Musterlösung

1. a) 2 b) 3 c) 2 d) 3 e) 2 f) 1 g) 2 h) 1 i) 3 j) 1

2. Für eine eingehende Mail benützen wir die Notationen

$$\begin{aligned} S &= \{\text{Mail ist Spam}\}, \\ A &= \{\text{Mail wird vom Filter A als 'Spam' klassiert}\}, \\ B &= \{\text{Mail wird vom Filter B als 'Spam' klassiert}\}, \\ C &= \{\text{Mail wird vom Filter C als 'Spam' klassiert}\}. \end{aligned}$$

a) Zu berechnen ist $P[S|A^c]$, wobei A^c das Komplement von A bezeichnet. Mit der Formel von Bayes folgt

$$P[S|A^c] = \frac{P[A^c|S]P[S]}{P[A^c|S]P[S] + P[A^c|S^c]P[S^c]} = \frac{0.1 \times 0.9}{0.1 \times 0.9 + 0.99 \times 0.1} = \frac{10}{21}.$$

b) i) Die erwarteten Kosten pro eingehende Mail beim Filter A sind

$$\begin{aligned} 0.1 \times P[S \cap A^c] + 1 \times P[S^c \cap A] &= 0.1 \times P[A^c|S]P[S] + 1 \times P[A|S^c]P[S^c] \\ &= 0.1 \times 0.1 \times 0.9 + 1 \times 0.01 \times 0.1 = 0.01. \end{aligned}$$

Analog sind die erwarteten Kosten pro eingehende Mail beim Filter B

$$0.1 \times 0.05 \times 0.9 + 1 \times 0.02 \times 0.1 = 0.0065$$

und beim Filter C

$$0.1 \times 0.02 \times 0.9 + 1 \times 0.03 \times 0.1 = 0.0048.$$

Somit fährt die Mail AG unter diesen Annahmen mit dem Filter C am besten.
ii) Wenn $q = P[S]$ den Spam-Anteil bezeichnet, belaufen sich die erwarteten Kosten beim Filter A auf

$$0.1 \times 0.1 \times q + 1 \times 0.01 \times (1 - q) = 0.01$$

und beim Filter B auf

$$0.1 \times 0.05 \times q + 1 \times 0.02 \times (1 - q) = 0.02 - 0.015q.$$

Gleichheit gilt für $q = \frac{0.01}{0.015} = \frac{2}{3}$.

c) i) Da X als Wartezeit auf den zehnten (Miss-)Erfolg betrachtet werden kann, hat X eine negativbinomiale Verteilung mit Parametern $r = 10$ und $p = P[S|A^c] = 10/21$ aus **a)**.
ii) Daher ist $E[X] = r/p = 21$.

d) Wir bezeichnen mit D das Ereignis, dass eine eingehende Mail durch das zweistufige Verfahren als 'Nicht-Spam' klassiert wird. Es gilt $D = A^c \cap B^c$.

i) Die Formel von Bayes und die Unabhängigkeitsannahme ergeben

$$\begin{aligned} P[S|A^c \cap B^c] &= \frac{P[A^c \cap B^c|S]P[S]}{P[A^c \cap B^c|S]P[S] + P[A^c \cap B^c|S^c]P[S^c]} \\ &= \frac{P[A^c|S]P[B^c|S]P[S]}{P[A^c|S]P[B^c|S]P[S] + P[A^c|S^c]P[B^c|S^c]P[S^c]} \\ &= \frac{0.1 \times 0.05 \times 0.9}{0.1 \times 0.05 \times 0.9 + 0.99 \times 0.98 \times 0.1} \\ &\left(= \frac{1}{1 + 22 \times 0.98} = \frac{25}{564} \approx 0.0443\right). \end{aligned}$$

ii) Die gesuchte bedingte Wahrscheinlichkeit $P[D^c|S^c]$ bestimmen wir durch

$$\begin{aligned} P[D^c|S^c] &= P\big[(A^c \cap B^c)^c\big|S^c\big] = P[A \cup B|S^c] = P[A|S^c] + P[A^c \cap B|S^c] \\ &= P[A|S^c] + P[A^c|S^c]P[B|S^c] = 0.01 + 0.99 \times 0.02 \ (= 0.0298). \end{aligned}$$

