Musterlösung zu Übungsblatt 6

erstellt von Pascal Neukirchner

21.S *Unabhängig? Unkorreliert?*
a) (i) Nach dem in Vorlesung V5b1 vorgestellten Kriterium muss nur überprüft werden, ob die einzelnen Zeilen bzw. Spalten der Matrix der Verteilungsgewichte proportional sind und hier ist dies nicht der Fall, denn wenn man die Gewichte der ersten Zeile verdoppelt und die einzelnen Werte mit der zweiten Zeile vergleicht, dann passt dies für die ersten beiden Einträge, aber für den Eintrag $(2, d)$ nicht. Mit dem gleichen Vorgehen findet man auch, dass der Eintrag $(3, d)$ das Kriterium verletzt. Somit sind X_1 und X_2 voneinander abhängig.

(ii) Hier soll eine Funktion h gefunden werden, sodass die Verarbeitung $h(X_2)$ der Zufallsvariable X_2 unabhängig von X_1 ist. Dafür kann jede beliebige konstante Funktion gewählt werden, also bspw. $h(x) = 0$ für $x \in \{b, c, d\}$, denn nun kann aus dem Wissen von $h(X_2)$ keine Information über X_1 gewonnen werden.
Dies zeigen wir auch noch formal. Als Wertebereich von $h(X_2)$ können wir die einelementige Menge $\{0\}$ wählen, denn $h(X_2)$ fällt nach Voraussetzung sicher auf 0. Es gilt $\{X_1 = a_1, h(X_2) = 0\} = \{X_1 = a_1\}$. Somit gilt für alle Paare $(a, a') \in \{1, 2, 3\} \times \{0\}$ die Produktformel:

$$\mathbf{P}(X_1 = a, h(X_2) = 0) = \mathbf{P}(X_1 = a) = \mathbf{P}(X_1 = a) \cdot \mathbf{P}(h(X_2) = 0).$$

(iii) Wie in Aufgabenteil (i) angesprochen, müssen die Werte für $(2, d)$ und $(3, d)$ angepasst werden, das heißt konkret:

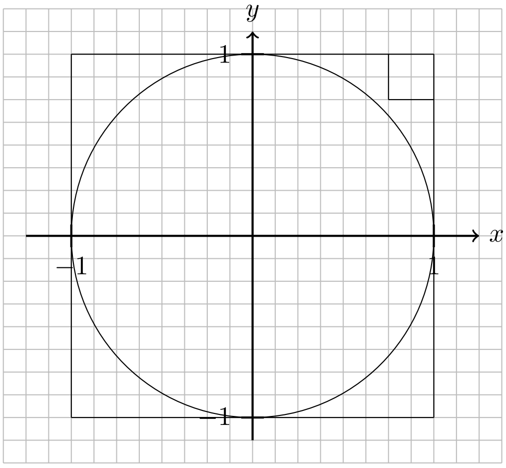
$$\mathbf{P}(X_1 = 2, X_2 = d) = 20\gamma$$

$$\mathbf{P}(X_1 = 3, X_2 = d) = 30\gamma$$

Nun entspricht die zweite Zeile genau dem Doppelten der Ersten und die dritte Zeile dem Dreifachen der Ersten.

(iv) Die Beantwortung der Fragestellung funktioniert analog zu Aufgabe 12b (iii), denn die Zufallsvariable X_2 ist hier erneut nicht reellwertig und daher ist ihr Erwartungswert nicht definiert. Dementsprechend kann auch nicht von der Kovarianz der Zufallsvariablen X_1 und X_2 gesprochen werden, denn für diese werden die einzelnen Erwartungswerte benötigt. Schließlich ist die Unkorreliertheit aber dadurch definiert, dass die Kovarianz 0 ist und somit ist die Fragestellung schlecht formuliert.

b) (i) X_1 und X_2 sind voneinander abhängig, was man leicht erkennen kann, wenn man sich ein Quadrat mit Seitenlängen 2 zentriert um den Urspung denkt. Der Einheitskreis liegt dann auf jeden Fall in diesem Quadrat, aber die Punkte nah an den Ecken werden nicht im Einheitskreis liegen, denn für diese gilt $a_1^2 + a_2^2 > 1$. Wenn wir jedoch die Koordinate a_1 fest halten, dann findet man ein a_2, sodass das Paar (a_1, a_2) in dem Einheitskreis liegt. Genauso funktioniert dies nun auch für alle Punkte, die in einem kleinen Quadrat um a_1 herum liegen und dies verletzt die Produktformel für unabhängige Ereignisse. Siehe dazu Folgende Skizze:



Das kann natürlich auch noch etwas formaler gezeigt werden, verfolgt aber im Grunde dieselbe Idee:
Betrachte die Menge $T := \{(b_1, b_2) : b_1, b_2 \in (0.75, 1)\}$. Es gilt $0.75^2 + 0.75^2 > 1$ und somit liegt keiner der Punkte aus T im Einheitskreis. Es folgt:

$$\mathbf{P}((X_1, X_2) \in T) = 0 \neq \mathbf{P}(X_1 \in (0.75, 1)) \cdot \mathbf{P}(X_2 \in (0.75, 1))$$

