist. Um ein asymptotisches Konfidenzintervall für $F(x)$ zu konstruieren setzen wir $Y_i := \mathbb{1}_{(-\infty, x]}(X_i)$ für jedes $i \in \{1, \dots, n\}$ und erhalten Y_1, \dots, Y_n i.i.d. mit $Y_1 \sim \text{Bin}(1, F(x))$. Anwendung von Satz 4.15 liefert daher sofort das asymptotisch exakte Konfidenzintervall $C_n^x(X_1, \dots, X_n)$ für $F(x)$ mit Überdeckungswahrscheinlichkeit $1 - \alpha$ gegeben durch

$$C_n^x(X_1, \dots, X_n) = \left[\hat{F}_n(x) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{F}_n(x)(1 - \hat{F}_n(x))}{n}}, \hat{F}_n(x) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{F}_n(x)(1 - \hat{F}_n(x))}{n}} \right].$$

Um ein asymptotisches Konfidenzintervall $C_n^y(X_1, \dots, X_n)$ für $F(y)$ herzuleiten können wir vollkommen analog vorgehen. Wenn wir $C_n^x(X_1, \dots, X_n) \times C_n^y(X_1, \dots, X_n)$ als (zweidimensionalen) Konfidenzbereich für $(F(x), F(y))$ interpretieren wird das eigentliche Problem dieses 'punktweisen' Zugangs deutlich: $\hat{F}_n(x), \hat{F}_n(y)$ sind nicht unabhängig, es gilt daher im Allgemeinen (wir schreiben kurzfristig $C_n = C_n^x(X_1, \dots, X_n)$)

$$P(F(x) \in C_n^x, F(y) \in C_n^y) \neq P(F(x) \in C_n^x) \cdot P(F(y) \in C_n^y).$$

Wenn wir daher konservativ vorgehen und das Worst-Case-Szenario betrachten, dann liefert die (scharfe!) Bonferroni Ungleichung sofort

$$P(F(x) \in C_n^x, F(y) \in C_n^y) \geq P(F(x) \in C_n^x) + P(F(y) \in C_n^y) - 1.$$

und damit

$$\liminf_{n \rightarrow \infty} P(F(x) \in C_n^x, F(y) \in C_n^y) \geq 1 - \alpha + 1 - \alpha - 1 = 1 - 2\alpha.$$

Betrachten wir allgemeiner k Punkte $z_1 < z_2 < \dots < z_k$ dann erhalten wir analog für das Worst-Case-Szenario

$$\liminf_{n \rightarrow \infty} P\left(F(z_i) \in C_n^{z_i}(X_1, \dots, X_n) \text{ für alle } i \in \{1, \dots, k\}\right) \geq 1 - k\alpha,$$

die rechte Seite - die Überdeckungswahrscheinlichkeit wird im Worst-Case-Szenario also beliebig klein, wenn k hinreichend groß wird. Für abzählbar viele Punkte z_1, z_2, \dots kann daher keine echt positive Überdeckungswahrscheinlichkeit sichergestellt werden, der punktweise Zugang liefert kein Konfidenzband.

Satz 2.1 bietet einen eleganten Ausweg zumindest für stetiges F - wir gehen wie folgt vor: Sei X_1, X_2, \dots , i.i.d. mit stetiger Verteilungsfunktion F . F_{KS} bezeichne wie gehabt die Kolmogorov-Smirnov Verteilungsfunktion, α liege in $(0, 1)$, und $x_{1-\alpha}$ sei definiert durch $x_{1-\alpha} := F_{KS}^-(1 - \alpha)$. Dann gilt

$$\lim_{n \rightarrow \infty} P(\sqrt{n} D_n \leq x_{1-\alpha}) = F_{KS}(x_{1-\alpha}) = F_{KS}(F_{KS}^-(1 - \alpha)) = 1 - \alpha.$$

Mit anderen Worten

$$\lim_{n \rightarrow \infty} P\left(\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \leq \frac{x_{1-\alpha}}{\sqrt{n}}\right) = 1 - \alpha,$$

also

$$\lim_{n \rightarrow \infty} P\left(\hat{F}_n(x) - x_{1-\alpha} \frac{1}{\sqrt{n}} \leq F(x) \leq \hat{F}_n(x) + x_{1-\alpha} \frac{1}{\sqrt{n}} \quad \forall x \in \mathbb{R}\right) = 1 - \alpha.$$

Wir haben damit folgenden Satz bewiesen:

Satz 4.19. *Sei $(X_n)_{n \in \mathbb{N}}$ i.i.d. mit $X_1 \sim F$ und F stetig. Die Funktionen $L_n, U_n : \mathbb{R} \rightarrow [0, 1]$ seien definiert durch*

$$L_n(x) := \max\left\{0, \hat{F}_n(x) - \frac{1}{\sqrt{n}} F_{KS}^-(1 - \alpha)\right\}, \quad U_n(x) = \min\left\{\hat{F}_n(x) + \frac{1}{\sqrt{n}} F_{KS}^-(1 - \alpha), 1\right\}. \quad (4.9)$$

und C_n durch

$$C_n = \{G \in \mathcal{F} : L_n(x) \leq G(x) \leq U_n(x) \quad \forall x \in \mathbb{R}\}.$$

Dann ist C_n ein asymptotisch exaktes Konfidenzband für F .

Wie am Beginn dieses Abschnitts erwähnt ist es für stetige Verteilungsfunktionen sogar möglich, mittels Satz 2.3 ein exaktes (also nicht nur ein asymptotisch exaktes) Konfidenzband zu konstruieren. F_{d_n} bezeichne die (stetige, strikt wachsende) Verteilungsfunktion von d_n gemäß Gleichung 2.1.

Satz 4.20. *Sei F eine stetige Verteilungsfunktion, $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge von Zufallsvariablen mit Verteilungsfunktion F , und d_n definiert gemäß Gleichung 2.1. Die Funktionen $L_n, U_n : \mathbb{R} \rightarrow [0, 1]$ seien definiert durch*

$$L_n(x) := \max \left\{ 0, \hat{F}_n(x) - F_{d_n}^-(1 - \alpha) \right\}, \quad U_n(x) = \min \left\{ \hat{F}_n(x) + F_{d_n}^-(1 - \alpha), 1 \right\}. \quad (4.10)$$

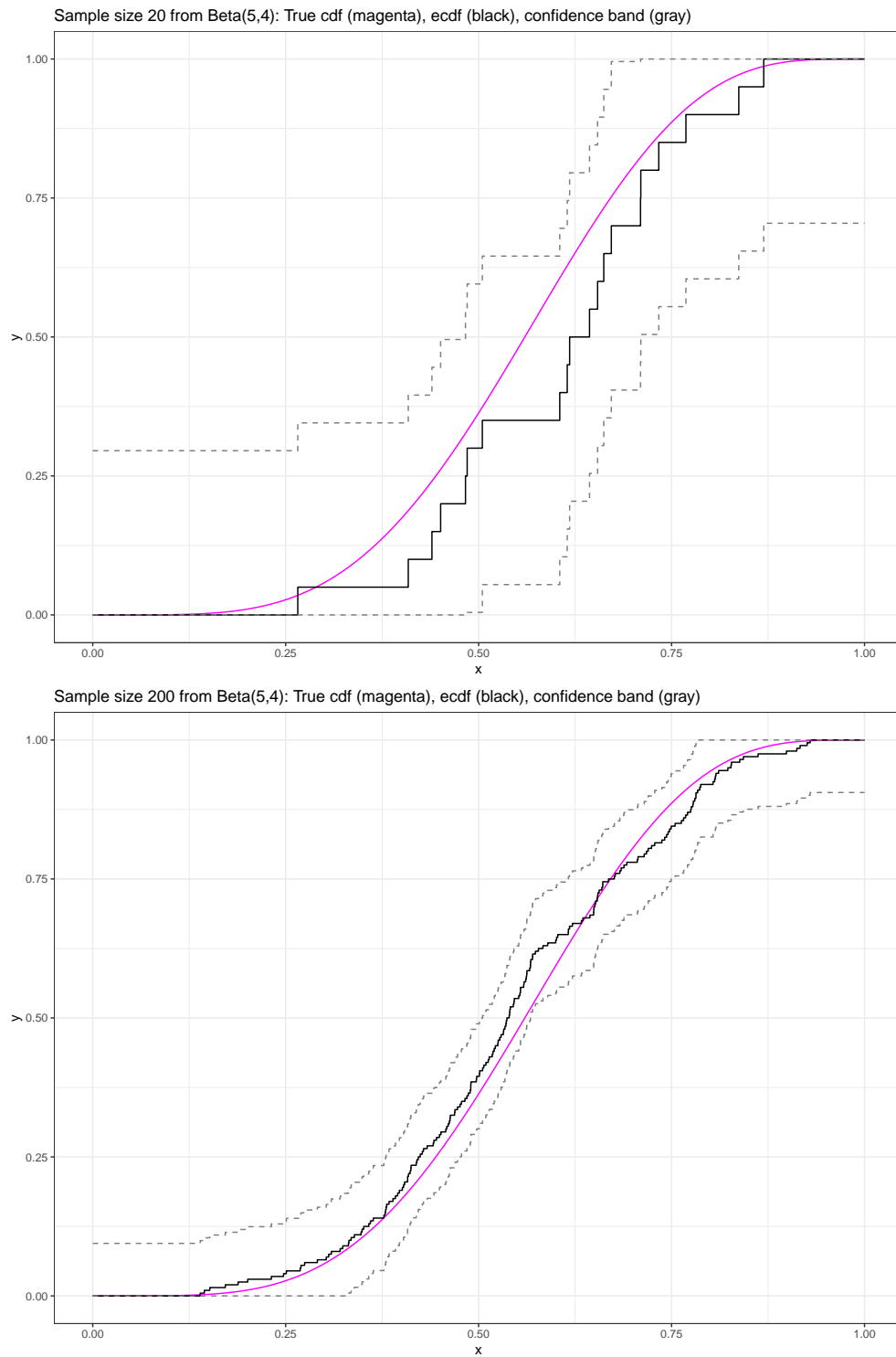
und \mathcal{C}_n durch

$$\mathcal{C}_n = \{G \in \mathcal{F} : L_n(x) \leq G(x) \leq U_n(x) \quad \forall x \in \mathbb{R}\}.$$

Dann ist \mathcal{C}_n ein exaktes Konfidenzband für F .

Bemerkung 4.21. Die Herleitung eines asymptotischen Konfidenzbands für nicht notwendigerweise stetiges F mittels Satz 2.2 sowie die Überprüfung der Güte dieses Bands ist eine Übungsaufgabe.

Beispiel 4.22. Wir betrachten X_1, X_2, \dots i.i.d. mit $X_1 \sim B(5, 4)$ und berechnen \mathcal{C}_n für $n \in \{20, 200\}$. Die erhaltenen Bänder sind in Abbildung 4.2 dargestellt. Dass das Band auch tatsächlich exakt ist kann mit Hilfe des hier verlinkten R-Codes illustriert werden, zusätzliche Beispiele folgen in den Übungen.

Abbildung 4.2: Beispiel eines mittels Satz 4.20 konstruierten Konfidenzbands für $X \sim B(5, 4)$.