Alternativ kann man rechnen

$$\begin{aligned} P[D^c|S^c] &= P\big[(A^c \cap B^c)^c\big|S^c\big] = 1 - P[A^c \cap B^c|S^c] \\ &= 1 - P[A^c|S^c]P[B^c|S^c] = 1 - 0.99 \times 0.98 \ (= 0.0298). \end{aligned}$$

3. **a)** Aus $\int_{-4}^{0}(4+x)\,dx+\int_0^8 \frac{1}{2}(8-x)\,dx = 16-8+32-16=24$ folgt $c=\frac{1}{24}$.

b) $E[B]=\int_{-4}^{0} xc(4+x)\,dx+\int_0^8 xc\frac{1}{2}(8-x)\,dx = c\left(-32+\frac{64}{3}+2\cdot 64-\frac{4}{3}64\right)=c\cdot 32=\frac{4}{3}$.

c) Rochus verpasst den Zug, falls der Bus mehr als 4 Minuten Verspätung hat, d.h. wir berechnen die Wahrscheinlichkeit des Ereignisses $\{B>4\}$:
$P[B>4]=\int_4^8 c\frac{1}{2}(8-x)\,dx=c(32-16-16+4)=c\cdot 4=\frac{1}{6}$.
Alternativ kann man rechnen:

$$\begin{aligned}1-P[B\le 4] &= 1-\left(\int_{-4}^{0} c(4+x)\,dx+\int_0^4 c\frac{1}{2}(8-x)\,dx\right)\\ &= 1-c\left(16-\frac{16}{2}\right)-c(16-4)\\ &= 1-20c\\ &= \frac{1}{6}.\end{aligned}$$

d) Rochus erreicht die ETH zu spät, falls der Zug mehr als 10 Minuten verspätet ist oder falls er den Zug verpasst. Dies gibt
$P[\text{Rochus ist zu spät an der ETH}]=P[Z>10]+P[B>4]\,P[Z\le 10]=\frac{1}{10}+\frac{1}{6}\frac{9}{10}=\frac{15}{60}=\frac{1}{4}$.
Alternativ kann man rechnen: $P[B>4]+P[B\le 4]\,P[Z>10]=\frac{1}{6}+\frac{5}{6}\frac{1}{10}=\frac{1}{4}$.
(Mit gegebenem Wert für $P[B\le 4]=\frac{2}{3}$ ergibt sich folgende Wahrscheinlichkeit:
$\frac{1}{10}+\frac{1}{3}\frac{9}{10}=\frac{12}{30}=\frac{2}{5}$.)

e) "Mindestens eines" kann durch "das Maximum von Z und B grösser als 4" ersetzt werden. Durch Komplementbildung kann dieses Ereignisses ohne max geschrieben werden. Schliesslich benutzen wir die Unabhängigkeit von Z und B.

$$\begin{aligned}P[\max(Z,B)>4] &= 1-P[\max(Z,B)\le 4]\\ &= 1-P[Z\le 4, B\le 4]\\ &\overset{\text{unab}}{=} 1-P[Z\le 4]\,P[B\le 4]\\ &= 1-\frac{4}{5}\cdot\frac{5}{6}\\ &= \frac{1}{3}.\end{aligned}$$

(Mit gegebenem Wert für $P[B\le 4]=\frac{2}{3}$ ergibt sich folgende Wahrscheinlichkeit:
$1-\frac{4}{5}\frac{2}{3}=1-\frac{8}{15}=\frac{7}{15}$.)

f) i) Es handelt sich um 50 unabhängige 0-1-Experimente mit Erfolgsparameter $p=0.03$. Folglich benutzen wir die Binomialverteilung für 3 Erfolge. Die Wahrscheinlichkeit beträgt $P[X=3]=\binom{50}{3}(0.03)^3(1-0.03)^{47}$.

ii) Die Binomialverteilung kann für grosses n und kleines p durch die Poissonverteilung approximiert werden. Dabei verwendet man den Parameter $\lambda=n\cdot p=50\cdot 0.03=3/2$.

