Langrock, Vienken, Heine, Müller-Hanssen, Lorch Wintersemester 2023/24

Statistik I — Aufgabenzettel 3

(Abgabe: 22.01.2024)

Hinweise zur Abgabe dieses Aufgabenzettels

- es handelt sich um eine Gruppenabgabe, wobei Gruppen mit 2-3 Mitgliedern zulässig sind — **Einzelabgaben sind nicht mehr zulässig**
- Gruppen können tutorienübergreifend gebildet und bei späteren Aufgabenzetteln jederzeit neu zusammengesetzt werden
- jede Person kann in nur einer Gruppe mitwirken und jede Gruppe nur eine Lösung abgeben
- die Abgabe ist bis zum 22.01.2024, 23:59 Uhr über Moodle möglich und von nur *einer* (beliebigen) Person aus jeder Gruppe durchzuführen
- geben Sie in Ihrer Abgabedatei auf der ersten Seite der Lösungen klar und deutlich alle Mitglieder der Gruppe an (Vorname, Name, Matrikelnummer)
- wir empfehlen, dass die abgebende Person sowohl die finale Abgabedatei als auch einen Screenshot der Abgabe in Moodle an die anderen Gruppenmitglieder schickt
- Lösungen können u.a. aus folgenden Komponenten bestehen: R Code und R Output, Screenshots, abfotografierte oder eingescannte handschriftliche Lösungen, mit einem Editor wie MS Word erstellte Lösungen — wir sind hier flexibel, alles was lesbar ist wird akzeptiert
- die einzelnen Komponenten sind für die Abgabe jedoch zu einer *einzigen* **PDF Datei** zusammenzufügen — hierfür gibt es kostenlose Online-Tools, z.B. Smallpdf
- der Lösungsweg muss immer klar ersichtlich und die Lösung vollständig sein — sollten beispielsweise Grafiken zu erstellen sein, so sollten diese auch Teil Ihrer Abgabe sein
- wir nutzen das folgende Punkteschema (pro Aufgabe):
  - gar nichts gemacht $\rightsquigarrow$ 0 Punkte
  - sich an der Aufgabe versucht, aber sehr wenig richtig gemacht $\rightsquigarrow$ 1 Punkt
  - teilweise richtige Lösungen/Lösungsansätze vorgelegt $\rightsquigarrow$ 2 Punkte
  - die Aufgabe gut bearbeitet, mit kleinen Schönheitsfehlern $\rightsquigarrow$ 3 Punkte
  - die Aufgabe vollständig zufriedenstellend bearbeitet $\rightsquigarrow$ 4 Punkte

  (bewertet wird nicht kleinteilig jede Teilaufgabe, sondern nach Gesamteindruck)
- nach Ende der Abgabefrist werden Musterlösungen bei Moodle bereitgestellt, eine Besprechung in den Tutorien wird es nicht geben
- falls während der Bearbeitungszeit Fragen auftreten sollten, dann melden Sie sich jederzeit gerne über die Pinnwand (eine Antwort kommt in der Regel innerhalb einiger Stunden)

Aufgabe 1: Fußgängerzone

In der Bahnhofstraße in Bielefeld spricht ein ehrenamtliches Mitglied eines Tierschutzvereins Passant*innen an und versucht, sie zu einer Tierpatenschaft zu überreden. Die allermeisten Passant*innen sind in erster Linie genervt — im Mittel führt nur jeder 80. Versuch zum Erfolg, d.h. zu einer abgeschlossenen Patenschaft.

a) Welche Verteilung beschreibt X = Anzahl der angesprochenen Passant*innen, bis die erste Patenschaft abgeschlossen wird? Was ist demnach $E(X)$?

b) Wie hoch ist die Wahrscheinlichkeit, dass auch nach 200 angesprochenen Passant*innen immer noch keine Patenschaft erfolgreich abgeschlossen werden kann?

c) Angenommen, in ganz Deutschland sind 1000 solche ehrenamtliche Helfer*innen unterwegs, wobei jede Person für genau ein Tier zuständig ist und solange Passant*innen anspricht, bis dieses Tier einen Paten/eine Patin hat. Simulieren Sie diese Situation in R, indem Sie die Anzahl der angesprochenen Passant*innen jedes ehrenamtlichen Helfers/jeder ehrenamtlichen Helferin gemäß der Verteilung von X generieren. Geben Sie dann an:

   - den Mittelwert der Anzahl angesprochener Passant*innen pro Helfer*in;
   - den Anteil an Helfer*innen, die mehr als 200 Passant*innen ansprechen mussten.

   Die Ergebnisse können Sie mit jenen aus a) und b) vergleichen.

d) Wie viele Passant*innen muss ein Helfer/eine Helferin ansprechen, damit mit mindestens 80% Wahrscheinlichkeit eine Patenschaft zustandekommt?