Kapitel 5

Hypothesentests

Vor diesem Kapitel lesen Sie bitte `testing_toy_example.pdf`, in dem die statistischen Hypothesentests zugrundeliegende Idee skizziert und grundlegende Konzepte definiert werden. Die folgenden Seite bauen auf `testing_toy_example.pdf` auf.

Wir entwickeln die klassischen t -Tests (Normalverteilungsannahme!), werden allgemeine Konzepte (wie Teststatistik, Hypothese, Hypothesentest und p -Wert) aber auch möglichst allgemein definieren.

Beispiel 5.1. Sei X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. Gesucht ist ein Hypothesentest für $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, also eine Entscheidungsregel, gemäß derer die Nullhypothese H_0 verworfen wird oder nicht. Lt. Satz 3.10 gilt unter $H_0 : \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}} \sim t_{n-1}$, zusätzlich wissen wir aus Kapitel 4, dass

$$C_n(X_1, \dots, X_n) = \left[\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}} \right]$$

ein exaktes Konfidenzintervall für μ_0 mit Überdeckungswahrscheinlichkeit $1 - \alpha$ ist. Zur Konstruktion eines Hypothesentests für $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ können wir daher insbesondere wie folgt vorgehen:

Idee 1: Wir verwerfen H_0 , falls $T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}}$ vom Betrag her 'zu groß' ist (woher kommt diese Idee?). Genauer: wir verwerfen H_0 dann und nur dann, wenn

$$|T_n| = \left| \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}} \right| > t_{n-1, 1-\frac{\alpha}{2}}$$

gilt. Bei dieser Entscheidungsregel erhalten wir sofort

$$P_{H_0}(|T_n| > t_{n-1, 1-\frac{\alpha}{2}}) = \alpha,$$

der resultierende Hypothesentest ist also exakt zum Niveau α .

Idee 2: Wir verwerfen H_0 dann und nur dann wenn

$$\mu_0 \notin C_n(X_1, \dots, X_n) = \left[\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}} \right].$$

Auch in diesem Fall erhalten wir sofort

$$P_{H_0}(\mu_0 \notin C_n(X_1, \dots, X_n)) = \alpha,$$

der resultierende Hypothesentest ist daher ebenfalls exakt zum Niveau α .

Beide Tests stimmen überein, denn es gilt offensichtlich:

$$\begin{aligned} \mu_0 \in C_n(X_1, \dots, X_n) &\iff |\mu_0 - \bar{X}_n| \leq t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_n^2}{n}} \\ &\iff \sqrt{n} \frac{|\mu_0 - \bar{X}_n|}{\sqrt{S_n^2}} \leq t_{n-1, 1-\frac{\alpha}{2}} \\ &\iff |T_n| \leq t_{n-1, 1-\frac{\alpha}{2}}. \end{aligned}$$

Mit anderen Worten:

$$\mu_0 \notin C_n \iff |T_n| > t_{n-1, 1-\frac{\alpha}{2}}.$$

Wir berechnen nun die Power π_n ⁱ des soeben hergeleiteten Hypothesentests: Gelte X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und $\mu \neq \mu_0$ (in diesem Fall ist also $H_1 : \mu \neq \mu_0$ richtig). Unter Verwendung von

$$\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}} = \sqrt{n} \frac{\bar{X}_n - \mu + \mu - \mu_0}{\sqrt{S_n^2}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} + \sqrt{n} \frac{\mu - \mu_0}{\sqrt{S_n^2}}$$

folgt

$$\mathbb{P}_\mu(|T_n| \leq t_{n-1, 1-\frac{\alpha}{2}}) = \mathbb{P}_\mu\left(-t_{n-1, 1-\frac{\alpha}{2}} \leq T_n \leq t_{n-1, 1-\frac{\alpha}{2}}\right) = \mathbb{P}_\mu\left(L_n \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \leq U_n\right),$$

wobei L_n, U_n ('Lower' and 'Upper') definiert sind durch

$$\begin{aligned} L_n &:= -t_{n-1, 1-\frac{\alpha}{2}} - \sqrt{n} \frac{\mu - \mu_0}{\sqrt{S_n^2}} \\ U_n &:= t_{n-1, 1-\frac{\alpha}{2}} - \sqrt{n} \frac{\mu - \mu_0}{\sqrt{S_n^2}}. \end{aligned}$$

(i) Für $\mu - \mu_0 > 0$ folgt sofort $L_n, U_n \xrightarrow{[P]} -\infty$ für $n \rightarrow \infty$ (warum?) und wir erhalten

$$\lim_{n \rightarrow \infty} \mathbb{P}_\mu\left(|T_n| \leq t_{n-1, 1-\frac{\alpha}{2}}\right) = \lim_{n \rightarrow \infty} \mathbb{P}_\mu\left(L_n \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \leq U_n\right) = 0.$$

(ii) Für $\mu - \mu_0 < 0$ folgt $L_n, U_n \xrightarrow{[P]} +\infty$ für $n \rightarrow \infty$ (warum?) und wir erhalten ebenfalls

$$\lim_{n \rightarrow \infty} \mathbb{P}_\mu\left(|T_n| \leq t_{n-1, 1-\frac{\alpha}{2}}\right) = \lim_{n \rightarrow \infty} \mathbb{P}_\mu\left(L_n \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \leq U_n\right) = 0.$$

Mit anderen Worten: Die Power $\pi_n(\mu)$ des Hypothesentests konvergiert für $n \rightarrow \infty$ gegen 1, für hinreichend großes n hat der Test also Power beliebig nahe bei 1.

ⁱWir schreiben in diesem Abschnitt manchmal π_n statt π um die Abhängigkeit von der Sample Size hervorzuheben

Definition 5.2. Sei $X \sim (P_\theta)_{\theta \in \Theta}$. Jedes nichtleere $\Theta_0 \subseteq \Theta$ heißt (parametrische) Hypothese. Besteht Θ_0 nur aus einem Element, i.e. gilt $\Theta_0 = \{\mu_0\}$, dann heißt Θ_0 einfach, ansonsten zusammengesetzt.

Definition 5.3. Sei X_1, \dots, X_n eine Zufallsstichprobe von X . Ein Hypothesentest ist eine messbare Abbildung $\varphi : \text{Rg}(X)^n \rightarrow \{0, 1\}$, die Menge $V := \varphi^{-1}(\{1\})$ heißt Verwerfungsbereich des Tests. Eine Teststatistik T_n ist eine messbare (reellwertige) Funktion von X_1, \dots, X_n .

Beispiel 5.4 (Fortsetzung Beispiel 5.1). Im Falle von Beispiel 5.1 ist, wie schon erwähnt, die Teststatistik T_n gegeben durch

$$T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}},$$

der Hypothesentest $\varphi(X_1, \dots, X_n) := \mathbf{1}_{(t_{n-1, 1-\frac{\alpha}{2}}, \infty)}(|T_n|)$ ist offensichtlich eine Funktion der Teststatistik, und der Verwerfungsbereich V_α ⁱⁱ ist gegeben durch

$$V_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \sqrt{n} \frac{|\bar{x}_n - \mu_0|}{\sqrt{s_n^2}} > t_{n-1, 1-\frac{\alpha}{2}} \right\}$$

In Beispiel 5.1 haben folgendes Resultat hergeleitet:

Satz 5.5 (Zweiseitiger t -test). Sei X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und T_n definiert durch $T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}}$. Dann ist $\varphi : \text{Rg}(X)^n \rightarrow \{0, 1\}$, definiert durch

$$\varphi(X_1, \dots, X_n) := \mathbf{1}_{(t_{n-1, 1-\frac{\alpha}{2}}, \infty)}(|T_n|)$$

ein exakter Test zum Niveau α für $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, i.e., es gilt

$$\mathbb{E}_{H_0}(\varphi(X_1, \dots, X_n)) = \alpha.$$

Die Power dieses Tests geht für $n \rightarrow \infty$ mit Wahrscheinlichkeit 1 gegen 1.

Möchten wir (wiederum im Normalverteilungsfall) $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$ testen, dann können wir vollkommen analog vorgehen:

Beispiel 5.6. Sei X_1, \dots, X_n i.i.d. $\mathcal{N}(\mu, \sigma^2)$, $\mu_0 \in \mathbb{R}$ fest und wir möchten einen Hypothesentest für $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$ herleiten. Wir betrachten wiederum die Teststatistik

$$T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}}.$$

und gehen in zwei Schritten vor:

1. Wir betrachten zuerst $H'_0 : \mu = \mu_0$ vs. $H'_1 : \mu > \mu_0$. Verwerfen wir H'_0 genau dann wenn $T_n > t_{n-1, 1-\alpha}$ dann erhalten wir $\mathbb{P}_{H'_0}(T_n > t_{n-1, 1-\alpha}) = \alpha$.

ⁱⁱWir schreiben oft V_α anstatt V um die Abhängigkeit von α klarzustellen.

2. Wenn H_0 gilt dann bedeutet dies, dass X_1, \dots, X_n i.i.d. $\mathcal{N}(\mu, \sigma^2)$ mit $\mu \leq \mu_0$ gilt. Betrachten wir wiederum

$$T_n = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}} = \sqrt{n} \frac{\bar{X}_n - \mu + \mu - \mu_0}{\sqrt{S_n^2}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} - \sqrt{n} \frac{\mu_0 - \mu}{\sqrt{S_n^2}}.$$

dann erhalten wir daher

$$\begin{aligned} \mathbb{P}_\mu(T_n > t_{n-1, 1-\alpha}) &= \mathbb{P}_\mu\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} > t_{n-1, 1-\alpha} + \sqrt{n} \frac{\mu_0 - \mu}{\sqrt{S_n^2}}\right) \\ &\leq \mathbb{P}_\mu\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} > t_{n-1, 1-\alpha}\right) = \alpha. \end{aligned}$$

Mit anderen Worten:

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu(T_n > t_{n-1, 1-\alpha}) = \mathbb{P}_{\mu_0}(T_n > t_{n-1, 1-\alpha}) = \alpha.$$

Verwerfen wir also H_0 genau dann wenn $T_n > t_{n-1, 1-\alpha}$ dann liefert dies einen exakten Test auf Niveau α für $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$.

Wir haben folgendes Resultat bewiesen:

Satz 5.7 (Einseitiger t -test). Sei X_1, \dots, X_n i.i.d. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ und T_n definiert durch $T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2}}$. Dann ist $\varphi : \text{Rg}(X)^n \rightarrow \{0, 1\}$, definiert durch

$$\varphi(X_1, \dots, X_n) := \mathbf{1}_{(t_{n-1, 1-\alpha}, \infty)}(T_n)$$

ein exakter Test zum Niveau α für $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$, i.e., es gilt

$$\mathbb{E}_{H_0}(\varphi(X_1, \dots, X_n)) = \alpha.$$

Die Power dieses Tests geht für $n \rightarrow \infty$ mit Wahrscheinlichkeit 1 gegen 1.

Angenommen x_1, \dots, x_n eine Stichprobe von $X \sim \mathcal{N}(\mu, \sigma)$ und wir möchten $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ mit Hilfe von Satz 5.7 testen. Anstatt nur die binäre Testentscheidung $\varphi(x_1, \dots, x_n) \in \{0, 1\}$ anzugeben hat es sich in den letzten Jahren durchgesetzt, den wesentlich informativeren p -Wert zu reporten, der das Ausmaß der Evidenz gegen H_0 quantifiziert oder, anders ausgedrückt, die Wahrscheinlichkeit 'unter H_0 etwas mindestens so Extremes wie das gerade Beobachtete zu erhalten' angibt. Wir werden in den Übungen einige konkrete Beispiele sehen, die diese Interpretation konkretisieren.

