Übungen zur Vorlesung

Mathematik 3: Stochastik

Sommersemester 2023

Universität Bremen

Institut für Statistik

Prof. Dr. Thorsten Dickhaus
MZH, Raum 7250
E-Mail: dickhaus@uni-bremen.de
Tel.: 0421/218-63651

Blatt 9

Abgabe bis Dienstag, 27. Juni 2023, 23:59 Uhr
Jede komplett richtig gelöste Aufgabe ergibt 4 Punkte.

Aufgaben

33. **Chebyshev'sche Ungleichung.**
Angenommen, auf zwei Gefäße A und B mit je einem halben Liter Fassungsvermögen werden $0{,}26 \cdot 10^{23}$ Gas-Moleküle so (zufällig) verteilt, dass jedes Molekül unabhängig von den anderen mit der Wahrscheinlichkeit $1/2$ in A bzw. B gelangt.

    Schätzen Sie vermittels der Chebyshev'schen Ungleichung die Wahrscheinlichkeit dafür ab, dass in eines der beiden Gefäße mindestens $0{,}13 \cdot 10^{23} \cdot \left(1 + 10^{-8}\right)$ Moleküle gelangen.

34. **Statistische Modellierung.**
Formalisieren Sie (mathematisch exakt) statistische Modelle für die folgenden beiden Situationen.

    (a) Wir betrachten die Ziehung der Lottozahlen „6 aus 49“ und haben Zweifel daran (Unsicherheit darüber), dass jede Zahl die gleiche Wahrscheinlichkeit hat, gezogen zu werden. Daher beobachten wir alle 520 Samstagsziehungen des Jahrzehnts $2021 - 2030$ und führen eine Strichliste, welche Zahl wie oft gezogen wird.

    (b) Es soll die Abhängigkeit des Typ I-Diabetes Risikos beim Menschen vom Geschlecht untersucht werden. Dazu werden natürliche Zahlen n_1 und n_2 fest vorgegeben, n_1 zufällig ausgewählte Frauen und n_2 zufällig ausgewählte Männer aus einer vorgegebenen Grundgesamtheit (z. B. der Bevölkerung Bremens) einem oralen Glukose-Toleranztest (OGTT) unterzogen und das Ergebnis (pro Versuchsperson: Typ I-Diabetes ja / nein) protokolliert. Der Einfachheit halber nehmen wir an, dass sich Typ I-Diabetes mit einem OGTT eineindeutig diagnostizieren lasse. Alle anderen Einflüsse (außer dem Geschlecht) auf das Typ I-Diabetes Risiko seien bei dieser Studie nicht von Interesse.

35. **Programmieraufgabe: Monte Carlo-Integration.**

    a) Berechnen Sie das zweite Moment

    $$\mathbb{E}[Y^2] = \int_0^\infty y^2 \exp(-y) dy \tag{1}$$

    einer standard-exponentialverteilten Zufallsvariable Y in `R` mit der `integrate`-Funktion. Benutzen Sie den Befehl `str`, um die Datentypen des Ergebnisses anzuzeigen und extrahieren Sie den Wert des Integrals mit Hilfe des Feldauswahloperators `$`. Stellen Sie den Integranden aus (1) auf dem Intervall $[0, 10]$ grafisch dar.

b) Berechnen Sie das Integral aus (1) näherungsweise mit Hilfe der sogenannten Monte-Carlo Integration, d. h., als empirisches zweites Moment für n generierte (Pseudo-) Zufallszahlen $(y_i)_{1 \leq i \leq n}$. Wählen Sie $n = 1.000, 10.000, 100.000$.

c) Zeigen Sie, dass

$$\int_0^\infty y^2 \exp(-y) dy = 2 \cdot \int_0^\infty y \exp(-y) dy$$

gilt und nutzen Sie dies, um das Integral durch $2n^{-1} \sum_{i=1}^n y_i$ näherungsweise zu berechnen. Vergleichen Sie die Ergebnisse der Methoden aus b) und c).

d) Schreiben Sie eine `R`-Funktion, die in B Simulationsläufen die beiden Methoden aus b) und c) jeweils für eine Pseudo-Stichprobe vom Umfang n durchführt und als Rückgabe die mittlere Abweichung vom wahren Wert für beide Methoden ausgibt. Welche Methode liefert genauere Ergebnisse?

Hinweis:

Bitte schicken Sie Ihren `R`-Code auch elektronisch an den für Sie zuständigen Übungsgruppenleiter.

36. **Multiple Select-Aufgabe.**
Betrachten Sie die folgenden Aussagen über empirische Verteilungen und empirische Verteilungsfunktionen. Dazu seien $Y_1 \ldots, Y_n$ reellwertige, stochastisch unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion F von Y_1.

a) Falls F stetig ist, so nimmt die empirische Verteilungsfunktion $\hat{F}_n$ von $Y_1, \ldots, Y_n$ mit Wahrscheinlichkeit 1 alle Werte $\{k/n : 0 \leq k \leq n\}$ an.

b) Für jede vorgegebene Zahl $t \in \mathbb{R}$ ist die Zufallsvariable $n\hat{F}_n(t)$ binomialverteilt mit Parametern n und $F(t)$.

c) Sei die Stichprobe $\mathbf{y} = (y_1, \ldots, y_n)^\top$ vorgegeben und sei X eine reellwertige Zufallsvariable, deren Verteilung die durch $\mathbf{y}$ induzierte empirische Verteilung ist. Sei ferner $g : \mathbb{R} \to \mathbb{R}$ eine vorgegebene Funktion. Dann ist der Erwartungswert von $g(X)$ gleich $n^{-1} \sum_{i=1}^n g(y_i)$.

d) Angenommen, $y_1, \ldots, y_n$ sind paarweise verschieden. Dann ist der Median der durch $y_1, \ldots, y_n$ induzierten empirischen Verteilung genau dann eindeutig bestimmt, wenn der Stichprobenumfang n eine ungerade Zahl ist.

Ermitteln Sie die richtige Kombination korrekter Aussagen.

Hinweise:

Das korrekte Ermitteln des Wahrheitsgehaltes der Aussagen ergibt jeweils einen Punkt. Um Raten nicht zu belohnen, werden nur begründete Antworten gewertet.