Institut für Angewandte Mathematik
Sommersemester 2024
Andreas Eberle, Stefan Oberdörster

3. Übungsblatt „Einführung in die Statistik"

Abgabe bis Dienstag 30.4 bei Ihrem Tutor.

---

1. (Ein Test auf Zufälligkeit) Daniela Düsentrieb hat einen neuen Zufallsgenerator entwickelt und möchte Ihnen diesen schmackhaft machen. Zur Illustration präsentiert sie Ihnen eine "rein zufällige" Sequenz $\omega \in \{0,1\}^{100}$ (zeilenweise zu lesen):

1	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	1	1	0	1	0	1	1	1
1	0	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	1	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0
0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0

Testen Sie die Nullhypothese, dass diese Sequenz rein zufällig erzeugt wurde, auf dem Niveau $\alpha = 5\%$, indem Sie für die Teststatistik $R(\omega) = \#\{i < n : \omega_i \neq \omega_{i+1}\}$ einen zweiseitigen p-Wert berechnen. Verwenden Sie dafür folgende Tabelle der Binomialverteilungsfunktion $F_{99,0.5}$ (auf vier Nachkommastellen gerundet):

x	35	36	37	38	39	40	41	42
$F_{99,0.5}(x)$	0.0023	0.0043	0.0077	0.0133	0.0219	0.0350	0.0537	0.0795

Beurteilen Sie auch folgende Sequenz ω, ohne eine andere Tabelle hinzuzuziehen:

1	0	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	0	0	1	1	0	0
1	0	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1	0	1
1	1	0	1	1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	0	0	0	0	1	1
1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	1	0	1	0	1	1

2. (Ein supereffizienter Schätzer) Es werden n Zufallszahlen $x_1, x_2, \dots, x_n$ aus dem reellen Intervall $(0, \theta)$ erzeugt, und Sie sollen den Wert von θ schätzen.

a) Geben Sie ein statistisches Modell an, dass das Experiment beschreibt.

b) Zeigen Sie, dass $\hat{\theta}_1 = 2\bar{X}_n$ ein erwartungstreuer Schätzer für θ mit Varianz $\frac{\theta^2}{3n}$ ist.

c) Zeigen Sie, dass $\hat{\theta}_2 = \max_{1 \le i \le n} X_i$ sowohl eine suffiziente Statistik als auch ein Maximum-Likelihood-Schätzer ist. Zeigen Sie weiter, dass dieser die Verteilungsfunktion $F(c) = c^n/\theta^n$ für $0 \le c \le \theta$ hat. Folgern Sie, dass $E_\theta[\hat{\theta}_2] = \frac{n}{n+1}\theta$.

d) Um einen erwartungstreuen Schätzer zu erhalten, betrachten wir schließlich den korrigierten Schätzer $\hat{\theta}_3 = \frac{n+1}{n}\hat{\theta}_2$. Zeigen Sie

$$\mathrm{MSE}_\theta(\hat{\theta}_3) \;=\; \frac{\theta^2}{n(n+2)} \;\in\; O(n^{-2}).$$

3. (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?

4. (Capture-Recapture-Methode) Ein Ökologe macht sich Sorgen, dass die Population einer bestimmten Schmetterlingsart in einem bestimmten Gebiet stark abgenommen hat. Um dies zu untermauern, führt er ein Capture-Recapture-Experiment mit $\ell = n = 50$ Schmetterlingen durch.

a) Formulieren Sie ein statistisches Modell, und bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{N}$ für die Gesamtzahl N aller Schmetterlinge. Ist für den Biologen eine untere oder eine obere Vertrauensschranke für N von Interesse?

b) Angenommen, er findet in der zweiten Runde $x = 5$ Tiere, die er in der ersten Runde markierte. Berechnen Sie $\hat{N}$, und bestimmen Sie eine 90%-Vertrauensschranke mit Hilfe der folgenden Tabelle mit Werten von $F_N(x) = F_{N,50,50}(x)$:

N	291	292	293	294	295	296	297	298	299
$F_N(5)$	.0973	.0993	.1014	.1035	.1056	.1077	.1098	.1120	.1141
N	991	992	993	994	995	996	997	998	999
$F_N(4)$	.8987	.8990	.8994	.8997	.9000	.9004	.9007	.9010	.9013

Da der Ökologe kein Statistiker ist, bittet er Sie, den Sachverhalt in ein oder zwei Sätzen prägnant zu formulieren.

c) Wie kann man in der folgenden Abbildung die untere bzw. die obere 90%-Vertrauensschranke ablesen?