Einführung in die Stochastik

Klausur am 01.02.13 Lösungen

Aufgabe 1:

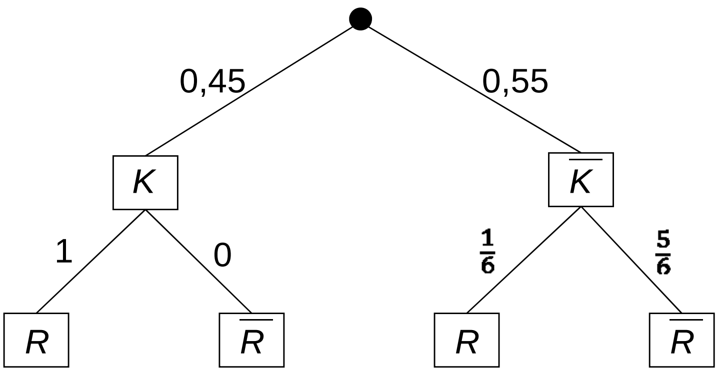
Es sei K das Ereignis, dass Karl die richtige Antwort für eine Aufgabe kennt, und R das Ereignis, dass Karl die richtige Antwort ankreuzt. Gegeben sind die folgenden Wahrscheinlichkeiten

$$P(K) = 0{,}45, \qquad P(R|K) = 1\,, \qquad P(R|\bar{K}) = \frac{1}{6},$$

und somit auch

$$P(\bar{K}) = 0{,}55, \qquad P(\bar{R}|K) = 0\,, \qquad P(\bar{R}|\bar{K}) = \frac{5}{6}.$$

Wahrscheinlichkeitsbaum



a) Die gesuchte Wahrscheinlichkeit für eine falsche Antwort, $P(\bar{R})$, ergibt sich mit Hilfe der Formel für die totale Wahrscheinlichkeit zu:

$$P(\bar{R}) = P(\bar{R}|K) \cdot P(K) + P(\bar{R}|\bar{K}) \cdot P(\bar{K})$$

$$= 0 \cdot 0{,}45 + \frac{5}{6} \cdot 0{,}55 = 0{,}458\bar{3} \approx 45{,}83\%$$

b) Die gesuchte Wahrscheinlichkeit, mit der Karl eine richtige Antwort gewusst und nicht „erwürfelt" hat, $P(K|R)$, ergibt sich zu:

$$P(K|R) = \frac{P(K \cap R)}{P(R)} = \frac{P(R|K) \cdot P(K)}{1 - P(\bar{R})} = \frac{1 \cdot 0{,}45}{1 - 0{,}458\bar{3}}$$

$$\approx 0{,}8308 = 83{,}08\%$$

c) Die Anzahl N der richtig beantworteten Aufgaben ist binomialverteilt mit der Versuchsanzahl $n = 8$ und der Trefferwahrscheinlichkeit

$$p = P(R) = 1 - P(\bar{R}) \approx 0{,}5417$$

Somit ergibt sich die gesuchte Wahrscheinlichkeit für genau vier richtige Antworten zu

$$P(N = 4) = \binom{8}{4} \cdot p^4 \cdot (1-p)^{8-4}$$

$$= \frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1} \cdot 0{,}5417^4 \cdot 0{,}4583^4$$

$$= 70 \cdot 0{,}5417^4 \cdot 0{,}4583^4 = 26{,}59\%$$

Aufgabe 2:

Die $\mathfrak{N}(\vartheta, \sigma^2 = 25)$-verteilte Zufallsgröße T beschreibe den Wert einer Messung der tatsächlichen Ofentemperatur ϑ mittels der Temperatursonde. Das Verfahren arbeitet einwandfrei genau dann, wenn $1064°C < \vartheta < 1083°C$. Wenn $\vartheta < 1064°C$, schmilzt das Gold nicht, und wenn $\vartheta > 1083°C$, schmilzt auch das Kupfer.

