ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik
Musterlösung 6

1. In einer Studie an Ehepaaren einer hessischen Landbevölkerung wurde untersucht, ob eine Beziehung zwischen Körperbau und Gattenwahl besteht. Es wurden bei Ehegatten die Körperbautypen *1 (leptosom, schmächtig), 2 (athletisch)* und *3 (pyknisch, fettleibig)* unterschieden. Die unten stehende Tabelle zeigt die entsprechende gemeinsame Verteilung von X (Körperbau Mann) und Y (Körperbau Frau).

Körperbautyp des Ehemannes	Körperbautyp der Ehefrau: 1 leptosom	2 athletisch	3 pyknisch
1 leptosom	11.6%	4.5%	7.6%
2 athletisch	7.1%	46.0%	7.6%
3 pyknisch	4.5%	a	9.8%

a) Bestimme den fehlenden Wert a.

b) Berechne die Randverteilungen, d.h. die Verteilung von X und die Verteilung von Y.

c) Berechne die bedingte Wahrscheinlichkeit $P[X \leq 2 | Y \leq 2]$.

d) Berechne die gemeinsame Verteilung von X und Y unter Annahme der Unabhängigkeit. Benutze dazu die Randverteilungen aus **b)** und vergleiche dann die Werte mit obiger Tabelle.

Lösung:

a) Alle Wahrscheinlichkeiten müssen sich zu 1 aufaddieren, also

$$1 = \sum_{i,j=1}^{3} \mathbb{P}\left(X = i, Y = j\right),$$

daraus folgt $a = 1.3\%$.

b) $\mathbb{P}\left(X = 1\right) = \sum_{j=1}^{3} \mathbb{P}\left(X = 1, Y = j\right) = 11.6\% + 4.5\% + 7.6\% = 23.7\%$.
Analog $\mathbb{P}\left(X = 2\right) = 60.7\%$, $\mathbb{P}\left(X = 3\right) = 15.6\%$.

Ebenso: $\mathbb{P}\left(Y = 1\right) = 23.2\%$, $\mathbb{P}\left(Y = 2\right) = 51.8\%$, $\mathbb{P}\left(Y = 3\right) = 25.0\%$.

Bitte wenden!

In Tabellenschreibweise:

Körperbautyp des Ehemannes	Körperbautyp der Ehefrau 1 leptosom	2 athletisch	3 pyknisch	insgesamt
1 leptosom	11.6%	4.5%	7.6%	23.7%
2 athletisch	7.1%	46.0%	7.6%	60.7%
3 pyknisch	4.5%	1.3%	9.8%	15.6%
insgesamt	23.2%	51.8%	25.0%	100.0%

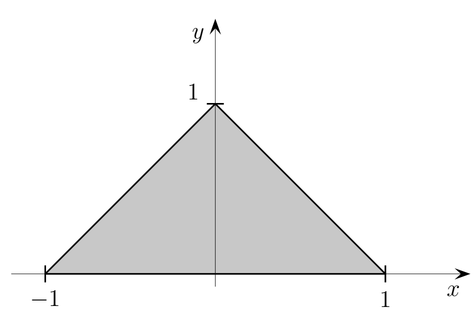
c) $P[X \leq 2 | Y \leq 2] = \frac{\mathbb{P}(X \leq 2, Y \leq 2)}{\mathbb{P}(Y \leq 2)} = \frac{11.6\% + 4.5\% + 7.1\% + 46.0\%}{23.2\% + 51.8\%} = 92.3\%.$

d) Multiplikation der Randverteilungen (Unabhängigkeitsannahme!) ergibt folgende Werte für die gemeinsame Verteilung:

Körperbautyp des Ehemannes	Körperbautyp der Ehefrau 1 leptosom	2 athletisch	3 pyknisch	insgesamt
1 leptosom	5.5%	12.3%	5.9%	23.7%
2 athletisch	14.1%	31.4%	15.2%	60.7%
3 pyknisch	3.6%	8.1%	3.9%	15.6%
insgesamt	23.2%	51.8%	25.0%	100.0%

