Konzentrationsungleichungen

Übungsblatt 1

TU Dortmund, Sommersemester 2021

Prof. Dr. Ivan Veselić

Übung 1 (4 Punkte)**.** *In der Vorlesung wurde in einer Bemerkung erwähnt, dass "Momentenabschätzungen im Allgemeinen besser sind als Chernoffabschätzungen". Wir wollen diese Aussage nun präzisieren.*
Sei X eine nichtnegative Zufallsvariable und $t > 0$. Beweisen Sie, dass die kleinste Momentenschranke *kleiner gleich der* kleinsten Chernoffschranke *ist, d.h.*

$$\inf_{q\in\mathbb{N}} \frac{\mathbb{E}(X^q)}{t^q} \leq \inf_{\lambda>0} \frac{\mathbb{E}(\mathrm{e}^{\lambda X})}{\mathrm{e}^{\lambda t}}.$$

Tipp: zeigen sie zuerst: Falls $a_1, ..., a_n, b_1, ..., b_n \geq 0$ (nicht alle $b_k = 0$), so gilt

$$\frac{a_1 + \cdots + a_n}{b_1 + \cdots + b_n} \geq \min_{1\leq q\leq n} \frac{a_q}{b_q}. \tag{1}$$

Beweis. Wir zeigen zunächst (1) per Induktion: Der Fall $n = 1$ gilt offensichtlich. Wir nehmen nun (1) für n an. Dann folgt

$$\begin{aligned}\frac{a_1 + \cdots + a_{n+1}}{b_1 + \cdots + b_{n+1}} &= \left(\frac{b_1 + \cdots + b_n}{b_1 + \cdots + b_{n+1}}\right) \cdot \frac{a_1 + \cdots + a_n}{b_1 + \cdots + b_n} + \left(\frac{b_{n+1}}{b_1 + \cdots + b_{n+1}}\right) \cdot \frac{a_{n+1}}{b_{n+1}} \\ &\geq \min_{1\leq q\leq n+1} \frac{a_q}{b_q} \left[\left(\frac{b_1 + \cdots + b_n}{b_1 + \cdots + b_{n+1}}\right) + \left(\frac{b_{n+1}}{b_1 + \cdots + b_{n+1}}\right)\right] = \min_{1\leq q\leq n+1} \frac{a_q}{b_q}\end{aligned}$$

und die Induktion liefert den Beweis. Nun können wir fortfahren und die Taylorentwicklung der Exponentialfunktion nutzen, da $X \geq 0$

$$\frac{\mathbb{E}\left[\mathrm{e}^{\lambda X}\right]}{\mathrm{e}^{\lambda t}} = \lim_{n\to\infty} \frac{\sum_{q=1}^n \mathbb{E}\left[\frac{(\lambda X)^q}{q!}\right]}{\sum_{q=1}^n \frac{(\lambda t)^q}{q!}}.$$

Jetzt nutzen wir (1) und erhalten

$$\frac{\mathbb{E}\left[\mathrm{e}^{\lambda X}\right]}{\mathrm{e}^{\lambda t}} \leq \lim_{n\to\infty} \min_{1\leq q\leq n} \frac{\mathbb{E}[X^q]}{t^q} = \inf_{q\in\mathbb{N}} \frac{\mathbb{E}[X^q]}{t^q}.$$

für alle λ. □

Übung 2 (4 Punkte)**.** *Für eine Zufallsvariable X hatten wir die* momentenerzeugende Funktion $M(\lambda) = \mathbb{E}\left[\mathrm{e}^{\lambda X}\right] \in [0, \infty]$*, die* kumulantenerzeugende Funktion $\psi(\lambda) := \ln M(\lambda)$ *und für $t > \mathbb{E}(X)$ die* Cramér-Transformierte $\psi^*(t) := \sup_{\lambda\geq 0} (\lambda t - \psi(\lambda))$ *definiert.*
Zeigen Sie:

a) Existiert ein $\lambda_0 > 0$ so dass $M(\lambda_0) < \infty$, so gilt auch $M(\lambda) < \infty$ für alle $\lambda \in [0, \lambda_0]$.