Wir definieren nun den p -Wert allgemein (keinerlei Verteilungsannahme) und formulieren dann seine wichtigsten Eigenschaften:

Definition 5.8. Sei $X \sim (P_\theta)_{\theta \in \Theta}$, $\emptyset \neq \Theta_0 \subseteq \Theta$. Weiters sei T_n eine Teststatistik für $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$. V_α bezeichne den Verwerfungsbereich auf Niveau α . Dann heißt

$$\hat{p}(\omega) := \inf \{ \alpha \in [0, 1] : (X_1, \dots, X_n) \in V_\alpha \}, \quad (\inf \emptyset := 1)$$

der p -Wert des Tests.

Bemerkung 5.9. Der p -Wert kann alternativ auch direkt über die Teststatistik T_n definiert werden, wobei in diesem Fall dann der Verwerfungsbereich entsprechend als Teilmenge von \mathbb{R} zu sehen ist. Bei den hier betrachteten Tests kann die Äquivalenz leicht überprüft werden.

Bemerkung 5.10. Der p -Wert ist also (Messbarkeit vorausgesetzt) selbst wieder eine Zufallsvariable mit Werten in $[0, 1]$ und ist (mehr oder weniger) definiert als das kleinste Niveau, zu dem noch verworfen wird.

Bemerkung 5.11. Für einen Test auf Niveau α können wir die Testentscheidung alternativ auch wie folgt ausdrücken: Verwerfe H_0 genau dann wenn $\hat{p} < \alpha$.

Satz 5.12 (Eigenschaften des p -Werts). *Sei $X \sim (P_\theta)_{\theta \in \Theta}$, $\emptyset \neq \Theta_0 \subseteq \Theta$ und T_n sei eine Teststatistik für $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$. V_α bezeichne den Verwerfungsbereich auf Niveau α , die Abbildung $\alpha \mapsto V_\alpha$ sei monoton wachsend. Dann gelten folgende Aussagen.*

(a) Falls $\sup_{\theta \in \Theta_0} P_\theta((X_1, \dots, X_n) \in V_\alpha) \leq \alpha$ für jedes $\alpha \in (0, 1)$ dann gilt

$$\mathbb{P}_\theta(\hat{p} \leq x) \leq x.$$

für jedes $\theta \in \Theta_0$ und für jedes $x \in (0, 1)$.

(b) Existiert zusätzlich (!) ein $\theta_0 \in \Theta_0$, sodass $\mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in V_\alpha) = \alpha$ für jedes $\alpha \in (0, 1)$, dann gilt sogar

$$\mathbb{P}_{\theta_0}(\hat{p} \leq x) = x$$

für alle $x \in (0, 1)$.

Beweis: Zum Beweis der erste Aussage sei $\theta \in \Theta_0$ und $x \in (0, 1)$ beliebig. Nachdem aus $\hat{p}(\omega) \leq x$ aus der Definition des p -Werts sofort $(X_1(\omega), \dots, X_n(\omega)) \in V_y$ für jedes $y \in (x, 1)$ folgt erhalten wir insgesamt

$$\mathbb{P}_\theta(\hat{p} \leq x) \leq \mathbb{P}_\theta((X_1, \dots, X_n) \in V_y) \leq y$$

und damit (via Infimumsbildung über alle $y \in (x, 1)$) die gewünschte Eigenschaft

$$\mathbb{P}_\theta(\hat{p} \leq x) \leq x.$$

Zum Beweis des zweiten Teils sei wiederum $x \in (0, 1)$. Aus $(X_1(\omega), \dots, X_n(\omega)) \in V_x$ folgt per definitionem $\hat{p}(\omega) \leq x$ und wir erhalten

$$x = \mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in V_x) \leq \mathbb{P}_{\theta_0}(\hat{p} \leq x) \leq x. \blacksquare$$

Bemerkung 5.13. Definieren wir unter den Voraussetzungen von Satz 5.12 einen Hypothesentest $\psi : Rg(X_1)^n \rightarrow \{0, 1\}$ durch $\psi(X_1, \dots, X_n) = \mathbf{1}_{[0, \alpha]}(\hat{p})$ dann erhalten wir im Falle von (a)

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta(\psi(X_1, \dots, X_n)) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\hat{p} < \alpha) \leq \alpha,$$

der Fehler 1. Art des Hypothesentests ψ ist also höchstens α . Gilt zusätzlich (b) dann ist der Fehler 1. Art sogar gleich α .

Der Großteil der in R implementierten Hypothesentests liefert sowohl den Wert der (dem Test zugrundeliegenden) Teststatistik als auch den p -Wert, in vielen Fällen aber nicht den Verwerfungsraum bzw. den Threshold. Für eine Testentscheidung ist der p -Wert aber (gemäß oder obigen Ausführungen) ausreichend, er enthält mehr Information als nur die binäre Entscheidung (verwerfe H_0 oder nicht).

Bemerkung 5.14. Nicht selten findet man in Online Skripten die Aussage ‘der p -Wert ist die Wahrscheinlichkeit, dass unter den gegebenen Daten H_0 richtig ist’. Diese Aussage macht nur im Sinne der hier nicht behandelten Bayes Statistik Sinn, in der frequentistischen Statistik (also dem hier studierten Zugang) ist die Aussage schlichtweg sinnbefreit.

Kapitel 6

A glimpse on Regression

Wir haben die Grundidee von Regression schon als Nebenprodukt/Werkzeug (Übungsaufgabe Flaschenpfand) in der UV kennengelernt. Im folgenden kurzen Kapitel skizzieren wir zuerst die der Regression zugrundeliegende Idee, diskutieren kurz das sog. R^2 , und fokussieren dann auf die Spezialfälle linearer und logistischer Regression.

Gegeben seien $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, wobei die \mathbf{x}_i mehrdimensional sein können, die y_i jedoch eindimensional sind, i.e., es gelte $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ für ein geeignetes $d \geq 1$. Wir nehmen weiters an, dass es eine unbekannte (!), der Einfachheit halber parametrische Funktion r_{θ^*} sowie zufällige i.i.d. Fehler $\varepsilon_1, \dots, \varepsilon_n$ mit $\mathbb{E}(\varepsilon_1) = 0$ gibt, sodass die folgende Gleichheit gilt:

$$y_i = r_{\theta^*}(\mathbf{x}_i) + \varepsilon_i, \quad i \in \{1, \dots, n\} \quad (6.1)$$

Zielsetzung ist es, aus den vorhandenen Daten die sog. *Regressionsfunktion* r_{θ^*} bestmöglich zu schätzen, letzteres ist äquivalent zur bestmöglichen Schätzung von θ^{*i} . Die einzelnen Koordinaten von \mathbf{x}_i heißen *erklärende Variablen oder Features*, Y_i die *erklärte Variable, Zielgröße oder Output*.

Formal sauber schreiben wir in der Folge oft Y_i statt y_i um anzudeuten, dass Y_i eine Zufallsvariable ist, die sich aus $Y_i = r_{\theta^*}(\mathbf{x}_i) + \varepsilon_i$ ergibt. Die \mathbf{x}_i nehmen wir als deterministisch anⁱⁱ. Für den Erwartungswert von Y_i erhalten wir daher sofort

$$\mathbb{E}(Y_i) = \mathbb{E}(r_{\theta^*}(\mathbf{x}_i) + \varepsilon_i) = r_{\theta^*}(\mathbf{x}_i) + \mathbb{E}(\varepsilon_i) = r_{\theta^*}(\mathbf{x}_i).$$

Naheliegende Vorgehensweise zur Schätzung von θ^* (und damit zur Schätzung von r_{θ^*}) ist es, unter allen möglichen Werten von $\theta \in \Theta$ jenen (oder einen von jenen) mit kleinstem quadratischen Fehler zu wählen, i.e., man betrachtet

$$\arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i - r_{\theta}(\mathbf{x}_i))^2. \quad (6.2)$$

Für jedes feste Modell r_{θ} heißen die Größen $f_i := y_i - r_{\theta}(\mathbf{x}_i)$ *Residuen* des Modells (hier f_i für 'Fehler'), das zuvor erwähnte Minimierungsproblem entspricht also genau der Minimierung der Quadratsumme der Residuen.

ⁱWir nehmen (wie gewohnt) an, dass der Parameter θ identifizierbar ist, dass also für $\theta \neq \psi$ auch die Funktionen r_{θ} und r_{ψ} unterschiedlich sind

ⁱⁱIm allgemeineren stochastischen Setting eines Zufallsvektors (\mathbf{X}, Y) mit Stichprobe $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ interessiert man sich beispielsweise für eine Funktion r , die $\mathbb{E}(Y - r(X))^2$ minimiert. Zur sauberen Herleitung dieser Funktion - der sog. bedingten Erwartung $\mathbb{E}(Y|X)$ von Y unter X - benötigen wir allerdings weitere Werkzeuge, siehe VO Abhängigkeitsmodellierung

Bemerkung 6.1. Die oben gemachte Annahme, dass die Fehler $(\varepsilon_i)_{i=1}^n$ i.i.d. sind, also insbesondere auch nicht vom Wert \mathbf{x}_i abhängen, ist einigermaßen restriktiv und in der Praxis nicht immer erfüllt. Für diese erste kurze Einführung gehen wir aber davon aus, dass die Bedingung erfüllt ist.

Ein in der Literatur sehr häufig verwendetes Maß für die Güte eines Regressionsmodells r_θ ist das sog. R^2 (Bestimmtheitsmaß, coefficient of determination). Grundidee von R^2 ist, die Größe $\sum_{i=1}^n (y_i - \bar{y}_n)^2$ der Größe $\sum_{i=1}^n f_i^2$ gegenüberzustellen. Genauer: R^2 ist definiert durch

$$R^2 = 1 - \frac{\sum_{i=1}^n f_i^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2}. \quad (6.3)$$

In zahlreichen Fällen, darunter die im folgenden Abschnitt erwähnte lineare Regression, fällt R^2 mit der Anteil der sog. erklärten Varianz, gegeben durch $\sum_{i=1}^n (r_\theta(\mathbf{x}_i) - \bar{y}_n)^2$, zusammen. In diesen Fällen gilt dann

$$R^2 = \frac{\sum_{i=1}^n (r_\theta(\mathbf{x}_i) - \bar{y}_n)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2}.$$

Beachten Sie, dass die Namensgebung Sinn macht, da im Falle von $\frac{1}{n} \sum_{i=1}^n r_\theta(\mathbf{x}_i) = \bar{y}_n$ der Ausdruck $\sum_{i=1}^n (r_\theta(\mathbf{x}_i) - \bar{y}_n)^2$ bis auf die normierende Konstante n bzw. $n - 1$ genau der Stichprobenvarianz der Werte $r_\theta(\mathbf{x}_1), \dots, r_\theta(\mathbf{x}_n)$ entspricht.

Bemerkung 6.2. Sobald ein Schätzer $r_{\hat{\theta}}$ für r_{θ^*} berechnet ist, können für neue Datenⁱⁱⁱ x_1^n, \dots, x_m^n die Prognosen y_1^n, \dots, y_m^n berechnet werden durch $y_i^n := r_{\hat{\theta}}(x_i^n)$.