Vergleich der beiden Tabellen: An den Randverteilungen haben wir nichts geändert. Dennoch weichen die einzelnen Wahrscheinlichkeiten der gemeinsamen Verteilungen stark voneinander ab. Hier wird offensichtlich, dass die Randverteilungen die gemeinsame Verteilung keineswegs eindeutig beschreiben.
Im vorliegenden Beispiel erkennen wir, dass in der hessischen Landbevölkerung Heiraten unter Personen mit gleichem Körperbau häufiger vorkommen, als man dies bei einer "zufälligen" Partnerwahl erwarten würde.

2. Die gemeinsame Dichte der beiden Zufallsvariablen X und Y sei konstant gleich c auf dem grauen Dreieck und 0 ausserhalb.



Siehe nächstes Blatt!

a) Bestimmen Sie die gemeinsame Dichtefunktion $f_{X,Y}$.

b) Bestimmen Sie die beiden Randdichten.

Tipp: Unterscheide bei der Rechnung zwischen den Regionen $\{x < -1\}$; $\{-1 \leq x < 0\}$; $\{0 \leq x < 1\}$ und $\{1 \leq x\}$.

c) Bestimmen Sie $E[X]$, $E[Y]$, $\mathrm{Var}(X)$ und $\mathrm{Var}(Y)$.

d) Bestimmen Sie die Kovarianz zwischen X und Y. Sind X und Y unabhängig? Begründen Sie Ihre Antwort.

e) Gegeben $X = 0.5$, was ist der bedingte Erwartungswert von Y?

Lösung:

a) Wir bezeichnen das Dreieck mit D, dann ist laut Angabe

$$f_{X,Y}(x,y) = \begin{cases} c & (x,y) \in D, \\ 0 & \text{sonst.} \end{cases}$$

Die Fläche von D ist 1, daraus folgt sofort $1 = \int\int_D c\,dxdy = c$, also $c = 1$.

b) Es ist

$$D = \{(x,y) \mid -1 \leq x \leq 1,\ 0 \leq y \leq 1,\ |x| + y \leq 1\}.$$

Die Randdichte von X ist

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \begin{cases} 0 & x < -1, \\ \int_0^{x+1} 1\,dy = x + 1 & -1 \leq x < 0, \\ \int_0^{-x+1} 1\,dy = 1 - x & 0 \leq x < 1, \\ 0 & 1 \leq x. \end{cases}$$

Die Randdichte von Y ist

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \begin{cases} 0 & y < 0, \\ \int_{-1+y}^{1-y} 1\,dy = 2(1-y) & 0 \leq y < 1, \\ 0 & 1 \leq y. \end{cases}$$

c) Es gilt

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-1}^{0} x^2 + x\,dx + \int_{0}^{1} x - x^2\,dx = 0,$$

Bitte wenden!

und

$$E[X^2] = \int\limits_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{1}{6}.$$

Also $\mathrm{Var}(X) = E[X^2] - E[X]^2 = E[X^2] = \frac{1}{6}$. Weiter gilt

$$E[Y] = \int\limits_{-\infty}^{\infty} y f_Y(y)\,dy = \int\limits_{0}^{1} 2y(1-y)\,dy = \frac{1}{3},$$

und

$$E[Y^2] = \int\limits_{-\infty}^{\infty} y^2 f_Y(y)\,dy = \int\limits_{0}^{1} 2y^2(1-y)\,dy = \frac{1}{6}.$$

Also ist $\mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$.

d) Aus **c)** folgt $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = E[XY]$. Es gilt

$$\begin{aligned} E[XY] &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dxdy \;=\; \int\int\limits_{D} xy\,dxdy \\ &= \int\limits_{0}^{1} y \left(\int\limits_{-1+y}^{1-y} x\,dx \right) dy \\ &= \int\limits_{0}^{1} y \left(x^2\Big|_{-1+y}^{1-y} \right) dy \;=\; \int\limits_{0}^{1} y((1-y)^2 - (-1+y)^2)\,dy \;=\; 0. \end{aligned}$$

Also ist $\mathrm{Cov}(X,Y) = 0$.