4. a) Seien die Zufallsvariablen X_i für $i = 1, \ldots, 9$ die Weiten der Trainingssprünge von Simon (in Metern). Dann gilt nach der Aufgabenstellung $X_i \sim N(\mu, \sigma^2)$ mit $\sigma^2 = 4$. Wegen des asymmetrischen Vorgehens bei statistischen Tests nehmen wir die Negation von Simons Aussage als Nullhypothese, d.h.,

$$\mathrm{H}_0 : \mu = \mu_0 = 65 \text{ (oder } \mu \leq \mu_0 = 65)$$
$$\mathrm{H}_A : \mu = \mu_A > 65.$$

Da σ^2 bekannt ist, führen wir einen z-Test durch. Die Teststatistik

$$T = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{3}{2}\big(\overline{X}_9 - 65\big)$$

ist unter der Nullhypothese standard-normalverteilt und wir erhalten mit Hilfe der Tabelle im Anhang der Prüfung

$$K = [1.645, \infty)$$

als den kritischen Bereich zum Niveau $\alpha = 5\%$. Für die beobachteten Weiten der Trainingssprünge ergibt dies

$$T(\omega) = \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{3}{2}(66 - 65) = \frac{3}{2}.$$

Da $\frac{3}{2} < 1.645$ ist, behalten wir die Nullhypothese bei. Da die Nullhypothese beibehalten wird, können wir anhand des Testergebnisses nicht bestätigen, dass der Erwartungswert der Weiten von Simons Sprüngen signifikant über $65m$ liegt.

b) Unter der Alternative $\mu = \mu_A = 66$ ist die Teststatistik

$$T = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X}_n - \mu_A}{\sigma/\sqrt{n}} + \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} = \frac{3}{2}(\overline{X}_9 - 66) + \frac{3}{2}$$

$N\left(\frac{3}{2}, 1\right)$-verteilt. Dies ergibt die Macht

$$P_{\mu_A}[T \in K] = P_{\mu_A}[T \geq 1.645] = P_{\mu_A}[T - 1.5 \geq 1.645 - 1.5] = 1 - \Phi(0.145) = 1 - 0.55765 = 0.44235.$$

Für den angegebenen kritischen Bereich $\tilde{K}$ erhalten wir

$$\begin{aligned} P_{\mu_A}[T \in \tilde{K}] &= P_{\mu_A}[T \leq -1.96] = P_{\mu_A}[T - 1.5 \leq -1.96 - 1.5] \\ &= \Phi(-3.46) = 1 - \Phi(3.46) = 1 - 0.99973 = 0.00027. \end{aligned}$$

c) Die Ereignisse, dass Simon im i-ten Sprung über $120m$ springt, können als eine Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p aufgefasst werden. Die Anzahl Sprünge, die Simon benötigt, um an einem der drei Trainingstage das erste Mal über $120m$ zu springen, ist dann die Wartezeit auf den ersten Erfolg und demnach *geometrisch-verteilt*, d.h., $N_i \overset{i.i.d.}{\sim} Geom(p)$ für $i = 1, 2, 3$. Sei $X_n \sim N(\mu, 4)$ die Weite des n-ten Sprungs für $n = 1, 2, \ldots$. Dann ist der Erfolgsparameter p in Abhängigkeit von μ gegeben durch

$$p = P_\mu[X_n \geq 120] = 1 - \Phi\left(\frac{120 - \mu}{2}\right) = 1 - \Phi\left(60 - \frac{\mu}{2}\right).$$

d) Nach Aufgabenteil **c)** sind N_1, N_2 und N_3 unabhängig und identisch $Geom(p)$-verteilt mit Gewichtsfunktion $p(n_i; p) = P[N_i = n_i] = (1-p)^{n_i - 1}p$ für $n_i = 0, 1 \ldots$ und $i = 1, 2, 3$. Hieraus ergeben sich die Likelihood-Funktion

$$L(n_1, n_2, n_3; p) = P[N_1 = n_1,\ N_2 = n_2,\ N_3 = n_3] = (1-p)^{(n_1 + n_2 + n_3 - 3)}p^3$$

und die log-Likelihood-Funktion

$$\ell(n_1, n_2, n_3; p) = \log\big(L(n_1, n_2, n_3; p)\big) = 3\log(p) + (n_1 + n_2 + n_3 - 3)\log(1-p).$$

Der Maximierer p_{max} von $\ell(n_1, n_2, n_3; p)$ bezüglich p ist dann bestimmt durch

$$\frac{\partial}{\partial p}\ell(n_1, n_2, n_3; p_{max}) = \frac{3}{p_{max}} - \frac{n_1 + n_2 + n_3 - 3}{1 - p_{max}} = 0.$$

Auflösen nach p_{max} und Einsetzen der Zufallsvariablen für die Beobachtungen ergibt

$$\widehat{p}_{MLE} = \frac{3}{N_1 + N_2 + N_3} = \big(\overline{N}_3\big)^{-1}.$$