6.1 Lineare Regression

Im univariaten Fall (also $d = 1$) sprechen wir von *univariater linearer Regression* falls die echte Regressionsfunktion r_{θ^*} von der Form

$$y_i = \theta_0^* + \theta_1^* x_i + \varepsilon_i, \quad i \in \{1, \dots, n\} \quad (6.4)$$

mit $\theta^* = (\theta_0^*, \theta_1^*)$ ist (also wenn r_{θ^*} affin in x ist). Unter der Annahme, dass nicht alle x_i gleich sind (Generalvoraussetzung in diesem Abschnitt), kann das Minimierungsproblem (6.2), in welchem wir als Klasse aller möglichen Regressionfunktionen alle Funktionen der Form $r_\theta(x) = \theta_0 + \theta_1 x$ betrachten, elementar gelöst werden^{iv} und wir erhalten die folgenden zwei Schätzer für θ_1^* und θ_0^* :

$$\hat{\theta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad \hat{\theta}_0 = \bar{y}_n - \hat{\theta}_1 \bar{x}_n \quad (6.5)$$

Beachten Sie, erstens, dass der Schätzer für θ_1^* genau der Quotient aus empirischen Kovarianz und Stichprobenvarianz $s_{n,x}^2$ der x -Werte entspricht und damit leicht zu merken ist; und, zweitens, dass $\hat{\theta}_1$ mittels empirischen Korrelationskoeffizienten ρ_n (Gleichung (5.20) im Skriptum Math. Statistik) und Stichprobenvarianzen $s_{n,x}^2, s_{n,y}^2$ alternativ ausgedrückt werden kann durch

$$\hat{\theta}_1 = \rho_n \sqrt{\frac{s_{n,y}^2}{s_{n,x}^2}}.$$

ⁱⁱⁱ das 'n' steht für 'neu'

^{iv} partielle Ableitungen nullsetzen, Definitheit checken, einfache Übungsaufgabe

Setzen wir $g_i = \frac{x_i - \bar{x}_n}{\sum_{j=1}^n (x_j - \bar{x}_n)^2}$ dann gilt offensichtlich $\sum_{i=1}^n g_i = 0$ und wir erhalten

$$\hat{\theta}_1 = \sum_{i=1}^n g_i y_i, \quad (6.6)$$

der Schätzer $\hat{\theta}_1$ ist also linear in y_1, \dots, y_n .

Die Schätzer $\hat{\theta}_1, \hat{\theta}_0$ haben auch ohne weitere Voraussetzungen gute Eigenschaften: Wegen

$$\mathbb{E}(\hat{\theta}_1) = \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}\right) = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(\mathbb{E}(Y_i - \bar{Y}_n))}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

und

$$\mathbb{E}(Y_i - \bar{Y}_n) = \theta_0 + \theta_1 x_i - \frac{1}{n} \sum_{i=1}^n (\theta_0 + \theta_1 x_i) = \theta_0 + \theta_1 x_i - \theta_0 - \theta_1 \bar{x}_n = \theta_1 (x_i - \bar{x}_n)$$

erhalten wir sofort

$$\mathbb{E}(\hat{\theta}_1) = \theta_1.$$

Anwendung von Gleichung (6.5) liefert daher

$$\mathbb{E}(\hat{\theta}_0) = \mathbb{E}(\bar{Y}_n - \hat{\theta}_1 \bar{x}_n) = \theta_0 + \theta_1 \bar{x}_n - \theta_1 \bar{x}_n = \theta_0.$$

Mit anderen Worten: Die Schätzer $\hat{\theta}_1, \hat{\theta}_0$ sind erwartungstreu.

Nehmen wir zusätzlich $\mathbb{V}(\varepsilon_1) < \infty$ an, dann folgt außerdem (Rechenregeln der Varianz)

$$\mathbb{V}(\hat{\theta}_1) = \frac{1}{(\sum_{i=1}^n (x_i - \bar{x}_n)^2)^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \underbrace{\mathbb{V}(Y_i)}_{=\mathbb{V}(\varepsilon_1)} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \mathbb{V}(\varepsilon_1),$$

die Varianz des Schätzers $\hat{\theta}_1$ ist also umso kleiner, je größer die Stichprobenvarianz $s_{n,x}^2$ ist (für den Schätzer $\hat{\theta}_0$ gilt eine analoge Aussage). Im Falle von

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \infty \quad (6.7)$$

erhalten wir weiters, dass $\mathbb{V}(\hat{\theta}_1)$ und $\mathbb{V}(\hat{\theta}_0)$ für $n \rightarrow \infty$ gegen 0 konvergiert. Zusammenfassend haben wir daher das folgende Resultat bewiesen:

Satz 6.3. *Wir betrachten Daten $(x_1, y_1), \dots, (x_n, y_n)$ des Regressionsmodells (6.4) und nehmen an, dass $\mathbb{E}(\varepsilon_1) = 0$, $\mathbb{V}(\varepsilon_1) < \infty$ gilt, dass die Fehler $(\varepsilon_i)_{i=1}^n$ i.i.d. sind, und dass Bedingung (6.7) erfüllt ist. Dann sind die Schätzer $\hat{\theta}_1, \hat{\theta}_0$ gemäß Gleichung (6.5) erwartungstreu und konsistent.*

Gilt sogar $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2)$ für ein $\sigma^2 \in (0, \infty)$, dann sind auch beide Schätzer $\hat{\theta}_1, \hat{\theta}_0$ normalverteilt.

Beweis: Der Großteil wurde schon bewiesen, die verbleibenden Aussagen sind eine einfache Übungsaufgabe.

Bemerkung 6.4. Unter den Voraussetzungen des ersten Teils von Satz 6.3 kann sogar gezeigt werden (Satz von Gauß-Markov), dass die Schätzer $\hat{\theta}_1, \hat{\theta}_0$ in der Klasse aller linearen erwartungstreuen Schätzer effizient ist (also minimale Varianz hat).

Mittels der Residuen kann die Varianz $\mathbb{V}(\varepsilon_1)$ geschätzt werden durch

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n f_i^2, \quad (6.8)$$

in den Übungen werden wir zeigen, dass $\hat{\sigma}^2$ erwartungstreu für $\mathbb{V}(\varepsilon_1)$ ist.^v

Bemerkung 6.5. Für das oben erwähnte R^2 lässt sich im Fall univariater linearer Regression die Identität $\rho_n^2 = R^2$ zeigen, die nicht zuletzt den Namen R^2 rechtfertigt.

Bemerkung 6.6. Für multivariate lineare Regression sei an dieser Stelle nur erwähnt, dass die Grundprinzipien zwar gleich sind, dass aber insbesondere durch etwaige Korrelationen erklärender Variablen Probleme auftreten (Stichwort: Multikollinearität). Wir werden einige multivariate lineare Regressionen zwar im Übungsteil in R rechnen, für mehr Infos muss an dieser Stelle aus Zeitgründen aber auf die VO Regression Methods & Computational Statistics verwiesen werden.

6.2 Logistische Regression

Wir betrachten ein einfaches Beispiel, das veranschaulicht, inwiefern logistische Regression ‘anders’ als lineare Regression ist und wie in diesem Fall vorgegangen werden kann.

Beispiel 6.7 (Bestehen einer Prüfung vs. Lernaufwand; Daten gemäß wikipedia). Eine Gruppe Studierender lernt zwischen 0 und 6 Stunden für eine (offensichtlich sehr einfache) Prüfung. Es ist zu erwarten, dass höherer Lernaufwand Hand in Hand mit einer höheren Bestehens-Wahrscheinlichkeit ist.

Die Zielvariable im vorliegenden Fall ist also ‘Bestehen’ oder ‘Durchfallen’, wir sprechen von einem binären Output und codieren selbigen meist mit 1 für Bestehen und 0 für Durchfallen. Lernzeit und Outcome für die 20 Studierenden sind, geordnet nach Lernaufwand, in nachfolgender Tabelle zusammengefasst. Nachdem die Daten sich offensichtlich nicht mittels linearer

	1	2	3	4	5	6	7	8	9	10
Stunden	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50
Outcome	0	0	0	0	0	0	1	0	1	0

	11	12	13	14	15	16	17	18	19	20
Stunden	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Outcome	1	0	1	0	1	1	1	1	1	1

Regression modellieren lassen (warum?) kann man versuchen, mittels sogenannter logistischer

^vDer Nenner $n-2$ ergibt sich auch aus dem Satz von Fisher-Cochran, siehe VO Wahrscheinlichkeitstheorie im Master.

Regression (in einer erklärenden Variablen) ein passendes Modell zu finden und geht dabei wie folgt vor: Man nimmt an, dass die echte Regressionsfunktion r_{θ^*} mit $\theta^* = (\theta_0^*, \theta_1^*)$ aus der logistischen Familie, i.e., aus der Familie aller Funktionen der Form

$$r_{\theta}(x) = l_{\theta}(x) := \frac{e^{\theta_0 + \theta_1 x}}{1 + e^{\theta_0 + \theta_1 x}} = \frac{1}{1 + e^{-\theta_0 - \theta_1 x}} \quad (6.9)$$

ist und versucht, θ_0^*, θ_1^* aus den Daten zu schätzen.

Das zugrundegelegte statistische Modell hat also die Form

$$\mathbb{P}(Y = 1 | X = x) = l_{\theta}(x), \quad (6.10)$$

hat die erklärende Variable X den Wert x dann ist die Zielvariable Y also alternativverteilt mit Parameter $l_{\theta}(x) \in (0, 1)$. Oft wird logistische Regression auch alternativ über den sog. ‘Logit’ beschrieben (logarithmierte Odds) beschrieben - eine einfache Umformung liefert sofort den Zusammenhang

$$\ln \left(\frac{\mathbb{P}(Y = 1 | X = x)}{\mathbb{P}(Y = 0 | X = x)} \right) = \theta_0 + \theta_1 x. \quad (6.11)$$

Beachten Sie, dass der Logit (also der linke Ausdruck in der obigen Formel) Werte in $(-\infty, \infty)$ annehmen kann. Wegen (6.11) fällt das logistische Modell daher auch in die Klassen der GLMs (generalized linear models)^{vi}.

Zur Schätzung von $\theta^* = (\theta_0^*, \theta_1^*)$ könnte man motiviert durch Gleichung (6.2) versuchen den Parameter $\theta = (\theta_0, \theta_1)$ so zu wählen, dass der Ausdruck

$$\sum_{i=1}^n (y_i - l_{\theta}(x_i))^2$$

minimal wird. Nachdem eine analytische Berechnung im Gegensatz zum linearen Fall aber i.A. nicht möglich ist, arbeitet man üblicherweise mit dem Likelihood und berechnet stattdessen

$$\arg \max_{\theta \in \mathbb{R}^2} \prod_{i=1}^n l_{\theta}(x_i)^{y_i} (1 - l_{\theta}(x_i))^{1-y_i}. \quad (6.12)$$

Beachten Sie, dass dies genau der Berechnung des Maximum Likelihood im diskreten Fall entspricht, den wir schon in Kapitel 1 kennengelernt haben. Im konkreten Fall erhalten wir unter Verwendung der R-Funktion ‘glm’^{vii} die Schätzer $\hat{\theta}_0 = -4.078$ und $\hat{\theta}_1 = 1.505$. Abbildung 6.2 zeigt die oben erwähnten Daten und die erhaltenen Schätzfunktion $r_{\hat{\theta}} = l_{\hat{\theta}}$.

