Wahrscheinlichkeit und Statistik
BSc D-INFK

1. **a)** (iii) **b)** (ii) **c)** (iii) **d)** (i) **e)** (ii) **f)** (i) **g)** (iii) **h)** (iii) **i)** (i) **j)** (ii)

2. **a)** Die Anzahl der bestellten Weine in einem Monat kann sehr gross sein. Die Wahrscheinlichkeit, dass ein bestellter Wein zurückgeschickt wird, ist jedoch sehr klein. In diesem Sinne ist das Zurückschicken einer Weinflasche also ein seltenes Ereignis und Z lässt sich sinnvoll durch eine Poisson-Verteilung modellieren. Sei also $Z \sim \mathcal{P}(\lambda)$. Dann wissen wir aus der Angabe, dass $1 = E[Z] = \lambda$, also ist $\lambda = 1$. Daher gilt für $k \in \mathbb{N}_0$,

$$P[Z = k] = e^{-1}\frac{1^k}{k!} = \frac{e^{-1}}{k!}.$$

b) A ist $Geom(\frac{1}{6})$- verteilt, d.h. $P[A = k] = \frac{1}{6}(1 - \frac{1}{6})^{k-1}$ für $k \in \mathbb{N}$. Insbesondere gilt

$$P[A = 2] = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}.$$

c) Bezeichne U das Ereignis, dass der ausgewählte Wein ungeniessbar ist, und N das Ereignis, dass der ausgewählte Wein aus der neusten Lieferung stammt. Dann ist

$$P[U] = P[U|N]P[N] + P[U|N^c]P[N^c] = 0.8 \cdot \frac{12}{72} + 0.04 \cdot \frac{60}{72} = \frac{1}{6}.$$

d) Mit der Bayes'schen Formel erhalten wir

$$P[N|U] = \frac{P[N \cap U]}{P[U]} = \frac{P[U|N]P[N]}{p} = \frac{0.8 \cdot \frac{12}{72}}{\frac{1}{6}} = 80\%.$$

Bitte wenden!

e) X ist hypergeometrisch verteilt mit Parametern $n = 72, r = 12, m = 2$ und

$$P[X = 1] = \frac{\binom{12}{1}\binom{60}{2-1}}{\binom{72}{2}} = \frac{12 \cdot 60}{\frac{72 \cdot 71}{2}} = \frac{20}{71}.$$

Bermerkung: Aufgrund der etwas unglücklichen Aufgabenstellung wird auch die Antwort akzeptiert, dass X binomialverteilt ist mit Parametern $n = 2$ und $p = 1/6$.

3. a) Es gilt

$$\frac{1}{c} f_X(x) \;=\; \int_0^x (x^2 + y^2) dy \;=\; x^3 + \frac{x^3}{3} \;=\; \frac{4x^3}{3},$$

also

$$f_X(x) = c\frac{4x^3}{3}, \quad 0 \leq x \leq 1.$$

Analog erhalten wir

$$\frac{1}{c} f_Y(y) \;=\; \int_y^1 (x^2 + y^2) dx \;=\; \frac{1}{3} - \frac{y^3}{3} + y^2(1 - y) \;=\; \frac{1}{3} + y^2 - \frac{4y^3}{3},$$

also

$$f_Y(y) = c\left(\frac{1}{3} + y^2 - \frac{4y^3}{3}\right), \quad 0 \leq y \leq 1.$$

Für c muss gelten, dass

$$\frac{1}{c} \stackrel{!}{=} \int_0^1 f_X(x) dx \;=\; \int_0^1 \frac{4x^3}{3} dx \;=\; \frac{1}{3},$$

bzw.

$$\frac{1}{c} \stackrel{!}{=} \int_0^1 f_Y(y) dy \;=\; \int_0^1 \left(\frac{1}{3} + y^2 - \frac{4y^3}{3}\right) dy \;=\; \frac{1}{3} + \frac{1}{3} - \frac{1}{3} \;=\; \frac{1}{3}.$$

Es folgt $c = 3$.

b) Für $E[X]$ erhalten wir

$$E[X] \;=\; \int_0^1 x f_X(x) dx \;=\; c\int_0^1 \frac{4x^4}{3} dx \;=\; c\frac{4}{15} \;=\; \frac{4}{5}.$$

Für die Varianz von X gilt $\mathrm{Var}[X] = E[X^2] - E[X]^2$, wobei

$$E[X^2] \;=\; \int_0^1 x^2 f_X(x) dx \;=\; c\int_0^1 \frac{4x^5}{3} dx \;=\; c\frac{2}{9} \;=\; \frac{2}{3}.$$

Siehe nächstes Blatt!

