Prof. Dr. U. Kamps
C. Empacher, D. Rauwolf RWTH Aachen, SoSe 2022

23. Juni 2022

# Einführung in die angewandte Stochastik

**5. Globalübung - Lösungen**

## Aufgabe 23

(a) Die vollständig ausgefüllte Tabelle sieht folgendermaßen aus:

| $P(X=i,Y=j)$ | | $j$ | | | | | $P(X=i)$ |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | |
| $i$ | 1 | 0 | 0,1 | 0 | 0,1 | 0,2 | 0,4 |
| | 2 | 0,1 | 0,1 | 0,1 | 0,2 | 0,1 | 0,6 |
| $P(Y=j)$ | | 0,1 | 0,2 | 0,1 | 0,3 | 0,3 | |

(b) Unter Benutzung der Randzähldichten aus der Tabelle in (a) sowie Definition C 5.1 für Erwartungswerte von diskret verteilten Zufallsvariablen erhalten wir

$$\mathrm{E}X = 1 \cdot P(X=1) + 2 \cdot P(X=2) = 1 \cdot 0{,}4 + 2 \cdot 0{,}6 = 1{,}6$$

und analog

$$\begin{aligned}\mathrm{E}Y &= 1 \cdot P(Y=1) + 2 \cdot P(Y=2) + 3 \cdot P(Y=3) + 4 \cdot P(Y=4) + 5 \cdot P(Y=5)\\ &= 1 \cdot 0{,}1 + 2 \cdot 0{,}2 + 3 \cdot 0{,}1 + 4 \cdot 0{,}3 + 5 \cdot 0{,}3 = 3{,}5.\end{aligned}$$

Für den Erwartungswert der Zufallsvariable $XY$ wenden wir Satz C 5.6 der Vorlesung auf den Zufallsvektor $(X,Y)$ und die Funktion

$$g : \mathbb{R}^2 \to \mathbb{R}, \quad g(x,y) = xy$$

an und erhalten mit Hilfe der gemeinsamen Zähldichte von $X$ und $Y$ gemäß der Tabelle aus (a)

$$\begin{aligned}\mathrm{E}(XY) &= \sum_{i=1}^{2}\sum_{j=1}^{5} g(i,j)P(X=i,Y=j) = \sum_{i=1}^{2}\sum_{j=1}^{5} ijP(X=i,Y=j)\\ &= 1\cdot 1\cdot 0 + 1\cdot 2\cdot 0{,}1 + 1\cdot 3\cdot 0 + 1\cdot 4\cdot 0{,}1 + 1\cdot 5\cdot 0{,}2\\ &\quad + 2\cdot 1\cdot 0{,}1 + 2\cdot 2\cdot 0{,}1 + 2\cdot 3\cdot 0{,}1 + 2\cdot 4\cdot 0{,}2 + 2\cdot 5\cdot 0{,}1\\ &= 5{,}4.\end{aligned}$$

(c) Mit den Erwartungswerten aus Teil (b) und der Formel für die Kovarianz aus Lemma C 5.16 der Vorlesung erhalten wir schließlich

$$\mathrm{Kov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y = 5{,}4 - 1{,}6 \cdot 3{,}5 = 5{,}4 - 5{,}6 = -0{,}2.$$

Unter erneuter Benutzung von Lemma C 5.16 folgern wir aus $\mathrm{Kov}(X,Y) \neq 0$, dass $X$ und $Y$ nicht stochastisch unabhängig sind.

**Bemerkung:** Die stochastische Abhängigkeit von $X$ und $Y$ hätte man auch einfacher ohne die Berechnung der Kovarianz $\mathrm{Kov}(X,Y)$ zeigen können, indem man Lemma C 1.10 benutzt und aus der Tabelle aus (a) abliest, dass z.B.

$$0 = P(X = 1, Y = 1) \neq P(X = 1)P(Y = 1) = 0{,}4 \cdot 0{,}1 = 0{,}04.$$

## Aufgabe 24

Zunächst gilt gemäß C 5.2 und C 5.13:

$$\mathrm{E}(X) = \frac{-2+4}{2} = 1 \quad \text{und} \quad \mathrm{E}(Y) = \frac{1}{0{,}5} = 2 \tag{1}$$

sowie

$$\mathrm{Var}(X) = \frac{(4-(-2))^2}{12} = \frac{36}{12} = 3 \quad \text{und} \quad \mathrm{Var}(Y) = \frac{1}{0{,}5^2} = 4\,. \tag{2}$$

