Michael Baake
Neil Mañibo

Fakultät für Mathematik
Universität Bielefeld

Sommersemester 2024
Bielefeld, 27.06.2024

Wahrscheinlichkeitstheorie und Statistik

Sommersemester 2024

Übungsblatt 12

(39) Die *Gumbel-Verteilung* ist die vielleicht wichtigste Verteilung der Bioinformatik: Sie spielt eine zentrale Rolle beim Sequenzalignment mit dem Algorithmus BLAST. Sie hat die Verteilungsfunktion

$$F(y) = \exp(-C\,\mathrm{e}^{-\vartheta y}), \quad y \in \mathbb{R},$$

mit Parametern $C, \vartheta > 0$.

(a) Weisen Sie nach, dass F tatsächlich eine Verteilungsfunktion ist, dass also

$$\lim_{y\to-\infty} F(y) = 0, \quad \lim_{y\to+\infty} F(y) = 1 \quad \text{und} \quad F \text{ monoton steigend für alle } y \in \mathbb{R}.$$

Hinweis: Erst nachdenken!

(b) Berechnen Sie die zugehörige Dichtefunktion.

(c) Betrachten Sie eine Zufallsvariable Y mit der oben definierten Verteilungsfunktion (also $\mathbb{P}(Y \leqslant y) = F(y)$) und berechnen Sie das zugehörige q-Quantil, also den Wert y_q, für den gilt

$$\mathbb{P}(Y \leqslant y_q) = q.$$

Warum ist dieser Wert eindeutig?

(d) Berechnen Sie nun das Intervall $[a, b]$ mit $a, b \in \mathbb{R}$ und $a < b$, so dass für ein vorgegebenes $c \in (0, 1)$ gilt

$$\mathbb{P}(Y \in [a, b]) = c \quad \text{sowie} \quad \mathbb{P}(Y < a) = \mathbb{P}(Y > b).$$

(1+1+2+2 Punkte)

(40) In zwei sehr großen Populationen soll jeweils der Anteil der Individuen mit Merkmal M untersucht werden; dieser (unbekannte) Anteil sei p_1 in Population 1 und p_2 in Population 2. Es wird erwartet, dass p_1 und p_2 beide etwa gleich $\frac{1}{2}$ sind.

Aus jeder Population wird eine Stichprobe der Größe n genommen; Schätzwert $\widehat{P}_i$ für p_i ist $\frac{N_i}{n}$, $i = 1, 2$, wobei N_i die Anzahl der Individuen mit Merkmal M in der Stichprobe aus Population i ist.

Geben Sie $\mathbb{V}(\widehat{P}_1 - \widehat{P}_2)$ an. Wie groß muss n mindestens sein, damit die Standardabweichung von $\widehat{P}_1 - \widehat{P}_2$ höchstens 0.02 beträgt? (Für letztere Überlegung dürfen Sie $p_1, p_2 \approx 0.5$ benutzen.)

(2 Punkte)

(41) Es ist schwierig, in Umfragen zuverlässige Antworten auf sensible Fragen zu erhalten, wie etwa „Haben Sie jemals in einer Klausur geschummelt?“. Diesem Problem kann man folgendermaßen begegnen:

Die befragte Person macht ein Zufallsexperiment mit zwei möglichen Ausgängen A und B (z.B. Ziehen einer Kugel aus einer Urne mit schwarzen und weißen Kugeln), wobei die Wahrscheinlichkeit p für Ausgang A *bekannt* ist, die fragende Person aber den jeweiligen Ausgang nicht sehen kann. Bei Ausgang A beantwortet die Person die Frage „Haben Sie jemals in einer Klausur geschummelt?“; bei Ausgang B beantwortet Sie die Frage „Haben Sie noch nie in einer Klausur geschummelt?“, jeweils wahrheitsgemäß mit „Ja“ oder „Nein“. (Zugrunde liegt die Erfahrung, dass es Menschen leichter fällt, wahrheitsgemäß zu antworten, wenn die fragende Person nicht weiß, welche Version der Frage beantwortet wird.)

Sei q die (*unbekannte*) Wahrscheinlichkeit dafür, dass eine Person schon einmal geschummelt hat.

(a) Geben Sie die Wahrscheinlichkeit r für die Antwort „Ja“ an (in Abhängigkeit von p und q).

(b) Sei R der zufällige Anteil an Probanden in einer Stichprobe, die „Ja“ antwortet. Es gilt $\mathbb{E}(R) = r$ (warum?).

Überlegen Sie sich (auf Basis von (a)) einen Schätzer Q für q und zeigen Sie, dass dieser erwartungstreu ist.

(c) Wie lauten $\mathbb{V}(R)$ und $\mathbb{V}(Q)$?

(1+1$\frac{1}{2}$+1$\frac{1}{2}$ Punkte)

Abgabe bis Donnerstag, 04.07.24, 12 Uhr, im Postfach des Tutors.