b) Setze $b := \sup\{\lambda \geq 0 \colon M(\lambda) < \infty\}$. Dann ist ψ auf dem Intervall $(0, b)$

  i) unendlich oft differenzierbar,

  ii) konvex

c) Ist X eine zentrierte Zufallsvariable, so gilt zudem

i) $\psi : [0, b) \to \mathbb{R}$ *ist* $C^1([0, b))$,

ii) $\psi^*(t) = \sup_{\lambda \in (0,b)} (\lambda t - \psi(\lambda))$.

Beweis. a) Wir können zunächst die Fälle $X \geq 0$ und $X < 0$ unterscheiden:

$$M(\lambda) = \mathbb{E}\left[\mathrm{e}^{\lambda X} \mathbb{1}_{X \geq 0} + \mathrm{e}^{\lambda X} \mathbb{1}_{X<0}\right].$$

Da $\mathbb{E}[\mathrm{e}^{\lambda X} \mathbb{1}_{X<0}] \leq \mathbb{E}[1] = 1$, reicht es $\mathbb{E}[\mathrm{e}^{\lambda X} \mathbb{1}_{X \geq 0}] < \infty$ zu zeigen. Jetzt können wir unser Wissen über λ_0 sowie die Abschätzung
$\mathrm{e}^{(\lambda - \lambda_0) X} \mathbb{1}_{X \geq 0} \leq 1$ nutzen, sodass

$$\begin{aligned} \mathbb{E}[\mathrm{e}^{\lambda X} \mathbb{1}_{X \geq 0}] &= \mathbb{E}(\mathrm{e}^{\lambda_0 X} \mathbb{1}_{X \geq 0} \cdot \mathrm{e}^{(\lambda - \lambda_0) X} \mathbb{1}_{X \geq 0}) \leq \mathbb{E}(\mathrm{e}^{\lambda_0 X} \mathbb{1}_{X \geq 0}) \cdot 1 \\ &\leq \mathbb{E}(\mathrm{e}^{\lambda_0 X}) = M(\lambda_0) < \infty. \end{aligned}$$

b) i) Da $0 < M(\lambda) < \infty$ und $\ln : (0, \infty) \to \mathbb{R}$ in C^∞ liegt, reicht es zu zeigen, dass $\lambda \mapsto M(\lambda)$ in C^∞ liegt.
Dazu zeigen wir zuerst das folgende Lemma:

Lemma 0.1. *Sei $I \subset \mathbb{R}$ offen und $\{Y(\lambda)\}_{\lambda \in I}$ eine Familie von $L^1(\Omega)$-Zufallsvariablen, sodass $\partial_\lambda^n Y(\lambda)$, $n = 1, \ldots, N$ für alle $\lambda \in I$ fast sicher existiert. Falls es $L^1(\omega)$-Zufallsvariablen $G_1, ..., G_N$ gibt, so dass fast sicher $\sup_{\lambda \in I} |\partial_\lambda^n Y_n(\lambda)| \leq G_n$, so gilt*

- $\partial_\lambda^N Y(\lambda) \in L^1(\omega)$,
- $\partial_\lambda^N \mathbb{E} Y(\lambda)$ *existiert und*
- $$\partial_\lambda^N \mathbb{E} Y(\lambda) = \mathbb{E}(\partial_\lambda^N Y(\lambda))$$

