

## Stochastik

### Serie 7

1. Die Randdichten eines zweidimensionalen Zufallsvektors  $Z = (X, Y)$  sind wie folgt definiert:

$$f_X(x) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases} \quad f_Y(y) = \begin{cases} \frac{3}{4}(2y - y^2) & 0 \leq y \leq 2 \\ 0 & \text{sonst} \end{cases}$$

Der Korrelationskoeffizient  $\rho_{X,Y}$  zwischen  $X$  und  $Y$  ist  $\rho_{X,Y} = \sqrt{\frac{1}{3}}$ .

- Berechne den Erwartungswert von  $6X - 4Y + 2$ .
  - Berechne die Kovarianz  $\text{Cov}(6X, 4Y)$ .
  - Berechne die Varianz von  $6X - 4Y + 2$ .
  - Berechne den Erwartungswert von  $6X^2 - 4Y^2$ .
2. Es seien  $\{X_i\}_{1 \leq i \leq 50}$  unabhängige und normalverteilte Zufallsvariablen mit Erwartungswert  $\mu = 1$  und Standardabweichung  $\sigma = 2$ . Darüber hinaus sind folgende Zufallsvariablen definiert:

$$S_n = X_1 + X_2 + \dots + X_n$$

und

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{S_n}{n}$$

Dabei ist  $n = 50$ .

- Bestimme die Parameter der Normalverteilung von  $S_n$  sowie  $\bar{X}_n$ .
- Berechne die Wahrscheinlichkeit  $P(E[X_1] - 1 \leq X_1 \leq E[X_1] + 1)$ .
- Berechne  $P(E[S_n] - 1 \leq S_n \leq E[S_n] + 1)$ .
- Berechne  $P(E[\bar{X}_n] - 1 \leq \bar{X}_n \leq E[\bar{X}_n] + 1)$ .

**Bitte wenden!**

3. Die erwartete Lebensdauer  $\mu$  eines Batterietyps ist unbekannt und soll durch das arithmetische Mittel der Lebensdauern von  $n$  unabhängigen Testbatterien dieses Typs geschätzt werden. Erfahrungsgemäss ist die Standardabweichung der Lebensdauer ungefähr 5 Stunden. Wie gross muss  $n$  mindestens sein, damit der Absolutbetrag der Differenz zwischen arithmetischem Mittel und  $\mu$  mit einer Wahrscheinlichkeit von mindestens 99% höchstens 1 Stunde beträgt?

**Tipp:** Zentraler Grenzwertsatz.

4. Die Krankenkasse easyHealth bietet einen sehr günstigen Studententarif für 160 Franken pro Jahr an. Sie hat pro Kunde einen jährlichen Administrationsaufwand von 100 Franken. Zusätzlich fallen die variierenden Gesundheitskosten (Arztrechnungen etc.) an. Zur Modellierung der *Gesamtkosten*  $X$  pro Kunde und Jahr benutzt die Krankenkasse die folgende Dichte

$$f(x) = \begin{cases} 0 & x \leq c \\ \frac{\alpha c^\alpha}{x^{\alpha+1}} & x > c \end{cases}$$

wobei  $c > 0, \alpha > 1$  Parameter sind.

easyHealth hat zur Zeit 250 studentische Kunden. Was ist (approximativ) die Wahrscheinlichkeit, dass easyHealth mit diesem Tarif in einem Jahr Gewinn macht, wenn wir annehmen, dass  $\alpha = 3, c = 100$ ?

**Tipp:** Die folgenden Formel können ohne Herleitung benutzt werden:

$$E[X] = c \frac{\alpha}{\alpha - 1}; \quad \text{Var}(X) = \frac{c^2 \alpha}{(\alpha - 1)^2 (\alpha - 2)}.$$

5. Auf [http://onlinestatbook.com/stat\\_sim/sampling\\_dist/index.html](http://onlinestatbook.com/stat_sim/sampling_dist/index.html) findest du ein Tool, das sehr hilfreich ist beim Verständnis des Zentralen Grenzwertsatzes (der Link ist auch auf der Homepage zur Vorlesung).

Gegeben seien i.i.d. Zufallsvariablen  $X_1, \dots, X_N$ . Wie üblich sei  $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$  das arithmetische Mittel.

**Eine kurze Anleitung:** Wähle 'Beginn' im Menü auf der linken Seite (um über weitere Funktionalitäten zu erfahren kannst du auch zuerst die 'Instructions' lesen).

- Der oberste Plot zeigt die Verteilung eines einzelnen  $X_i$  (bzw. ein Histogramm davon). Diese Verteilung ist zu Beginn auf 'Normal' eingestellt, kann aber in einem Dropdown-Menü gewählt oder per Mausklick auf den Plot manuell geändert werden.

**Siehe nächstes Blatt!**

- Per Default ist die Anzahl an Beobachtungen (die 'Sample Size') auf  $N = 5$  gestellt, im Dropdown-Menü weiter unten kann diese erhöht werden. Ein Klick auf 'Animated' generiert nun eine Beobachtung von  $X_1, \dots, X_N$  und zeichnet alle in den Plot 'Sample Data'. Dann wird der Mittelwert  $\bar{X}_N$  dieses Samples im Plot 'Distribution of Means' eingezeichnet.
- Ein weitere Klick auf 'Animated' wiederholt diesen Vorgang. Wiederhole dies einige Male um sicherzugehen, dass du den vorherigen Punkt verstanden hast. Ein Klick auf 5, 10'000 oder 100'000 wiederholt den Vorgang dann 5, 10'000 bzw. 100'000 Mal. Dabei werden aber die Beobachtungen nicht mehr im Plot 'Sample Data' eingezeichnet, sondern für jede Ziehung wird direkt das arithmetische Mittel der  $N$  Beobachtungen im Plot "Distribution of Means" eingezeichnet.
- Dieser Plot zeigt nun die empirische Verteilung von  $\bar{X}_N$ . Wenn die Sample Size  $N$  gross genug ist, sollte diese laut dem Zentralen Grenzwertsatz ungefähr wie eine Normalverteilung aussehen. Indem du ein Häkchen bei 'Fit normal' setzt, kannst du überprüfen wie gut dies stimmt.

Experimentiere nun mit diesem Tool um zu bestimmen wie gross die Sample Size  $N$  in etwa sein muss, damit die Verteilung von  $\bar{X}_N$  ungefähr wie eine Normalverteilung aussieht, wenn  $X_i$  folgende Verteilung hat:

- a) Normal
- b) Uniform
- c) 'Skewed' (siehe Dropdown Menü).

**Abgabe:** 11. oder 12. November.

**Präsenz:** Montag und Donnerstag, 12:00-13:00 Uhr im HG G 32.6.

**Homepage:** <https://metaphor.ethz.ch/x/2019/hs/401-0603-00L/>