(a) Mit der Linearität des Erwartungswertes gemäß Lemma C 5.7 erhält man

$$\mathrm{E}(U) = \mathrm{E}(X - 2\,Y) \overset{\text{Lin.}}{=} \mathrm{E}(X) - 2\,\mathrm{E}(Y) \overset{(1)}{=} 1 - 4 = -3\,,$$

$$\mathrm{E}(V) = \mathrm{E}(3\,X + Y - 5) \overset{\text{Lin.}}{=} 3\,\mathrm{E}(X) + \mathrm{E}(Y) - 5 \overset{(1)}{=} 3 + 2 - 5 = 0.$$

(b) Zunächst folgt aus der Unkorreliertheit von $X$ und $Y$ gemäß Bezeichnung C 5.17, dass $\mathrm{Kov}(X,Y) = 0$. Des Weiteren folgt mit den Lemmas C 5.12, C 5.15 und C 5.16, dass

$$\begin{aligned}
\mathrm{Var}(U) = \mathrm{Var}(X - 2\,Y) &\overset{\text{C 5.15}}{=} \mathrm{Var}(X) + \mathrm{Var}(-2Y) + 2\,\mathrm{Kov}(X, -2Y) \\
&\overset{\text{C 5.12}}{=} \mathrm{Var}(X) + 4\,\mathrm{Var}(Y) + 2\,\mathrm{Kov}(X, -2Y) \\
&\overset{\text{C 5.16}}{=} \mathrm{Var}(X) + 4\,\mathrm{Var}(Y) - 4\,\mathrm{Kov}(X,Y) \\
&\underset{\text{unkorr.}}{\overset{X,Y}{=}} \mathrm{Var}(X) + 4\,\mathrm{Var}(Y) \overset{(2)}{=} 3 + 16 = 19
\end{aligned}$$

und analog

$$\begin{aligned}
\mathrm{Var}(V) = \mathrm{Var}(3\,X + Y - 5) &\overset{\text{C 5.12}}{=} \mathrm{Var}(3\,X + Y) \\
&\overset{\text{C 5.15}}{=} \mathrm{Var}(3X) + \mathrm{Var}(Y) + 2\,\mathrm{Kov}(3X,Y) \\
&\overset{\text{C 5.12}}{=} 9\,\mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Kov}(3X,Y) \\
&\overset{\text{C 5.16}}{=} 9\,\mathrm{Var}(X) + \mathrm{Var}(Y) + 6\,\mathrm{Kov}(X,Y) \\
&\underset{\text{unkorr.}}{\overset{X,Y}{=}} 9\,\mathrm{Var}(X) + \mathrm{Var}(Y) = 27 + 4 = 31.
\end{aligned}$$

(c) Mit der Linearität der Kovarianz aus Lemma C 5.16 erhalten wir schließlich

$$\mathrm{Kov}(U,X) = \mathrm{Kov}(X - 2\,Y,X) \overset{\text{Lin.}}{=} \underbrace{\mathrm{Kov}(X,X)}_{=\mathrm{Var}(X)} - 2 \cdot \underbrace{\mathrm{Kov}(Y,X)}_{=0,\text{ da } X,Y \text{ unkorreliert}} = \mathrm{Var}(X) \overset{(2)}{=} 3\,.$$

## Bemerkung zur Linearität in (c):

Aus Lemma C 5.16 folgt, dass allgemein für Zufallsvariablen $X_1, \dots, X_n$ , $Y_1, \dots, Y_m$ und $a_1, \dots ,a_n$ , $b_1, \dots ,b_m \in \mathbb{R}$

$$\operatorname{Kov}\left(\sum_{i=1}^n a_i X_i \, , \, \sum_{j=1}^m b_j Y_j\right) \;=\; \sum_{i=1}^n \sum_{j=1}^m a_i \, b_j \operatorname{Kov}(X_i,Y_j)$$

gilt, falls die einzelnen Kovarianzen alle (als reelle Zahlen) existieren.

