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

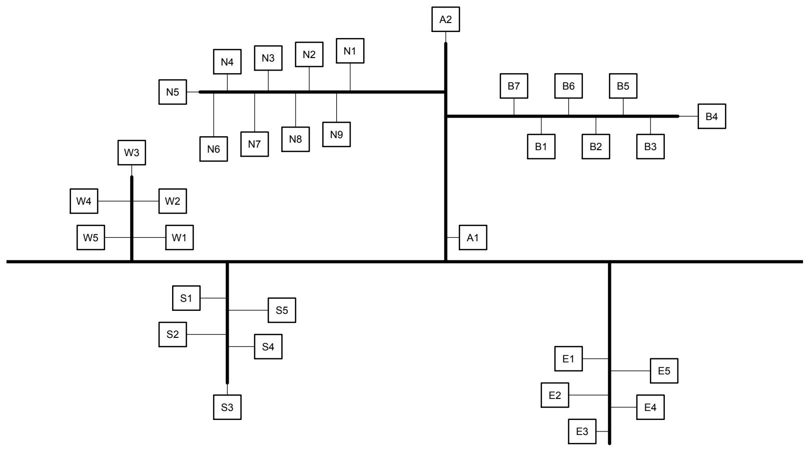
8. Übungsblatt „Einführung in die Statistik"

Abgabe bis Freitag 3.6., 16 Uhr.

---

1. (Zwei Optimierungsprobleme)

a) In einem Dorf mit 33 Anwesen soll ein Briefkasten so aufgestellt werden, dass die Summe aller Entfernungen von einem Haus zum Briefkasten minimal wird. Gemeint ist hierbei die Entfernung entlang der Straßen wie im Plan, der in der Abbildung unten gezeigt wird. Zeigen Sie, dass es genau eine optimale Position für den Briefkasten gibt. (Hierzu muss man keine Entfernungen ausmessen!)



b) Den Bäckermeister von Schilda kostet die Herstellung eines Hefezopfes einen Betrag $h > 0$, und er bietet ihn für den Betrag $v > h$ zum Verkauf an. Nach seinen Erfahrungen in der Vergangenheit geht er davon aus, dass die Nachfrage X nach Hefezöpfen (Anzahl potenziell verkaufter Zöpfe) am kommenden Samstag eine bestimmte Verteilung P auf $\mathbb{N}_0$ hat. Die Frage ist nun, wie viele Zöpfe er backen sollte, damit sein erwarteter Nettogewinn möglichst hoch ist. (Schildbürger sind übrigens 'schnäderfrässig' und kaufen nur frische Hefezöpfe.) Das Ergebnis hängt von der Verteilung von X und dem Quotienten h/v ab.

Zu welchem Ergebnis kommen Sie im Falle von $P = \text{Poiss}(40)$ und $h/v = 0.3$?

2. (Anpassungstests) Die folgende Tabelle enthält die Anzahl von Todesfällen in den USA in den 12 Monaten des Jahres 1966:

Januar	166 761	Mai	156 455	September	141 164
Februar	151 296	Juni	149 251	Oktober	154 777
März	164 804	Juli	159 924	November	150 678
April	158 973	August	145 184	Dezember	163 882

Die Frage ist nun, ob die Todesfallrate eines Monats proportional zu seiner zeitlichen Länge ist. Man kann begründen, dass sich die Sterbemonate X_1, X_2, ..., X_N der im Jahre 1966 verstorbenen US-Amerikaner nach Bedingen auf N wie unabhängige und identisch verteilte Zufallsvariablen verhalten, und wir interessieren uns für die unbekannten Wahrscheinlichkeiten $p_j = \mathbb{P}(X_i = \text{Monat Nr. } j)$. Formulieren und überprüfen Sie eine Nullhypothese mit den beiden in der Vorlesung beschriebenen Methoden, also dem χ^2-Anpassungstest auf dem Niveau $\alpha = 0.01$ bzw. simultanen 99%-Konfidenzintervallen für die p_j. Wie interpretieren Sie die Ergebnisse?

3. (Monte-Carlo-p-Werte) Für eine Teststatistik $t = T(\text{Daten})$ betrachten wir den p-Wert

$$\pi \;:=\; 1 - F_0(t-)$$

für eine gegebene Verteilungsfunktion F_0 sowie den Monte-Carlo-p-Wert

$$\hat{\pi} \;:=\; \frac{\#\bigl\{s \in \{1,\ldots,m\} : T_s \geq t\bigr\} + 1}{m+1}.$$

Dabei sind $T_1, T_2, \ldots, T_m$ untereinander (und von den Daten) unabhängige, nach F_0 verteilte Zufallsvariablen. Nun vergleichen wir π und $\hat{\pi}$ bei gegebenen Daten, berücksichtigen also nur den Zufall in den (simulierten) Variablen $T_1, \ldots, T_m$. Zeigen Sie, dass

$$\mathbb{E}\left[(\hat{\pi} - \pi)^2\right] \;\leq\; \frac{1}{4m+1} \quad \text{falls } m \geq 2.$$

4. (Eine Ungleichung für Student-Quantile) Wir betrachten unabhängige Zufallsvariablen Z und Y, wobei Z standardnormalverteilt ist, und $Y > 0$ mit $\mathbb{E}(Y) = 1$.

a) Zeigen Sie, dass
$$\mathbb{P}\left(Z/\sqrt{Y} > t\right) \;=\; \mathbb{E}\Phi\left(-t\sqrt{Y}\right)$$
für beliebige $t \in \mathbb{R}$. *(Verwenden Sie hierfür den Satz von Fubini, oder betrachten Sie nur den Spezialfall, dass Y eine diskrete Verteilung hat.)*

b) Angenommen, $\mathbb{P}(Y \neq 1) > 0$. Zeigen Sie, dass
$$\mathbb{P}\left(Z/\sqrt{Y} > t\right) \;>\; \Phi(-t)$$
für jedes $t > 0$. *(Betrachten Sie hierfür die Funktion $[0,\infty) \ni y \mapsto \Phi(-t\sqrt{y})$, und verwenden Sie die Jensensche Ungleichung.)*

c) Zeigen Sie nun, dass
$$t_{k;\beta} \;>\; \Phi^{-1}(\beta)$$
für beliebige $k \in \mathbb{N}$ und $\beta \in (1/2, 1)$.