Statistik für Informatiker **SS 2017**

Prof. Dr. Matthias Birkner
M.Sc. Frederik Klement

Blatt 5

Aufgabe 1: (1 + 2 + 1 + 2 Punkte)
Betrachten Sie für $K > 0$ die folgende Funktion:

$$f_K(x, y) = K(x^2 + y^3), \quad -1 \leq x \leq 1, \quad 0 \leq y \leq 1.$$

a): Bestimmen Sie $\tilde{K}$, so dass $f_{\tilde{K}}$, die Dichte eines zweidimensionalen Zufallsvektor ist.

Es seien X und Y zwei stetige Zufallsvariablen, deren gemeinsame Dichte gegeben ist durch $f_{\tilde{K}}$.

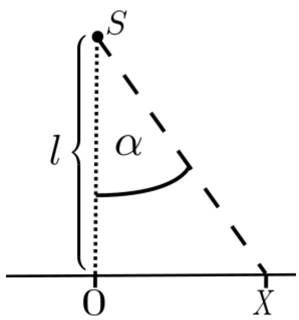
b): Bestimmen Sie die Randverteilung von X und Y, d.h. berechnen Sie die Dichten von X und Y, welche gegeben sind durch:

$$f_X(x) = \int_0^1 f_{\tilde{K}}(x, y)dy, \quad -1 \leq x \leq 1, \quad f_Y(y) = \int_{-1}^1 f_{\tilde{K}}(x, y)dx, \quad 0 \leq y \leq 1.$$

c): Sind X und Y unabhängig voneinander?

d): Berechnen Sie $P(Y < 0.5)$ und $P(Y < 0.5, X < 0)$.

Aufgabe 2 (1 + 2 Punkte)
Wir wollen folgendes Zufallsexperiment simulieren. Wir denken uns die reellen Zahlen als eine horizontale Linie. Weiter denken wir uns einen Punkt S, der sich senkrecht über der 0 mit einem vertikalen Abstand von $l > 0$ befindet. Von diesem Punkt geht ein Strahl aus, wobei der Winkel α des Strahls zum Lot auf die 0 zufällig und uniform über $(-\frac{\pi}{2}, \frac{\pi}{2})$ verteilt ist. Mit X bezeichnen wir die Position des Schnittpunktes zwischen Strahl und der Linie der reellen Zahlen (siehe Bild).



a): Drücken Sie X in Abhängigkeit von α aus.

b): Zeigen Sie, dass die Dichte der Zufallsvariable X gegeben ist durch

$$f_X(x) = \frac{1}{\pi(l + \frac{x^2}{l})}.$$

Aufgabe 3 (1 + 1 + 1 Punkte)

a): Schreiben Sie eine Funktion `lot(l)`, welche das Experiment aus Aufgabe 2 für eine Lotlänge von `l`$=l > 0$ simuliert.

b): Simulieren Sie 10^6 mal mit `lot` die Zufallsvariable X für $l = 0.5$. Erstellen Sie ein Histogramm mit den relativen Häufigkeiten, wobei nur die Simulationen von X betrachtet werden sollen, deren Wert im Intervall $[-10, 10]$ liegt. Verwenden Sie `breaks=200`.

c): Durch das Abschneiden verändert sich die Dichte von X zu

$$\tilde{f}_X(x) = \frac{1}{M} \frac{1}{\pi(l + \frac{x^2}{l})} \mathbf{1}_{[-10,10]}(x),$$

mit $M := \int_{-10}^{10} f_X(x)dx$. Ergänzen Sie das Histogramm um die passende Dichte. Sie können die Funktionen `dcauchy` und `pcauchy` verwenden. Nutzen Sie `help`, um die Parameter der Funktionen richtig zu setzen.

Hinweis: Die Cauchy-Verteilung ist eine Heavy-Tailed-Verteilung, dies bedeutet, dass die Dichte für extreme hohe bzw. niedrige Werte nur sehr langsam gegen 0 konvergiert. Wenn Sie also 10^6 Stichproben erzeugen, befinden sich darunter mit hoher Wahrscheinlichkeit einige wenige, sehr starke Ausreißer. Unter den Standardeinstellungen von `hist` führt dies zur einer unschönen Balkeneinteilung, daher schneiden wir die Ausreißer ab. Das Abschneiden ist äquivalent dazu, dass wir auf das Ereignis $\{X \in [-10, 10]\}$ bedingen, dies wiederum verändert die Dichte.

Aufgabe 4 (1 + 3 Punkte)
Die Box-Muller-Methode ist ein Verfahren, um aus zwei auf $[0, 1]$ gleichverteilten, unabhängigen Zufallsvariablen U, V einen zweidimensionalen standardnormalverteilte Zufallsvektor $Z = (X, Y)$ zu erzeugen. Hierzu wählt man

$$X = \sqrt{-2\log(U)} \cos(2\pi V) \text{ und } Y = \sqrt{-2\log(U)} \sin(2\pi V).$$

a) Erstellen Sie eine Funktion `box_muller(n)`, die unter Verwendung der Box-Muller-Methode aus $2n$ unabhängigen, uniform Verteilten Zufallsvariablen n unabhängige, zweidimensionale standardnormalverteilte Zufallsvektoren $Z_1, Z_2, ..., Z_n$ erzeugt.

b) Simulieren Sie mit Hilfe der Funktion aus **a)** 10^5 Paare (X, Y) unabhängiger, normalverteilter Zufallsvariablen. Erstellen Sie ein Histogramm mit den Simulationen von X und ein weiteres mit denen von Y. Um die Abhängigkeit zwischen X und Y zu untersuchen, erstellen Sie ein Histogramm aller Simulationen von X mit $Y < 0$ und ein Histogramm aller Simulationen von X mit $Y \geq 0$. Zeichnen Sie alle vier Histogramme in ein Bild mit `par(mfrow=c(2,2))`.

Achtung: Zusätzlich zur elektronischen Abgabe per Mail, drucken Sie bitte Ihren Code aus und reichen Sie diesen zusammen mit Ihrer restlichen Abgabe ein.

Abgabe: 10 Uhr 02.06.2017