Es seien $T_1, \dots, T_n$ die unabhängigen Stichprobenvariablen einer Stichprobe von T. Dann gilt

$$\frac{\overline{T}_n - \vartheta}{\sigma/\sqrt{n}} \sim N(0,1)$$

Als $(1-\alpha)$ −Konfidenzintervall wählen wir

$$I(T_1, \dots, T_n) = \left\{ \vartheta \,\middle|\, -c \leq \frac{\overline{T}_n - \vartheta}{\sigma/\sqrt{n}} \leq c \right\}$$

$$= \left[\ \overline{T}_n - c \cdot \frac{\sigma}{\sqrt{n}}, \overline{T}_n + c \cdot \frac{\sigma}{\sqrt{n}}\ \right]$$

c wird festgelegt durch die Konfidenzwahrscheinlichkeit $1 - \alpha = 0{,}95$:

$$P_\vartheta\big(\vartheta \in I(T_1, \dots, T_n)\big) = P\left(-c \leq \underbrace{\frac{\overline{T}_n - \vartheta}{\sigma/\sqrt{n}}}_{\sim N(0,1)} \leq c \right)$$

$$= \Phi(\mathrm{c}) - \Phi(-\mathrm{c}) = \Phi(\mathrm{c}) - \big(1 - \Phi(\mathrm{c})\big)$$

$$= 2\Phi(\mathrm{c}) - 1 = 1 - \alpha \qquad (\text{mit } \alpha = 0{,}05)$$

$$\Longrightarrow \quad \Phi(\mathrm{c}) = 1 - \frac{\alpha}{2} = 0{,}975$$

$$\Longrightarrow \quad c = u_{1-\frac{\alpha}{2}} = u_{0{,}975} = 1{,}96$$

a) Der aktuelle Wert einer Messung ergab $T = \overline{T}_1 = t = 1073°C$. Hieraus ergibt sich als zweiseitiges 95%-Konfidenzintervall für ϑ mit $n = 1, \sigma = 5$:

$$I(t) = [\ 1073 - 1{,}96 \cdot 5\ ; 1073 + 1{,}96 \cdot 5\ ]$$

$$= [\ 1063{,}2\ ;\ 1082{,}8\ ]$$

Mit einer statistischen Sicherheit von 95% liegt die tatsächliche Temperatur in diesem Intervall.

Da weder 1083°C noch eine höhere Temperatur im Konfidenzbereich liegt, kann mit einer statistischen Sicherheit von 95% davon ausgegangen werden, dass das Kupfer nicht schmilzt, und, da sowohl niedrigere als auch höhere Temperaturwerte als 1064°C im Konfidenzbereich liegen, kann über das Schmelzverhalten des Goldes und somit die einwandfreie Arbeitsweise des Verfahrens keine Aussage mit einer statistischen Sicherheit von 95% gemacht werden, das Gold kann, muss aber nicht schmelzen.

b) Aus dem Konfidenzintervall kann mit einer statistischen Sicherheit von 95% auf die einwandfreie Arbeitsweise des Verfahrens geschlossen werden, wenn $I(t) \subseteq \,]1064{,}1083[$.

Somit muss, damit prinzipiell auf die einwandfreie Arbeitsweise des Verfahrens geschlossen werden kann, die Länge des Konfidenzintervalls kleiner als 1083-1064=19 $[°C]$ sein.

Die Länge des Intervalls basierend auf *n* Beobachtungen beträgt

$$l_n = 2c\frac{\sigma}{\sqrt{n}}$$

sie ist kleiner als $19[°C]$ genau dann, wenn

$$l_n = 2c\frac{\sigma}{\sqrt{n}} = 2 \cdot 1{,}96 \cdot \frac{5}{\sqrt{n}} < 19$$

$$\Leftrightarrow \sqrt{n} > \frac{19{,}6}{19} \Leftrightarrow n > 1{,}06$$

Somit sind mindestens 2 Temperaturmessungen erforderlich.

Aufgabe 3:

Die $\mathfrak{N}(\vartheta = 3, \sigma^2 = 9)$-verteilte Zufallsgröße T beschreibe die sich nach der Reinigung einstellende Temperatur.

a) Die gesuchte Wahrscheinlichkeit $P(T > 9)$ ergibt sich wie folgt:

$$P(T > 9) = P\left(\frac{T-\vartheta}{\sigma} > \frac{9-3}{3}\right) = P\left(\overbrace{\frac{T-\vartheta}{\sigma}}^{\sim N(0,1)} > 2\right) = 1 - \Phi(2)$$

$$= 1 - 0{,}97725$$

$$= 0{,}02275 = 2{,}275\%$$

b) Die gesuchte Wahrscheinlichkeit $P(T < 9)$ ergibt sich wie folgt:

$$P(T < 0) = P\left(\frac{T-\vartheta}{\sigma} < \frac{0-3}{3}\right) = P\left(\overbrace{\frac{T-\vartheta}{\sigma}}^{\sim N(0,1)} < -1\right) = \Phi(-1)$$

$$= 1 - \Phi(1)$$

$$= 1 - 0{,}84134 = 0{,}15866 = 15{,}866\%$$

c) Gesucht wird die Temperatur $t_{0,99}$ mit $P\big(T \le t_{0,99}\big) = 0{,}99$. Diese ergibt sich wie folgt:

