Wahrscheinlichkeit und Statistik

Musterlösung Serie 8

1. Seien $X_2, X_3, \ldots$ unabhängige Zufallsvariablen. Wir nehmen an, dass für alle $n \in \{2, 3, \ldots\}$
$$\mathbb{P}(X_n = \pm n) = \frac{1}{2n \log n} \quad \text{und} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n \log n}.$$
Zeigen Sie, dass die Folge $(X_n)_{n\geq 2}$ das schwache, aber nicht das starke Gesetz der grossen Zahlen erfüllt.

Lösung: Wir zeigen zuerst, dass das starke Gesetz der grossen Zahlen nicht gilt. Wir betrachten die Ereignisse $A_n = \{|X_n| \geq n\}, n \geq 2$. Dann gilt
$$\mathbb{P}[A_n] = \frac{1}{n \log n} \Rightarrow \sum_{n=2}^{\infty} \mathbb{P}[A_n] = \infty.$$
Aus der Divergenz der Reihe und der Unabhängigkeit der X_i, folgt mit dem zweiten Teil des Borel-Cantelli Lemmas, dass die Ereignisse A_n unendlich oft eintreten mit Wahrscheinlichkeit 1. Wenn das starke Gesetz gelten würde, müsste $S_n/n \to 0, \mathbb{P}$-f.s. Dann müsste es zu fast jedem ω und jedem $\epsilon > 0$ ein $n_0(\omega, \epsilon)$ geben, so dass wenn $n \geq n_0$, dann $|S_n(\omega)| \leq n\epsilon$. Dann folgt aber, für $n \geq n_0$: $|X_{n+1}| \leq |S_{n+1}| + |S_n| \leq (n+1)\epsilon + n\epsilon \leq 2(n+1)\epsilon$. Dies ist aber ein Widerspruch zu dem unendlich oft Eintreten der A_n.

Wir zeigen jetzt, dass das schwache Gesetz der grossen Zahlen gilt. Es gilt: $\mathrm{Var}(X_k) = k/\log k$. Da die Funktion $x/\log x$ ein lokales Minimum bei $x = e$ hat, erhalten wir mit der Chebychev Ungleichung für $\epsilon > 0$:
$$\begin{aligned}\mathbb{P}(|S_n/n| \geq \epsilon) &\leq \frac{1}{\epsilon^2 n^2} \sum_{k=2}^{n} \mathrm{Var}(X_k) \leq \frac{1}{\epsilon^2 n^2}\Big(\frac{2}{\log 2} + \sum_{k=3}^{n} (k/\log k)\Big) \\ &\leq \frac{2}{\epsilon^2 n^2 \log 2} + \frac{(n-3)n}{\epsilon^2 n^2 \log(n)} \to 0,\ n \to \infty.\end{aligned}$$

2. Wir betrachten eine standard-normalverteilte Zufallsvariable $X \sim \mathcal{N}(0,1)$.

   a) Zeigen Sie, dass die Dichte von $Y := X^2$ von der Form $f_Y(y) = c\, e^{-y/2}\, y^{-1/2}\, \mathbf{1}_{\{y>0\}}$ für ein $c > 0$ ist.
   Bemerkung: Die Verteilung Y heisst **Chiquadrat-Verteilung** mit einem Freiheitsgrad und man schreibt $Y \sim \chi_1^2$.

   b) Seien Y_i, $i \in \{1, \ldots, n\}$, i.i.d. χ_1^2-verteilt. Zeigen Sie, dass die Dichte f_2 der Zufallsvariable $Y_1 + Y_2$ die Form $f_2(x) = c_2\, e^{-x/2}\, \mathbf{1}_{\{x>0\}}$ für ein $c_2 > 0$ hat.
   Bemerkung: Die Verteilung von $Y_1 + Y_2$ heisst Chiquadrat-Verteilung mit zwei Freiheitsgraden und man schreibt $Y \sim \chi_2^2$.

   c) Beweisen Sie via Induktion, dass die Summe $Y_1 + \cdots + Y_n$ die Dichte
   $$f_n(x) = c_n\, x^{n/2-1}\, e^{-x/2}\, \mathbf{1}_{\{x>0\}}$$
   für ein $c_n > 0$ hat.
   Bemerkung: Die Verteilung von $Y_1 + \cdots + Y_n$ heisst Chiquadrat-Verteilung mit n Freiheitsgraden und man schreibt $Y \sim \chi_n^2$.

