Musterlösung

1. (9 Punkte)

a) (1 Pt) Nach Definition der bedingten Wahrscheinlichkeit gilt:

$$\mathcal{P}[V \cap A] = \mathcal{P}[V|A]\mathcal{P}[A] = 0.97 * 0.6 = 0.582$$

b) (1 Pt) Wir kennen die Wahrscheinlichkeit des Komplements:

$$\mathcal{P}[V^c|A] = 1 - \mathcal{P}[V|A] = 1 - 0.97 = 0.03$$

c) (2 Pt) Wir zerlegen in disjunkte Ereignisse und benutzen dann die Definition der bedingten Wahrscheinlichkeit:

$$\mathcal{P}[V^c] = \mathcal{P}[V^c \cap A] + \mathcal{P}[V^c \cap B] = \mathcal{P}[V^c|A]\mathcal{P}[A] + \mathcal{P}[V^c|B]\mathcal{P}[B] = 0.03 * 0.6 + 0.05 * 0.4 = 0.038$$

1 Punkt für $\mathcal{P}[V^c] = \mathcal{P}[V^c|A]\mathcal{P}[A] + \mathcal{P}[V^c|B]\mathcal{P}[B]$, 1 Punkt für das richtige Ergebnis.

d) (1 Pt) Mit den obig ausgerechneten Werten:

$$\mathcal{P}[V^c] \neq \mathcal{P}[V^c|A]$$

Nein, per Definition sind V^c und A nicht unabhängig.

e) (2 Pt) Wir benutzen den Satz von Bayes oder zweimal die Definition der bedingten Wahrscheinlichkeit:

$$\mathcal{P}[A|V^c] = \frac{\mathcal{P}[V^c \cap A]}{\mathcal{P}[V^c]} = \frac{\mathcal{P}[V^c|A]\mathcal{P}[A]}{\mathcal{P}[V^c]} = \frac{0.03 * 0.6}{0.038} = 0.4736842$$

1 Punkt für $\mathcal{P}[A|V^c] = \frac{\mathcal{P}[V^c|A]\mathcal{P}[A]}{\mathcal{P}[V^c]}$, 1 Punkt für das richtige Ergebnis.

f) (2 Pt) Bestellt Ruben n Tomaten, erhält er im Erwartungswert $n\mathcal{P}[V]$ verwendbare Tomaten. Wir wissen aus obiger Aufgabe:

$$\mathcal{P}[V] = 1 - \mathcal{P}[V^c] = 1 - 0.038 = 0.962$$

Gesucht ist also die kleinste ganze Zahl n sodass $n\mathcal{P}[V] \geq 1000$. Wir rechnen $1000/0.962 = 1039.501$. Folglich muss er mindestens 1040 Tomaten bestellen.
1 Punkt für $n\mathcal{P}[V]$, 1 Punkt für das richtige Ergebnis.

2. (8 Punkte)
Wenn Christina für Prüfungen lernen muss, ist sie nach dem Aufwachen um punkt 7.00 Uhr immer mindestens 2 Stunden wach, wird danach aber schnell wieder müde und macht täglich einen "power nap". Die Zeit (in Stunden) bis zum Beginn des "power naps" sei durch die Zufallsvariable T mit kumulativer Verteilungsfunktion

$$F_T(t) = \begin{cases} 1 - \left(\frac{2}{t}\right)^2 & \text{für } t \geq 2 \\ 0 & \text{für } t < 2. \end{cases}$$

beschrieben.

a) (1 Punkt) Berechnen Sie die Dichte der Zufallsvariable T.
Die Dichte erhält man durch Ableiten der kumulativen Verteilungsfunktion. Für $t < 2$ gilt $f_T(t) = 0$. Für $t \geq 2$ erhält man

$$\frac{d}{dt}(1 - \left(\frac{2}{t}\right)^2) = \frac{d}{dt}\frac{-4}{t^2} = \frac{8}{t^3}.$$

