Institut für Angewandte Mathematik
Sommersemester 2022
Andreas Eberle, Luis Aniello La Rocca

5. Übungsblatt „Einführung in die Statistik"

Abgabe bis Montag 9.5., 16 Uhr.

---

1. (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 $(x_1, \ldots, x_n) \sim \mathrm{Mult}(s, \mathbf{p})$, wobei $\mathbf{p} = (p_k)_{k=1,\ldots,n}$ mit $p_k = 1/n$ die Gleichverteilung auf der der Menge $\{1, \ldots, n\}$ ist.

2. (Beispiele für Maximum-Likelihood-Schätzer)

a) Im *Hardy-Weinberg-Gleichgewicht* der Populationsgenetik treten in einer Population die Genotypen aa, aA und AA mit den relativen Häufigkeiten

$$p_{aa} = \theta^2, \quad p_{aA} = 2\theta(1-\theta), \quad p_{AA} = (1-\theta)^2, \qquad \theta \in [0,1]$$

auf. In einer Stichprobe der Größe n aus der Population beobachten wir h_t Individuen von Typ t, wobei $t \in \{aa, aA, AA\}$ und $h_{aa} + h_{aA} + h_{AA} = n$. Formulieren Sie ein statistisches Modell, und bestimmen Sie den Maximum-Likelihood-Schätzer für θ.

b) Seien $x_1, \ldots, x_n$ unabhängige Stichproben von der Doppelexponentialverteilung mit Dichte

$$f_\theta(x) \;=\; \frac{1}{2}\, e^{-|x-\theta|}\,.$$

Bestimmen Sie einen Maximum-Likelihood-Schätzer für θ. Ist dieser eindeutig?

3. (Konfidenzschranken für Binomialparameter)

a) Um zu klären, ob bei Neugeborenen die relativen Anteile von Mädchen und Jungen unterschiedlich sind, wurden die Daten von $n = 429.440$ Neugeborenen ausgewertet. Darunter waren 221.023 Jungen. Berechnen Sie nun (mit Wilsons Methode oder exakt) ein 99%-Vertrauensintervall für die Wahrscheinlichkeit p, dass ein Neugeborenes ein Junge ist. Wie beantworten Sie die Ausgangsfrage?

b) Berechnen Sie mit den Daten aus Teil a) ein 99%-Vertrauensintervall für p, welches 0.5 enthält, und leiten Sie hieraus eine präzise obere Schranke für $|p - 0.5|$ ab.

c) Die Konfidenzschranken $a_\alpha(h)$ und $b_\alpha(h)$ für einen Binomialparameter p lassen sich in der Regel nur numerisch bestimmen. Ausnahmen sind die Schranken $a_\alpha(n)$ und $b_\alpha(0)$. Leiten Sie konkrete Ausdrücke hierfür ab.

d) Der Hersteller eines gewissen Gerätes geht davon aus, dass die Ausfallwahrscheinlichkeit dieses Gerätes bei einem bestimmten Belastungstest nahezu 0 ist. Um dies zu untermauern, plant er eine entsprechende Testserie mit n gleichartigen Geräten durchzuführen und die Zahl der Ausfälle h zu bestimmen. (Dabei geht er davon aus, dass $h = 0$.) Wie groß muss der Stichprobenumfang n sein, damit er im Falle von $h = 0$ mit einer Sicherheit von $1-\alpha$ behaupten kann, p sei kleiner als ein vorgegebener Wert $p_o > 0$?

4. (Stichprobenumfänge bei Schätzung eines Binomialparameters)
Bisher haben wir den Stichprobenumfang n als fest vorgegeben betrachtet. Mitunter kann man vor der Datenerhebung überlegen, wie groß die Stichprobe eigentlich sein sollte. Als Beispiel betrachten wir $H \sim \mathrm{Bin}(n, p)$ und das $(1 - \alpha)$-Vertrauensintervall für p nach der Wilson-Methode.

a) Beweisen Sie die Umformung aus der Vorlesung: Für $p, \hat{p} \in [0, 1]$ und $c > 0$ gilt

$$\hat{p} \ \le \ p + c\sqrt{p(1-p)} \qquad \text{bzw.} \qquad \hat{p} \ \ge \ p - c\sqrt{p(1-p)}$$

genau dann, wenn

$$p \ \ge \ \frac{\hat{p} + c^2/2 \ - \ c\sqrt{\hat{p}(1-\hat{p}) + c^2/4}}{1 + c^2} \ \ \text{bzw.} \ \ p \ \le \ \frac{\hat{p} + c^2/2 \ + \ c\sqrt{\hat{p}(1-\hat{p}) + c^2/4}}{1 + c^2}.$$

b) Wie groß muss der Stichprobenumfang sein, damit die Länge des Vertrauensintervalls garantiert kleiner oder gleich $\delta > 0$ ist? Zu welchem Ergebnis gelangen Sie für $\alpha = 0.05$ und $\delta = 0.03$?

c) Von zwei vorgegebenen Werten $0 < p_1 < p_2 < 1$ soll das Vertrauensintervall höchstens einen enthalten. Wie groß muss n sein, damit dies gewährleistet ist?

d) Für die FDP ist ein Wähleranteil von $p_1 = 5\%$ oder darunter wegen der 5%-Hürde verheerend, ein Wähleranteil von $p_2 = 10\%$ oder darüber ist schon ein Anlass zum Feiern. Wie groß muss der Stichprobenumfang sein, damit man mindestens einen dieser Fälle mit einer Sicherheit von ca. 95% ausschließen kann?