Bemerkung 6.8. Gegeben die geschätzte Regressionsfunktion $r_{\hat{\theta}} = l_{\hat{\theta}}$ können für neue Daten x_1^n, \dots, x_m^n Prognosen y_1^n, \dots, y_m^n in zwei Schritten berechnet werden:

- (i) Berechne $l_{\hat{\theta}}(x_i^n)$
- (ii) Falls $l_{\hat{\theta}}(x_i^n) \geq \frac{1}{2}$ setze $y_i^n := 1$, andernfalls setze $y_i^n := 0$.

Bemerkung 6.9. Analog zum linearen Fall können auch im Falle logistischer Regression mehrere erklärende Variable betrachtet werden. Im einfachsten Fall^{viii} sprechen wir dann von multivariater logistischer Regression. Genauso wie bei linearer Regression können hoch-korrelierte Variablen Probleme machen (Stichwort: Multikollinearität).

^{vi}nach einer Transformation ist das Modell linear in den Parametern

^{vii}und der Auswahl ‘family=binomial’

^{viii}ohne Interaktion der Variablen

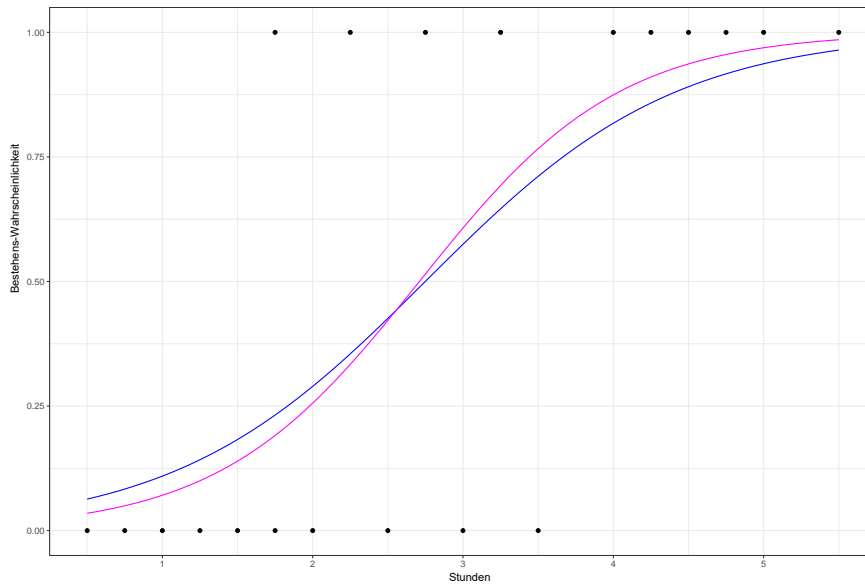


Abbildung 6.1: Daten aus Beispiel 6.7 Mittels ML geschätzte logistische Regressionsfunktion $r_{\hat{\theta}}$ (magenta) und mittels ‘nls’ berechnete logistische Funktion (blau)

6.3 Overfitting, Cross-Validation, Feature Selection

In der Regel wird es so sein, dass (sowohl im linearen als auch im logistischen Setting) das geschätzte Modell für die vorliegenden Daten (also die Daten, an die es angepasst wurde^{ix}) besser performt also für neue Daten. Würden wir beispielsweise für Daten $(x_1, y_1), \dots, (x_n, y_n)$ statt linearer Regression polynomiale Regression hinreichend hohen Grades betrachten, dann wäre für den Fall paarweise verschiedener x_i der Fehler sogar 0 - für neue Daten wird dies nicht der Fall sein. Bei Betrachtung zu komplexer Modelle (Bsp: zu viele erklärende Variablen, zu hoher Grad des Polynoms) sprechen wir von Overfitting.

Um eine realistische Einschätzung der Güte eines betrachteten Modells zu ermöglichen, kann beispielsweise mit Cross-Validation gearbeitet werden: Man betrachtet (i) ein (zufällig gewähltes) Subsample der vorliegenden Daten, (ii) schätzt dafür die Regressionsfunktion, und betrachtet (iii) $\sum_{i \in I} (y_i - r_{\hat{\theta}}(x_i))^2$, wobei I jene Elemente aus $\{1, \dots, n\}$ bezeichnet, die nicht in (i) gezogen würden. Obige Schritte werden mehrmals wiederholt.

In der Praxis ist es häufig so, dass für die Erklärung einer Zielvariable Y zahlreiche Features zur Verfügung stehen, a-priori aber vollkommen unklar ist, welche Features überhaupt relevant sind. Die Auswahl der relevanten Features ist oftmals ein nicht-triviales Problem - man kann schließlich nicht alle möglichen Modelle durchprobieren (siehe die vorigen zwei Absätze). Mit modernen, erst in den letzten Jahren entwickelten nichtparametrischen Methoden ist es überraschenderweise aber doch möglich, relevante Features relativ gut herauszufiltern. Mit *FOCI* (siehe [1]) und *qmd* (siehe [2]) stehen zwei R-packages genau dafür zur Verfügung - wir werden die Performanz der packages kurz in den Übungen diskutieren.

^{ix}Im Machine Learning Kontext üblicheres Wording: Daten anhand derer das Modell trainiert wurde

Literatur

- [1] M. Azadkia, S. Chatterjee, N. Matloff (2021). FOCI: Feature Ordering by Conditional Independence, R package version 0.1.3. <https://cran.r-project.org/web/packages/FOCI/index.html>
- [2] F. Griessenberger, R.R. Junker, W. Trutschnig (2021). qmd: Quantification of Multivariate Dependence, R package version 1.0.0. <https://cran.r-project.org/web/packages/qmd/index.html>
- [3] P. Massart: The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality, *Annals of Probability* 18(3): 1269–1283 (1990)
- [4] I. Molchanov: The Theory of Random Sets, Springer, New York (2005)
- [5] A. Vaart (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press (1998)

Kapitel 7

Übungsblätter

01. Übung am 11. März 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 1. Sei $X \sim \mathcal{U}(a, b)$ mit $a < b$. Wir betrachten die beiden Schätzer $\hat{a}_n = \min\{X_1, \dots, X_n\}$ und $\hat{b}_n = \max\{X_1, \dots, X_n\}$. Berechnen Sie die Erwartungswerte von \hat{a}_n und \hat{b}_n für $n \in \{1, 2\}$ - sind \hat{a}_n und \hat{b}_n erwartungstreu? Sind die Schätzer stark konsistent? Bestätigen Sie Ihre Antwort mittels Simulationen in R.

Übungsaufgabe 2. Sei X exponentialverteilt und X_1, \dots, X_n eine Stichprobe von X . Aus der VO wissen wir, dass $\hat{\theta}_n = \frac{1}{X_n}$ ein stark konsistenter Schätzer für θ ist. Beweisen Sie, dass auch der Schätzer $\hat{\psi}_n := -\log(1 - F_n(1))$ stark konsistent für θ ist und überlegen Sie sich, wie man auf den Schätzer $\hat{\psi}_n$ kommt.

Finden Sie weiters mit Hilfe von R heraus, welcher der beiden Schätzer besser ist, indem Sie für ein festes, von Ihnen gewähltes $\theta \in (0, \infty)$ insgesamt $R = 1000$ Mal Stichproben der Größe $n = 100$ generieren, beide Schätzer berechnen, und dann deren Fehler vergleichen (Boxplot oder Ähnliches).

Übungsaufgabe 3. Sei $\hat{\theta}_n$ ein erwartungstreuer Schätzer von $\theta \in \Theta \subseteq \mathbb{R}$. Ist dann $(\hat{\theta}_n)^2$ auch erwartungstreuer Schätzer von θ^2 ? Wenn nein, wie groß ist die Verzerrung?

(R) Betrachten Sie $X \sim \mathcal{N}(\theta, 2^2)$ und den erwartungstreuer Schätzer $\hat{\theta}_n = \bar{X}_n$ für θ . Wählen Sie $n = 100$, erzeugen Sie für von Ihnen fest gewähltes θ Stichproben x_1, \dots, x_n und berechnen Sie $z = \bar{x}_n$ sowie $w = (\bar{x}_n)^2$. Wiederholen Sie den Vorgang $R = 100000$ mal, plotten Sie Histogramme und berechnen Sie den Mittelwert der so erhaltenen Werte $(z_i)_{i=1}^R$ bzw. $(w_i)_{i=1}^R$ - was ist zu beobachten?

Übungsaufgabe 4. Sei $X \sim \mathcal{U}(-a, a)$. Berechnen Sie sowohl mittels Maximum Likelihood Methode als auch mittels Momentenmethode Schätzer von a . Sind die erhaltenen Schätzer erwartungstreu? Sind sie (stark) konsistent?

Übungsaufgabe 5. Beweisen Sie Satz 1.15.

Übungsaufgabe 6. Eine consulting Firma sucht neue Mitarbeiter mit Universitätsabschluss. Die Einstellungskriterien inkludieren eine Punktezahl von mindestens 120 bei einem (in der Firma zu absolvierenden) IQ-Test, wobei der IQ als $\mathcal{N}(\theta, 5^2)$ -verteilt mit unbekanntem θ angenommen wird. Von 500 Bewerbern schaffen nur 25 den Test - schätzen Sie θ . Überprüfen Sie Güte des erhaltenen Schätzers mit Hilfe von Simulationen in R.

2. Übung am 18. März 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 7. Beweisen Sie, dass D_n gemäß Gleichung (2.1) tatsächlich eine Zufallsvariable (also messbar) ist.

Übungsaufgabe 8. Zu zeigen, dass F_{KS} gemäß Satz 2.1 eine Verteilungsfunktion ist, ist nicht einfach. Zeigen Sie, dass F_{KS} zumindest zwei der drei Eigenschaften von Verteilungsfunktionen erfüllt.

Übungsaufgabe 9. Beweisen Sie die zwei im Beweis von Satz 2.3 am Ende behaupteten Eigenschaften.

Übungsaufgabe 10. Berechnen Sie die Verteilungsfunktion von d_n gemäß Gleichung (2.2) approximativ mittels Simulationen in R für die Samples Sizes $n \in \{20, 50, 100, 500, 1000\}$.

Übungsaufgabe 11. Gehen Sie Beispiel 2.5. Schritt für Schritt durch und berechnen Sie approximativ mit R die Überdeckungswahrscheinlichkeit des Intervalls (2.3) für verschiedene Sample Sizes n der Poisson Verteilung.

3. Übung am 8. April 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 12 (Verallgemeinerung von Aufgabe 6). Eine consulting Firma sucht neue Mitarbeiter mit Universitätsabschluss. Die Einstellungskriterien inkludieren eine Punktezahl von mindestens 120 bei einem (in der Firma zu absolvierenden) IQ-Test, wobei der IQ als $\mathcal{N}(\theta, \sigma^2)$ -verteilt mit unbekanntem θ und $\sigma > 0$ angenommen wird. Von 5000 Bewerbern schaffen 630 den Test, 4000 haben ein Testergebnis von mindestens 100. Schätzen Sie θ und σ . Überprüfen Sie Güte der erhaltenen Schätzers mit Hilfe von Simulationen in R.