Die Dichte ist also

$$f_T(t) = \begin{cases} \frac{8}{t^3} & \text{für } t \geq 2 \\ 0 & \text{für } t < 2. \end{cases}$$

(1 Punkt für richtige Antwort)

b) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass Christina den „power nap" vor 10.00 Uhr beginnt (d.h. $T \leq 3$)?
Der Zeitunterschied zwischen 7.00 Uhr und 10.00 Uhr beträgt 3 Stunden, sodass

$$P[T \leq 3] = F_T(3) = 1 - \left(\frac{2}{3}\right)^2 \approx 0.5556 \quad \textbf{(1 Punkt)}$$

c) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass Christina den „power nap" zwischen 12.00 und 13.00 Uhr beginnt?
Der Zeitunterschied zwischen 7.00 Uhr und 13.00 Uhr beträgt 6 Stunden bzw. zwischen 7.00 Uhr und 12.00 Uhr sind es 5 Stunden, sodass

$$P[5 < T \leq 6] = P[T \leq 6] - P[T \leq 5] = F_T(6) - F_T(5) = 1 - \left(\frac{2}{6}\right)^2 - \left(1 - \left(\frac{2}{5}\right)^2\right) \approx 0.0489 \quad \textbf{(1 Punkt)}$$

d) (2 Punkte) Berechnen Sie den Median m von T.
Für den Median m gilt: $F_T(m) = P[T \leq m] = 0.5$. Daher folgt:

$$\begin{aligned} & 1 - \left(\frac{2}{t}\right)^2 = 0.5 \quad \textbf{(1 Punkt)} \\ \Rightarrow & \left(\frac{2}{t}\right)^2 = 0.5 \\ \Rightarrow & \frac{2}{t} = \sqrt{0.5} \\ \Rightarrow & t = \frac{2}{\sqrt{0.5}} = 2 \cdot \sqrt{2}. \quad \textbf{(1 Punkt)} \end{aligned}$$

Christinas Mitbewohnerin Claudia arbeitet in der Gastronomie und kommt abends oft erst spät nach Hause. Daher schläft sie morgens immer länger als Christina. Die Zeit (in Stunden) zwischen Christinas Aufwachen um 7.00 Uhr bis zu Claudias Aufwachen sei durch die Zufallsvariable D mit kumulativer Verteilungsfunktion

$$F_D(t) = \begin{cases} 1 - \exp\left(-\frac{t}{4}\right) & \text{für } t \geq 0 \\ 0 & \text{für } t < 0. \end{cases}$$

beschrieben. Wir nehmen an, dass die Zufallsvariablen T und D unabhängig sind.

e) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass an einem Tag Christina den "power nap" nach 12 Uhr beginnt und Claudia vor 12 Uhr aufwacht?
Wegen der Unabhängigkeit von T und D folgt:

$$P[T > 5 \text{ und } D < 5] = P[T > 5] \cdot P[D < 5] = \left(\frac{2}{5}\right)^2 \cdot (1 - \exp\left(-\frac{5}{4}\right)) \approx 0.1142.$$

f) (2 Punkte) Wenn Christina einen "power nap" von 30 Minuten macht, wie gross ist dann die Wahrscheinlichkeit, dass um 13.30 Uhr mindestens eine der beiden schläft?

Wenn Christina um 13.30 Uhr schläft, dann beginnt sie ihren "power nap" nach 13.00 Uhr und vor oder um 13.30 Uhr. Wenn Claudia um 13.30 Uhr schläft, heisst dies, dass sie bis 13.30 Uhr noch nicht aufgewacht ist.

$$\begin{aligned}
&P[6 < T \leq 6.5 \text{ oder } D > 6.5] \\
&= P[((6 < T \leq 6.5)^c \text{ und } (D > 6.5)^c)^c] \\
&= 1 - P[(6 < T \leq 6.5)^c \text{ und } D \leq 6.5] \\
&= 1 - P[(6 < T \leq 6.5)^c] \cdot P[D \leq 6.5] \\
&= 1 - ((1 - P[6 < T \leq 6.5]) \cdot P[D \leq 6.5]) \\
&\approx 0.2101.
\end{aligned}$$