Beweis des Lemma. Wir nutzen den Satz von der majorisierten Konvergenz und machen uns für die Konstruktion der Majorante den Mittelwertsatz zunutze.
Wir werden nur den Fall $n = 1$ zeigen, die übrigen Fälle folgen per Induktion.
Nach Voraussetzung existiert $\partial_\lambda Y(\lambda) = Y'(\lambda)$. Aus dem Mittelwertsatz folgt, dass für jedes $h > 0$, dass klein genug ist, damit $[\lambda, \lambda + h] \subset I$ gilt, ein $\theta \in [\lambda, \lambda + h]$ existiert, sodass

$$|\frac{Y(\lambda + h) - Y(\lambda)}{h}| = |Y'(\theta)| \leq \sup_{t \in I} |Y'(t)| \leq G_1.$$

Somit haben wir eine Majorante für

$$\lim_{h \to 0} \frac{Y(\lambda + h) - Y(h)}{h}$$

und der Satz von der dominierten Konvergenz liefert die Aussagen. □

Es reicht nun, die Differenzierbarkeit auf Intervallen $(c, d) \subset (0, b)$ mit $d < b$ zu zeigen. Es gilt

$$\partial_\lambda^n \mathrm{e}^{\lambda X} = X^n \mathrm{e}^{\lambda X}.$$

Als Majorante G_n auf (c, d) wählen wir daher

$G_n := |X|^n \mathrm{e}^{dX} \mathbb{1}_{\{X \geq 0\}} + |X|^n \mathrm{e}^{cX} \mathbb{1}_{\{X<0\}}$ und müssen somit nur $G_n \in L^1$ zeigen. Dazu machen wir uns

$$\forall \epsilon > 0,\ x \in \mathbb{R}:\ |x| \leq \frac{\mathrm{e}^{\epsilon|x|}}{\epsilon}$$

zu nutze.

$$\begin{aligned}
\mathbb{E}\left[|X|^n \mathrm{e}^{dX}\right] &= \mathbb{E}\left[|X|^n \mathrm{e}^{dX} \mathbb{1}_{\{X\geq 0\}}\right] + \mathbb{E}\left[|X|^n \mathrm{e}^{dX} \mathbb{1}_{\{X<0\}}\right] \\
&\leq \mathbb{E}\left[\left(\frac{\mathrm{e}^{\epsilon X}}{\epsilon}\right)^n \mathrm{e}^{dX} \mathbb{1}_{\{X\geq 0\}}\right] + \mathbb{E}\left[\left(\frac{\mathrm{e}^{\epsilon |X|}}{\epsilon}\right)^n \mathrm{e}^{dX} \mathbb{1}_{\{X<0\}}\right] \\
&= \mathbb{E}\left[\frac{\mathrm{e}^{(d+n\epsilon)X}}{\epsilon^n} \mathbb{1}_{\{X\geq 0\}}\right] + \mathbb{E}\left[\left(\frac{\mathrm{e}^{-\epsilon X}}{\epsilon}\right)^n \mathrm{e}^{dX} \mathbb{1}_{\{X<0\}}\right] \\
&= \mathbb{E}\left[\frac{\mathrm{e}^{(d+n\epsilon)X}}{\epsilon^n} \mathbb{1}_{\{X\geq 0\}}\right] + \mathbb{E}\left[\frac{\mathrm{e}^{(d-n\epsilon)X}}{\epsilon^n} \mathbb{1}_{\{X<0\}}\right] < \infty
\end{aligned}$$

durch die Wahl $0 < \epsilon < \min\left(\frac{b-d}{n}, \frac{d}{n}\right)$. Diese Rechnung lässt sich genauso für c durchführen, und da außerdem

$$\forall a \in (c,d) : |X|^n \mathrm{e}^{aX} \leq |X|^n \mathrm{e}^{dX} \mathbb{1}_{X\geq 0} + |X|^n \mathrm{e}^{cX} \mathbb{1}_{\{X<0\}}$$

gilt und beide Summanden in der Majorante in L^1 liegen existieren aufgrund des Satz von der dominierten Konvergenz alle Ableitungen und da unsere Majorante nicht nur für die Differnzenquotienten sondern auch für die Ableitungen selbst gilt, sind diese auch stetig.