Übungsaufgabe 13. Konstruieren Sie (i) eine diskrete und (ii) eine absolut stetige Zufallsvariable X , die $\mathbb{P}^X = \mathbb{P}^{T \circ X}$ mit $T(x) = \frac{1}{x}$ für jedes $x \in \mathbb{R} \setminus \{0\}$ erfüllt.

Übungsaufgabe 14. Sei X_1, X_2, \dots eine Folge unabhängiger, $\mathcal{N}(\mu, \sigma^2)$ -verteilter Zufallsvariable; $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ bezeichne die Stichprobenvarianz. Für jedes $n \geq 2$ sei die Zufallsvariable T_n definiert durch

$$T_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}}.$$

Beweisen Sie, dass T_n (für $n \rightarrow \infty$) schwach gegen $Z \sim \mathcal{N}(0, 1)$ konvergiert und illustrieren Sie das Resultat mittels Simulationen in R.

Hinweis: Lesen Sie den Satz von Slutsky auf Wikipedia (English Version!) und wenden Sie ihn dann auf obige Situation an.

Übungsaufgabe 15. Wir werden kommende Woche eine auf den ersten Blick sehr überraschende Eigenschaft der Normalverteilung kennenlernen: Für eine Stichprobe X_1, \dots, X_n von $X \sim \mathcal{N}(\mu, \sigma^2)$ sind das Stichprobenmittel \bar{X}_n und die Stichprobenvarianz S_n^2 unabhängig. Zeigen Sie die Existenz einer Verteilung F mit $\mathbb{V}(X) > 0$ für $X \sim F$ und folgender Eigenschaft: für eine Stichprobe X_1, \dots, X_n von $X \sim F$ ist S_n^2 sogar eine Funktion von \bar{X}_n .

Übungsaufgabe 16. Sei $X = (X_1, X_2, \dots, X_d)$ multivariat normalverteilt mit Mittelwert $\mu \in \mathbb{R}^d$ und Kovarianzmatrix $\Sigma \in \mathbb{R}^{d \times d}$. Weiters sei $A \in \mathbb{R}^{d \times d}$ regulär. Beweisen Sie, dass $Y = AX$ ebenfalls multivariat normalverteilt ist und berechnen Sie Mittelwert und Kovarianzmatrix. Welche spezielle Eigenschaft hat Y für den Fall, dass A sogar orthogonal ist? Illustrieren Sie das Resultat in Dimension $d = 2$ mit Hilfe von Simulationen in R (zur Erzeugung von Stichproben der multivariaten Normalverteilung kann die Funktion 'mvrnorm' verwendet werden).

Übungsaufgabe 17. X, Y seien Zufallsvariable, A, B Borelmengen, und es gelte

$$\mathbb{P}(X \in A) = \mathbb{P}(Y \in B) = \alpha$$

für ein $\alpha \in (\frac{1}{2}, 1]$. Beweisen Sie die Ungleichung

$$2\alpha - 1 \leq \mathbb{P}(X \in A, Y \in B) \leq \alpha$$

und konstruieren Sie dann Beispiele, die zeigen, dass sowohl die linke als auch die rechte Ungleichung (wenn auch nicht simultan) zu Gleichungen werden können.

4. Übung am 22. April 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 18. Angenommen, für jedes $n \in \mathbb{N}$ gilt $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$, wobei $\lim_{n \rightarrow \infty} \sigma_n^2 = 0$ und $\lim_{n \rightarrow \infty} \mu_n = \mu$. Konvergiert dann $(X_n)_{n \in \mathbb{N}}$ schwach und, wenn ja, wogegen?

Übungsaufgabe 19 (Flaschenrückgabe). Der Datensatz in `and_out.RData` enthält die Anzahl der pro Tag verkauften und retournierten Pfandflaschen einer Supermarktketteⁱ, wobei bekannt ist, dass die Zeit D zwischen dem Verkauf und der Rückgabe einer Flasche geometrisch verteilt ist. Schätzen Sie $\mathbb{E}(D)$.

Hinweis: Falls ein Link nicht funktioniert, laden Sie bitte die aktuelle Version des Skriptums herunter und klicken Sie dort auf den Link.

Übungsaufgabe 20 (Logistik). Das R-Snippet `logistics_simulation.R` gibt mittels Simulationen eine Antwort auf die folgende Frage: Angenommen, ein Unternehmen produziert Bauteile eines festen Typs, die in LKWs zum Kunden transportiert werden. In einem LKW haben bis zu 40 Bauteile Platz. Je voller der LKW, desto geringer die Lieferkosten pro Bauteil. Zur Abschätzung der Kosten für das kommende Jahr wird aus den Daten der letzten Jahre die mittlere Anzahl von Bauteilen pro LKW berechnet und damit dann die Kosten für das kommende Jahr (unter Kenntnis der zu erwartenden Absatzzahlen) geschätzt - ist diese Vorgangsweise clever?

Übungsaufgabe 21. Beantworten Sie Frage 3.13.

Übungsaufgabe 22. Beweisen Sie Lemma 3.6.

ⁱam ersten Tag wurden als 5668 verkauft und schon 61 (davon) retourniert; vor dem ersten Tage wurden keine Pfandflaschen verkauft

5. Übung am 29. April 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 23. Gehen Sie Satz 3.14 samt Beweis durch, beweisen Sie damit Satz 4.7, und zeigen Sie dann mittels Simulationen in R, dass das Intervall in Satz 4.7 tatsächlich das tut, was es tun soll, also dass es den Parameter mit Wahrscheinlichkeit $1 - \alpha$ enthält.

Übungsaufgabe 24 (Fortsetzung von Aufgabe 19). Gehen Sie wie in der UV besprochen vor, um den Parameter p der geometrischen Verteilung (unter Verwendung aller - bzw. zumindest vieler - Daten) mittels minimalem L_2 -Fehler zu schätzen.

Übungsaufgabe 25 (Fortsetzung von Aufgabe 31). Überprüfen Sie die Güte des in der vorigen Aufgabe erhaltenen Schätzers \hat{p}_n für p mittels Simulationen und gehen Sie dafür wie folgt vor:

1. Wählen Sie ein festes $p \in (0, 0.5)$ und generieren Sie (unter Verwendung der Geometrischen Verteilung) einen Datensatz wie in `and_out.RData`.
2. Berechnen Sie (ausgehend von dem generierten Datensatz) den Schätzer \hat{p}_n für p wie in der vorigen Aufgabe.
3. Vergleichen Sie \hat{p}_n und p .

Übungsaufgabe 26. Beweisen Sie: für $X \sim Ex(\theta)$ gilt $e^{-\theta X} \sim \mathcal{U}(0, 1)$ und damit $2\theta \sum_{i=1}^n X_i \sim \chi_{2n}^2$ ⁱⁱ. Verwenden Sie diesen Zusammenhang dann, um ein exaktes Konfidenzintervall C_n für den Parameter θ der Exponentialverteilung herzuleiten und verifizieren Sie dessen Güte mittels Simulationen in R.

Übungsaufgabe 27 (Fortsetzung zu Aufgabe 21 betreffend Frage 3.13). Wie in der letzten UV kurz diskutiert, kann die Unabhängigkeit von \bar{X}_n und S_n^2 auch illustriert werden, indem überprüft wird, ob $H_n(x, y) \approx F_n(x)G_n(y)$ für alle $x, y \in \mathbb{R}$ gilt. Dabei bezeichnet H_n die zweidimensionale Verteilungsfunktion der Stichprobe $(\bar{x}_{n,1}, s_{n,1}^2), (\bar{x}_{n,2}, s_{n,2}^2), \dots, (\bar{x}_{n,R}, s_{n,R}^2)$ von (\bar{X}_n, S_n^2) und F_n, G_n die entsprechenden eindimensionalen empirischen Verteilungsfunktionen.

Gehen Sie dazu wie folgt vor:

1. Wählen Sie $n = 20$ und erzeugen Sie eine Stichprobe X_1, \dots, X_n von $X \sim \mathcal{N}(0, 1)$ und berechnen Sie $(\bar{x}_{n,1}, s_{n,1}^2) := (\bar{X}_n, S_n^2)$.
2. Wiederholen Sie den obigen Vorgang $R = 10.000$ Mal, $(\bar{x}_{n,1}, s_{n,1}^2), (\bar{x}_{n,2}, s_{n,2}^2), \dots, (\bar{x}_{n,R}, s_{n,R}^2)$ bezeichnen die erhaltenen Paare.

ⁱⁱder zweite Teil muss nicht bewiesen werden, siehe Additionstheorem für die Gamma-Verteilung im Skriptum Mathematische Statistik

-
3. Berechnen Sie das Maximum der Funktion $(x, y) \mapsto \Delta_n(x, y) := |H_n(x, y) - F_n(x)G_n(y)|$ auf einem feinen, äquidistanten Gitter G in $[-5, 5]^2$. Zusätzlich kann die Funktion auch als heatmap geplottet werden (ggplot command 'geom_tile').
 4. Wiederholen Sie obige Schritte für $n \in \{50, 100, 1000\}$.

6. Übung am 06. Mai 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 28. Angenommen, die Verteilungsfunktionen F, F_1, F_2, \dots sind absolut stetig mit Wahrscheinlichkeitsdichten f, f_1, f_2, \dots und es gilt

$$\lim_{n \rightarrow \infty} \|f - f_n\|_1 = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} |f(x) - f_n(x)| d\lambda(x) = 0.$$

Folgt dann, dass $(F_n)_{n \in \mathbb{N}}$ schwach gegen F konvergiert? Wie lautet Ihre Antwort, wenn stattdessen $\lim_{n \rightarrow \infty} |f(x) - f_n(x)| = 0$ für λ -fast jedes $x \in \mathbb{R}$ gilt?

Übungsaufgabe 29. Geben Sie ein Beispiel dafür, dass die erste Aussage des Satzes von Slutsky i.A. nicht gilt, wenn $Y_n \xrightarrow{w} Y$, die Zufallsvariable Y aber nicht konstant $[\mathbb{P}]$ ist. Verifizieren Sie Ihr Beispiel mittels Simulationen in R.

Übungsaufgabe 30. Es gibt mittlerweile zahlreiche Hypothesentests auf Unabhängigkeit von Zufallsvariablen X und Y . Ein relativ neuer Zugang arbeitet mit Copulas und testet nicht nur auf Unabhängigkeit, sondern schätzt auch die Stärke der Abhängigkeit. Gehen Sie die Beschreibung auf <http://www.trutschnig.net/software.html> bis inkl. Example 2 durch (es reicht, die Grundidee zu verstehen), installieren Sie 'qad' und schätzen Sie dann mittels qad die Abhängigkeit von \bar{X}_n und S_n^2 basierend auf den in Schritt 2 von Aufgabe 27 generierten Daten.

Übungsaufgabe 31 (Fortsetzung Flaschenpfand). Versuchen Sie, $\mathbb{E}(D)$ ohne dem Wissen, dass die Zeit D zwischen Kauf und Rückgabe geometrisch verteilt ist, zu schätzen. Lassen Sie Ihren Ideen dabei freien Lauf (in der Praxis ist die zugrundeliegende Verteilung a priori so gut wie nie bekannt).