X und Y sind nicht unabhängig, da $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ für gewisse x und y (beispielsweise gilt $f_{X,Y}\left(\frac{1}{2},\frac{1}{2}\right) = 1 \neq f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right) = \frac{1}{2}$). Das ist aber auch intuitiv klar: abhängig von der Beobachtung von Y kann man gewisse Werte für X sofort ausschließen; wenn Y z.B. $\frac{1}{2}$ ist, dann kann der Betrag von X nicht mehr grösser als $\frac{1}{2}$ sein, was a priori (also bevor man Y kennt) nicht auszuschliessen ist. Aus der Beobachtung von Y gewinnt man also Information über X (bedingte Verteilung von X gegeben Y).

e) Die bedingte Dichte von Y gegeben $X = 0.5$ ist

$$f_{Y|X=0.5}(y) \;=\; \frac{f_{X,Y}(0.5,y)}{f_X(0.5)} \;=\; 2f_{X,Y}(0.5,y).$$

Siehe nächstes Blatt!

Also $f_{Y|X=0.5}(y) = 2$ für $0 \leq y \leq 0.5$ und $f_{Y|X=0.5}(y) = 0$ sonst. Somit ist

$$E[Y \mid X = 0.5] \;=\; \int_{-\infty}^{\infty} y f_{Y|X=0.5}(y) dy \;=\; 2 \int_0^{0.5} y dy \;=\; 0.25.$$

3. Wenn an der Tramhaltestelle "ETH/Universitätsspital" je ein in Richtung Bahnhof fahrendes Tram der Linien 6 und 10 mit weniger als $\tau = 30s$ zeitlichem Abstand eintreffen, wird eines ausgebremst und muss auf offener Strecke warten; wir bezeichnen dies als "Kollision".

Im abendlichen Stossverkehr fahren Trams nicht mehr nach Fahrplan, sondern zufällig. Wir nehmen an, dass zwei Trams (eines von Linie 6, eines von Linie 10) unabhängig und über die Zeit $[0, T]$ ($T = 5\text{min}$) uniform verteilt eintreffen. Wie gross ist die Wahrscheinlichkeit einer Kollision?

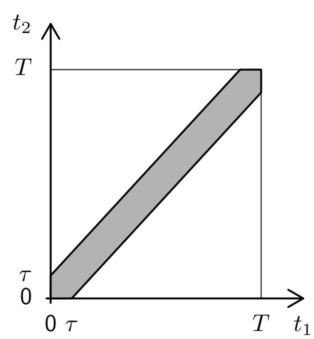
Tipp: Zeichne ein $[0, T] \times [0, T]$ Viereck wo jede Achse der Ankunftszeit eines Trams entspricht. Finde das Gebiet auf dem Viereck wo es zur Kollision kommt.

Lösung:

Wir nennen die beiden Ankunftszeiten T_1 und T_2. Da T_1 und T_2 unabhängig sind, ist ihre gemeinsame Dichte das Produkt der Dichten von T_1 und T_2:

$$f(t_1, t_2) = \frac{1}{T^2} \;;$$

mit anderen Worten, (T_1, T_2) ist uniform verteilt auf $[0, T] \times [0, T]$. Die Wahrscheinlichkeit einer Kollision ist die Wahrscheinlichkeit der schattierten Fläche im Diagramm unten.



Die Fläche der nicht-schattierten Dreiecke beträgt je $\frac{1}{2}(T - \tau)^2$, der schattierte Streifen hat daher eine Fläche von $T^2 - (T - \tau)^2$. Dividieren durch die Fläche T^2 des gesamten Quadrats ergibt die gesuchte Wahrscheinlichkeit: $P[\text{Kollision}] = 1 - (1 - \tau/T)^2 = 0.19$.

Bitte wenden!

