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

6. Übungsblatt „Einführung in die Statistik"

Abgabe bis Montag 16.5., 16 Uhr.

---

1. (Beispiele zu Hypothesentests)

a) Um zu überprüfen, ob die Kommunikation mit einem Satelliten funktioniert, wird von der Erde aus ein Signal gesendet. Empfängt der Satellit das Signal, antwortet er, indem ein Signal der Intensiät m für n Sekunden ausgestrahlt wird. Wegen des allgemeinen Hintergrundrauschens schwanken die auf der Erde empfangenen Signale zufällig, unabhängig davon, ob der Satellit sendet oder nicht. Für jede der n Sekunden wird die mittlere Spannung des auf der Erde empfangenen Signals aufgezeichnet. Sei x_i die mittlere Spannung des in der i-ten Sekunde empfangenen Signals abzüglich der erwarteten mittleren Spannung aufgrund des Hintergrundrauschens. Wir nehmen an, dass die Messwerte x_i Realisierungen von unabhängigen Zufallsvariablen X_i mit Verteilung $\mathrm{N}(\theta, v)$ sind, wobei $\theta = m$ ist, falls der Satellit funktioniert, und $\theta = 0$ sonst. Sie möchten nun eines von zwei Systemen kaufen. Das *signal-to-noise-ratio* $m/\sqrt{v}$ beträgt 2 für das erste System, und 1 für das zweite. Das erste System kostet 10^6€, das andere 10^5€. Eine Sekunde Übertragung mit einem der beiden Systeme kostet jeweils 10^3€. Unabhängig davon, welches System Sie kaufen, wollen Sie den Satelliten im Laufe eines Jahres 100 mal testen. Wenn Sie bei jedem Test eine ausreichende Anzahl von Sekunden benötigen, um sicherzustellen, dass alle Fehlerwahrscheinlichkeiten $\leq 0,05$ sind, welches System ist dann auf der Grundlage eines Jahresbetriebs billiger?

b) Wie schon in Aufgabe 2 a) von Blatt 5 betrachten wir erneut das *Hardy-Weinberg-Gleichgewicht* der Populationsgenetik mit unbekanntem Parameter $\theta \in [0, 1]$. Bestimmen Sie den mächtigsten Test für das Testproblem $H_0 \colon \theta = \theta_0$ versus $H_1 \colon \theta = \theta_1$.

2. (Ein gezinkter Würfel?) Ein Spieler beobachtet ein Spiel, bei dem ein einzelner Würfel wiederholt geworfen wird. Dabei gewinnt er den Eindruck, dass die 6 bei etwa 18% der Würfe und die 5 bei etwa 14% der Würfe fällt, während die anderen vier Augenzahlen gleich häufig auftreten. Nach der Aufforderung mitzuspielen bittet der Spieler darum, seine Hypothese zunächst testen zu dürfen, indem er den Würfel n-mal wirft.

a) Welche Teststatistik sollte er verwenden, wenn die einzige Alternative, die er in Betracht zieht, ist, dass der Würfel fair ist?

b) Zeigen Sie, dass für $n = 2$ der mächtigste Test zum Niveau 0,0196 die Nullhypothese genau dann verwirft, wenn zweimal die 5 fällt.

c) Ist $(N_1, \ldots, N_k)$ eine multinomialverteilte Zufallsvariable mit Parametern $n, \theta_1, \ldots, \theta_k$, dann hat die Summe $a_1 N_1 + \ldots + a_k N_k$ nach dem mehrdimensionalen zentralen Grenzwertsatz annähernd die Verteilung $N(nm, nv)$ mit $m = \sum_{i=1}^k a_i \theta_i$ und $v = \sum_{i=1}^k (a_i - m)^2 \theta_i$. Benutzen Sie dies, um den Schwellenwert des mächtigsten Tests zum Niveau α für obiges Problem näherungsweise zu berechnen.

3. (Öffnungszeiten eines Serviceschalters) Sei x_i die Anzahl der Ankünfte an einem Serviceschalter am i-ten von n Tagen. Ein mögliches Modell für diese Daten ist die Annahme, dass die Kunden nach einem homogenen *Poisson-Prozess* ankommen. In diesem Fall ist x_i eine Stichprobe einer Poisson-Verteilung mit dem Parameter θ, der erwarteten Anzahl der Ankünfte pro Tag. Angenommen, es lohnt sich nicht, den Serviceschalter offen zu halten, wenn $\theta \leq \theta_0$ ist.

a) Bestimmen Sie die gleichmäßig mächtigste Teststatistik für $H_0 \colon \theta \leq \theta_0$ vs. $H_1 \colon \theta > \theta_0$. Für welche Niveaus α gibt es einen gleichmäßig mächtigsten nichtrandomisierten Test?

b) Welche Verteilungstabelle würden Sie benötigen, um die Macht des gleichmäßg mächtigsten Tests zu berechnen?

c) Sie möchten sicherstellen, dass bei einer Ankunftsrate von $\theta \leq 10$ die Wahrscheinlichkeit, dass Sie sich für die Öffnung entscheiden, $\leq 0,01$ ist, und dass bei einer Ankunftsrate $\theta \geq 15$ die Wahrscheinlichkeit, dass Sie sich für die Schließung entscheiden, ebenfalls $\leq 0,01$ ist. Wie viele Tage müssen Sie beobachten, um sicherzustellen, dass der Test diese Bedingung erfüllt? *(Sie können die Normalapproximation verwenden.)*

4. (Einkommensverteilung) Seien $x_1, \ldots, x_n$ die Einkommen von n zufällig aus einer bestimmten Grundgesamtheit ausgewählten Personen. Ein mögliches Modell für die Einkommensverteilung ist die Pareto-Verteilung mit Parametern $\theta > 1$ und $c > 0$. Diese hat die Dichte

$$f(x, \theta) \;=\; c^\theta \theta x^{-(1+\theta)} \, \mathbb{1}_{x>c}.$$

a) Geben Sie das mittlere Einkommen m in Abhängigkeit von θ an.

b) Finden Sie die optimale Teststatistik für den Test $H_0 \colon m = m_0$ versus $H_1 \colon m > m_0$.

c) Verwenden Sie den zentralen Grenzwertsatz, um eine Normalapproximation für den Schwellenwert des Tests in Teil b) zu finden. *Hinweis: Verwenden Sie die Formel für den Mittelwert und die Varianz in exponentiellen Familien.*