(ii) Hier ist eine Symmetrie-Überlegung der Schlüssel zum Erfolg. Wenn man die rechte Hälfte des Einheitskreises an der y-Achse spiegelt, dann erhält man exakt die linke Hälfte, d.h. für den jeden Punkt mit einem positiven Wert für X_1 finden wir einen Punkt für X_1 mit genau diesem Wert, nur negativ. Demnach ist (X_1, X_2) genauso verteilt wie $(-X_1, X_2)$.
Mit dieser Überlegung können wir die einzelnen Erwartungswerte für die Kovarianz bequem berechnen, indem wir X_1 wie folgt aufteilen:

$$\begin{aligned}
\mathbf{E}[X_1] &= \mathbf{E}[X_1 \cdot \mathbf{1}_{X_1 > 0} + X_1 \cdot \mathbf{1}_{X_1 \leq 0}] \\
&= \mathbf{E}[X_1 \cdot \mathbf{1}_{X_1 > 0}] + \mathbf{E}[X_1 \cdot \mathbf{1}_{X_1 \leq 0}] \\
&= \mathbf{E}[X_1 \cdot \mathbf{1}_{X_1 > 0}] - \mathbf{E}[-X_1 \cdot \mathbf{1}_{X_1 \leq 0}] \\
&= \mathbf{E}[X_1 \cdot \mathbf{1}_{X_1 > 0}] - \mathbf{E}[X_1 \cdot \mathbf{1}_{X_1 > 0}] = 0
\end{aligned}$$

Wobei der Fall $X_1 = 0$ im rechten Term weggelassen wurde, denn für kontinuierliche Verteilungen ist die Wahrscheinlichkeit einen Punkt zu treffen gleich null. Der Erwartungswert von $X_1 \cdot X_2$ folgt nun analog und somit erhält man schließlich:

$$\mathbf{Cov}[X_1, X_2] = \mathbf{E}[X_1 \cdot X_2] - \mathbf{E}[X_1] \cdot \mathbf{E}[X_2] = 0 - 0 \cdot \mathbf{E}[X_2] = 0$$

Das heißt trotz ihrer Abhängigkeit sind X_1 und X_2 unkorreliert.

22. *Paarweise Unabhängigkeit von Ereignissen.*
a) Insgesamt liegen 6 Punkte vor und diese werden rein zufällig gewählt, es liegt also eine uniforme Verteilung vor und es kann die bekannte Formel $\frac{\#\text{günstige}}{\#\text{mögliche}}$ verwendet werden. Somit ergeben sich folgende Wahrscheinlichkeiten:

$$\mathbf{P}(E_1 \cap E_2) = \frac{1}{6}$$
$$\mathbf{P}(E_1^c \cap E_2) = \frac{2}{6}$$
$$\mathbf{P}(E_1 \cap E_2^c) = \frac{2}{6}$$
$$\mathbf{P}(E_1^c \cap E_2^c) = \frac{1}{6}$$

b) Die Verteilungsgewichte von (X_1, X_2) wurden bereits in Aufgabenteil a berechnet, somit folgt als Matrix:

X_1, X_2	oben	unten
links	$\frac{2}{6}$	$\frac{1}{6}$
rechts	$\frac{1}{6}$	$\frac{2}{6}$

Die Ereignisse E_1 und E_2 sind abhängig, denn die Produktformel ist verletzt:

$$\mathbf{P}(E_1 \cap E_2) = \frac{1}{6} \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$$

Zudem sind die Ereignisse negativ korreliert, denn sie treten mit größerer Wahrscheinlichkeit nicht zusammen ein, als zusammen einzutreten. Formal heißt das:

$$\mathbf{P}(E_1 \cap E_2) < \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$$

c) Auch hier kann man wie bei Aufgabenteil a vorgehen. Es ergibt sich:

$$\mathbf{P}(E_1 \cap E_2 \cap E_3) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3} = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2) \cdot \mathbf{P}(E_3)$$

Aber Achtung, die Ereignisse sind natürlich dennoch nicht unabhängig, dafür muss die Produktformel auch für jegliche Paare (E_i, E_j) gelten und nach Aufgabenteil b ist bereits bekannt, dass dem für (E_1, E_2) nicht so ist.