4. Seien X und Y die Lebensdauer zweier Maschinen, "Maschine 1" bzw. "Maschine 2", in Monaten. Die beiden Variablen sind unabhängig und exponentialverteilt:

$$X \sim \text{Exp}(\lambda_1), \quad Y \sim \text{Exp}(\lambda_2)$$

a) Was ist die Wahrscheinlichkeit, dass Maschine 1 maximal einen Monat länger funktioniert als Maschine 2, wenn $\lambda_1 = \frac{1}{10}$ und $\lambda_2 = \frac{1}{15}$?

b) Sei weiterhin $\lambda_1 = \frac{1}{10}$ und $\lambda_2 = \frac{1}{15}$. Wenn man weiss, dass Maschine 1 nach 4 Monaten kaputt war, was ist dann die erwartete Lebensdauer der Maschine 2?

c) (**Zusatz**) Wir nehmen nun an, dass $\lambda_1 = \lambda_2 = \frac{1}{10} := \lambda$. Mit welcher Wahrscheinlichkeit gehen die Maschinen im gleichen Monat kaputt?

Sei $N(t)$ die Anzahl der Produkte die Maschine 1 im Zeitintervall $[0, t]$, $t > 0$ (in Stunden), herstellt. Wir modellieren $N(t)$ mit einem *homogenen Poissonprozess* mit *Intensität* λ, das bedeutet für jedes $t > 0$ gilt

$$N(t) \sim \text{Poisson}(\lambda t).$$

d) Angenommen $\lambda = 8.1$, was ist der erwartete Wert an Produkten die innerhalb von einem Tag (10 Stunden) hergestellt werden? Berechne auch die Standardabweichung davon.

e) Sei T_1 der Zeitpunkt an dem das erste Produkt hergestellt wird. Bestimme die kummulative Verteilungsfunktion von T_1 und berechne den Erwartungswert und die Varianz. Wie heisst die Verteilung von T_1?

Tipp: beschreibe das Ereignis $\{T_1 > t\}$ mithilfe von $N(t)$.

Lösung:

a) Es ist $P(X \leq Y + 1)$ gefragt.
Wir wissen $f(x) = \lambda_1 e^{-\lambda_1 x}$, $x > 0$ und $f(y) = \lambda_2 e^{-\lambda_2 y}$, $y > 0$.
X und Y sind unabhängig $\Rightarrow f(x, y) = \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 y}$, $x, y \geq 0$.
Wir sind an der Fläche $A = \{(x, y) : \ y \geq 0, 0 \leq x \leq y + 1\}$ interessiert.

Siehe nächstes Blatt!

$$
\begin{aligned}
P(X \leq Y+1) &= \int\int_A f(x,y)dxdy = \int_0^\infty \Big[\int_0^{y+1} \lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y} dx\Big] dy \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 y}\Big[\int_0^{y+1} \lambda_1 e^{-\lambda_1 x} dx\Big] dy = \int_0^\infty \lambda_2 e^{-\lambda_2 y}\left(1 - e^{-\lambda_1 (y+1)}\right) dy \\
&= \underbrace{\int_0^\infty \lambda_2 e^{-\lambda_2 y} dy}_{=1,\text{ da Dichte}} - \int_0^\infty \lambda_2 e^{-\lambda_2 y} e^{-\lambda_1 y - \lambda_1} dy = 1 - \lambda_2 e^{-\lambda_1}\int_0^\infty e^{-(\lambda_1+\lambda_2)y} dy \\
&= 1 - \frac{\lambda_2 e^{-\lambda_1}}{\lambda_1 + \lambda_2} = 0.6381
\end{aligned}
$$

b) Wegen der Unabhängigkeit von X und Y kann man den bedingten Erwartungswert direkt berechnen, d.h. $E(Y|X = x) = E(Y) = 1/\lambda_2 = 15$. Die erwartete Lebensdauer ist also 15 Monate.

c) Wir bezeichnen mit $\lfloor x \rfloor$ die Abrundung von x. Gesucht ist $P[\lfloor X \rfloor = \lfloor Y \rfloor]$.