$$P\big(T \le t_{0,99}\big) = P\left(\overbrace{\frac{T-\vartheta}{\sigma}}^{\sim N(0,1)} \le \frac{t_{0,99}-3}{3}\right) = \Phi\left(\frac{t_{0,99}-3}{3}\right) = 0{,}99$$

$$\Leftrightarrow \quad \frac{t_{0,99}-3}{3} = u_{0,99} = 2{,}3263$$

$$\Leftrightarrow \quad t_{0,99} = 3 \cdot u_{0,99} + 3 = 3 \cdot 2{,}3263 + 3 = 9{,}9789\ [°C]$$

Aufgabe 4:

a) Das Gerät muss höchstens einmal pro Jahr repariert werden genau dann, wenn $X_1 + X_2 \le 1$. Die gesuchte Wahrscheinlichkeit ergibt sich somit wie folgt:

$$P(X_1 + X_2 \le 1)$$

$$= P(X_1 = 0, X_2 = 0) + P(X_1 = 1, X_2 = 0) + P(X_1 = 0, X_2 = 1)$$

$$\underbrace{=}_{X_1, X_2 \text{ unabh.}} P(X_1 = 0) \cdot P(X_2 = 0) + P(X_1 = 1) \cdot P(X_2 = 0)$$

$$+ P(X_1 = 0) \cdot P(X_2 = 1)$$

$$= 0{,}1 \cdot 0{,}1 + 0{,}1 \cdot 0{,}3 + 0{,}6 \cdot 0{,}1 = 0{,}1$$

b)

$$Y_1 = 3X_1\,, \qquad Y_2 = 2X_2 + 1\,, \qquad Z = Y_1 + Y_2$$

$$E[X_1] = 0 \cdot 0{,}1 + 1 \cdot 0{,}6 + 2 \cdot 0{,}3 = 1{,}2$$

$$E[X_2] = 0 \cdot 0{,}1 + 1 \cdot 0{,}3 + 2 \cdot 0{,}5 + 3 \cdot 0{,}1 = 1{,}6$$

$$E[Z] = E[3X_1 + 2X_2 + 1] \underbrace{=}_{E[.]\text{ linear}} 3E[X_1] + 2E[X_2] + 1$$

$$= 3 \cdot 1{,}2 + 2 \cdot 1{,}6 + 1 = 7{,}8$$

c)

$$Var(Y_1) = Var(3X_1) = 9Var(X_1) = 9 \cdot (E[X_1^2] - E[X_1]^2)$$

$$= 9 \cdot \left((0^2 \cdot 0{,}1 + 1^2 \cdot 0{,}6 + 2^2 \cdot 0{,}3) - 1{,}2^2\right)$$

$$= 9 \cdot (1{,}8 - 1{,}44) = 9 \cdot 0{,}36 = 3{,}24$$

Aufgabe 5:

a) $f_{X,Y}$ ist eine Dichte genau dann, wenn $f_{X,Y} \geq 0$ und $\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy = 1$. Da

$$\iint\limits_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy = \int\limits_0^1 \int\limits_0^1 cxy\,dx\,dy = c\int_0^1 x\,dx \int_0^1 y\,dy$$

$$= c \cdot \frac{x^2}{2}\bigg|_0^1 \cdot \frac{y^2}{2}\bigg|_0^1 = \frac{c}{4} \quad ,$$

liegt eine Dichte genau dann vor, wenn $c = 4$.

b) Für die marginale Dichte von X gilt:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \underbrace{=}_{0 \leq x \leq 1} \int_0^1 4xy\,dy = 4x \cdot \frac{y^2}{2}\bigg|_0^1 = 2x$$

Somit lautet die marginale Dichte von X:

$$f_X(x) = \begin{cases} 2x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

c) Die marginale Verteilungsfunktion von X ergibt sich zu:

$$F_X = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx = \begin{cases} 0 & \text{für } x < 0 \\ x^2 & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}$$

Aufgabe 6:

