Übungen zur Angewandten Stochastik

Blatt 2

1. Aufgabe **a)** Unter P_ϑ seien $X_1, \dots, X_n$ unabhängig, normalverteilt mit (festem) Mittelwert m und Varianz ϑ. Geben Sie einen besten erwartungstreuen Schätzer für ϑ an, was ist seine Varianz?

b) Seien S und T beides erwartungstreue beste Schätzer für ein Parametermerkmal $\tau(\vartheta)$ in einem einparametrigen statistischen Modell. Zeigen Sie, dass dann $P_\vartheta(S = T) = 1$ für alle ϑ gilt. (Hinweis: Betrachten Sie beispielsweise $(S + T)/2$, das ebenfalls ein erwartungstreuer Schätzer ist, und folgern Sie, dass der Korrelationskoeffizient von S und T unter P_ϑ gleich 1 ist).

2. Aufgabe (Randomized response) Um den Anteil $\vartheta \in [0, 1]$ von Personen in einer (großen) Gruppe, die eine „heikle“ Eigenschaft haben, zu schätzen, betrachten wir folgendes Verfahren: Als Teilnehmer erhalten Sie die Frage, beispielsweise „Haben Sie schon einmal Übungsaufgaben abgeschrieben?“ und zwei Sechser-Würfel. Sie würfeln verdeckt: Wenn die Augensumme 5 oder 6 ist, antworten Sie „ja“, wenn sie 8 oder 9 ist, antworten Sie „nein“, ansonsten beantworten Sie die Frage. Formulieren Sie ein statistisches Modell für die Befragung von n Personen, bestimmen Sie darin einen besten erwartungstreuen Schätzer für ϑ und dessen Varianz.

3. Aufgabe (Ewens'sche Stichprobenverteilung als exponentielle Familie) Die Ewens'sche Stichprobenverteilung beschreibt die Verteilung der Typhäufigkeiten in einer Stichprobe aus einer (idealisierten) biologischen Population unter dem sogenannten Infinitely-many-alleles-Mutationsmodell (in dem angenommen wird, dass jede Mutation auf einen völlig neuen Typ führt): Sei B_j die Anzahl Typen, die j-mal in einer Stichprobe der Größe n vertreten sind ($j = 1, \dots, n$), dann ist für $(b_1, \dots, b_n) \in \mathbb{N}_0^n$ mit $\sum_{j=1}^n j b_j = n$

$$\rho_{\vartheta,n}\big((b_1, \dots, b_n)\big) = P_{\vartheta,n}\big((B_1, \dots, B_n) = (b_1, \dots, b_n)\big) = \frac{n!}{\vartheta(\vartheta+1)\cdots(\vartheta+n-1)} \prod_{j=1}^n \frac{(\vartheta/j)^{b_j}}{b_j!},$$

wobei $\vartheta \in (0, \infty)$ der (skalierte) Mutationsparameter ist.

a) Für festes n ist $\{\rho_{\vartheta,n} : \vartheta \in (0, \infty)\}$ eine exponentielle Familie. Bestimmen Sie Erwartungswert und Varianz von $K_n := \sum_{j=1}^n B_j$, der Gesamtanzahl unterschiedlicher Typen in der Stichprobe, unter $P_{\vartheta,n}$, sowie die Fisher-Information $I(\vartheta)$.

b) Zeigen Sie: Für $\vartheta > 0$ gilt $\mathbb{V}_{\vartheta,n}(K_n)/\log n \to \vartheta$, und $K_n/\log n$ ist eine asymptotisch erwartungstreue und konsistente Schätzfolge für ϑ.

c) Wie lautet der Maximum-Likelihood-Schätzer für ϑ?

d)* Zeigen Sie (beispielsweise durch Induktion über n), dass $\rho_{\vartheta,n}$ die Verteilung der Farbenhäufigkeiten in der sogenannten Hoppe-Urne nach n Zügen beschreibt: In einer Urne befindet sich anfangs nur die „Mutationskugel“ mit Gewicht $\vartheta > 0$. In jedem Zug wird eine Kugel aus der Urne (mit Wahrscheinlichkeit proportional zu ihrem Gewicht) aus der Urne gezogen: wenn es die Mutationkugel ist, wird sie zusammen mit einer neuen Kugel vom Gewicht 1, die eine neue, noch nicht benutzte Farbe hat, zurückgelegt, andernfalls wird die gezogene Kugel zusammen mit einer neuen Kugel vom Gewicht 1 mit derselben Farbe zurückgelegt. Nach n Zügen enthält die Urne n farbige Kugeln (mit einer zufäligen Anzahl K_n unterschiedlicher Farben), und B_j verschiedene Farben kommen j-mal vor.
Können Sie aus dieser Darstellung folgern, dass $(K_n - \vartheta \log n)/\sqrt{\log n}$ mit $n \to \infty$ in Verteilung gegen $\mathcal{N}(0, \vartheta)$ konvergiert?

4. Aufgabe Schreiben Sie eine `R`-Routine, die bei gegebenem ϑ die Verteilung von K_n aus Aufgabe 3 simuliert. Prüfen Sie damit empirisch, beispielsweise mittels eines Normalplots (siehe die `R`-Hilfe zu `qqnorm`), wie gut die Normalapproximation aus Teil d) für verschiedene n, z.B. $n = 10, 50, 100, 1000$, paßt.