$$
\begin{aligned}
P[\lfloor X \rfloor = \lfloor Y \rfloor] &= \sum_{i=0}^\infty P[i \leq X < i+1, i \leq Y < i+1] \\
&= \sum_{i=0}^\infty P[i \leq X < i+1] P[i \leq Y < i+1] \\
&= \sum_{i=0}^\infty \int_i^{i+1} \lambda e^{-\lambda x} dx \int_i^{i+1} \lambda e^{-\lambda y} dy \\
&= \sum_{i=0}^\infty \left(-e^{-\lambda(i+1)} + e^{-\lambda i}\right)^2 \\
&= \sum_{i=0}^\infty \left(e^{-\lambda i}\right)^2 \left(-e^{-\lambda} + 1\right)^2 \\
&= \left(-e^{-\lambda} + 1\right)^2 \sum_{i=0}^\infty \left(e^{-2\lambda}\right)^i \\
&= \left(-e^{-\lambda} + 1\right)^2 \frac{1}{1 - e^{-2\lambda}} \\
&= \left(-e^{-\frac{1}{10}} + 1\right)^2 \frac{1}{1 - e^{-\frac{2}{10}}} \approx 0.05
\end{aligned}
$$

In der zweiten Gleichung haben wir die Unabhängigkeit verwendet, in der zweitletzten die Summenformel für die geometrische Reihe.

Bitte wenden!

d) Es gilt $N(10) \sim \text{Poisson}(8.1 \cdot 10)$, daher gilt (siehe Poissonverteilung) $\mathbb{E}[N(10)] = 81 = \text{Var}[N(10)]$. Die Standardabweichung ist also $\sigma = 9$.

e) Bemerke dass $T_1 > t$ genau dann wenn $N(t) = 0$ gilt. Damit folgt

$$P[T_1 \leq t] = 1 - P[T_1 > t] = 1 - P[N(t) = 0] = 1 - \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = 1 - e^{-\lambda t}.$$

Also gilt, dass T_1 einer Exponentialverteilung mit Parameter λ folgt. Somit gilt $\mathbb{E}[T_1] = 1/\lambda = \frac{10}{81}$ und $\text{Var}[T_1] = 1/\lambda^2 = \frac{100}{6561}$.

5. (leicht modifizierte Prüfungsaufgabe - Sommer 2014) In einem Labor werden Nanodrähte von 10μm Länge erzeugt, die $1\text{pg} = 10^{-12}\text{g}$ wiegen und aus Platin und Nickel bestehen. Aufgrund des Herstellungsverfahrens sind der Nickelinhalt X (in pg) und die Dauer der Funktionsfähigkeit T (in Sekunden) der Drähte zufällig, mit gemeinsamer Dichte

$$f_{X,T}(x,t) = \begin{cases} cx\cos\left(\frac{\pi}{2}x\right)\mathrm{e}^{-xt}, & \text{falls } x \in (0,1),\ t > 0, \\ 0, & \text{sonst.} \end{cases}$$

a) Zeigen Sie, dass die Randdichte $f_X(x)$ von X gegeben ist durch

$$f_X(x) = \begin{cases} c\cos\left(\frac{\pi}{2}x\right), & \text{falls } x \in (0,1), \\ 0, & \text{sonst.} \end{cases}$$

b) Berechnen Sie c, so dass $f_{X,T}$ eine Wahrscheinlichkeitsdichte ist.

c) Berechnen Sie die bedingte Dichte von T gegeben, dass der Nickelinhalt $X = x$ eines Nanodrahts bekannt ist. Wie heisst diese Verteilung?

d) Sind X und T unabhängig? Begründe.

e) Berechnen Sie die Wahrscheinlichkeit, dass der Nickelinhalt weniger als 0.5pg beträgt.

f) Berechnen Sie den Erwartungswert und die Varianz von X (ohne den numerischen Wert von π einzusetzen).