Wir erhalten $\mathrm{Var}[X] = \frac{2}{3} - \frac{16}{25} = \frac{2}{75}$.

Falls **a)** nicht gelöst wurde, erhalten wir für die alternative Randdichte für X,

$$E[X] = \int_0^1 x f_X(x) dx = 4 \int_0^1 (x^2 - x^4) dx = 4\left(\frac{1}{3} - \frac{1}{5}\right) = \frac{8}{15}.$$

Für die Varianz von X gilt $\mathrm{Var}[X] = E[X^2] - E[X]^2$, wobei

$$E[X^2] = \int_0^1 x^2 f_X(x) dx = 4 \int_0^1 (x^3 - x^5) dx = 4\left(\frac{1}{4} - \frac{1}{6}\right) = \frac{1}{3}.$$

Wir erhalten $\mathrm{Var}[X] = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}$.

c) Es gilt

$$f_X(x) f_Y(y) = c^2 \frac{4x^3}{3}\left(\frac{1}{3} + y^2 - \frac{4y^3}{3}\right) = 4x^3\left(1 + 3y^2 - 4y^3\right).$$

Ein Koeffizientenvergleich zeigt, dass $f_X f_Y \neq f_{X,Y}$. Beispielsweise ist für $x = 1/2$ und $y = 1/3$

$$f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{3}\right) = \frac{1}{2}\left(1 + \frac{1}{3} - \frac{4}{27}\right) = \frac{16}{27} \neq \frac{39}{36} = 3\left(\frac{1}{4} + \frac{1}{9}\right) = f_{X,Y}\left(\frac{1}{2}, \frac{1}{3}\right).$$

Somit sind X und Y nicht unabhängig.

Falls **a)** nicht gelöst wurde, erhalten wir für die alternativen Randdichten für X und Y,

$$f_X(x) f_Y(y) = \frac{16}{5}(x - x^3)(3y^3 + y).$$

Für $x = 1/2$ und $y = 1/3$ gilt mit $c = 2$,

$$\begin{aligned} f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{3}\right) &= \frac{16}{5}\left(\frac{1}{2} - \frac{1}{8}\right)\left(\frac{1}{9} + \frac{1}{3}\right) = \frac{16}{5} \cdot \frac{3}{8} \cdot \frac{4}{9} \\ &= \frac{8}{15} \neq \frac{13}{18} = 2\left(\frac{1}{4} + \frac{1}{9}\right) = f_{X,Y}\left(\frac{1}{2}, \frac{1}{3}\right). \end{aligned}$$

Somit sind X und Y nicht unabhängig.

d) Sei Z_i, $i = 1, 2, \ldots$, das Abfüllvolumen des i-ten Beutels. Nach Annahme gilt, dass die Z_i's i.i.d. sind. $S_n = Z_1 + \ldots + Z_n$ beschreibt dann das gesamte Abfüllvolumen von (den ersten) n Beuteln. Die Chebyshev-Ungleichung besagt, dass

$$P\left[|S_n - E[S_n]| > 0.1 \cdot E[S_n]\right] \leq \frac{\mathrm{Var}[S_n]}{(0.1 \cdot E[S_n])^2} = \frac{100\mathrm{Var}[S_n]}{(E[S_n])^2}.$$

Bitte wenden!

Wegen der Unabhängigkeit gilt

$$\frac{\mathrm{Var}[S_n]}{(E[S_n])^2} = \frac{n\mathrm{Var}[Z_1]}{(nE[Z_1])^2} = \frac{\mathrm{Var}[Z_1]}{n(E[Z_1])^2} = \frac{1}{8n}.$$

Es folgt, dass

$$P\left[|S_n - E[S_n]| > 0.1 \cdot E[S_n]\right] \;\leq\; 100 \cdot \frac{1}{8n} \;=\; \frac{100}{8n}.$$

Laut Aufgabenstellung soll gelten, dass

$$P\left[|S_n - E[S_n]| > 0.1 \cdot E[S_n]\right] \;\leq\; \frac{1}{10}.$$

Mit Hilfe der Chebyshev-Ungleichung folgt, dass n mindestens so gross sein muss, so dass

$$\frac{100\mathrm{Var}[Z_1]}{nE[Z_1]^2} \;=\; \frac{100}{8n} \;=\; \frac{50}{4n} \;\leq\; \frac{1}{10},$$

also

$$n \;\geq\; \frac{500}{4} \;=\; 125.$$

Es folgt, dass $n = 125$ Beutel sicher genügen.

