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

Koordinator
Florian Krach

Stochastik

Interaktive Aufgabe für die Übung 6

1. Seien X und Y Zufallsvariablen.

a) Gib die Definition an, dass X und Y unabhängig sind.

Angenommen X und Y sind Zufallsvariablen mit gemeinsamer Dichtefunktion $f_{X,Y}$ und Randdichten f_X bzw. f_Y.

b) Gib eine equivalente Eigenschaft zur Unabhängigkeit von X und Y an.

c) Zeige dass die Eigenschaft aus **b)** wirklich die Unabhängigkeit von X und Y gemäss **a)** impliziert.

Lösung:

a) (Siehe Buch S. 26) X und Y sind unabhängig, falls für alle $A, B \subset \mathbb{R}$ gilt

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B].$$

b) (Siehe Buch S. 83) X und Y sind unabhängig genau dann wenn für alle $x, y \in \mathbb{R}$ gilt

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

c) Seien $A, B \subset \mathbb{R}$, dann gilt

$$\begin{aligned} P[X \in A, Y \in B] &= \int\limits_{A \times B} f_{X,Y}(x, y) d(x, y) \\ &= \int\limits_{A \times B} f_X(x) f_Y(y) d(x, y) \\ &= \int\limits_A f_X(x) dx \int\limits_B f_Y(y) dy \\ &= P[X \in A]P[Y \in B]. \end{aligned}$$

Dabei haben wir die Eigenschaft aus **b)** für die zweite und den Satz von Fubini für die dritte Zeile verwendet.

Bitte wenden!

2. Seien X sowie U uniform auf dem Intervall $[-1,1]$ verteilt, wobei X und U unabhängig sind.

a) Was sind der Erwartungswert und die Varianz von X?

b) Zeige, dass $E[X^k] = 0$ für k ungerade.

c) Sei $Y = 4\left(X^2 - 1/2\right)^2 + U/3$. Berechnen Sie die Korrelation zwischen X und Y.

d) Simulieren Sie mit R $n = 1000$ Realisationen von X mit `runif(1000,-1,1)`. Gegeben diese Realisierungen, generieren Sie $n = 1000$ Realisationen von Y. Erstellen Sie eine Grafik, welche die generierten Datenpunkte zeigt.

Berechnen Sie mit `cor` die empirische Korrelation der Realisierungen und vergleichen Sie diesen Wert mit **c)**.

e) Sind X und Y unabhängig?

Lösung:

a) Die Lösung kann man aus dem Anhang zum Vorlesungsskript direkt ablesen: Man setzt $a = -1, b = 1$ in die Formel zur Uniformverteilung ("Uni") und erhält $E[X] = 0$ beziehungsweise $Var(X) = 1/3$.

Alternativ kann man diese Werte auch direkt berechnen. Die Dichte von X ist $f_X(t) = 1_{[-1,1]}(t)1/2$, daher gilt es

$$E[X] = \frac{1}{2}\int_{-1}^{1} x dx = \left.\frac{x^2}{4}\right|_{-1}^{1} = 0,$$

und

$$\mathrm{Var}(X) = E[X^2] = \frac{1}{2}\int_{-1}^{1} x^2 dx = \left.\frac{x^3}{6}\right|_{-1}^{1} = \frac{1}{3}.$$

b) Ähnlich zu der ersten Teilaufgabe haben wir für k ungerade

$$E[X^k] = \frac{1}{2}\int_{-1}^{1} x^k dx = \left.\frac{x^{k+1}}{4}\right|_{-1}^{1} = 0.$$

Die letzte Gleichung hierbei folgt aus der Symmetrie von x^{k+1}. Hier wird auch die Notwendigkeit von k ungerade ersichtlich.

c) Wir berechnen die Kovarianz zwischen X und Y.

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = E[XY] = E\left[4X\left(X^2 - 1/2\right)^2 + XU/3\right].$$

Siehe nächstes Blatt!

Aus der Linearität des Erwartungswertes folgt

$$\begin{aligned}\mathrm{Cov}(X,Y) &= 4E\left[XX^4 - X^3 + X/4\right] + E\left[XU\right]/3 \\ &= 4E\left[X^5\right] - 4E\left[X^3\right] + E\left[X\right] + E\left[X\right]E\left[U\right]/3,\end{aligned}$$

wobei wir im zweiten Schritt die Unabhängigkeit von X und U verwendet haben. Da $E\left[X^k\right] = 0$ für k ungerade, folgt

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

und somit sind X und Y unkorreliert.

