Wahrscheinlichkeit und Statistik

Lösungen Serie 7

Lösung 7-1. (a) Die Dichte muss zu eins aufintegrieren, deshalb gilt:

$$\int_7^{10} c\,(x-7)^2\,dx \overset{!}{=} 1$$

1.Version: Polynom ausmultiplizieren:

$$\begin{aligned}
\int_7^{10} c\,(x-7)^2\,dx &= c \cdot \int_7^{10} x^2 - 14x + 49dx \\
&= c \cdot \left(\frac{1}{3}x^3 - \frac{14}{2}x^2 + 49x\right)\Big|_7^{10} \\
&= \left(\frac{1}{3}10^3 - 7 \cdot 10^2 + 49 \cdot 10 - \frac{1}{3}7^3 + 7 \cdot 7^2 - 49 \cdot 7\right) \\
&= c \cdot 9
\end{aligned}$$

2.Version: Mit Substitution:

$$\begin{aligned}
\int_7^{10} c\,(x-7)^2\,dx &\overset{y:=x-7}{=} c \cdot \int_0^3 y^2 dx \\
&= c \cdot \frac{1}{3}y^3\Big|_0^3 \\
&= c \cdot \frac{1}{3}\left(3^3 - 0\right) \\
&= c \cdot 3^2 \\
&= c \cdot 9
\end{aligned}$$

d.h. $c = \frac{1}{9}$.

(b) $F(x) = P[X \leq x] = \int_7^x c(z-7)^2 dz = c\frac{(x-7)^3}{3} = \frac{(x-7)^3}{27}$ $(7 \leq x \leq 10)$.

(c) $P[8 < X < 9] = F(9) - F(8) = \frac{2^3}{3c} - \frac{1^3}{3c} = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$.

Lösung 7-2. (a) Die Fläche von Q ist $4 \cdot \frac{1 \cdot 1}{2} = 2$.

Dann ist die gemeinsame Dichte $f_{X,Y}(x,y)$:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{falls } (x,y) \in Q \\ 0 & \text{sonst} \end{cases}$$

(b) Randdichte $f_X(x)$: 2 Fälle zu unterscheiden.

- Für $-1 \leq x \leq 0$ gilt

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-1-x}^{1+x} \frac{1}{2}dy = \frac{1}{2}y\Big|_{-1-x}^{1+x} = \frac{1}{2}(1+x+1+x) = 1+x$$

Siehe nächste Seite!

- Für $0 \leq x \leq 1$ gilt

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-1+x}^{1-x} \frac{1}{2}dy = \frac{1}{2}y\Big|_{-1+x}^{1-x} = \frac{1}{2}(1-x+1-x) = 1-x$$

D.h.:

$$f_X(x) = \begin{cases} 1-|x| & \text{falls } -1 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Randdichte $f_Y(y)$: analog dank der Symmetrie.
Oder berechne wie früher (2 Fälle zu unterscheiden).

- Für $-1 \leq y \leq 0$ gilt

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-1-y}^{1+y} \frac{1}{2}dx = \frac{1}{2}x\Big|_{-1-y}^{1+y} = \frac{1}{2}(1+y+1+y) = 1+y$$

- Für $0 \leq y \leq 1$ gilt

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-1+y}^{1-y} \frac{1}{2}dx = \frac{1}{2}x\Big|_{-1+y}^{1-y} = \frac{1}{2}(1-y+1-y) = 1-y$$

D.h.:

$$f_Y(y) = \begin{cases} 1-|y| & \text{falls } -1 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

(c) Wir benutzen, dass gilt $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-1}^{1} x(1-|x|)dx = 0,$$

denn $x(1-|x|)$ ist eine ungerade Funktion. Analog gilt $E[Y] = 0$.

$$\begin{aligned}
E[XY] &= \int\int_Q xy f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\frac{1}{2}\mathbb{1}_{\{(x,y)\in Q\}}dxdy \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\mathbb{1}_{\{-1\leq x\leq 1 \text{ und } -(1-|x|)\leq y\leq(1-|x|)\}}dxdy \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\mathbb{1}_{\{-1\leq x\leq 1\}}\mathbb{1}_{\{-(1-|x|)\leq y\leq(1-|x|)\}}dxdy \\
&= \frac{1}{2}\int_{-\infty}^{\infty} x\mathbb{1}_{\{-1\leq x\leq 1\}}dx\left(\int_{-\infty}^{\infty} y\mathbb{1}_{\{-(1-|x|)\leq y\leq(1-|x|)\}}dy\right) \\
&= \frac{1}{2}\int_{-1}^{1} xdx\underbrace{\left(\int_{-(1-|x|)}^{1-|x|} ydy\right)}_{=0},
\end{aligned}$$

denn y ist eine ungerade Funktion. Damit gilt $\mathrm{Cov}(X,Y) = 0$.

