ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik

Musterlösung 4

1. Die Zufallsvariable, die die Anzahl eingehender Telefonanrufe in einer Telefonzentrale innerhalb von 10 Minuten beschreibt, nennen wir X. Sie folge einer Poissonverteilung mit Erwartungswert $\lambda = 2$, d.h. $X \sim \text{Pois}(\lambda)$.

a) Die (sehr kleine) Telefonzentrale ist überlastet, wenn es in einer bestimmten 10-Minuten-Periode mehr als drei Telefonanrufe gibt. Wie gross ist die Wahrscheinlichkeit, dass sie in einer bestimmten 10-Minuten-Periode überlastet ist?

b) Wir nehmen an, dass die Anzahl Anrufe in einer 10-Minuten-Periode von der Anzahl Anrufe in einer anderen 10-Minuten-Periode unabhängig ist. Die Zufallsvariable, welche die Anzahl Anrufe in einer Stunde beschreibt, bezeichnen wir mit Y. Welcher Verteilung folgt Y?

Lösung:

a) Mit der Formel für Komplementärereignisse erhalten wir $P(X > 3) = 1 - P(X \leq 3) = 1 - 0.857 \approx 0.143$.

b) Die Anzahl Anrufe in einer Stunde lässt sich beschreiben als sechsmal die Anzahl Anrufe in zehn Minuten, da wir zwischen 10-Minuten-Perioden nicht unterscheiden (gleiche Verteilung). Die Anzahl Anrufe in diesen sechs 10-Minuten-Periode sind zudem laut Annahmen Poisson-verteilt und unabhängig. Da die Summe von unabhängigen Poisson-verteilten Zufallsvariablen wieder Poisson-verteilt ist mit der Summe der einzelnen Parameter, haben wir $Y \sim \text{Poisson}(6 \cdot \lambda) = \text{Poisson}(12)$.

2. Es wird so lange mit einem fairen Würfel gewürfelt, bis jede der Zahlen $1, ..., 6$ mindestens einmal erschienen ist.

a) Sei X_i die Anzahl Würfe bis die i-te verschiedene Zahl, $i = 1, ..., 6$ geworfen ist. Begründe, dass $X_1 = 1$.

b) Sei $Y_i = X_i - X_{i-1}$, $i = 2, ..., 6$. Beschreibe Y_i in Wörter. Was ist die Verteilung von Y_i?

Bitte wenden!

c) Wie gross ist der Erwartungswert der Anzahl der benötigten Würfe bis jede der Zahlen $1, ..., 6$ mindestens einmal erschienen ist?

Lösung:

a) X_1 ist die Anzahl Würfe bis die erste (verschiedene) Zahl gewürfelt wird. Das wird immer beim ersten Wurf passieren, somit ist $X_1 = 1$.

b) Für $i = 2, \ldots, 6$ beschreibt $Y_i := X_i - X_{i-1}$ die Anzahl der Würfe, welche es nach dem Erscheinen der $(i-1)$-ten verschiedenen Zahl braucht, bis die i-te Zahl erscheint. Die Wahrscheinlichkeit, dass nach dem Erscheinen der $(i-1)$-ten verschiedenen Zahl die i-te verschiedene Zahl erscheint, ist gegeben durch $p_i := (6-(i-1))/6$, das heisst $P[Y_i = 1] = p_i$. So folgern wir $P[Y_i = 2] = (1-p_i)p_i$, $P[X_i = 3] = (1-p_i)^2 p_i$ und so weiter. Die Y_i-s sind daher geometrisch verteilt mit Erfolgsparameter p_i.

c) Wir wissen dass die Erwartung von einer geometrisch verteilten Zufallsvariable mit Erfolgsparamter p ist $1/p$, und somit $E[Y_i] = 1/p_i$. Wegen der Linearität des Erwartungswertes gilt daher

$$E[X_6] = E[X_1] + E[Y_2] + \ldots + E[Y_6] \;=\; 1 + \frac{6}{5} + \frac{6}{4} + \ldots + \frac{6}{1} \;=\; 14.7.$$