23. *Die Standardabweichung der zufälligen Trefferquote.*
$\hat{p}$ gibt die relative Treffzahl von F an, d.h. $\frac{\#\text{Treffer}}{\#\text{Versuche}}$. Sei K die Anzahl der Treffer, nun ist K binomialverteilt mit Parametern n und $p = \frac{1}{5}$, denn einen Treffer zu landen entspricht dem Münzwurf mit Erfolgswahrscheinlichkeit $\frac{1}{5}$. Somit kann bei der Berechnung der Varianz die Varianz der Binomialverteilung verwendet werden (Diese lautet $n \cdot p \cdot (1-p)$):

$$\mathbf{Var}[\hat{p}] = \mathbf{Var}\left[\frac{K}{n}\right] = \frac{1}{n^2} \cdot \mathbf{Var}[K] = \frac{1}{n^2} \cdot n \cdot \frac{1}{5} \cdot \frac{4}{5} = \frac{4}{25n}$$

Die Standardabweichung ist nun die Wurzel aus der Varianz, also $\sigma = \frac{2}{5\sqrt{n}}$. Dies ergibt letzendlich die gesuchten Ergebnisse. Für $n = 100$ ist $\sigma = 0,04$, für $n = 400$ ist $\sigma = 0,02$ und für $n = 1600$ ist $\sigma = 0,01$. Das heißt, wenn sich n vervierfacht, dann halbiert sich die Standardabweichung.

24.S *Ziehen so lange das Zeug hält.*
a) Mit der Formel für die Varianz und der Transformationsformel für den Erwartungswert folgt:

$$\begin{aligned}\mathbf{Var}[X_1] &= \mathbf{Var}[h(T_1)] \\ &= \mathbf{E}[h(T_1)^2] - \mathbf{E}[h(T_1)]^2 \\ &= 0,4 \cdot 1^2 + 0,3 \cdot 2^2 + 0,3 \cdot 3^2 - 1,9^2 = 0,69\end{aligned}$$

Wobei $\mathbf{E}[h(T_1)]^2$ bereits aus Aufgabe 11 bekannt war.

b) Dies ist eine kleine Fangfrage, da wir laut Aufgabenstellung nur 100 Bäume vorliegen haben, werden hier alle gezogen und somit ist $Y = 40 \cdot 1 + 30 \cdot 2 + 30 \cdot 3 = 190$.

c) Ähnlich zu Aufgabe 13 ist auch hier das Schlagwort die Austauschbarkeit. Es liegt ein Ziehen ohne Zurücklegen vor. Die 100 Bäume können zunächst mit den Zahlen von 1 bis 100 durchnummeriert werden. Wenn nun alle der Reihe nach gezogen werden, erhält man eine rein zufällige Permutation. Somit ist (vgl. Aufgabe 12a) jegliches T_i uniform verteilt auf $\{1, \dots, 100\}$. Für $i \neq j$ und $a, b \in \{1, \dots, 100\}$ mit $a \neq b$ gilt $\mathbf{P}((T_i, T_j) = (a, b)) = \frac{1}{100 \cdot 99}$, also sind alle Paare (T_i, T_j) mit $i \neq j$ identisch verteilt. Die Verarbeitungen von zwei identisch verteilten Zufallsvariablen (hier (T_5, T_{26}) und (T_1, T_2)) mit ein-und derselben Abbildung (hier $(t_1, t_2) \mapsto (h(t_1), h(t_2))$ sind wiederum identisch verteilt. So wie der Erwartungswert einer Zufallsvariablen durch deren Verteilung bestimmt ist, ist auch die Kovarianz von zwei reellwertigen Zufallsvariablen durch deren gemeinsamer Verteilung bestimmt. Daraus folgt $\mathbf{Cov}[X_i, X_j] = \mathbf{Cov}[X_1, X_2]$ für $i \neq j$.
d) Weil die X_i identisch verteilt sind, sind ihre Varianzen gleich. Zudem benötigen wir noch, über wieviele Kovarianzen hier summiert wird, das ist aber nicht sonderlich schwer, denn es gibt $\binom{g}{2}$ 2-elementige Teilmengen von g Elementen. Dies kann nun nach und nach alles verwendet werden:

$$\begin{aligned}\sum_{i=1}^{g} \mathbf{Var}[X_i] + 2 \cdot \sum_{1 \leq i < j \leq g} \mathbf{Cov}[X_i, X_j] &= g \cdot \mathbf{Var}[X_1] + 2 \cdot \binom{g}{2} \cdot \mathbf{Cov}[X_1, X_2] \\ &= g \cdot \sigma^2 + g \cdot (g-1) \cdot \mathbf{Cov}[X_1, X_2]\end{aligned}$$

e) In Aufgabenteil d haben wir die Varianz von $\sum_{i=1}^{g} X_i = \mathbf{Var}[Y]$ berechnet. Nach Aufgabenteil b ist diese aber nun 0, weil Y sicher den Wert 190 annimmt (es werden ja alle Bäume gezogen und somit kann der Wert für Y nicht variieren). Demnach gilt:

$$0 = g \cdot \sigma^2 + g \cdot (g-1) \cdot \mathbf{Cov}[X_1, X_2] \Leftrightarrow \frac{-\sigma^2}{g-1} = \mathbf{Cov}[X_1, X_2].$$