Übungsaufgabe 32. Seien $x \geq 0, t > 0, m \in \mathbb{N}$ und X poissonverteilt mit Parameter mt (wir schreiben $X \sim \text{Pois}(mt)$). Berechnen Sie

$$\lim_{m \rightarrow \infty} \mathbb{P}(X \leq mx).$$

Hinweis: Nach dem Additionstheorem der Poisson-Verteilung existieren für jedes $m \in \mathbb{N}$ unabhängige $\text{Pois}(t)$ -verteilte Zufallsvariable X_1, \dots, X_m mit $\sum_{i=1}^m X_i = X$, obiger Grenzwert lässt sich daher durch geschickte Anwendung des CLTs berechnen.

7. Übung am 13. Mai 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 33. Für Wahrscheinlichkeitsmaße μ, ν auf $\mathcal{B}(\mathbb{R})$ ist die sog. *total variation* Metrik definiert durch

$$TV(\mu, \nu) := \sup\{|\mu(B) - \nu(B)| : B \in \mathcal{B}(\mathbb{R})\}.$$

Angenommen, F, F_1, F_2, \dots sind absolut stetige Verteilungsfunktionen mit Wahrscheinlichkeitsdichten f, f_1, f_2, \dots ; weiters bezeichne μ, μ_1, μ_2, \dots die entsprechenden Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$. Zeigen Sie: Wenn $\lim_{n \rightarrow \infty} \|f_n - f\|_1 = 0$, dann folgt sogar $\lim_{n \rightarrow \infty} TV(\mu_n, \mu) = 0$.

Übungsaufgabe 34. Angenommen $X \sim F$ mit F absolut stetig und X_1, \dots, X_n ist eine Stichprobe von X . Dann wissen wir aus dem Satz von Glivenko-Cantelli, dass F_n gleichmäßig gegen F konvergiert. Konvergieren die entsprechenden Wahrscheinlichkeitsmaße $\mu_1 := \mu_{F_1}, \mu_2 := \mu_{F_2}, \dots$ sogar in total variation gegen $\mu := \mu_F$?

Übungsaufgabe 35. Angenommen $(x_1, y_1), \dots, (x_n, y_n)$ ist eine Stichprobe von (X, Y) , wobei wir annehmen, dass die x_i paarweise verschieden sind. Wir möchten die dadurch entstehende Punktwolke möglichst gut durch eine Gerade der Form $y = ax + b$ beschreiben. Berechnen Sie dafür das Minimum (und die Minimierer \hat{a}, \hat{b}) der Funktion $L : \mathbb{R}^2 \rightarrow [0, \infty)$, definiert durch

$$L(a, b) := \sum_{i=1}^n (y_i - (ax_i + b))^2.$$

Generieren Sie dann Daten $(x_1, y_1), \dots, (x_n, y_n)$ aus dem Modell $y = 2x + 1 + \varepsilon$, in dem Sie die x -Werte zufällig aus der stetigen Gleichverteilung auf $[-5, 5]$ und die Fehlerwerte ε aus der Standardnormalverteilung ziehen, und dann $y_i := 2x_i + 1 + \varepsilon_i$ setzen. Berechnen Sie für diese Daten die Werte der Minimierer \hat{a}, \hat{b} von L und zeichnen Sie die Punkte $(x_1, y_1), \dots, (x_n, y_n)$ zusammen mit der Geraden $y = \hat{a}x + \hat{b}$ in \mathbb{R}^2 .

Übungsaufgabe 36. Gehen Sie den unter Beispiel 4.22 verlinkten R-Code Zeile für Zeile durch und überlegen Sie, was hier genau passiert. Modifizieren Sie den Code dann so, dass Sie die Überdeckungswahrscheinlichkeit für $X_1 \sim \mathcal{N}(0, 1)$ und eine weitere, von Ihnen selbst gewählte stetige Verteilung statt $B(5, 4)$ approximieren.

Übungsaufgabe 37. Verwenden Sie den Satz von Massart, um ein asymptotisches Konfidenzband für beliebige (nicht notwendigerweise stetige) Verteilungsfunktionen herzuleiten. Überprüfen Sie mittels Simulationen die Überdeckungswahrscheinlichkeit des erhaltenen Konfidenzbands für mindestens zwei verschiedene unstetige Verteilungenⁱⁱⁱ.

ⁱⁱⁱi.e. für Zufallsvariable deren Verteilungsfunktion mindestens eine Unstetigkeitsstelle hat

8. Übung am 27. Mai 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 38. Lösen Sie Exercise 42 in den Slides.

Übungsaufgabe 39. Lösen Sie Exercise 43 in den Slides.

Übungsaufgabe 40. Lesen Sie Beispiel 5.1 im Skriptum bis inkl. Idee 1. Erika Musterfrau schlägt basierend darauf den folgenden Test auf Normalverteilung^{iv} vor: Ausgehend von einer Stichprobe X_1, \dots, X_n von $X \sim F$ setzen wir $\mu_0 = \bar{X}_n$ und $\sigma_0 = \sqrt{S_n^2}$ und testen dann (mittels Idee 1) $H_0 : F = F_{\mu_0, \sigma_0^2}$, wobei F_{μ_0, σ_0^2} die Verteilungsfunktion von $\mathcal{N}(\mu_0, \sigma_0^2)$ bezeichnet. Überprüfen Sie mittels Simulationen, dass der resultierende Test für kleine Sample Sizes ($n = 10, 20, 30$) sehr geringe Power hat und überlegen Sie sich, warum dem so ist.

Übungsaufgabe 41. Sei $T : [0, 1] \rightarrow [0, 1]$ messbar und es gelte $\lambda(T^{-1}(\{a\})) = 0$ für jedes $a \in [0, 1]$. (i) Beweisen Sie, dass sich T dann nicht-fallend umordnen lässt, i.e., dass es eine nicht-fallende Transformation $T^* : [0, 1] \rightarrow [0, 1]$ und eine λ -treue Transformation $\varphi : [0, 1] \rightarrow [0, 1]$ gibt, sodass

$$T(x) = T^*(\varphi(x))$$

für jedes $x \in [0, 1]$ gilt.

(ii) Nehmen wir weiters an, dass T, T_1, T_2, \dots die obige Bedingung erfüllen und, dass $(T_n)_{n \in \mathbb{N}}$ λ -fast überall gegen T konvergiert. Konvergiert dann auch die entsprechende Folge $(\varphi_n)_{n \in \mathbb{N}}$ λ -fast überall gegen φ ?

Hinweis: @ (i): Betrachten Sie die Funktion $\varphi = F_T \circ T$ (wobei F_T die Verteilungsfunktion von T aufgefasst als Zufallsvariable auf $([0, 1], \mathcal{B}([0, 1]), \lambda)$ bezeichnet) und setzen Sie $T^* := F_T^-$.

@ (ii) Wenn Zufallsvariablen fast überall konvergieren, dann konvergieren die entsprechenden Verteilungsfunktionen schwach. Ist die Ziel-Verteilungsfunktion stetig, dann folgt aus der schwachen Konvergenz der Verteilungsfunktionen sogar deren gleichmäßige Konvergenz.

Übungsaufgabe 42 (Schadenssummen). Der Datensatz Insurance.RData enthält für 100.000 Sturmschäden einer Versicherung die ausbezahlte Summe, wobei selbige maximal 5.000 Euro (Deckungssumme im Vertrag) beträgt obwohl die echte Schadenshöhe höher sein kann^v. Angenommen, wir wissen, dass die echte Schadenshöhe S logarithmisch normalverteilt mit Parameter μ, σ^2 ist. Überlegen Sie sich zwei verschiedene Zugänge, um in diesem Fall μ und σ^2 und damit $F_S(10000)$ und $F_S(100000)$ zu schätzen, wobei F_S wie gewohnt die Verteilungsfunktion von S bezeichnet.

^{iv}die H_0 ist also 'X ist normalverteilt'

^vdiese Fälle sind im Datensatz mit 'censored=1' markiert, d.h. in diesen Fällen ist nur bekannt, dass die echte Schadenshöhe mehr als 5.000 Euro betrug

9. Übung am 03. Juni 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 43 (Warm-up). Sei ϑ ein Wahrscheinlichkeitsmaß mit $\vartheta((0, 1)) = 1$ (also volle Masse in $(0, 1)$). Weiter sei $a \in (0, 1)$ beliebig aber fest. Beweisen Sie, dass ein eindeutiges $\alpha \in [0, \infty)$ mit $\int_{(0,1)} x^\alpha d\vartheta(x) = a$ existiert.

Übungsaufgabe 44. Sei $F \in \mathcal{F}$ absolut stetig. Weiters sei X_1, X_2, \dots eine Stichprobe von $X \sim F$ und \hat{F}_n bezeichne die entsprechende empirische Verteilungsfunktion. Beweisen Sie, dass die an der Tafel skizzierte, absolut stetige Modifikation der empirischen Verteilungsfunktion $\hat{\hat{F}}_n$ sogar in der total variation Metrik gegen F konvergiert^{vi}.

Übungsaufgabe 45. Studieren Sie Abschnitt 5 bis inkl. Satz 5.7 im Skriptum und entwickeln Sie dann unter den selben Annahmen wie in Satz 5.7 einen exakten Test zum Niveau α für $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$.

Übungsaufgabe 46 (Sample Size Berechnung). Berechnen Sie ausgehend von Satz 5.5., wie groß die Sample Size n in dieser Situation sein muss, damit die Power $\pi(\mu_0 + 1)$ bzw. $\pi(\mu_0 + 2)$ mindestens 99% ist.

Übungsaufgabe 47. Angenommen, $X \sim (P_\theta)_{\theta \in \Theta}$ und C_n ist ein Konfidenzintervall mit Überdeckungswahrscheinlichkeit $1 - \alpha$ für den eindimensionalen Parameter θ . Kann (analog zu Beispiel 5.1) daraus sofort (und ohne Annahme der Normalverteilung) ein exakter Hypothesentest $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ zum Niveau α konstruiert werden?

^{vi}genauer: die entsprechenden Wahrscheinlichkeitsmaße $\mu_{\hat{\hat{F}}_n}$ konvergieren in total variation gegen μ_F

10. Übung am 10. Juni 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 48 (Dichteschätzung). Sei K eine eindimensionale Wahrscheinlichkeitsdichte und die sog. Bandbreite $h > 0$ beliebig, aber fest. Weiters sei x_1, \dots, x_n eine Stichprobe von $X \sim F$ und $\hat{f}_n : \mathbb{R} \rightarrow [0, \infty)$ definiert durch

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right).$$

Im Folgenden setzen wir immer $h = \frac{1}{\sqrt{n}}$.

1. Beweisen Sie, dass \hat{f}_n eine Wahrscheinlichkeitsdichte ist.
2. Wählen Sie $K = \frac{1}{2} \mathbf{1}_{[-1,1]}$, betrachten Sie $\alpha = 2, \beta = 4$, erzeugen Sie Stichproben x_1, \dots, x_n von X , wobei X Beta-verteilt mit den Parameters α, β ist, und plotten Sie die Funktion \hat{f}_n auf dem Intervall $[-0.5, 1.5]$ zusammen mit der Dichte von X .
Was passiert mit wachsender Sample Size n - nähert sich \hat{f}_n immer mehr der echten Dichte an?
3. Ersetzen Sie in Teil 2. K durch die Dichte der Standardnormalverteilung und wiederholen Sie die Simulation. Wird die Performanz des Schätzers \hat{f}_n besser oder schlechter?

