Stochastik - Lösung
(BSc D-MAVT / BSc D-MATH / BSc D-MATL)

1. (jeweils: 1 Punkt für richtig, 0 Punkt für falsch)

a) 2.

b) 3.

c) 3.

d) 1.

e) 1.

f) 2.

g) 3.

h) 1.

i) 2.

j) 1.

2. a) Wir bezeichnen das Ereignis, dass der Widerstand in Fabrik A (B) hergestellt wird mit "A" ("B"). Dann gilt $P[X = 1|A] = 0.2$ und $P[X = 1|B] = 0.1$. Ferner

$$P[X = 1] = P[X = 1|A]P[A] + P[X = 1|B]P[B] = 0.2 \cdot 0.7 + 0.1 \cdot 0.3 = 0.17.$$

daraus folgt $E[X] = 0.17$ **(1 Punkt)**. X ist also Bernoulli-verteilt mit Parameter 0.17 **(1/2 Punkt)**. Wir berechnen $Var(X) = 0.17 \cdot 0.83 = 0.1411$ **(1/2 Punkt)**.

b) Wir möchten $P[A|\{X = 1\}]$ berechnen. Nach Bayes haben wir

$$P[A|\{X = 1\}] = \frac{P[\{X = 1\}|A]P[A]}{P[\{X = 1\}]} \quad \textbf{(1/2 Punkt)} = \frac{0.2 \cdot 0.7}{0.17} = 0.824 \quad \textbf{(1/2 Punkt)}.$$

c) Als Summe von 1000 i.i.d. $Ber(0.17)$-verteilten Zufallsvariablen, ist Z Binomial-verteilt **(1/2 Punkt)**, mit Parametern $Z \sim Bin(1000, 0.17)$ **(1/2 Punkt)**.

d) *Lösung 1:* Sei Z wie in d). Laut CLT und a) ist $Z \approx N(1000 \cdot 0.17, 1000 \cdot 0.1411)$ **(1 Punkt)**. Das ergibt mithilfe der Normalverteilungstabelle

$$P[S \geq 200] = P\left[\frac{S-170}{\sqrt{141.1}} \geq \frac{200-170}{\sqrt{141.1}}\right] = P\left[\frac{S-170}{\sqrt{141.1}} \geq 2.526\right] \quad \textbf{(1/2 Punkt)}$$
$$= 1 - 0.9943 = 0.0057 \quad \textbf{(1/2 Punkt)}$$

die Wahrscheinlichkeit, dass die Produktion des Unternehmens an einem Tag stillsteht.

Lösung 2: Die Binomialverteilung für grosses n lässt sich gut durch die Normalverteilung mit der gleichen Erwartung und Varianz approximieren. Das heisst $Z \approx N(1000 \cdot 0.17, 1000 \cdot 0.1411)$. Der Lösungsweg ist ab hier das gleiche wie in Lösung 1.

e) Die Wahrscheinlichkeit, dass es fünf Tage lang keinen Produktionsausfall gibt, ist 0.9943^5. Somit ist $1 - 0.9943^5 = 0.028$ die Wahrscheinlichkeit, dass es mindestens einmal in einer Woche die Produktion ausfällt **(1 Punkt)**.

3. **a)** Gefragt ist $P[X \geq 3 | Y = 1.5]$.

$$P[X \geq 3|Y = 1.5] = \int_3^\infty 1.5 e^{-1.5x} dx = e^{-4.5} \quad \textbf{(1 Punkt)}.$$

b) Nach dem Bayes'schen Gesetz können wir die gemeinsame Dichte berechnen:

$$f_{X,Y}(x,y) = f_X(x|Y=y) f_Y(y) = y e^{-yx} 1_{[0,\infty)}(x) \frac{1}{2} 1_{[1,3]}(y) \quad \textbf{(1 Punkt)}.$$

c) Mit der gemeinsamen Dichte aus b) haben wir

$$P[X \geq 3] = \int_1^3 \int_3^\infty y e^{-yx} \frac{1}{2} dx dy \ \textbf{(1 Punkt)} = \int_1^3 y \frac{0 - e^{-3y}}{-y} \frac{1}{2} dy$$
$$= \int_1^3 e^{-3y} \frac{1}{2} dy = -\frac{1}{6} e^{-9} + \frac{1}{6} e^{-3} \approx 0.00828 \quad \textbf{(1 Punkt)}.$$

d) Damit f_Z eine Dichte ist, gilt

$$\int_3^\infty c \lambda e^{-\lambda x} dx = 1 \quad \textbf{(1/2 Punkt)}.$$

Daraus berechnen wir

$$c = \left(\int_3^\infty \lambda e^{-\lambda x} dx \right)^{-1} = \left(e^{-3\lambda}\right)^{-1} = e^{3\lambda} \quad \textbf{(1/2 Punkt)}.$$

e) Da die Realisierungen unabhängig sind, können wir die Likelihoodfunktion wie folgt berechnen:

$$L(\lambda) = \Pi_{i=1}^{6} f_Z(z_i, \lambda) = \Pi_{i=1}^{6} \lambda e^{\lambda(3-z_i)} \quad \textbf{(1 Punkt)}.$$

Als Nächstes berechnen wir die log-Likelihoodfunktion:

$$l(\lambda) = \log L(\lambda) = 6 \log \lambda + \lambda(18 - \sum_{i=1}^{6} z_i) \quad \textbf{(1/2 Punkt)}.$$

Nun setzten wir die Ableitung der log-Likelihoodfunktion gleich 0:

$$0 = \frac{d}{d\lambda} l(\lambda) = \frac{6}{\lambda} + (18 - \sum_{i=1}^{6} z_i) \quad \textbf{(1/2 Punkt)}.$$

Daraus ergibt sich

$$\hat{\lambda} = \frac{6}{\sum_{i=1}^{6} z_i - 18} = \frac{1}{\sum_{i=1}^{6} z_i/6 - 3}$$

(1 Punkt, nur 1/2 Punkt wenn λ statt $\hat{\lambda}$, λ^{MLE}, etc.).