Lösung:

a) Via Verteilungsfunktion

$$F_Y(y) = \mathbb{P}[Y \le y] = \mathbb{P}[-\sqrt{y} \le X \le \sqrt{y}] = 2F_X(\sqrt{y}) - 1$$

berechnet man

$$f_Y(y) = f_X(\sqrt{y})\, y^{-1/2} = c\, e^{-y/2}\, y^{-1/2}$$

für $y > 0$.

b) Aus der Faltungsformel erhalten wir für $x > 0$

$$\begin{aligned}
f_2(x) &= \int_0^x f_Y(x-y)\, f_Y(y)\, dy = \int_0^1 f_Y(x(1-z))\, f_Y(xz)\, x\, dz \\
&= \int_0^1 k\, e^{-x(1-z)/2}\, e^{-xz/2}\, (x(1-z))^{-1/2}\, (xz)^{-1/2}\, x\, dz \\
&= k\, e^{-x/2} \int_0^1 (1-z)^{-1/2}\, z^{-1/2}\, dz.
\end{aligned}$$

Das Integral hängt nicht von x ab. Somit ist die Behauptung gezeigt.

c) Aus a) und b) wissen wir, dass die Behauptung für $n = 1, 2$ stimmt. Für $n > 1$ gilt (für $x > 0$):

$$\begin{aligned}
f_{n+1}(x) &= \int_0^x f_n(y)\, f_Y(x-y)\, dy \\
&= \int_0^x c_n y^{n/2-1}\, e^{-y/2}\, c\, (x-y)^{-1/2}\, e^{-(x-y)/2}\, dy \\
&= c\, c_n\, e^{-x/2} \int_0^x y^{n/2-1}\, (x-y)^{-1/2}\, dy \\
&= c\, c_n\, e^{-x/2} \int_0^1 (xz)^{n/2-1}\, (x(1-z))^{-1/2}\, x\, dz \\
&= c\, c_n\, e^{-x/2}\, x^{(n+1)/2-1} \int_0^1 z^{n/2-1}\, (1-z)^{-1/2}\, dz.
\end{aligned}$$

Das Integral ist wiederum unabhängig von x, womit die Behauptung gezeigt ist.

3. Die Zufallsvariable X_ν sei χ^2_ν-verteilt mit $\nu \in \mathbb{N}$. Das heisst insbesondere, dass $\mathbb{E}[X_\nu] = \nu$ und $\mathrm{Var}\,(X_\nu) = 2\nu$ gilt.

   a) Geben Sie mit Hilfe der Chebyshev-Ungleichung eine untere Schranke für die Wahrscheinlichkeit

   $$\mathbb{P}\left(\left| \frac{X_\nu}{\nu} - 1 \right| \le 0.5 \right)$$

   an. Welchen Wert hat die Schranke für $\nu = 40$?

   b) Berechnen Sie für $\nu = 40$ die obige Wahrscheinlichkeit mit Hilfe des Zentralen Grenzwertsatzes. Vergleichen Sie diesen Wert mit der in Teilaufgabe a) erhaltenen Schranke.

Lösung:

a) Chebyshev-Ungleichung (Slides S. 206):

$$\mathbb{P}(|X_\nu - \nu| > c) \leq \mathrm{Var}(X_\nu)/c^2 = \frac{2\nu}{c^2}.$$

Also folgt:

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{X_\nu}{\nu} - 1\right| \leq 0.5\right) &= \mathbb{P}\left(\left|\frac{X_\nu - \nu}{\nu}\right| \leq 0.5\right) = 1 - \mathbb{P}\left(|X_\nu - \nu| > \frac{\nu}{2}\right) \\
&\geq 1 - \frac{2\nu}{\nu^2/4} = \frac{\nu - 8}{\nu} \quad (= 0.8 \text{ für } \nu = 40)
\end{aligned}$$

b) Y_i i.i.d. $\sim \chi_1^2$, $\mathbb{E}[Y_i] = 1$, $\mathrm{Var}[Y_i] = 2$, $i = 1, \ldots, \nu$. Der zentrale Grenzwertsatz für $X_\nu = \sum_{i=1}^{\nu} Y_i$ lautet:

$$Z = \frac{X_\nu - \nu\mathbb{E}[Y_i]}{\sqrt{\nu \mathrm{Var}[Y_i]}} = \frac{X_\nu - \nu}{\sqrt{2\nu}} \xrightarrow[\nu \to \infty]{\mathcal{D}} N(0,1)$$