## Aufgabe 25

(a) Laut Bezeichnung B 2.5 besitzt $X$ die Zähldichte $p^X$ gegeben durch

$$p_k^X \;=\; e^{-\lambda} \frac{\lambda^k}{k!}$$

für $k \in \mathbb{N}_0$. Gemäß Definition C 6.1 bzw. Bemerkung C 6.2 ist die wahrscheinlichkeitserzeugende Funktion $g_X$ von $X$ für $t \in \mathbb{R}$ definiert durch

$$\begin{aligned} g_X(t) &\overset{\text{C 6.1}}{=} \mathrm{E}t^X \overset{\text{C 6.2}}{=} \sum_{k=0}^\infty t^k p_k^X \overset{\text{B 2.5}}{=} \sum_{k=0}^\infty t^k \cdot e^{-\lambda} \frac{\lambda^k}{k!} \;=\; \sum_{k=0}^\infty e^{-\lambda} \frac{(\lambda t)^k}{k!} \\ &= e^{-\lambda} \underbrace{\sum_{k=0}^\infty \frac{(\lambda t)^k}{k!}}_{= e^{\lambda t} \text{ (Exponentialreihe)}} \;=\; e^{-\lambda} \cdot e^{\lambda t} \;=\; e^{-\lambda + \lambda t} \;=\; e^{\lambda(t-1)}, \end{aligned}$$

da die obige Exponentialreihe für alle $t \in \mathbb{R}$ konvergiert.

(b) (i) Wir bestimmen die momenterzeugende Funktion $h_Y$ gemäß Vorlesung (vgl. C 6.5). Für $t \in \mathbb{R}$ erhalten wir demnach

$$\begin{aligned} h_Y(t) &= \mathrm{E}(e^{tY}) \;=\; \sum_{k=0}^n e^{tk} \cdot \binom{n}{k} p^k (1-p)^{n-k} \;=\; \sum_{k=0}^n \binom{n}{k} \left(e^t p\right)^k (1-p)^{n-k} \\ &= \left(e^t p + 1 - p\right)^n . \end{aligned} \tag{1}$$

(ii) Gemäß (1) ist $h_Y$ (beliebig oft) differenzierbar bezüglich $t \in \mathbb{R}$ mit

$$h_Y'(t) \;=\; n\left(e^t p + 1 - p\right)^{n-1} e^t p, \quad t \in \mathbb{R}, \tag{2}$$

$$h_Y''(t) \;=\; n(n-1)\left(e^t p + 1 - p\right)^{n-2} \left(e^t p\right)^2 + n\left(e^t p + 1 - p\right)^{n-1} e^t p, \quad t \in \mathbb{R}. \tag{3}$$

Es folgt mit Satz C 6.7 und wegen $e^0 = 1$, dass

$$\mathrm{E}(Y) \;=\; h_Y'(0) \overset{(2)}{=} n(p+1-p)^{n-1} p \;=\; np, \tag{4}$$

$$\begin{aligned} \mathrm{E}(Y^2) &= h_Y''(0) \overset{(3)}{=} n(n-1)\,(p+1-p)^{n-2} p^2 + n\,(p+1-p)^{n-1}\, p \\ &= n(n-1)p^2 + np. \end{aligned} \tag{5}$$

Die Varianz von $Y$ bestimmt man daraus gemäß Lemma C 5.12 zu

$$\operatorname{Var}(Y) = \mathrm{E}Y^2 - \mathrm{E}^2 Y \overset{(4),(5)}{=} n^2p^2 - np^2 + np - n^2p^2 \;=\; np(1-p).$$