3. (**8 Punkte**)
Jan hat einen Zufallszahlengenerator programmiert. Er sagt, dass dieser (i.i.d.) uniform-verteilte Zufallszahlen generiert auf dem Intervall $[0, c]$. Die obere Schranke c verrät er Ihnen aber nicht. Alles, was wir haben, ist folgende (i.i.d.) Stichprobe:

x_1	x_2	x_3	x_4	x_5
1.3	2.4	5.0	8.5	11.9

a) (3 Punkte) Ermitteln Sie den Momentenschätzer für c. Ist dieser für die konkret vorliegende Stichprobe sinnvoll?

Sei $X \sim \text{Unif}([0, c])$. Berechnen wir zuerst den Erwartungswert:

$$\mathcal{E}[X] = \int_0^c \frac{x}{c} dx = \frac{c}{2}$$

(1 Punkt)
Folglich ist der Momentenschätzer für c:

$$\frac{2}{n} \sum_{i=1}^{n} x_i = 11.64$$

(1 Punkt) Falls die Studenten den Erwartungswert nicht explizit berechnen und den Momentenschätzer direkt richtig hinschreiben, gibt es dafür auch zwei Punkte.
Für die konkret vorliegende Stichprobe ist der Momentenschätzer nicht sinnvoll, da die Likelihood für die Stichprobe 0 ist ($x_5 > 11.64$). (1 Punkt)

b) (2 Punkte) Bestimmen Sie den Maximum-Likelihood-Schätzer für c (für die konkret vorliegende Stichprobe). *Hinweis:* Überlegen Sie sich, welche Werte von c überhaupt sinnvoll sind. Nichts ableiten.

Bestimmen wir zuerst die Likelihood:

$$\text{Likelihood} = \begin{cases} \frac{1}{c^5} & \text{falls } 0 \leq x_i \leq c, i = 1, ..., 5; \\ 0 & \text{sonst.} \end{cases}$$

(1 Punkt)
Dies lässt sich für die gegebenen Beobachtungen leicht umformen:

$$\text{Likelihood} = \begin{cases} \frac{1}{c^5} & \text{falls } \max_{i=1,...,5} x_i \leq c; \\ 0 & \text{sonst.} \end{cases}$$

Die Likelihood ist daher maximal für $c = \max_{i=1,...,5} x_i$. Der Maximum-Likelihood-Schätzer für c ist folglich $\max_{i=1,...,5} x_i = 11.9$. (1 Punkt) Falls die Studenten die Zwischenschritte nicht machen und den Maximum-Likelihood-Schätzer für die konkrete Stichprobe direkt richtig hinschreiben, gibt es dafür auch zwei Punkte.

Die folgenden Teilaufgaben sind unabängig untereinander und unabhängig von von a) und b).

c) (1 Punkt) Die Zufallsvariable X habe die Dichte

$$f(x) = \frac{1}{\lambda\pi}\left(\frac{1}{1+\frac{x^2}{\lambda^2}}\right),\ x \in \mathbb{R}$$

wobei $\lambda > 0$ ein Parameter ist.
Bestimmen Sie die Likelihoodfunktion basierend auf n unabhängigen, identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer Zufallsvariable mit obiger Dichte.

$$\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\lambda\pi}\left(\frac{1}{1+\frac{x_i^2}{\lambda^2}}\right) = \frac{1}{(\lambda\pi)^n}\prod_{i=1}^{n}\left(\frac{1}{1+\frac{x_i^2}{\lambda^2}}\right)$$

d) (2 Punkte) Sie schätzen den Modellparameter θ mit dem Schätzer $\widehat{\theta}$. Von der Theorie her weiss man, dass

$$\widehat{\theta} \sim N(\theta, \theta/n),$$

wobei n die Stichprobengrösse ist. Geben Sie damit ein (approximatives) 95%-Vertrauensintervall für θ an, falls der realisierte Wert $\widehat{\theta} = 19.2$ und $n = 16$ ist.
Sei $z_{0.975}$ das 0.975-Quantil der Standardnormalverteilung. Dann nehmen wir als 95%-Vertrauensintervall:

$$[\widehat{\theta} - z_{0.975}\sqrt{\frac{\theta}{n}}, \widehat{\theta} + z_{0.975}\sqrt{\frac{\theta}{n}}].$$