Übungsaufgabe 49 (Dichteschätzung). Wiederholen Sie die vorigen Aufgabe für zwei andere absolut stetige Verteilungen (ersetzt wird also die Beta-Verteilung und das Plot-Intervall, nicht aber die zwei konkreten Kerne).

Übungsaufgabe 50. Verwenden Sie Satz 4.20 um ausgehend von einer Stichprobe $X_1, \dots, X_n \sim F$ (mit F stetig) einen Test für $H_0 : F = F_0$ versus $H_1 : F \neq F_0$ herzuleiten. Wie kann die Power dieses Tests mittels Simulationen überprüft werden?

Übungsaufgabe 51. Denken Sie Abschnitt 5 ab dem Paragraphen nach Satz 5.7 Schritt für Schritt durch.

Übungsaufgabe 52. Das R-Snippet `R-Codes_permtest.R` implementiert einen auf den ersten Blick etwas ungewöhnlichen Hypothesentest. Finden Sie heraus, was der R-Code macht, und berechnen Sie approximativ die power-Funktion des Tests. Vergleichen Sie die Powerfunktion mit jener des entsprechenden t -tests. Adaptieren Sie weiters den Code so, dass im Falle zweier Exponentialverteilung auf Gleichheit der Parameter getestet wird, und berechnen Sie approximativ den Fehler erster Art.

11. Übung am 17. Juni 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 53. Finden Sie heraus, was die Funktion `power.t.test` macht und lösen Sie damit abermals Übungsaufgabe 46. Um wie viel muss n vergrößert werden, wenn die Varianz verdoppelt/verdreifacht wird?

Übungsaufgabe 54. X_1, \dots, X_n sei eine Stichprobe von $X \sim \mathcal{N}(\mu, \sigma^2)$, die Zufallsvariable T_n sei definiert durch $T_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}}$. Weiters gelte $Y \sim t_{n-1}$ und F_Y bezeichne die Verteilungsfunktion von Y . Berechnen Sie die Verteilung der Zufallsvariable \hat{p} , definiert durch

$$\hat{p}(\omega) = \mathbb{P}(|Y| \geq |T_n(\omega)|) = 2(1 - F_Y(|T_n(\omega)|)).$$

NB: $\hat{p}(\omega)$ ist die Wahrscheinlichkeit dafür, dass eine t_{n-1} -verteilte Zufallsvariable einen (vom Absolutbetrag her) mindestens so großen Wert wie $T_n(\omega)$ annimmt. Welche in Abschnitt 5 im Skriptum getätigte Aussage verifiziert diese Übungsaufgabe?

Definition 7.1. Wir betrachten wie bisher $\emptyset \neq \Theta_0 \subset \Theta$ und eine Stichprobe $\mathbf{X} = (X_1, \dots, X_n)$ von $X \sim (P_\theta)_{\theta \in \Theta}$. Mit $\mathcal{X} := Rg(X)^n$ bezeichnen wir den Stichprobenraum. Weiters sei \mathcal{G} eine endliche Gruppe von Transformationen $g : \mathcal{X} \rightarrow \mathcal{X}$. Wir sagen, dass *unter H_0 die Randomisierungseigenschaft* gilt, genau dann, wenn für jedes $X \sim P_\theta$ mit $\theta \in \Theta_0$, jede Stichprobe $\mathbf{X} = (X_1, \dots, X_n)$ von X , und jedes $g \in \mathcal{G}$ gilt: \mathbf{X} und $g \circ \mathbf{X}$ haben die selbe Verteilung.

Übungsaufgabe 55. Sei (X_1, \dots, X_m) eine Stichprobe von $X \sim \mathcal{N}(\mu_1, \sigma^2)$, (Y_1, \dots, Y_n) eine dazu unabhängige Stichprobe von $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, und gelte $H_0 : \mu_1 = \mu_2$. Wir setzen $\mathcal{X} = \mathbb{R}^{m+n}$ und $\mathbf{X} = (X_1, \dots, X_m, Y_1, \dots, Y_n)$. Für jede Permutation π von $\{1, \dots, m+n\}$ sei $g_\pi : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^{m+n}$ definiert durch

$$g_\pi(z_1, \dots, z_{m+n}) = (z_{\pi(1)}, \dots, z_{\pi(m+n)}),$$

\mathcal{G} bezeichne die Gruppe all dieser Transformationen. Zeigen Sie, dass in diesem Setting die Randomisierungseigenschaft gilt.

Übungsaufgabe 56. Wir betrachten das Setting von Definition 7.1 und schreiben $M := \#G$ für die Kardinalität von G . Zusätzlich sei T eine beliebige Teststatistik, für jede Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ bezeichne $T_{(1)}(\mathbf{x}) \leq T_{(2)}(\mathbf{x}) \leq \dots \leq T_{(n)}(\mathbf{x})$ die Ordnungsstatistik der Werte $\{T \circ g(\mathbf{x}) : g \in G\}$. Für jedes $\alpha \in (0, 1)$ setzen wir $k := M - \lfloor M\alpha \rfloor$ sowie

$$\begin{aligned} M^0(\mathbf{x}) &= \#\{j \in \{1, \dots, M\} : T_{(j)}(\mathbf{x}) = T_{(k)}(\mathbf{x})\} \\ M^+(\mathbf{x}) &= \#\{j \in \{1, \dots, M\} : T_{(j)}(\mathbf{x}) > T_{(k)}(\mathbf{x})\}, \end{aligned}$$

und definieren einen ‘drei’-wertigen Test φ durch

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) > T_{(k)}(\mathbf{x}), \\ \frac{M\alpha - M^+(\mathbf{x})}{M^0(\mathbf{x})} & \text{if } T(\mathbf{x}) = T_{(k)}(\mathbf{x}), \\ 0 & \text{if } T(\mathbf{x}) < T_{(k)}(\mathbf{x}). \end{cases}$$

Beweisen Sie, dass dann für jedes $\theta \in \Theta_0$ die folgende Gleichheit gilt:

$$\mathbb{E}_\theta(\varphi \circ \mathbf{X}) = \alpha.$$

Übungsaufgabe 57. Die Zufallsvariable U und V seien unabhängig und stetig gleichverteilt auf $[0, 1]$. Berechnen Sie die Verteilung des Zufallsvektors (X, Y) gegeben durch

$$X = \sqrt{-2 \ln(U)} \cos(2\pi V), \quad Y = \sqrt{-2 \ln(U)} \sin(2\pi V).$$

Hinweis: Ein Scatterplot von Stichproben von (X, Y) zeigt sofort, auf welche Verteilung es hinausläuft.

12. Übung am 24. Juni 2024

UV Angewandte Statistik (405.170)

Link Ankreuzliste: siehe www.trutschnig.net/courses

Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.

Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 58. Zeigen Sie die in Bemerkung 6.5 behauptete Darstellung von R^2 im Falle univariater linearer Regression.

Übungsaufgabe 59 (Fortsetzung von Aufgabe 35). Generieren Sie Stichproben wie in Aufgabe 35 beschrieben und berechnen Sie dann die Schätzer $\hat{\theta}_0$ und $\hat{\theta}_1$ mit Hilfe der R-Funktion `lm`. Wiederholen Sie den Vorgang $R = 1000$ und plotten Sie die erhaltenen Werte $(\hat{\theta}_0^i, \hat{\theta}_1^i), i = 1 \dots R$ - was ist zu beobachten? Was ändert sich, wenn (i) die Stichprobengröße n erhöht wird, und wenn (ii) die Varianz von ε verringert wird?

Übungsaufgabe 60. Laden Sie den Lebenserwartungs-Datensatz der Weltbank herunter (<https://data.worldbank.org/indicator/SP.DYN.LE00.IN>) und lesen Sie ihn in R ein.

1. Filtern Sie auf die USA und bringen Sie den Datensatz in die folgende Form:

	country	year	life_expectancy
1	United States	1960	69.77
2	United States	1961	70.27
3	United States	1962	70.12
4	United States	1963	69.92
5	United States	1964	70.17
6	United States	1965	70.21

2. Plotten Sie die Zeitreihe in `ggplot2`.
3. Passen Sie mit Hilfe von `lm` ein lineares Modell an und prognostizieren damit Sie die Lebenserwartung in den USA im Jahr 2025 und im Jahr 2035. Verwenden Sie dafür die Funktion `predict`.
4. Wiederholen Sie die ersten drei Schritte für Österreich und für Japan. Was ist zu erkennen?

Übungsaufgabe 61. In der Praxis ist oft nicht einmal die parametrischer Form der Regressionsfunktion gegeben, nichtsdestotrotz gibt es nichtparametrische Methoden, die dann (zumindest im niedrigdimensionalen Setting) eingesetzt werden können. Für $(x_1, y_1), \dots, (x_n, y_n)$ kann beispielsweise die sog. Kernregression (Nadaraya Watson Regression) berechnet werden, i.e. als Schätzer für die Regressionsfunktion $r^*(x)$ an der Stelle x betrachtet man

$$\hat{r}(x) = \frac{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)}, \quad (7.1)$$

wobei als K beispielsweise die Dichte von $\mathcal{N}(0, 1)$ verwendet werden kann. Illustrieren Sie die Performanz des obigen Schätzers durch eine einfache Simulationsstudie, in dem Sie wie folgt vorgehen:

- Wählen Sie eine konkrete stetige^{vii} Regressionfunktion $r^* : [0, 10] \rightarrow \mathbb{R}$, betrachten Sie $(\varepsilon_i)_{i=1}^n$ i.i.d. mit $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2)$ für ein von Ihnen gewähltes $\sigma^2 > 0$ sowie eine Stichprobe x_1, \dots, x_n von $\mathcal{U}(0, 10)$ und setzen Sie $y_i := r^*(x_i) + \varepsilon_i$.
- Berechnen Sie dann \hat{r} gemäß Gleichung (7.1) auf einem äquidistanten Gitter in $[0, 10]$ und vergleichen Sie r^* und \hat{r} für kleines und für großes n . Was ist zu beobachten?

Übungsaufgabe 62. Lesen Sie Abschnitt 6.2 im Skriptum. Wählen Sie dann $\theta_1^*, \theta_0^* \in \mathbb{R}$ und generieren Sie Daten $(x_1, y_1), \dots, (x_n, y_n)$ des logistischen Modells $\mathbb{P}(Y = 1|X = x) = l_{\theta^*}(x)$. Verwenden Sie *glm* um $l_{\hat{\theta}}$ zu berechnen. Wiederholen Sie den obigen Vorgang $R = 1000$ mal und plotten Sie die erhaltenen R Paare $(\hat{\theta}_1^j, \hat{\theta}_0^j)$, $j \in 1, \dots, R$, gemeinsam mit (θ_1^*, θ_0^*) . Vergleichen Sie die erhaltenen Scatterplots für $n = 20$, $n = 200$ und $n = 2000$ - was ist zu sehen?

Übungsaufgabe 63. Importieren Sie den sog. Titanic Datensatz (abrufbar unter <https://github.com/datasciencedojo/datasets/blob/master/titanic.csv>) in R. Finden Sie mittels (univariater) logistischer Regression heraus, ob das Feature ‘age’ einen Einfluss auf die Überlebenswahrscheinlichkeit hat. Wie sieht es mit der Variable ‘fare’ aus?

^{vii}aber nicht notwendigerweise lineare oder polynomiale