## Aufgabe 26

(a) Für $y > 0$ gilt (analog zu Aufgabe 22 (b)(i)):

$$f^Y(y) \overset{\text{C 3.12}}{=} \int_{-\infty}^{\infty} f^{(X,Y)}(x,y)\, dx \overset{\text{Vor.}}{=} \int_0^{\infty} \frac{y^s}{(s-1)!} \underbrace{e^{-(x+1)y}}_{=e^{-xy}\, e^{-y}} dx$$

$$= \underbrace{\frac{y^s\, e^{-y}}{(s-1)!}}_{\text{unabh. von } x} \int_0^{\infty} e^{-xy}\, dx \overset{y \geq 0}{=} \frac{y^s\, e^{-y}}{(s-1)!} \frac{1}{y} \underbrace{\int_0^{\infty} \underbrace{y\, e^{-yx}}_{\text{Dichte zu Exp}(y)} dx}_{=1}$$

$$= \frac{1}{(s-1)!} y^{s-1} e^{-y} \overset{\Gamma(s) = (s-1)!}{=} \frac{1}{\Gamma(s)} y^{s-1} e^{-y} . \tag{1}$$

Und für $y \leq 0$ gilt:

$$f^Y(y) \overset{\text{C 3.12}}{=} \int_{-\infty}^{\infty} \underbrace{f^{(X,Y)}(x,y)}_{=0 \text{ nach Vor.}} dx = 0 . \tag{2}$$

Insgesamt erhält man aus (1),(2):

$$f^Y(y) = \begin{cases} \frac{1}{\Gamma(s)} y^{s-1} e^{-y} & , \quad y > 0 , \\ 0 & , \quad y \leq 0 . \end{cases} \tag{3}$$

Aus (3) folgt mit B 3.9, dass $Y$ einer Gammaverteilung folgt: $Y \sim \Gamma(1,s)$ .

**Alternativ-Lösung zu** (1)

Zunächst gilt:

$$\int_0^{\infty} e^{-xy}\, dx = \Big[ -\frac{1}{y} e^{-yx} \Big]_{x=0}^{x=\infty} = \frac{1}{y} \big( - \underbrace{\lim_{x\to\infty} e^{-yx}}_{=0 \text{ mit } y>0} + e^0 \big) = \frac{1}{y} .$$

Hiermit erhält man:

$$f^Y(y) \overset{\text{s.o.}}{=} \frac{y^s\, e^{-y}}{(s-1)!} \int_0^{\infty} e^{-xy}\, dx = \frac{y^s\, e^{-y}}{(s-1)!} \frac{1}{y} = \frac{1}{(s-1)!} y^{s-1} e^{-y} .$$

(b) Sei $y > 0$. Dann gilt gemäß (1) bzw. (3):

$$f^Y(y) = \underbrace{\frac{1}{(s-1)!}}_{>0} \underbrace{y^{s-1}}_{>0 \text{ für } y>0} \underbrace{e^{-y}}_{>0} > 0 . \tag{4}$$

Mit der Definition C 7.1 von bedingten Dichten folgt für $x > 0$

$$f^{X|Y=y}(x) \overset{\text{C 7.1,(4)}}{=} \frac{f^{(X,Y)}(x,y)}{f^Y(y)} \overset{\text{Vor.,(1)}}{=} \frac{\frac{y^s\, e^{-xy}\, e^{-y}}{(s-1)!}}{\frac{y^{s-1}\, e^{-y}}{(s-1)!}} = y\, e^{-xy} \tag{5}$$

und für $x \le 0$

$$f^{X|Y=y}(x) \overset{\text{C 7.1,(4)}}{=} \frac{f^{(X,Y)}(x,y)}{f^Y(y)} \overset{\text{Vor.}}{=} \frac{0}{f^Y(y)} = 0\,. \tag{6}$$

Insgesamt erhält man aus (5),(6):

$$f^{X|Y=y}(x) = \begin{cases} y\,e^{-yx} & , \quad x > 0\,, \\ 0 & , \quad x \le 0\,. \end{cases} \tag{7}$$

Gemäß B 3.7 ist dies die Dichte einer Exponentialverteilung mit Parameter $y$.