Da wir θ nicht kennen schätzen wir es durch $\widehat{\theta}$. Da $z_{0.975} \approx 2$, ergibt das

$$\widehat{\theta} \pm z_{0.975}\sqrt{\frac{\theta}{n}} \approx \widehat{\theta} \pm 2\sqrt{\frac{\widehat{\theta}}{n}} = 19.2 \pm 2\sqrt{\frac{19.2}{16}}.$$

Folglich ist $[17.01, 21.39]$ ein approximatives 95%-Vertrauensintervall für θ.

4. a) t-Test für $U_i \sim \mathcal{N}(\mu, \sigma^2)$:

- $H_0 : \mu = 0, \quad H_A : \mu \neq 0$ (0.5 Pt)
- Teststatistik $T = \frac{\bar{U}_n}{S_n/\sqrt{n}} \sim t_{n-1}$ unter H_0 (0.5 Pt)
- Verwerfungsbereich $K = (-\infty, -t_{9,0.975}] \cup [t_{9,0.975}, \infty) = (-\infty, -2.26] \cup [2.26, \infty)$ (0.5 Pt)
- Testentscheid: $t = 0.3 \notin K$, also wird H_0 nicht verworfen. (0.5 Pt)

b) Notwendige Annahme: die Daten sollten normalverteilt sein (sonst stimmt die Verteilung der Teststatistik und damit Verwerfungsbereich und Niveau des Tests nicht). (1 Pt)

c) Das zweiseitige Vertrauensintervall für den t-Test ist gegeben durch $I = \bar{u}_{10} \pm \frac{s_{10}}{\sqrt{10}} t_{9,0.995}$ (0.5 Pt). Also $I = [-8.53, 10.3]$ (0.5 Pt).

d) t-Test für Y_i i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$ und W_i i.i.d. $\sim \mathcal{N}(\mu_W, \sigma^2)$:

- $H_0 : \mu_Y = \mu_W, \quad H_A : \mu_Y < \mu_W$ (0.5 Pt)
- Teststatistik $T = \frac{\bar{Y}_n - \bar{W}_m}{S_{pool}\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2} = t_{16}$ unter H_0 (0.5 Pt)
- $s_{pool} = \sqrt{\frac{1}{n+m-2}\left((n-1)s_Y^2 + (m-1)s_W^2\right)} = 5.87$ (0.5 Pt)
- $t = -1.893$ (0.5 Pt)
- Verwerfungsbereich $K = (-\infty, -t_{16,0.95}] = (-\infty, -1.75])$ (0.5 Pt)
- Testentscheid: $t \in K$, also wird H_0 verworfen. (0.5 Pt)

e) Die beiden Stichproben haben unterschiedlich grosse Streuungen. Dies spricht dafür, dass die Annahme $\sigma_Y^2 = \sigma_W^2$ nicht erfüllt ist und wir daher eine Verallgemeinerung des Zwei-Stichproben t-Tests für ungleiche Varianzen wählen sollten. (1 Pt)

5. (**10 Punkte**) Entscheiden Sie für jede Aussage, ob sie richtig oder falsch ist. Markieren Sie Ihre Antwort auf dem separaten Multiple Choice Antwortblatt.

a) (2 Punkte) R, R, R, F

b) (2 Punkte) R, F, F, F

c) (2 Punkte) R, F, F, F

d) (2 Punkte) F, R, F, F

e) (2 Punkte) F, R, F, F