e) Sei $X \sim \mathcal{N}(0,1)$-verteilt. Wegen dem Zentralen Grenzwertsatz gilt

$$\begin{aligned} P\left[S_n \geq x\right] \;&=\; P\left[\frac{S_n - nE[Z_1]}{\sqrt{n\mathrm{Var}[Z_1]}} \geq \frac{x - nE[Z_1]}{\sqrt{n\mathrm{Var}[Z_1]}}\right] \;\approx\; P\left[X \geq \frac{x - nE[Z_1]}{\sqrt{n\mathrm{Var}[Z_1]}}\right] \\ &=\; 1 - \Phi\left(\frac{x - nE[Z_1]}{\sqrt{n\mathrm{Var}[Z_1]}}\right). \end{aligned}$$

Für $n = 50$, $x = 111$ und $E[Z_1] = 2$, $\mathrm{Var}[Z_1] = 1/2$ gilt

$$P\left[S_{50} \geq 111\right] \;\approx\; 1 - \Phi\left(\frac{111 - 100}{5}\right) \;=\; 1 - \Phi\left(2.2\right) \;=\; 1 - 0.9861 \;=\; 0.0139.$$

4. **a)** Da die Daten nach Annahme normalverteilt sind mit unbekannter Varianz und unbekanntem Mittelwert, bietet sich ein t-Test an. Da die Hedgefonds behaupten, dass ihre mittlere Rendite *genau* 7% ist, müssen wir Abweichungen in beide Richtungen untersuchen. Der Test ist also zweiseitig.

b) Nach Annahme sind die $X_1, \ldots, X_{25}$ unter $P_{(\mu,\sigma)}$ $\mathcal{N}(\mu, \sigma^2)$ verteilt mit unbekanntem Mittelwert $\mu \in \mathbb{R}$ und unbekannter Varianz $\sigma^2 > 0$.

(i) Die Nullhypothese und die Alternative lauten

$$H_0 : \mu = \mu_0 := 7 \quad \text{und} \quad H_A : \mu \neq \mu_0.$$

(ii) Die Teststatistik des t-Tests lautet $T = \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} = \frac{\overline{X}_{25} - 7}{S/5}$. Unter H_0 ist T t-verteilt mit 24 Freiheitsgraden.

(iii) Der Verwerfungsbereich hat die Form $K = (-\infty, -c] \cup [c, \infty)$ für ein $c \geq 0$. Damit der Test das Niveau $\alpha = 10\%$ einhält, muss man für c das $(1 - \frac{\alpha}{2})$-Quantil der t_{24}-Verteilung wählen. Aus der Tabelle finden wir $t_{24,0.95} = 1.711$. Somit lautet der gesuchte Verwerfungsbereich $K = (\infty, -1.711] \cup [1.711, \infty)$.

c) Der beobachtete Wert der Teststatistik beträgt $T(\omega) = \frac{\overline{x}_{25} - 7}{s/5} = \frac{9-7}{4/5} = \frac{5}{2} \in K$. Der Testentscheid lautet also, die Nullhypothese auf dem 10%-Niveau zu verwerfen.

d) Da die Varianz $\sigma^2 = 25$ nun bekannt ist, führen wir einen (zweiseitigen) z-Test durch mit der selben Nullhypothese und Alternative wie oben. Die Teststatistik lautet nun $T' = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X}_{25} - 7}{5/5} = \overline{X}_{25} - 7$. Der Verwerfungsbereich ist analog wie oben $K' = (-\infty, -c'] \cup [c', \infty)$ wobei nun $c' = z_{1-\frac{\alpha}{2}} = z_{0.95} \approx 1.645$ das 95%-Quantil der Standardnormalverteilung ist. Da der beobachtete Wert der Teststatistik $T'(\omega) = \overline{x}_{25} - 7 = 2$ im Verwerfungsbereich K' liegt, lautet der Testentscheid wieder, die Nullhypothese auf dem 10%-Niveau zu verwerfen.

e) Der P-Wert ist das kleinste Niveau, auf dem der Test die Nullhypothese gerade noch verwirft. Mit anderen Worten muss man c gerade so wählen, dass $c = |T'(\omega)| = 2 =: t_*$. Das zu diesem Test gehörige Niveau ist dann gerade

$$\begin{aligned} P_{\mu_0}[|T'| \geq t_*] &= 1 - \Phi(t_*) + \Phi(-t_*) = 2(1 - \Phi(t_*)) \approx 2(1 - 0.97725) \\ &= 2 \cdot 0.02275 = 0.0455. \end{aligned}$$