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

Koordinator
Florian Krach

Stochastik

Musterlösung 3

1. Wir betrachten eine Krankheit, zu der es einen Test beim Arzt gibt. Wir wissen, dass 0.1% der Bevölkerung die Krankheit haben. Bei erkrankten Personen ist der Test ist zu 98% positiv. Bei gesunden Personen ist die Wahrscheinlichkeit das der Test fälschlicherweise positiv ist 1%. Sei

- $A :=$ "Der Test ist positiv"
- $B :=$ "Man ist krank".

Wie oben beschrieben, wissen wir

$$\begin{aligned} P[B] &= 0.001, \\ P[A|B] &= 0.98 \\ P[A|B^c] &= 0.01. \end{aligned}$$

Peter hat einen Test beim Arzt gemacht. Dieser ist positiv ausgefallen. Wie hoch ist die Wahrscheinlichkeit, dass Peter tatsächlich die Krankheit hat?

a) Raten Sie die Antwort.

b) Berechnen Sie die Antwort mittels Bayes Formel

c) Die Antwort in Teil b) war eventuell niedriger als in Teil a). Füllen Sie die folgende Tabelle für 100'000 Menschen gemäss den gegebenen Wahrscheinlichkeiten aus, um die Antwort aus b) zu verifizieren:

	B	B^c	Summe
A			
A^c			
Summe			100'000

Lösung:

a) jede Antwort ist richtig.

b) $P[B|A] = 0.09$

Bitte wenden!

c)

	B	B^c	Summe
A	98	999	1'097
A^c	2	98'901	98'903
Summe	100	99'900	100'000

In der Tat haben 1'097 Personen einen positiven Test, aber nur 98 davon haben die Krankheit wirklich. $98/1'097 = 0.09$

2. **a)** Sei X eine diskrete Zufallsvariable, welche nur die Werte $W = \{1, \pi, 7, 8.6, 11\}$ annimmt. Die folgenden Wahrscheinlichkeiten sind gegeben:

$$P[X = 1] = \frac{1}{16}, \quad P[X = \pi] = \frac{3}{8}, \quad P[X = 7] = \frac{1}{8}, \quad P[X = 8.6] = \frac{1}{4}.$$

(i) Wie gross ist die Wahrscheinlichkeit $P[X = 11]$?

(ii) Wie gross ist die Wahrscheinlichkeit, dass X einen ganzzahligen Wert annimmt?

(iii) Berechne den Erwartungswert $E(X)$.

(iv) Berechne die Standardabweichung σ_X.

(v) Berechne den Erwartungswert $E(\ln X)$.

b) Es sei $X \sim \mathrm{Bin}(10, 1/3)$, P_X die Wahrscheinlichkeitsfunktion von X und F die kumulative Verteilungsfunktion von X.

(i) Berechne $P_X(1)$, $F(1.5)$ und $F(2)$.

(ii) Ist F stetig? Begründe.

Lösung:

a) (i) Für die Wahrscheinlichkeitsfunktion $p : W \to [0, 1]$ einer Zufallsvariablen gilt

$$\sum_{x \in W} p(x) = \sum_{x \in W} \mathbb{P}(X = x) = 1.$$

Also ist

$$\begin{aligned}\mathbb{P}(X = 11) &= 1 - \mathbb{P}(X = 1) - \mathbb{P}(X = \pi) - \mathbb{P}(X = 7) - \mathbb{P}(X = 8.6) \\ &= 1 - \frac{1}{16} - \frac{3}{8} - \frac{1}{8} - \frac{1}{4} = \frac{3}{16}.\end{aligned}$$

(ii) $\mathbb{P}(X = 1) + \mathbb{P}(X = 7) + \mathbb{P}(X = 11) = \frac{1}{16} + \frac{1}{8} + \frac{3}{16} = \frac{3}{8}$.

(iii) $\mathbb{E}(X) = 1 \cdot \frac{1}{16} + \pi \cdot \frac{3}{8} + 7 \cdot \frac{1}{8} + 8.6 \cdot \frac{1}{4} + 11 \cdot \frac{3}{16} = 6.3281$.

Siehe nächstes Blatt!

(iv) $\mathbb{E}(X^2) = 1^2 \cdot \frac{1}{16} + \pi^2 \cdot \frac{3}{8} + 7^2 \cdot \frac{1}{8} + 8.6^2 \cdot \frac{1}{4} + 11^2 \cdot \frac{3}{16} = 51.0661$,
$\mathbb{E}(X)^2 = (6.3281)^2 = 40.0441$, also
$\sigma_X = \sqrt{\mathbb{V}(X)} = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2} = \sqrt{51.0661 - 40.0441} = \sqrt{11.0213} = 3.31983$. Beachte: die Berechnung der Varianz als $V(X) = E((X - E(X))^2)$ wäre wesentlich komplizierter.