3. In der Stadt Zürich gibt es bekanntlich viele Baustellen. Die Dauer X der Arbeiten bei einer Baustelle liege zwischen 0 und 20 Wochen. Die Dichte $f(x)$ habe die folgende Form.

$$f(x) = \begin{cases} c - \frac{c}{20}x & \text{falls } x \in [0, 20] \\ 0 & \text{sonst.} \end{cases}$$

a) Was ist c und warum?

b) Berechne die kumulative Verteilungsfunktion und skizziere diese.

c) Bestimme mittels der kumulativen Verteilungsfunktion die Wahrscheinlichkeit, dass die Bauzeit X *(i)* maximal 5 Wochen beträgt, *(ii)* zwischen 5 und 10 Wochen beträgt.

d) Berechne den Erwartungswert, den Median und die Standardabweichung der Dauer X.

e) Welche Dauer wird nur mit einer Wahrscheinlichkeit von 10% überschritten?

f) $K = 40'000 \cdot \sqrt{X}$ entspreche dem Betrag in Franken, den die Arbeiten bei einer Baustelle kosten. Wie gross ist die Wahrscheinlichkeit, dass die Arbeiten bei einer Baustelle höchstens $120'000$ Franken kosten?

Siehe nächstes Blatt!

Lösung:

a) Damit $f(x)$ eine Dichte ist, muss die Fläche unter der Funktionskurve (das Integral der Funktion) gleich 1 sein. Wir berechnen das Integral

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{20} c - \frac{c}{20}xdx = 20c - \frac{c}{20}\frac{20^2}{2} = 10c.$$

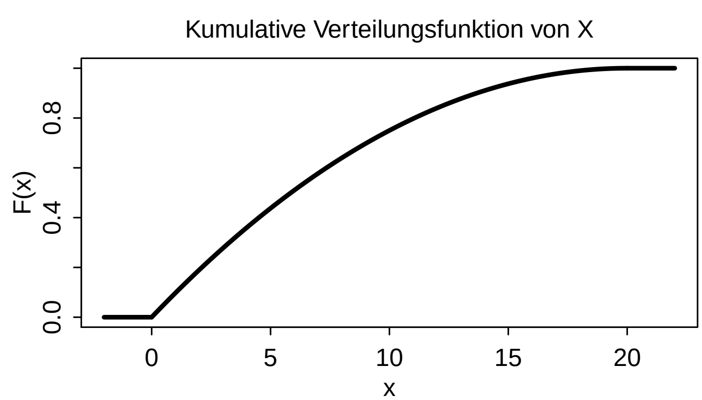
Es muss also gelten $10c = 1$, das heisst $c = \frac{1}{10}$.

b) Die kumulative Verteilungsfunktion von X lässt sich durch Integration der Dichtefunktion berechnen: Für $0 \leq x \leq 20$ gilt:

$$F(x) = P(X \leq x) = \int_{0}^{x} f(t)\,dt = \int_{0}^{x} \left(\frac{1}{10} - \frac{t}{200}\right)\,dt = \frac{x}{10} - \frac{x^2}{400}.$$

Für $x \leq 0$ ist $F(x) = 0$ und für $x \geq 20$ gilt $F(x) = 1$.

Skizze von $F(x)$:



c) Es gilt: $P[X \leq 5] = F(5) = 0.4375$ und, da die Verteilung stetig ist, $P[5 \leq X \leq 10] = P[5 < X \leq 10] = F(10) - F(5) = 0.3125$.

