Institut für Angewandte Mathematik
Sommersemester 2024
Andreas Eberle, Stefan Oberdörster

4. Übungsblatt „Einführung in die Statistik"

Abgabe bis Dienstag 7.5, 13 Uhr, bei Ihrem Tutor.

Ab jetzt werden wir R *als „Taschenrechner" verwenden. Installieren Sie RStudio auf Ihrem Computer, und machen Sie sich damit vertraut, wie man statistische Datensätze darstellt (z.B. in einer Excel-Datei), in* R *einliest (z.B. als Datei-Import oder über die Zwischenablage), modifiziert (z.B. mit* fix()*), und wie man auf einzelne Variablen zugreift und diese graphisch darstellt. Literatur siehe Vorlesungshomepage, zum Beispiel Kapitel 2 im Skript „Datenanalyse mit* R*" von T. Petzoldt.*

---

1. (Dualität von Konfidenzbereichen und Hypothesentests)
Sei $(\Omega, \mathcal{A}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $X : \Omega \to S$ die Stichprobe, und $g : \Theta \to \mathbb{R}$.

a) Zeigen Sie: Ist $C(x)$ ein Konfidenzbereich für $g(\vartheta)$ zum Niveau $1-\alpha$, dann ist

$$A(m_0) \;=\; \{x \in S : m_0 \in C(x)\}$$

für $m_0 \in \mathbb{R}$ der Akzeptanzbereich eines Tests zum Niveau α der Hypothese $g(\theta) = m_0$.

b) Umgekehrt sei $A(m_0)$ für jedes m_0 der Akzeptanzbereich eines Tests der Hypothese $g(\theta) = m_0$ zum Signifikanzniveau α. Konstruieren Sie einen Konfidenzbereich für $g(\theta)$ zum Niveau $1-\alpha$.

c) Illustrieren Sie die Aussagen anhand von Beispielen.

2. (Parameterschätzung im Poisson-Modell) Seien $x_1, x_2, \ldots, x_n$ unabhängige Stichproben von der Poisson-Verteilung mit unbekannter Intensität λ. Zum Beispiel beschreibt x_i die Anzahl der Schadensfälle, die bei einer Versicherung am Tag i gemeldet werden.

a) Geben Sie ein statistisches Modell an, und zeigen Sie, dass die Statistik $S = X_1 + X_2 + \ldots + X_n$ suffizient für λ ist.

b) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\lambda}$.

c) Zeigen Sie, dass n unabhängige Stichproben von Poisson(λ) auch mit dem folgenden Zweistufenmodell erzeugt werden können:

   – Ziehe eine Stichprobe s von der Poisson-Verteilung mit Parameter $n\lambda$.
   – Gegeben s, ziehe $(x_1, \ldots, x_n) \sim \text{Mult}(s, \mathbf{p})$, wobei $\mathbf{p} = (p_k)_{k=1,\ldots,n}$ mit $p_k = 1/n$ die Gleichverteilung auf der der Menge $\{1, \ldots, n\}$ ist.

3. (Minimale Suffizienz) Eine Statistik $T(X)$ heißt *minimal suffizient*, falls sie suffizient ist, und sich als Funktion von jeder anderen suffizienten Statistik darstellen lässt. Für ein reguläres statistisches Modell definieren wir eine Äquivalenzrelation auf dem Stichprobenraum S durch $x \sim y$ genau dann, wenn die Likelihood-Funktionen $\theta \mapsto L(\theta; x)$ und $\theta \mapsto L(\theta; y)$ zueinander proportional sind. Zeigen Sie:

a) Eine Statistik $T(X)$ ist genau dann suffizient, wenn für alle $x, y \in S$ gilt:

$$T(x) = T(y) \qquad \Longrightarrow \qquad x \sim y\,.$$

b) $T(X)$ ist genau dann minimal suffizient, wenn für alle $x, y \in S$ gilt:

$$T(x) = T(y) \qquad \Longleftrightarrow \qquad x \sim y\,.$$

c) Welche der folgenden Statistiken im Gaußschen Produktmodell sind suffizient bzw. minimal suffizient?

$$T_1(X) = (X_1, \ldots, X_n), \quad T_2(X) = (\bar{X}_n, V_n^*), \quad T_3(X) = \bar{X}_n, \quad T_4(X) = (\bar{X}_n, V_n, X_1).$$

4. (Beispiele zu Konfidenzbereichen für einen Binomialparameter) Definieren Sie für die folgenden Anwendungssituationen jeweils einen geeigneten Wahrscheinlichkeitsparameter p und überlegen Sie, ob hierfür eine untere Konfidenzschranke, eine obere Konfidenzschranke oder ein Konfidenzintervall besonders geeignet wäre. Berechnen Sie jeweils approximative und exakte Konfidenzbereiche zum Niveau $1 - \alpha$ mit $\alpha = 5\%$.

a) Wie verbreitet ist Flugangst? Anlässlich eines spektakulären "Fluchtversuches" eines Flugpassagiers kurz vor dem Start äußerten sich 240 Schweizer zu der Frage, ob sie unter Flugangst leiden. Ergebnis: 41 Personen antworteten mit 'ja'.

b) Ein Anbieter eines WLAN-Routers möchte untermauern, dass die meisten Kunden mit der neuen Installationssoftware gut zurechtkommen. Zu diesem Zweck recherchiert er über sein Callcenter, wie viele von 1400 Neukunden die Service-Hotline in Anspruch nahmen. Ergebnis: 11 Kunden ließen sich wegen Problemen bei der Installation beraten.

c) Möchte die Mehrheit der Wahlberechtigten gerne per Internet abstimmen? Man fragte 88 Personen, ob sie den Gang zur Urne, eine Briefwahl oder eine Online-Wahl bevorzugen würden. Ergebnis: 60 Personen bevorzugten die Online-Wahl.

Hinweis zu R*: Mit den Befehlen*

$$\mathsf{binom.test}(\mathsf{x} = H, \mathsf{n} = n), \quad \mathsf{binom.test}(\mathsf{x} = H, \mathsf{n} = n, \mathsf{conf.level} = 1 - \alpha)$$

erhalten Sie das $(1-\alpha)$-Konfidenzintervall $[a_{\alpha/2}(H), b_{\alpha/2}(H)]$ für den Parameter p im Modell $H \sim Bin(n, p)$. Mit

$$\mathsf{binom.test}(\ldots, \mathsf{alternative} = '\mathsf{less}'), \quad \mathsf{binom.test}(\ldots, \mathsf{alternative} = '\mathsf{greater}')$$

erhalten Sie das Intervall $[0, b_\alpha(H)]$ bzw. $[a_\alpha(H), 1]$.