Die Zufallsgröße Y beschreibe den Bluttyp eines zufällig aus der Population ausgewählten Individuums. Es sei $Y = 1$, wenn MM der Bluttyp des Individuums ist, $Y = 2$ für NM und $Y = 3$ für NN. Die Wahrscheinlichkeitsfunktion von Y lautet dann:

$$p_1 = P(Y = 1) = \theta^2, \;\; p_2 = P(Y = 2) = 2\theta(1 - \theta)\,,$$

$$p_3 = P(Y = 3) = (1 - \theta)^2$$

$Y_1,\ldots,Y_n$ seien die Stichprobenvariablen einer Stichprobe von Y vom Umfang n. Die gemeinsame Wahrscheinlichkeitsfunktion von $Y_1,\ldots,Y_n$ lautet:

$$\begin{aligned} P(Y_1 = y_1, \dots, Y_n = y_n) &= P(Y_1 = y_1) \cdot \dots \cdot P(Y_n = y_n) \\ &= p_{y_1} \cdot \dots \cdot p_{y_n} = p_1^{x_1} \cdot p_2^{x_2} \cdot p_3^{x_3} \\ &= (\theta^2)^{x_1} \big(2\theta(1-\theta)\big)^{x_2} ((1-\theta)^2)^{x_3} \end{aligned}$$

wobei x_1, x_2, x_3 die absoluten Häufigkeiten der drei Bluttypen in der konkreten Stichprobe $y_1, \dots, y_n$ sind. Die Likelihoodfunktion ergibt sich somit zu

$$\begin{aligned} L(\theta; y_1, \dots, y_n) &= \theta^{2x_1} \cdot 2^{x_2} \cdot \theta^{x_2} \cdot (1-\theta)^{x_2} (1-\theta)^{2x_3} \\ &= 2^{x_2} \cdot \theta^{2x_1 + x_2} \cdot (1-\theta)^{x_2 + 2x_3} \to \max_\theta \end{aligned}$$

Für $\theta \neq 0; 1$ kann die Log-Likelihood-Funktion gebildet werden:

$$\begin{aligned} l(\theta; y_1, \dots, y_n) &= \log L(\theta; y_1, \dots, y_n) \\ &= x_1 \log 2 + (2x_1 + x_2) \log \theta + (x_2 + 2x_3) \log(1-\theta) \to \max_\theta \end{aligned}$$

$$\frac{d}{d\theta} l(\theta; y_1, \dots, y_n) = \frac{2x_1 + x_2}{\theta} - \frac{x_2 + 2x_3}{1-\theta} \overset{!}{=} 0$$

$$\Leftrightarrow \ (2x_1 + x_2) \cdot (1-\theta) - \ (x_2 + 2x_3) \cdot \theta = 0$$

$$\Leftrightarrow \ \theta = \frac{2x_1 + x_2}{2x_1 + 2x_2 + 2x_3} = \frac{2x_1 + x_2}{2n}$$

$$\frac{d^2}{d\theta^2} l(\theta; y_1, \dots, y_n) = -\frac{2x_1 + x_2}{\theta^2} - \frac{x_2 + 2x_3}{(1-\theta)^2} < 0 \quad \text{für } 0 < \theta < 1$$

Also ist

$$\hat{\theta} = \frac{2x_1 + x_2}{2n}$$

für $\hat{\theta} \neq 0; 1$ eine lokale Maximalstelle und die einzige lokale Extremumstelle der Log-Likelihoodfunktion und somit auch der Likelihoodfunktion in $]0; 1[$. Da die Likelihoodfunktion auf $[0; 1]$ stetig ist, folgt hieraus, dass

$$\hat{\theta} = \frac{2x_1 + x_2}{2n}$$

für $\hat{\theta} \neq 0; 1$ globale Maximalstelle der Likelihoodfunktion ist.

Für $\hat{\theta} = \frac{2x_1 + x_2}{2n} = 0$, das heißt $x_1 = x_2 = 0$ und $x_3 = n$, ergibt sich

$$L(\theta; y_1, \dots, y_n) = (1-\theta)^{2n} < L(0; y_1, \dots, y_n) = 1 \text{ für } 0 < \theta \leq 1$$

und für $\hat{\theta} = \frac{2x_1 + x_2}{2n} = 1$, das heißt $x_2 = x_3 = 0$ und $x_1 = n$,

$$L(\theta; y_1, \dots, y_n) = \theta^{2n} < L(1; y_1, \dots, y_n) = 1 \text{ für } 0 \leq \theta < 1$$

somit ist $\hat{\theta} = \frac{2x_1 + x_2}{2n}$ auch für $\hat{\theta} = 0; 1$ globale Maximalstelle der Likelihhoodfunktion.