(v) $\mathbb{E}(\ln X) = \ln 1 \cdot \frac{1}{16} + \ln \pi \cdot \frac{3}{8} + \ln 7 \cdot \frac{1}{8} + \ln 8.6 \cdot \frac{1}{4} + \ln 11 \cdot \frac{3}{16} = 1.66006$.

b) (i) Es ist $P_X(1) = \binom{10}{1}\frac{1}{3}\left(\frac{2}{3}\right)^9$. Da

$$F(1.5) = P(X \leq 1.5) = P(X = 0) + P(X = 1) = P_X(0) + P_X(1),$$

erhalten wir $F(1.5) \approx 0.1040$. Analog ist $F(2) = P_X(0) + P_X(1) + P_X(2) \approx 0.2991$.

(ii) F ist nicht stetig. Es gilt z.B. $F(x) = F(1) \approx 0.1040$ für alle $x \in [1, 2)$, aber $F(2) \approx 0.2991$, d.h. F hat einen Sprung bei 2.

3. Bei einer Untersuchung werden Wasserproben (10ml) auf Verunreinigungen untersucht. Da nur 2 Prozent aller Proben verunreinigt sind, wird vorgeschlagen, von 10 Einzelproben jeweils die Hälfte (5ml) der Probe zu einer Sammelprobe (50ml) zusammenzumischen und zunächst nur die Sammelprobe zu untersuchen. Wird in der Sammelprobe keine Verunreinigung festgestellt, so ist die Untersuchung für die 10 Einzelproben beendet. Im anderen Fall werden alle 10 übriggebliebenen Hälften in 10 Einzeluntersuchungen geprüft.

a) Wie gross ist die Wahrscheinlichkeit, in der Sammelprobe keine Verunreinigung zu finden (unter der Annahme, dass die Einzelproben unabhängig voneinander sind)?

b) Sei die Zufallsvariable Y die Gesamtzahl benötigter Analysen. Welche Werte kann Y annehmen, und mit welchen Wahrscheinlichkeiten treten sie auf?

c) Wieviele Analysen werden "im Durchschnitt" für die gesamte Untersuchung benötigt (d.h. wie gross ist $E[Y]$)? Wieviele Analysen werden durch die Bildung von Sammelproben "im Durchschnitt" eingespart?

Lösung:

a) Sei X die Anzahl der verunreinigten Einzelproben in einer Sammelprobe. Die Erfolgswahrscheinlichkeit p, dass eine Einzelprobe verunreinigt ist, beträgt 0.02. Unter der Annahme, dass die Einzelproben voneinander unabhängig sind, gilt $X \sim \mathrm{Bin}(n = 10, p = 0.02)$.

Die Wahrscheinlichkeit, in der Sammelprobe keine Verunreinigung zu finden, ist gegeben durch

$$P(X=0) = \binom{10}{0} 0.02^0 \cdot 0.98^{10} = 0.98^{10} = 0.817\,.$$

Anderer Lösungsweg: Jede einzelne Probe ist unabhängig von den anderen Proben mit 98% Wahrscheinlichkeit sauber. Also gilt **(Multiplikationssatz für unabhängige Ereignisse)**

$$P(\text{alle Proben sauber}) = \prod_{i=1}^{10} P(i\text{-te Probe sauber}) = 0.98^{10} = 0.817\,.$$

b) Die Zufallsvariable Y kann nur die Werte 1 oder 11 annehmen, denn:

- falls alle Proben sauber sind, ist man nach einer Untersuchung fertig: $Y{=}1$
- sonst muss man jede Probe einzeln untersuchen (man darf nicht stoppen, wenn man die erste verunreinigte Probe gefunden hat, denn es könnte ja noch mehr geben!), also $Y{=}11$.

Folglich

$$\begin{aligned} P(Y=1) &= P(\text{alle Proben sauber}) = 0.817 \\ P(Y=11) &= 1 - P(Y=1) = 0.183 \end{aligned}$$

c) Die durchschnittliche Anzahl Analysen pro Sammelprobe ist gegeben durch den Erwartungswert der Zufallsvariablen Y, also

$$E(Y) = \sum_{k=0}^{\infty} kP(Y=k) = 1{\cdot}P(Y=1) + 11{\cdot}P(Y=11) = 1{\cdot}0.817 + 11{\cdot}0.183 = 2.83.$$

"Im Durchschnitt" spart man also $10 - 2.83 = 7.17 \approx 7$ Untersuchungen ein.