(c) Für $y > 0$ folgt mit der Definition C 7.4 des bedingten Erwartungswertes

$$\mathrm{E}(X|Y=y) \overset{\text{C 7.4}}{=} \int_{-\infty}^{\infty} x\, f^{X|Y=y}(x)\,dx \overset{(7)}{=} \underbrace{\int_0^{\infty} x\,y\,e^{-yx}\,dx}_{=\mathrm{E}(Z) \text{ mit } Z\sim\mathrm{Exp}(y)} \overset{\text{C 5.2,(iv)}}{=} \frac{1}{y}\,. \tag{8}$$

**Alternativ-Lösung zum letzten Schritt in (8)**
(mittels partieller Integration)

$$\begin{aligned}\int_0^{\infty} \underbrace{x}_{=v(x)}\; \underbrace{y\,e^{-yx}}_{=u'(x)}\,dx \overset{\text{Part. Int.}}{=} & \Big[\underbrace{x}_{=v(x)}\,\underbrace{\big(-e^{-yx}\big)}_{=u(x)}\Big]_{x=0}^{x=\infty} - \int_0^{\infty} \underbrace{1}_{=v'(x)}\,\underbrace{\big(-e^{-yx}\big)}_{=u(x)}\,dx \\ = & -\lim_{x\to\infty}\big(x\,e^{-yx}\big) + \int_0^{\infty} e^{-yx}\,dx \\ \overset{\text{L' Hosp.}}{=} & \;0 + \Big[-\frac{1}{y}\,e^{-yx}\Big]_{x=0}^{x=\infty} = \frac{1}{y}\Big(-\underbrace{\lim_{x\to\infty} e^{-yx}}_{=0 \text{ mit } y>0} + e^0\Big) = \frac{1}{y}\,.\end{aligned}$$

## Aufgabe 27

**Modell:**

Die Störungen können beschrieben werden durch Zufallsvariablen $X_1,\dots,X_{200}$ mit folgender Interpretation für $i \in \{1,\dots,200\}$:

$$X_i = \begin{cases} 1 & , \quad i\text{-ter Automat fällt mindestens einmal aus}\,, \\ 0 & , \quad \text{sonst}\,. \end{cases}$$

Gemäß Aufgabenstellung sind die Zufallsvariablen $X_1,\dots,X_{200}$ stochastisch unabhängig mit

$$X_i \sim \mathrm{bin}(1, 0{,}05) \quad \text{für} \quad i \in \{1,\dots,200\}\,.$$

Damit sind $X_1,\dots,X_{200}$ stochastisch unabhängige, identisch verteilte (kurz: i.i.d.) Bernoulli-Zufallsvariablen mit „Trefferwahrscheinlichkeit“ $p = 0{,}05$.

Die gesamte Anzahl der Geldautomaten, die innerhalb einer Woche mindestens eine Störung aufweisen, ist dann gegeben durch

$$S_n = \sum_{i=1}^{n} X_i \quad \text{mit} \quad n = 200\,.$$

Wir berechnen zunächst Erwartungswert und Varianz von $S_n$ gemäß Vorlesung zu

$$\mathrm{E}(S_n) \overset{\text{Lin.}}{=} \sum_{i=1}^{n} \underbrace{E(X_i)}_{=p} \overset{\text{C 5.2,1.}}{=} n\,p \;=\; 200 \cdot \frac{1}{20} \;=\; 10\,, \tag{1}$$

$$\mathrm{Var}(S_n) \underset{\text{C 5.19}}{\overset{\text{Unabh.}}{=}} \sum_{i=1}^{n} \underbrace{\mathrm{Var}(X_i)}_{=p\,(1-p)} \overset{\text{C 5.13,1.}}{=} n\,p\,(1-p) \;=\; 200 \cdot \frac{1}{20} \cdot \frac{19}{20} \;=\; \frac{19}{2}\,. \tag{2}$$

Gesucht (gemäß Aufgabenstellung): $P\big(5 \le S_n \le 15\big)$.

(a) Wir nutzen die Tschebyscheff-Ungleichung aus Satz C 5.23 und erhalten mit $n = 200$:

$$\begin{aligned}
P\big(5 \le S_n \le 15\big) \;&=\; P\big(5 - \mathrm{E}(S_n) \le S_n - \mathrm{E}(S_n) \le 15 - \mathrm{E}(S_n)\big) \\
&\overset{(1)}{=}\; P\big(-5 \le S_n - \mathrm{E}(S_n) \le 5\big) \;=\; P\big(|S_n - \mathrm{E}(S_n)| \le 5\big) \\
&=\; 1 - P\big(|S_n - \mathrm{E}(S_n)| > 5\big) = 1 - \underbrace{P\big(|S_n - \mathrm{E}(S_n)| \ge 6\big)}_{\underset{\text{Ungl.}}{\overset{\text{Tscheb.-}}{\le}} \frac{\mathrm{Var}(S_n)}{6^2}} \\
&\ge\; 1 - \frac{\mathrm{Var}(S_n)}{6^2} \overset{(2)}{=} 1 - \frac{19}{72} \;=\; \frac{53}{72} \;\approx\; 0{,}74\,.
\end{aligned}$$