Bitte wenden!

d)

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{20} x \left[\frac{1}{10}\left(1-\frac{x}{20}\right)\right] dx = \frac{1}{10}\left(\frac{x^2}{2}-\frac{x^3}{60}\right)\Bigg|_0^{20} = \frac{20}{3} \\
Var(X) &= E(X^2) - (E(X))^2 \\
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{0}^{20} x^2 \frac{1}{10}\left(1-\frac{x}{20}\right) dx = \frac{1}{10}\left(\frac{x^3}{3}-\frac{x^4}{80}\right)\Bigg|_0^{20} = \frac{200}{3} \\
Var(X) &= E(X^2) - (E(X))^2 = \frac{200}{3} - \left(\frac{20}{3}\right)^2 = \frac{200}{9}
\end{aligned}
$$

Also ist die Standardabweichung $\sigma_X = \sqrt{Var(X)} = \sqrt{2} \cdot 10/3 \approx 4.714$.
Für den Median m muss gelten: $F(m) = 0.5$. Der Median liegt sicher im Intervall [0,20] und somit haben wir

$$\frac{m}{10} - \frac{m^2}{400} = 0.5 \qquad \Longrightarrow \qquad m = 20 - 10\sqrt{2} \approx 5.858.$$

e) Gesucht ist x, so dass $P(X > x) = 0.1$. Es gilt:

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x) = 0.1.$$

Ähnlich wie in Teilaufgabe d) ergibt sich:

$$F(x) = \frac{x}{10} - \frac{x^2}{400} = 0.9 \qquad \Longrightarrow \qquad x = 13.6754,$$

da x im Intervall [0,20] liegen muss.
Hinweis: Beim gesuchten x handelt es sich um das 90%-Quantil von X.

f)

$$
\begin{aligned}
P(K \leq 120'000) &= P(40'000 \cdot \sqrt{X} \leq 120'000) = P(\sqrt{X} \leq 3) \\
&= P(X \leq 9) = F(9) = \frac{9}{10} - \frac{9^2}{400} = 0.6975.
\end{aligned}
$$

4. Wir betrachten eine Messsonde an einem Vulkankrater, welche den bevorstehenden Ausbruch beobachten soll. Ab Beginn der Messungen gehen wir davon aus, dass die Sonde in jeder Sekunde mit einer Wahrscheinlichkeit von $p = 5\%$ wegen zu grosser Beschädigung ausfällt. Die Zufallsvariable Y bezeichne die Lebensdauer der Sonde in Sekunden.

a) Wie gross ist die Wahrscheinlichkeit, dass die Sonde mehr als 10 Sekunden überlebt?

Siehe nächstes Blatt!

b) Wir wissen, dass die Sonde schon mehr als 20 Sekunden überlebt hat. Wie gross ist dann die Wahrscheinlichkeit, dass die Sonde nochmals 10 Sekunden überlebt?

c) Diese Sonde sende nun jede Sekunde bis zum Ausfall ein Datenpaket an die Empfangsstation, wobei wir annehmen, dass jedes Paket unabhängig von den anderen mit einer Wahrscheinlichkeit von 90% empfangen wird. Die Zufallsvariable Z bezeichne die Anzahl der erhaltenen Datenpakete. Nehmen wir an, dass die Sonde nach 30 Sekunden ausfällt. Wie gross ist dann die Wahrscheinlichkeit, dass während dieser Zeit mehr als 27 Datenpakete erhalten werden?

Lösung:

a) Wir modellieren $Y \sim Geom(p)$ und daher gilt $P[Y > k] = (1-p)^k$, $k \in \mathbb{N}$. Somit gilt also

$$P[Y > 10] = (1 - 0.05)^{10} = 0.95^{10} \approx 59.9\%.$$

b) Gesucht ist $P[Y > 30|Y > 20]$. Nach Definition der bedingten Wahrscheinlichkeit gilt

$$\begin{aligned} P[Y > 30|Y > 20] &= \frac{P[Y > 30, Y > 20]}{P[Y > 20]} = \frac{P[Y > 30]}{P[Y > 20]} = \frac{(1-0.05)^{30}}{(1-0.05)^{20}} \\ &= (1-0.05)^{10} \\ &= 0.95^{10} \approx 59.9\%. \end{aligned}$$