Der Maximum-Likelihood-Schätzer für θ ist somit

$$\hat{\theta}_{ML} = \frac{2x_1 + x_2}{2n}.$$

Aufgabe 7:

Herr Huber erhält n=6 Karten, die rein zufällig aus 12 Karten ausgewählt werden, von denen $r = 4$ vom Typ „Parkett“ und $s = 8$ vom Typ „zweite Reihe“ sind. Die Zufallsgröße X beschreibe die Anzahl der Karten für das Parkett, die Herr Huber in Händen hält. X ist hypergeometrisch verteilt mit der Wahrscheinlichkeitsfunktion

$$P(X = k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} = \frac{\binom{4}{k}\binom{8}{6-k}}{\binom{12}{6}}$$

Die gesuchte Wahrscheinlichkeit ist $P(X \leq 1)$. Für sie ergibt sich.

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= \frac{\binom{4}{0}\binom{8}{6}}{\binom{12}{6}} + \frac{\binom{4}{1}\binom{8}{5}}{\binom{12}{6}} = \frac{\frac{8\cdot 7}{2} + 4 \cdot \frac{8\cdot 7\cdot 6}{3\cdot 2}}{\frac{12\cdot 11\cdot 10\cdot 9\cdot 8\cdot 7}{6\cdot 5\cdot 4\cdot 3\cdot 2}}$$

$$= \frac{4 \cdot 7 + 4 \cdot 8 \cdot 7}{11 \cdot 3 \cdot 4 \cdot 7} = \frac{9}{33} \approx 0{,}2727$$

Aufgabe 8:

Die Dicke einer 1 € Münze wird beschrieben durch eine Zufallsgröße X mit Erwartungswert $E[X] = 1{,}5mm$ und Standardabweichung $\sigma_X = 0{,}2mm$. Die Höhe der 100.000 €-Säule wird beschrieben durch die Zufallsgröße H mit

$$H = \sum_{i=1}^{100.000} X_i \quad ,$$

wobei $X_1, \dots, X_{100.000}$ unabhängige und wie X verteilte Zufallsgrößen sind.

a) Da der Erwartungswertoperator additiv ist, ergibt sich für den Erwartungswert von H

$$\mu_H = E[H] = \sum_{i=1}^{100.000} E[X_i] = 100.000 \cdot 1{,}5mm = 150m \quad ,$$

und da die Zufallsgrößen $X_1, \dots, X_{100.000}$ unabhängig sind, folgt für die Varianz

$$\sigma_H^2 = Var(H) = \sum_{i=1}^{100.000} Var(X_i) = 100.000 \cdot (0{,}2mm)^2$$

$$= 4000mm^2 = 4000 \cdot 10^{-6} m^2 = 0{,}004m^2$$

und für die Standardabweichung

$$\sigma_H = \sqrt{Var(H)} = \sqrt{0{,}004}\, m \approx 0{,}06324555\, m$$

b) Die Tschebyscheffsche Ungleichung liefert für H die Abschätzung

$$P(|H-\mu_H| \geq c) \leq \frac{\sigma_H^2}{c^2}$$

Hieraus folgt

$$P(\mu_H - c < H < \mu_H + c) = 1 - P(|H-\mu_H| \geq c) \geq 1 - \frac{\sigma_H^2}{c^2}$$

Ein Intervall mit

$$P(\mu_H - c < H < \mu_H + c) \geq 0{,}9$$

ergibt sich, wenn c so gewählt wird, dass $1 - \frac{\sigma_H^2}{c^2} \geq 0{,}9$. Da

$$1 - \frac{\sigma_H^2}{c^2} \geq 0{,}9 \quad \Leftrightarrow \quad \frac{\sigma_H^2}{c^2} \leq 0{,}1 \quad \Leftrightarrow \quad c^2 \geq 10\sigma_H^2 = 0{,}04\,m^2$$

Mit $c = 0{,}2m$ ergibt sich ein möglichst kleines Intervall, dieses ist $[149{,}8; 150{,}2]$.

c) Gesucht wird der Wert h_{max} mit $P(H \leq h_{max}) = 0{,}9999$.
Aufgrund des zentralen Grenzwertsatzes gilt

$$\frac{H-\mu_H}{\sigma_H} \underbrace{\sim}_{as} \mathfrak{N}(0,1)$$

Somit ergibt sich

$$0{,}9999 = P(H \leq h_{max}) \underbrace{=}_{\text{Norm.}} P\left(\frac{H-\mu_H}{\sigma_H} \leq \frac{h_{max}-\mu_H}{\sigma_H}\right)$$

$$\approx \Phi\left(\frac{h_{max}-\mu_H}{\sigma_H}\right)$$

Hieraus folgt

$$\frac{h_{max}-\mu_H}{\sigma_H} = u_{0,9999} = 3{,}719$$

$$h_{max} = \mu_H + u_{0,9999}\sigma_H = 150m + 3{,}719 \cdot \sqrt{0{,}004}m$$
$$= 150{,}235m$$