(b) Es bezeichne $\Phi$ die Verteilungsfunktion der Standard-Normalverteilung $\mathrm{N}(0,1)$. Dann gilt unter Benutzung des zentralen Grenzwertsatzes C 8.6 wiederum mit $n = 200$:

$$\begin{aligned}
P\big(5 \le S_n \le 15\big) \;&=\; P\big(S_n \le 15\big) - P\big(S_n < 5\big) \\
&\overset{S_n \in \mathbb{N}_0}{=}\; P\big(S_n \le 15\big) - P\big(S_n \le 4\big) \\
&\overset{\text{Stand.}}{=}\; P\left(\frac{S_n - n\,p}{\sqrt{n\,p\,(1-p)}} \le \frac{15 - n\,p}{\sqrt{n\,p\,(1-p)}}\right) \\
&\qquad - P\left(\frac{S_n - n\,p}{\sqrt{n\,p\,(1-p)}} \le \frac{4 - n\,p}{\sqrt{n\,p\,(1-p)}}\right) \\
&\underset{\text{(C 8.6)}}{\overset{\text{ZGWS}}{\approx}}\; \Phi\left(\frac{15 - n\,p}{\sqrt{n\,p\,(1-p)}}\right) - \Phi\left(\frac{4 - n\,p}{\sqrt{n\,p\,(1-p)}}\right) \\
&\overset{(1),(2)}{=}\; \Phi\left(\frac{15 - 10}{\sqrt{19/2}}\right) - \Phi\left(\frac{4 - 10}{\sqrt{19/2}}\right) \\
&\approx\; \Phi(1{,}62) - \Phi(-1{,}95) \overset{\text{Symm.}}{=} \Phi(1.62) - \big(1 - \Phi(1{,}95)\big) \\
&\overset{\text{Tab.}}{\approx}\; 0{,}947 - 1 + 0{,}974 \;=\; 0{,}921\,.
\end{aligned}$$

**Bemerkungen zu Aufgabe 27:**

(i) Bei der praktischen Anwendung des Zentralen Grenzwertsatzes stellt sich (ebenso wie bei anderen Grenzwertsätzen) die Frage, ab welchem Stichprobenumfang $n$ die Approximation gerechtfertigt, d.h. hinreichend genau ist.

Für die hier betrachtete Situation mit stochastisch unabhängigen, identisch verteilten Bernoulli-Zufallsvariablen mit „Trefferwahrscheinlichkeit" $p$ (Zentraler Grenzwertsatz von de Moivre-Laplace) findet man in der Literatur häufig das folgende Kriterium:

Die Approximation ist hinreichend genau, falls folgende Bedingung erfüllt ist:

$$n\,p\,(1-p) \;\geq\; 9\,.$$

Hier gilt:

$$n\,p\,(1-p) \;=\; 200\cdot\frac{1}{20}\cdot\frac{19}{20} \;=\; \frac{19}{2} \;=\; 9{,}5 \;\geq\; 9\,.$$

Somit ist die Anwendung des Zentralen Grenzwertsatzes in Aufgabe 27 gerechtfertigt.

(ii) Die Tschebyscheff-Ungleichung liefert oftmals nur eine sehr grobe Abschätzung, d.h. eine sehr ungenaue Approximation für die wirkliche Wahrscheinlichkeit.

Dies zeigt auch der Vergleich des in Aufgabe 27 (a) berechneten Zahlenwerts 0,74 mit der exakten Wahrscheinlichkeit

$$P(5 \leq S_n \leq 15) \;=\; 0{,}929\,.$$

Der in Aufgabe 27 (b) mit Hilfe des Zentralen Grenzwertsatzes berechnete Zahlenwert 0,921 liefert hingegen eine sehr gute Näherung der exakten Wahrscheinlichkeit.