4. Für eine externe Benutzergruppe, der auch Franz angehört, stehen an der ETH vier Rechner zur Verfügung, wobei jedem Benutzer beim Einloggen jeweils mit Wahrscheinlichkeit $\frac{1}{4}$ unabhängig von den anderen einer der Rechner zugeteilt wird. Es kann also durchaus vorkommen, dass mehrere Leute auf dem gleichen Rechner arbeiten.

a) Wie gross ist die Wahrscheinlichkeit, dass höchstens eine Person auf dem gleichen Rechner arbeitet wie Franz, wenn ausser ihm genau 10 Personen eingeloggt sind?

Siehe nächstes Blatt!

b) Wieviele Leute dürfen sich ausser Franz noch einloggen, damit er mindestens mit Wahrscheinlichkeit $\frac{1}{2}$ allein auf einem Rechner arbeitet?

c) (**Zusatz**) Wir nehmen nun an, dass jeder Rechner unabhängig von den anderen nur mit Wahrscheinlichkeit 0.9 funktioniert, wobei die Benutzer analog wie vorher auf die funktionierenden Rechner "verteilt" werden (d.h. wenn gewisse Rechner nicht funktionieren, werden die Benutzer auf die Verbliebenen "verteilt"). Gegeben, dass ausser Franz drei Leute eingeloggt sind und genau zwei davon auf dem gleichen Rechner arbeiten wie er, wie gross ist dann die Wahrscheinlichkeit, dass alle Rechner funktionieren?

Tipp: Sei X die Anzahl funktionierender Rechner und S_3 die Anzahl der Personen, die ausser Franz auf seinem Rechner arbeiten. Bestimme für $k = 2, 3, 4$ die W'keiten $P(S_3 = 2|X = k)$ aus der Annahme wie die Benutzer auf die Rechner "verteilt" werden und benutze diese W'keiten um mit dem Satz der totale Wahrscheinlichkeit $P(S_3 = 2)$ zu berechnen. Mit dem Satz von Bayes erhält man dann die gesuchte Wahrscheinlichkeit.

Lösung: Es sei S_n die Anzahl Personen, die auf Franz' Rechner arbeiten (ausser ihm), wenn n Personen (ausser ihm) eingeloggt sind. Dann ist S_n Binomial$(n, \frac{1}{4})$ verteilt.

a) Da $S_{10} \sim Bin(10, \frac{1}{4})$ erhalten wir

$$\begin{aligned} P[S_{10} \leq 1] &= \left(\frac{3}{4}\right)^{10} + 10\left(\frac{3}{4}\right)^{9}\frac{1}{4} \\ &= \frac{3^9 \cdot 13}{4^{10}} = 0.244\ldots \end{aligned}$$

b) Gesucht wird das grösste n mit $P[S_n = 0] = \left(\frac{3}{4}\right)^n \geq \frac{1}{2}$. Da $\left(\frac{3}{4}\right)^2 = \frac{9}{16} > \frac{1}{2}$ und $\left(\frac{3}{4}\right)^3 = \frac{27}{64} < \frac{1}{2}$, lautet die Antwort "höchstens 2".

c) Sei X die Anzahl der funktionierenden Rechner, dann ist $X \sim Bin(4, 0.9)$, das heisst $P[X = 4] = 0.9^4$, $P[X = 3] = 4 \cdot 0.9^3 \cdot 0.1$, $P[X = 2] = 6 \cdot 0.9^2 \cdot 0.1^2$. Weiter gilt

$$\begin{aligned} P[S_3 = 2|X = 4] &= \binom{3}{2}\left(\frac{1}{4}\right)^2\frac{3}{4} = \frac{9}{64} \\ P[S_3 = 2|X = 3] &= \binom{3}{2}\left(\frac{1}{3}\right)^2\frac{2}{3} = \frac{2}{9} \\ P[S_3 = 2|X = 2] &= \binom{3}{2}\left(\frac{1}{2}\right)^2\frac{1}{2} = \frac{3}{8} \end{aligned}$$

Bitte wenden!

(da nämlich $S_n|X = k \sim Bin(n, 1/k)$). Der Satz von Bayes liefert nun die gesuchte Wahrscheinlichkeit:

$$
\begin{aligned}
&P[X = 4|S_3 = 2] \\
&= \frac{P[S_3 = 2|X = 4]P[X = 4]}{P[S_3 = 2|X = 2]P[X = 2] + P[S_3 = 2|X = 3]P[X = 3] + P[S_3 = 2|X = 4]P[X = 4]} \\
&= 0.526\ldots
\end{aligned}
$$