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{X_\nu}{\nu} - 1\right| \leq 0.5\right) &= \mathbb{P}\left(\frac{\nu}{2} \leq X_\nu \leq \frac{3\nu}{2}\right) = \mathbb{P}\left(20 \leq X_{40} \leq 60\right) \qquad (\nu = 40) \\
&= \mathbb{P}\left(\frac{20 - 40}{\sqrt{80}} \leq \frac{X_{40} - 40}{\sqrt{80}} \leq \frac{60 - 40}{\sqrt{80}}\right) \\
&\approx \Phi\left(\frac{20}{\sqrt{80}}\right) - \Phi\left(\frac{-20}{\sqrt{80}}\right) = 0.975\,.
\end{aligned}$$

In a) hatten wir $\mathbb{P}\left(\left|\frac{X_\nu}{\nu} - 1\right| \leq 0.5\right) \geq 1 - \frac{8}{\nu} = 0.8$ für $\nu = 40$. Die Chebyshev-Ungleichung ist allgemein gültig und daher meist relativ grob ("schlecht").

4. Berechnen Sie mit Hilfe des zentralen Grenzwertsatzes

$$\lim_{n \to \infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

Hinweis: Betrachten Sie i.i.d. Zufallsvariablen $X_1, X_2, \ldots$ mit $X_i \sim \mathrm{Poi}(1)$ und setzen Sie $S_n = X_1 + \cdots + X_n$.

Lösung: Wir betrachten $X_1, X_2, \ldots$ i.i.d. mit $X_i \sim \mathrm{Poi}(1)$. Es gilt $\mathbb{E}[X_1] = 1 = \mathrm{Var}[X_1]$. Ausserdem ist $S_n = X_1 + \cdots + X_n$ wieder Poisson verteilt mit Parameter n. Aus dem zentralen Grenzwertsatzes folgt

$$\begin{aligned}
e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} &= \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!} = \sum_{k=0}^{n} \mathbb{P}[S_n = k] = \mathbb{P}[S_n \leq n] \\
&= \mathbb{P}\left[\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{n - n\mu}{\sqrt{n}\sigma}\right] = \mathbb{P}\left[\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq 0\right] \stackrel{n \to \infty}{\longrightarrow} \Phi(0) = \frac{1}{2}.
\end{aligned}$$

5. Eine Cauchy-verteilte Zufallsvariable, $X \sim$ Cauchy, hat die Dichte

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \qquad x \in \mathbb{R}.$$

Seien nun $X, Y \sim$ i.i.d. Cauchy. Zeigen Sie, dass $(X+Y)/2 \sim$ Cauchy.

Hinweis: Benutze den Residuensatz, um charakteristische Funktionen zu berechnen.

Lösung: Wir betrachten $X, Y \sim$ i.i.d. Cauchy.

Aus der Dichte folgt mit dem Residuensatz (siehe `http://en.wikipedia.org/wiki/Residue_theorem#Example`) folgt:

$$\mathbb{E}[e^{iuX}] = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{iux} \frac{1}{1+x^2} \, dx = e^{-|u|}.$$

Daraus kann nun die charakteristische Funktion von $(X+Y)/2$ berechnet werden:

$$\mathbb{E}[e^{iu(X+Y)/2}] = \mathbb{E}[e^{iuX/2} \, e^{iuY/2}] = \mathbb{E}[e^{iuX/2}] \, \mathbb{E}[e^{iuY/2}] = e^{-|u|}.$$

Aus der Eindeutigkeit der charakteristischen Funktion folgt nun, dass $(X+Y)/2 \sim$ Cauchy-verteilt ist.