

Wahrscheinlichkeit & Statistik

Serie 12

Abgabe bis Mittwoch (25.05.2022) um 10:15 Uhr

Diese Serie beschäftigt sich mit approximativen Konfidenzintervallen. Ausserdem geht es um Tests und die grundlegenden Definitionen in diesem Kontext.

Weitere Informationen und Instruktionen zur Abgabe unter
<https://metaphor.ethz.ch/x/2022/fs/401-0614-00L/>

Aufgabe 12.1 [Approximatives Konfidenzintervall I: Geometrische Verteilung]

Sei $\Theta = [1/2, 1]$. Wir betrachten die Modellfamilie $(\mathbb{P}_\theta)_{\theta \in \Theta}$, wobei X_1, \dots, X_n unter \mathbb{P}_θ unabhängig, identisch verteilt sind mit $X_1 \sim \text{Geom}(\theta)$. In Quiz 10 haben wir bereits gesehen, dass der Maximum-Likelihood Schätzer für θ gegeben ist durch

$$T_{ML} = \frac{n}{\sum_{i=1}^n X_i}.$$

Bestimme ein approximative Konfidenzintervall für θ mit Niveau 95%.

Hinweis: Für alle $\theta \in [1/2, 1]$ gilt $\frac{\sqrt{1-\theta}}{\theta} \leq \sqrt{2}$. Weiterhin wurde in den Serien 5 und 6 berechnet, dass $\mathbb{E}_\theta[X_1] = 1/\theta$ und $\text{Var}_\theta[X_1] = (1-\theta)/\theta^2$.

Aufgabe 12.2 [Approximatives Konfidenzintervall II: Binomialverteilung]

Um die Anzahl N der Forellen in einem See zu bestimmen, wird folgendermassen vorgegangen (Capture-Recapture-Methode): In einem ersten Schritt werden 500 Forellen gefangen, markiert und wieder ausgesetzt. In einem zweiten Schritt werden nochmals 200 Forellen gefangen und die Anzahl X der markierten Forellen bestimmt.

- Für X wird oft eine Binomialverteilung angenommen, $X \sim \text{Bin}(n, \theta)$ (θ bezeichne die Wahrscheinlichkeit, dass ein im zweiten Schritt gefangener Fisch markiert ist).
Wie gross ist n ? Wie gross ist der Parameter θ , wenn die Gesamtzahl der Forellen im See $N = 2000$ bzw. $N = 5000$ ist?
- Die tatsächliche Beobachtung für X ergibt den Wert 40. Gebe eine vernünftige Schätzung für den Parameter θ an, und leite daraus eine Schätzung für die Gesamtzahl N der Forellen im See ab.
- Bestimme ein approximatives Konfidenzintervall für θ mit Level 95% und daraus ein approximatives Konfidenzintervall für N mit Level 95%.

Hinweis: Benutze für den Schätzer von θ den zentralen Grenzwertsatz.

Aufgabe 12.3 [Test I: Gezinkter Würfel]

Bei einem Würfel mit sechs Seiten soll getestet werden, ob der Würfel gezinkt ist und eher auf der Sechs landet. Hierzu wird ein Experiment durchgeführt, bei dem zehn Mal gewürfelt und jeweils die beobachtete Augenzahl notiert wird. Wir gehen davon aus, dass alle Würfe unabhängig voneinander sind und die Wahrscheinlichkeit, eine 1, 2, 3, 4 oder 5 zu würfeln, gleich ist. Wir modellieren die Ausgänge der Würfe als eine Stichprobe X_1, \dots, X_{10} , wobei $X_i = 1$ bedeutet, dass der i -te Wurf eine Sechs ist, und $X_i = 0$ sonst. Wir erhalten folgende Resultate:

| x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |

- Bestimme ein geeignetes Modell $(\mathbb{P}_\theta)_{\theta \in \Theta}$, d.h. einen Parameterraum und die Verteilungen von X_1, \dots, X_{10} unter jedem \mathbb{P}_θ .
- Bestimme eine geeignete Hypothese H_0 und Alternative H_A .
- Sei $T = \sum_{i=1}^{10} X_i$ die Teststatistik. Welcher Verteilung folgt T ?
- Sei $K = (4, 10]$ der Verwerfungsbereich. Berechne die Wahrscheinlichkeit für einen Fehler 1. Art.
- Beschreibe den Testentscheid basierend auf den obigen Resultaten.

Aufgabe 12.4 [Test II: Normalverteilung]

Seien X_1, \dots, X_{12} unabhängig und je $\mathcal{N}(\mu, \sigma^2)$ -verteilt unter \mathbb{P}_θ , wobei $\theta = \mu$ ein unbekannter Parameter ist. Die Standardabweichung $\sigma = 0.0499$ ist bekannt. Wir haben folgende Daten für die Stichprobe gegeben:

| x_1 | x_2 | x_3 | x_4 | x_5 | x_6 |
|---------|---------|---------|----------|----------|----------|
| 1.00781 | 1.00646 | 1.00801 | 1.00833 | 1.00738 | 1.00687 |
| x_7 | x_8 | x_9 | x_{10} | x_{11} | x_{12} |
| 1.00783 | 1.00936 | 1.00564 | 1.00543 | 1.00794 | 1.01060 |

Wir testen die Hypothese $H_0: \mu = \mu_0 = 1.0085$ gegen die Alternative $H_1: \mu \neq \mu_0$.

- Bestimme a und b so, dass die Teststatistik $T := \frac{\sum_{i=1}^{12} X_i + b}{a} \sim \mathcal{N}(0, 1)$ -verteilt ist unter \mathbb{P}_{μ_0} .
- Wähle $K := (-\infty, -c_\neq) \cup (c_\neq, \infty)$ als Verwerfungsbereich für ein zu bestimmendes $c_\neq \geq 0$. Teste H_0 gegen H_A für das Signifikanzniveau 5%.
- Berechne die Macht des oben durchgeführten Tests an der Stelle $\mu = 1.008$.