Alternativ kann man die Rechenregeln von Buch Seite 91, gleich zu Beginn der Rechnung benutzen und sie damit etwas vereinfachen.

d) Siehe separate Seite am Ende.

e) Nein, X und Y sind nicht unabhängig, denn der Wert von Y hängt von dem Wert von X ab. Das sieht man auf der Grafik zur Teilaufgabe **d)** sofort, wenn man bedenkt, dass die Unabhängigkeit von zwei *uniformverteilten* Zufallsvariablen eine homogene Verteilung von Punkten auf dem ganzen Rechteck voraussetzen würde. Dies ist hier nicht der Fall, denn es gibt offenbar Regionen, wo es keine Punkte gibt. (Beachte dass diese Argumentation nur bei uniformverteilten Zufallsvariablen funktioniert.)

Man kann das auch präzise beweisen, indem man zwei Mengen A, B findet, so dass $P[X \in A] > 0$ und $P[Y \in B] > 0$ aber $P[X \in A, Y \in B] = 0$, was der Unabhängigkeit von X und Y widerspricht.

Die Wahrscheinlichkeit von $Y \leq 0$ ist nicht Null (siehe auch Grafik zur Teilaufgabe **d)**), denn zum Beispiel $P[X \in [1/2, 1/\sqrt{2}], U \in [-1, -3/4]] > 0$ und wenn $X \in [1/2, 1/\sqrt{2}]$ und $U \in [-1, -1/2]$, dann ist $Y \leq 0$. Tatsächlich, für $X = 1/2$ und $U = -3/4$

$$Y = 4\left(\left(\frac{1}{2}\right)^2 - \frac{1}{2}\right)^2 + \frac{\frac{-3}{4}}{3} = 4\frac{1}{16} - \frac{1}{4} = 0$$

und für die restlichen Werte von X, U in diesen Intervallen wird Y nur kleiner.

Wenn wir aber wissen dass $-1/\sqrt{6} \leq X \leq 1/\sqrt{6}$, dann ist $Y \leq 0$ unmöglich (siehe auch Grafik zu **d)**). Das sieht man indem man $X = \pm 1/\sqrt{6}$ und $U = -1$ setzt, Y berechnet

$$Y = 4\left(\left(\pm\frac{1}{\sqrt{6}}\right)^2 - \frac{1}{2}\right)^2 + \frac{-1}{3} = 4\frac{1}{9} - \frac{1}{3} > 0$$

und argumentiert, dass für die restlichen Werte von $X \in [-1/\sqrt{6}, 1/\sqrt{6}]$ und $U \in [-1, 1]$,
Y nur noch grösser wird.

Bitte wenden!

Dadurch sieht man, dass mit der Wahl $A := [-1/\sqrt{6}, 1/\sqrt{6}]$ und $B := (-\infty, 0]$, $P[X \in A] > 0$ und $P[Y \in B] > 0$ aber $P[X \in A, Y \in B] = 0$.

Siehe nächstes Blatt!

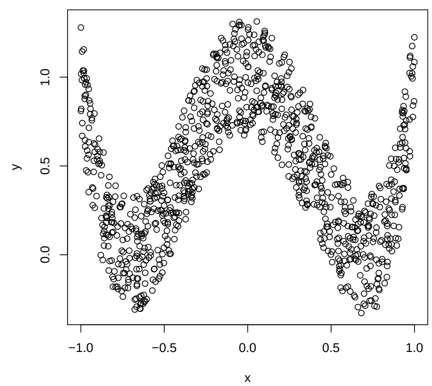
d) Folgender Code führt zum gewünschten Resultat.

```r
# Generiere n=1000 Realisationen von (X,Y)
n = 1000
x = runif(n, -1, 1)
y = 4 * (x^2 - 1/2)^2 + runif(n, -1, 1)/3


# Plotte die Datenpunkte (x,y)
plot(x,y)

# Berechne die empirische Korrelation
cor(x,y)
```

Wir erhalten eine Grafik folgender Form.



Als empirische Korrelation für die gezeigten Datenpunkte erhalten wir $s = -0.012$. Dieser Wert ist verschieden von 0, da es sich um einen *empirischen* und nicht um einen theoretischen Wert handelt.