Lösung:

Siehe nächstes Blatt!

a) Falls $x \in (0,1)$,

$$
\begin{aligned}
f_X(x) = \int_{-\infty}^{\infty} f_{X,T}(x,t)\mathrm{d}t &= \int_0^{\infty} cx\cos\left(\frac{\pi}{2}x\right)\mathrm{e}^{-xt}\mathrm{d}t \\
&= c\cos\left(\frac{\pi}{2}x\right)\int_0^{\infty} x\mathrm{e}^{-xt}\mathrm{d}t \\
&= c\cos\left(\frac{\pi}{2}x\right).
\end{aligned}
$$

Für $x \notin (0,1)$ ist $f_X(x) = 0$, da in diesem Fall $f_{X,T}(x,t) = 0$.

b)

$$
1 \overset{!}{=} \iint_{\mathbb{R}^2} f_{X,T}(x,t)\mathrm{d}t\,\mathrm{d}x = \int_{\mathbb{R}} f_X(x)\mathrm{d}x = \int_0^1 c\cos\left(\frac{\pi}{2}x\right)\mathrm{d}x = \left[\frac{2c}{\pi}\sin\left(\frac{\pi}{2}x\right)\right]_0^1 = \frac{2c}{\pi}.
$$

Somit $c = \pi/2$.

c) Falls $t > 0$, und $x \in (0,1)$

$$
f_{T|X=x}(t) = f_{X,T}(x,t)/f_X(x) = x\mathrm{e}^{-xt},
$$

sonst $f_{T|X=x}(t) = 0$. Also $(T|X = x) \sim \mathrm{Exp}(\lambda = x)$.

d) Da die bedingte Wahrscheinlichkeit von T gegeben $X = x$ von x abhängt, sind die Zufallsvariablen nicht unabhängig. Wären sie unabhängig, so würde gelten $f_{T[X=x}(t) = f_T(t)$, was hier offensichtlich nicht der Fall sein kann.

e)

$$
\begin{aligned}
\mathbb{P}[X < 0.5] = \int_0^{1/2} f_X(x)\mathrm{d}x &= \int_0^{1/2} \frac{\pi}{2}\cos\left(\frac{\pi}{2}x\right)\mathrm{d}x = \left[\sin\left(\frac{\pi}{2}x\right)\right]_0^{1/2} \\
&= \sin\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}} \approx 70.7\%
\end{aligned}
$$

f)

$$
\begin{aligned}
\mathbb{E}[X] = \int_{-\infty}^{\infty} xf_X(x)\mathrm{d}x &= \int_0^1 \frac{\pi}{2}x\cos\left(\frac{\pi}{2}x\right)\mathrm{d}x \\
&= \left[x\sin\left(\frac{\pi}{2}x\right)\right]_0^1 - \int_0^1 \sin\left(\frac{\pi}{2}x\right)\mathrm{d}x \\
&= 1 + \left[\frac{2}{\pi}\cos\left(\frac{\pi}{2}x\right)\right]_0^1 = 1 - \frac{2}{\pi}
\end{aligned}
$$

Bitte wenden!

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)\mathrm{d}x = \int_0^1 \frac{\pi}{2}x^2 \cos\left(\frac{\pi}{2}x\right) \mathrm{d}x \\
&= \left[x^2 \sin\left(\frac{\pi}{2}x\right)\right]_0^1 - \int_0^1 2x \sin\left(\frac{\pi}{2}x\right) \mathrm{d}x \\
&= 1 + \frac{2\cdot 2}{\pi}\left(\left[x\cos\left(\frac{\pi}{2}x\right)\right]_0^1 - \int_0^1 \cos\left(\frac{\pi}{2}x\right)\mathrm{d}x\right) \\
&= 1 + 0 + \frac{4}{\pi}\left[-\frac{2}{\pi}\sin\left(\frac{\pi}{2}x\right)\right]_0^1 \\
&= 1 - \frac{8}{\pi^2}
\end{aligned}
$$

$$
\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - \frac{8}{\pi^2} - \left(1 - \frac{2}{\pi}\right)^2 = \frac{4}{\pi} - \frac{12}{\pi^2} = \frac{4(\pi - 3)}{\pi^2}
$$