Vergleichen wir mit Teilaufgabe **a)** so sehen wir dass $P[Y > 30|Y > 20] = P[Y > 10]$. Die Wahrscheinlichkeit, dass die Sonde 10 Sekunden überlebt ist gleich gross wie die Wahrscheinlichkeit, dass die Sonde, gegeben sie hat schon 20 Sekunden überlebt, weitere 10 Sekunden überlebt. D.h. die Information, dass die Sonde vorher schon 20 Sekunden überlebt hat, spielt keine Rolle. Die Vergangenheit wird "vergessen". Dies ist kein Zufall, sondern aufgrund der sogenannten "forgetfulness property", welche die geometrische Verteilung besitzt.

c) Die Lebensdauer der Sonde sei also 30 Sekunden. Während dieser Zeit erhält die Empfangsstation jede Sekunde mit einer Wahrscheinlichkeit von 90% ein Datenpaket. In jeder Sekunde haben wir ein Bernoulli-Experiment, d.h. die Station empfängt oder sie empfängt nicht. Demzufolge ist also die Anzahl Datenpakete in 30 Sekunden binomialverteilt, $Z \sim Bin(n = 30, p = 0.9)$ und gesucht ist $P[Z > 27|Y = 30]$. Somit gilt

$$\begin{aligned} P[Z > 27|Y = 30] &= P[Z = 28|Y = 30] + P[Z = 29|Y = 30] + P[Z = 30|Y = 30] \\ &= \binom{30}{28} p^{28}(1-p)^2 + \binom{30}{29} p^{29}(1-p) + \binom{30}{30} p^{30} \approx 0.411. \end{aligned}$$

Bitte wenden!

5. Eine Fluggesellschaft weiss aus Erfahrung, dass durchschnittlich 4% der reservierten Plätze nicht eingenommen werden. Für ein Flugzeug mit 127 Sitzen werden daher in der Regel 129 Tickets verkauft.

a) Wie viele Personen werden in so einem Fall im Schnitt erscheinen? Wie gross ist die Wahrscheinlichkeit, dass mindestens eine Person keinen Platz findet? Welche vernünftige Annahme wird man treffen müssen?

b) Berechne exakt und approximativ die Wahrscheinlichkeit, dass 5 Personen nicht erscheinen?

Lösung: Sei X die Anzahl der Passagiere, die tatsächlich zum Flug erscheinen. Wenn wir davon ausgehen, dass sich die einzelnen Passagiere unabhängig voneinander entscheiden, zu fliegen oder nicht zu fliegen, und sich jeweils mit Wahrscheinlichkeit 0.96 für den Flug entscheiden, so ist X binomialverteilt mit Parametern $n = 129$ und $p = 0.96$, d.h.

$$\mathbb{P}(X = k) = \binom{129}{k} 0.96^k 0.04^{129-k} \qquad (k = 0, 1, \ldots, 129).$$

a) $\mathbb{E}(X) = np = 129 \times 0,96 = 123,84$. In so einem Fall im Schnitt werden $123,84$ Personen erscheinen. Gesucht wird die Wahrscheinlichkeit

$$\begin{aligned} \mathbb{P}(X \geq 128) &= \mathbb{P}(X = 128) + \mathbb{P}(X = 129) \\ &= 0.028 + 0.005 \\ &= 0.033\,. \end{aligned}$$

b) Sei Y die Anzahl der Personen, die nicht erscheinen. Y ist binomialverteilt mit Parametern $n = 129$ und $p = 0.04$ d.h.

$$\mathbb{P}(Y = k) = \binom{129}{k} 0.04^k 0.96^{129-k} \qquad (k = 0, 1, \ldots, 129).$$

Gesucht wird die Wahrscheinlichkeit

$$\mathbb{P}(Y = 5) = \binom{129}{5} 0.04^5 0.96^{129-5} = 0.178\,.$$

Die Poissonverteilung ist eine Approximation der Binomialverteilung für grosses n und kleines p mit $\lambda = np$, d.h.

$$\mathbb{P}(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{mit } \lambda = np\,.$$

Eine Approximation der gesuchten Wahrscheinlichkeit ist

$$\mathbb{P}(Y = 5) \approx e^{-5.16} \frac{5.16^5}{5!} = 0.175\,.$$