Aufgabe 2: Eurostar

Im Eurostar von Brüssel nach London dürfen maximal so viele Personen mitfahren wie der Zug Sitzplätze hat, nämlich 894. Nun könnte der Betreiber einfach genau die 894 Tickets entsprechend der Kapazität der Züge verkaufen. Allerdings weiß der Betreiber basierend auf historischen Daten, dass lediglich 95% aller Ticketkäufer*innen ihre Reise auch tatsächlich antreten, so dass die Chance groß ist, dass viele Plätze leer bleiben.

a) Wir nehmen einmal an, dass der Betreiber einfach 935 Tickets verkauft, also mehr als er Sitzplätze anbieten kann. Wie hoch ist dann die Wahrscheinlichkeit, dass alle ihre Reise auch tatsächlich antretenden Ticketkäufer*innen letztlich in den Zug passen? Wir nehmen hier der Einfachheit halber an, dass die Passagiere unabhängig voneinander agieren.

b) Wie viele Tickets könnte der Betreiber maximal verkaufen, wenn die Vorgabe ist, dass mit mindestens 99.9% Wahrscheinlichkeit alle ihre Reise antretenden Ticketkäufer*innen ihre Reise wie geplant durchführen können, d.h. in den Zug passen?

c) Wir nehmen nun einmal an, dass jedes Ticket genau 90 Euro kostet. Falls ein Passagier seine Reise wegen Überbuchung nicht antreten kann, dann muss der Betreiber ihm 1000 Euro Entschädigung zahlen. Schreiben Sie ein R-Skript, welches

   - für eine gegebene Anzahl verkaufter Tickets n eine zufällige Realisation der Anzahl ihre Reise antretender Ticketkäufer*innen generiert und
   - den Gesamtgewinn des Betreibers für die zufällig realisierte Anzahl zu befördernder Passagiere ermittelt.

Schreiben Sie dann einen Loop, mit Hilfe dessen Sie 1 000 000 Fahrten und den resultierenden mittleren Gewinn des Betreibers pro Fahrt realisieren. Nutzen Sie schließlich Ihren Code, um per Ausprobieren zu ermitteln, was das kostenoptimale n ist (also das n, welches zu dem höchsten erwarteten Gewinn führt).

Aufgabe 3: Dichtefunktion & Verteilungsfunktion

In dieser Aufgabe betrachten wir die Dichtefunktion

$$f(x) = \begin{cases} \frac{3}{8}\left(1 - \frac{1}{4}x^2\right) & \text{falls } -2 \leq x \leq 2; \\ 0 & \text{sonst.} \end{cases}$$

a) Zeichnen Sie mit R die Dichtefunktion auf dem Intervall $[-3, 3]$, z.B. durch Nutzung des `curve(...)`–Befehls (vgl. Übungsblatt 9). Falls Sie Probleme mit den R-Befehlen haben sollten, so fragen Sie ruhig nach, wir können hier gerne helfen.

b) Leiten Sie nun schriftlich die zu $f(x)$ gehörige Verteilungsfunktion $F(x)$ her. Zeichnen Sie dann $F(x)$ in R (vgl. Übungsblatt 9).

c) Sei X nun eine stetige Zufallsvariable mit der obigen Dichte. Berechnen Sie $P\big(0 < X < 1\big)$:

   1. über $f(x)$, und zwar mit dem `integrate(...)`–Befehl in R;
   2. mit Hilfe der in b) bestimmten Verteilungsfunktion.

   Für 1. müssen Sie die Dichte zunächst als Funktion in R definieren (vgl. Übungsblatt 9).

Aufgabe 4: Erwartungswert, Varianz & Quantile

Wir betrachten weiterhin eine Zufallsvariable X mit der in Aufgabe 3 definierten Dichte.

a) Ermitteln Sie $E(X)$.

b) Bestimmen Sie die Varianz von X.

   Hinweis: Es kommt ein rundes Ergebnis heraus (nur eine Nachkommatastelle) — falls Sie also eine sehr krumme Zahl herausbekommen, so haben Sie sich verrechnet.

c) Ermitteln Sie das 0.8-Quantil der Verteilung von X.

   Hinweis: Es gibt verschiedene Lösungswege. Sie dürfen das Quantil insbesondere auch durch Ausprobieren in R ermitteln — volle Punktzahl gibt es, wenn die ersten zwei Nachkommastellen stimmen.

Aufgabe 5: Kaffee bei Hasan

Sei

$$X = \text{Wartezeit auf Bedienung bei Hasan (in Sekunden)} \sim Exp(0.016).$$

Zeichnen Sie wie folgt die Dichtefunktion von X in R:

```
curve(dexp(x,0.016),0,200,lwd=2,bty="l",ylim=c(0,0.018),
      xlab="Wartezeit in Sekunden",ylab="Dichte")
```

Bestimmen Sie dann $P(X < 50)$ auf drei verschiedenen Wegen:

a) durch eine grobe Abschätzung der zu berechnenden Fläche, analog zu Aufgabenteil c) von Übungsblatt 9;

b) durch exakte Berechnung per Integration;

c) durch Nutzung von `pexp()` in R.