(d) Wegen $f_{X,Y}(x,y) \neq f_X(x)\cdot f_Y(y)$ sind die Zufallsvariablen X und Y nach **c)** zwar unkorreliert aber dennoch abhängig.

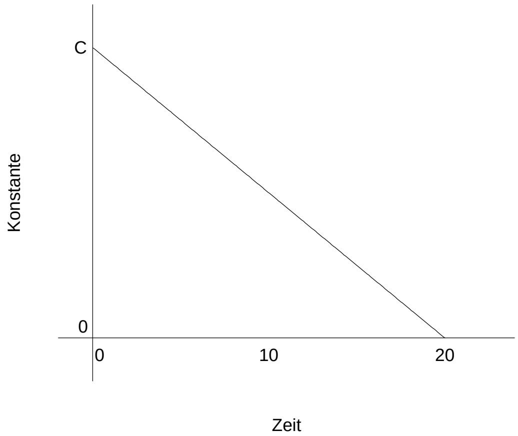
Siehe nächstes Blatt!

(e) Dank der Symmetrie kann man schliessen, dass in diesem Fall die Randdichten gleich und konstant auf dem Intervall $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ sind.
D.h.:

$$f_X(x) = f_Y(x) = \begin{cases} \frac{1}{\sqrt{2}} & \text{falls } -\frac{1}{\sqrt{2}} \leq x \leq \frac{1}{\sqrt{2}} \\ 0 & \text{sonst} \end{cases}$$

Wegen $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ sind die Zufallsvariablen X und Y unabhängig (und somit auch unkorreliert).

Lösung 7-3. (a) Gesucht ist die Dichte der Wartezeit T:



Da f eine lineare Funktion ist, machen wir den folgenden Ansatz

$$f(t) = \begin{cases} at+b, & 0 \leq t \leq 20 \\ 0, & \text{sonst} \end{cases} = \begin{cases} -\frac{c}{20}t + c & 0 \leq t \leq 20 \\ 0 & \text{sonst.} \end{cases}$$

Als nächstes bestimmen wir die Konstante c, sodass $\int_{-\infty}^{\infty} f(t)\,dt = 1$ gilt. Also:

$$\int_{-\infty}^{\infty} f(t)\,dt = \int_0^{20} \left(-\frac{c}{20}t + c\right)\,dt = \left(-\frac{c}{40}t^2 + tc\right)\Big|_0^{20} = c \cdot 10.$$

Also bekommen wir $c = \frac{1}{10}$.
Oder direkte Berechnung von c: Die Fläche unter der Kurve muss gleich 1 sein:

$$c \cdot 20 \cdot \frac{1}{2} = 1 \Rightarrow c = \frac{1}{10}$$

Als nächstes berechnen wir den Erwartungswert der Wartezeit T:

$$\begin{aligned} E[T] &= \int_0^{20} t\left(-\tfrac{1}{200}\,t + \tfrac{1}{10}\right)dt &&= \int_0^{20}\left(-\tfrac{1}{200}\,t^2 + \tfrac{1}{10}\,t\right)dt \\ &= \left(-\tfrac{1}{600}\,t^3 + \tfrac{1}{20}\,t^2\right)\big|_0^{20} &&= 6\tfrac{2}{3} \end{aligned}$$

$$\mathrm{Var}[T] = E[T^2] - E[T]^2$$

$$\begin{aligned} E[T^2] &= \int_0^{20} t^2\left(-\tfrac{1}{200}t + \tfrac{1}{10}\right)dt &&= \int_0^{20}\left(-\tfrac{1}{200}t^3 + \tfrac{1}{10}t^2\right)dt \\ &= \left(-\tfrac{1}{800}\,t^4 + \tfrac{1}{30}\,t^3\right)\big|_0^{20} &&= 66\tfrac{2}{3} \end{aligned}$$

Siehe nächste Seite!