4. Unterschiede:

Fahrradfahrer Nr.	1	2	3	4	5	6
alte Bekleidung (x_i)	52.6	47.1	48.1	50.7	55.5	48.0
neue Bekleidung (y_i)	48.1	47.3	48.6	47.1	49.2	46.7
Unterschiede ($u_i = x_i - y_i$)	4.5	-0.2	-0.5	3.6	6.3	1.3

Berechnungen:

$$\begin{aligned}
\bar{x} &= \sum x_i = 50.33 \\
\bar{y} &= \sum y_i = 47.83 \\
\bar{u} &= \sum u_i = 2.5 \\
s_X^2 &= \frac{1}{n-1} \sum (x_i - \bar{x})^2 = 10.571 \\
s_Y^2 &= \frac{1}{n-1} \sum (y_i - \bar{y})^2 = 0.927 \\
s_{X-Y}^2 &= \frac{1}{n-1} \sum (u_i - \bar{u})^2 = \frac{1}{5}(2^2 + 2.7^2 + 3^2 + 1.1^2 + 3.8^2 + 1.2^2) = 7.476
\end{aligned}$$

a) Die Reihenfolge, in welcher die beiden Bekleidungen getestet werden, soll bei jedem einzelnen Fahrradfahrer per Zufall bestimmt werden; i.e. Randomisierung **(1/2 Punkt)**. Somit können unerwünschte Nebenefekte (Wetter, Grippewelle, etc.) beseitigt werden **(1/2 Punkt)**.

b) (i) Der Test ist gepaart, weil die Versuchsbedingungen (Bekleidung) jeweils an derselben Veruchseinheit (Fahrradfahrer) eingesetzt werden **(1/2 Punkt)**.

(ii) Der Test ist einseitig, weil wir nur an einer Verbesserung der Fahrleistung interessiert sind **(1/2 Punkt)**.

c) (i) Den Vorzeichen-Test **(1/2 Punkt)**, da alle andere Tests bestimmte Bedingungen an die Verteilung voraussetzen (e.g. Symmetrie, Normalverteilung) **(1/2 Punkt)**.

(ii) Sei μ_{med} der Median der Zufallsvariable U_1 **(1/2 Punkt)**.
H_0: $\mu_{med} = 0$.
H_A: $\mu_{med} > 0$ **(1/2 Punkt)**.

(iii) Unsere Teststatistik, die Anzahl $X_i - Y_i$ die positive sind, i.e. $T = \sum 1_{[0,\infty)}(X_i - Y_i)$ ist unter der Nullhypothese $Bin(6, 0.5)$-verteilt **(1/2 Punkt)**. Die Realisierung der Teststatistik ist $t = \sum 1_{[0,\infty)}(x_i - y_i) = 4$ **(1/2 Punkt)**. Daraus können wir den P-Wert berechnen:

$$P\text{−Wert} = P[T \geq t] = P[T \geq 4] = 1/64+6/64+15/64 = 22/64 = 0.344 \quad \textbf{(1/2 Punkt)}.$$

Der P-Wert ist (um einiges) grösser als das Signifikanzniveau 0.05. Deswegen behalten wir die Nullhypothese **(1/2 Punkt)**.

d) (i) Unter Annahme der Normalverteilung soll man den t-Test wählen **(1/2 Punkt)**, da er eine bessere Macht hat (unter dieser Annahme) als der Wilcoxon-Test oder der Vorzeichen-Test. Der z-Test ist keine Option, denn die Varianz ist nicht bekannt **(1/2 Punkt)**.

(ii) Sei $\mu := E[U_1]$ die Erwartung der Zufallsvariable U_1.
H_0: $\mu = \mu_0 := 0$.
H_A: $\mu > 0$ **(1 Punkt)**.

(iii) Die Teststatistik, $T = \frac{\bar{U}_n - \mu_0}{S_n/\sqrt{n}}$ ist unter der Nullhypothese t_5-verteilt **(1/2 Punkt)**. Die Realisierung ist $t = \frac{\bar{u} - \mu_0}{s_{X-Y}/\sqrt{n}} = \frac{2.5}{\sqrt{7.476}/\sqrt{6}} = 2.240$ **(1/2 Punkt)**. Aus der Tabelle können wir den einseitigen Verwerfungsbereich für das 5% Niveau ablesen: $\mathcal{K} = [2.015, \infty)$ **(1/2 Punkt)**. Die Realisierung der Teststatistik fällt in den Verwerfungsbereich, darum verwerfen wir die Nullhypothese **(1/2 Punkt)**.