Also bekommen wir: $\mathrm{Var}[T] = 66\frac{2}{3} - \left(6\frac{2}{3}\right)^2 = 22.22.$

(b) Wir haben 10 Baustellen mit den Wartezeiten $T_1, T_2, \ldots, T_{10}$, wobei T_i, $i = 1, \ldots, 10$, verteilt ist wie T. Der totale mittlere Zeitverlust (=totale mittlere Wartezeit) ist also:

$$\begin{aligned} E[T_1 + T_2 + \cdots + T_{10}] &= E[T_1] + E[T_2] + \cdots + E[T_{10}] \\ &= 10 \cdot 6\tfrac{2}{3} = 66\tfrac{2}{3}. \end{aligned}$$

Im Gegensatz zum Erwartungswert kann man über die Varianz von $T_1 + T_2 + \cdots + T_{10}$ ohne zusätzliche Annahmen, wie zum Beispiel Unabhängigkeit der T_i's, nichts aussagen.

Lösung Challenge Serie 7. Nehmen wir der Einfachheit halber an, dass die Insassen die Kennzahlen von 1 bis 100 besitzen. Wir bezeichnen mit τ die Permutation der Kennnummern in den Boxen. D.h. dass z.B. in der k-ten Kiste die Kennnummer $\tau(k)$ enthalten ist.

Die Idee der Strategie ist folgende: Die Insassen einigen sich auf eine zufällige Scheinnummerierung der Kisten. D.h. man einigt sich auf eine Permutation σ der Zahlen von 1 bis 100 und ordnet der k-ten Kiste die Zahl $\sigma(k)$ zu, $k = 1, \ldots, 100$.

Der Insasse k folgt nun folgendem Suchalgorithmus:

- Gehe zur Kiste mit der Scheinnummer k (das wäre dann die $\sigma^{-1}(k)$-te Kiste) und schaue ob deine Nummer drin ist.
- Falls ja, nimm sie und gehe raus. Andernfalls, gehe zur Kiste mit der Scheinnummer, die in der Kiste drin war.
- Wiederhole den Vorgang maximal 50 mal.

Das ist die Suchstrategie und nun berechnen wir die Wahrscheinlichkeit für eine Freilassung.

Jeder Insasse folgt einem Zyklus einer Permutation beginnend mit seiner Kennnummer. Falls der Zyklus der Permutation nicht länger als 50 ist, dann wird der Insasse seine Kennnummer finden.

Wir berechnen nun die Wahrscheinlichkeit, dass der Zyklus C von τ länger als 50 ist. Nehme $50 < k \leq 100$. Es gibt genau $\binom{100}{k}$ Möglichkeiten im Zyklus C zu beginnen, $(k-1)!$ Möglichkeiten sie zu ordnen und $(100-k)!$ Möglichkeiten um den Rest zu permutieren. Das Produkt

$$\binom{100}{k}(k-1)!(100-k)! = \frac{100!}{k}$$

entspricht dann der Anzahl Permutationen bei denen ein Zyklus der Länge k existiert. Die Wahrscheinlicheit ist dementsprechend gleich mit $\frac{1}{k}$, da es insgesamt 100! Permutationen gibt.

Daraus forlgt, dass die Wahrscheinlichkeit für einen kurzen Zyklus (≤ 50) ist gleich

$$\begin{aligned} P\left[\text{Insassen kommen frei}\right] &= 1 - P\left[\text{Insassen kommen nicht frei}\right] \\ &= 1 - P\left[\bigcup_{k=51}^{100} \{\text{Es existiert ein Zyklus der Länge } k\}\right] \\ &= 1 - \sum_{k=51}^{100} P\left[\text{Es existiert ein Zyklus der Länge } k\right] \\ &= 1 - \frac{1}{51} - \frac{1}{52} - \quad \cdots \quad - \frac{1}{100}. \end{aligned}$$

Siehe nächstes Blatt!

Im letzten Schritt verwenden wir die Approximationsformel wie im Coupon Sammelproblem in Skript S. 75/76:

$$P\,[\text{Insassen kommen frei}] = 1 - \frac{1}{51} - \frac{1}{52} - \quad \cdots \quad - \frac{1}{100} \approx 1 - \log(2) \approx 0.307.$$