ii) Da ψ zwei mal stetig differnzierbar ist, können wir zur Überprüfung der Konvexität einfach überprüfen, ob die zweite Ableitung positiv ist:

$$\psi''(\lambda) = \partial_\lambda \left(\frac{M'(\lambda)}{M(\lambda)}\right) = \frac{M(\lambda)M''(\lambda) - M'(\lambda)^2}{M(\lambda)^2} \geq 0$$

denn

$$\begin{aligned}
M'(\lambda) &= \mathbb{E}(X\mathrm{e}^{\lambda X}) = \mathbb{E}(\mathrm{e}^{\lambda X/2} \cdot X\mathrm{e}^{\lambda X/2}) \\
&\leq \mathbb{E}(\mathrm{e}^{\lambda X})^{1/2} \cdot \mathbb{E}(X^2\mathrm{e}^{\lambda X})^{1/2} = M(\lambda)^{1/2} M''(\lambda)^{1/2}.
\end{aligned}$$

durch Cauchy-Schwarz.

c) i) Für $I = [0,b)$ funktioniert unsere Rechnung aus Teil b) i) nicht mehr ganz, aber wir können sie zumindest im Fall $n = 1$ anpassen. Wir können weiterhin das Lemma nutzen, wir müssen jedoch unsere Majorante ändern. Nun gilt

$$\begin{aligned}
\forall a \in [0,d) : |X| \mathrm{e}^{aX} &\leq |X| \mathrm{e}^{dX} \mathbb{1}_{X\geq 0} + |X| \mathrm{e}^{0X} \mathbb{1}_{\{X<0\}} \\
&= |X| \mathrm{e}^{dX} \mathbb{1}_{X\geq 0} + |X| \mathbb{1}_{\{X<0\}}
\end{aligned}$$

Der Erwartungswert des ersten Summanden lässt sich wie gehabt abschätzen, für den zweiten sehen wir, dass

$$\mathbb{E}\left(|X| \mathbb{1}_{\{X<0\}}\right) = \mathbb{E}\left(-X\mathbb{1}_{\{X<0\}}\right) < \infty$$

da X zentriert und somit integrierbar ist, weshalb auch das Integral des Negativteils $\mathbb{E}\left(-X\mathbb{1}_{\{X<0\}}\right)$ nach Definition existieren und endlich sein muss. Dominierte

Konvergenz liefert anschließend wieder Differnzierbarkeit und Stetigkeit der Ableitung.
Bemerkung: Wie an der Herleitung zu sehen, ist der Erwartungswert des zweiten Summanden im Fall $n > 1$ auch endlich, falls $X \geq 0$ oder falls alle absoluten Momente endlich sind. In diesem Fall ist ψ sogar in $C^\infty([0,b))$.

ii) Es reicht, das Supremum über $[0,b)$ zu bilden, da $\lambda t - \psi(\lambda)$ sonst $-\infty$ ist. Wir können weiterhin $\lambda = 0$ ausschließen, da wir bei $\lambda = 0$

$$\lambda t - \psi(\lambda) = 0 \quad \text{und} \quad \partial_\lambda(\lambda t - \psi(\lambda)) = t - \psi'(0) = t > 0,$$

haben, sodass die Funktion rechts von 0 größer als bei 0 sein muss.

□

Übung 3 (4 Punkte)**.** *Sei X eine Zufallsvariable aus $L^2(\mathbb{P})$ und $\mathrm{Me}(X) \in \mathbb{R}$ ein Median (d.h. $\mathbb{P}(X \geq \mathrm{Me}(X)) \geq 1/2$ und $\mathbb{P}(X \leq \mathrm{Me}(X)) \geq 1/2$).*

i) Zeigen Sie, dass $\mathrm{Me}(X)$ die Funktion

$$\mathbb{R} \ni t \mapsto \mathbb{E}|X - t|$$

minimiert.

ii) Zeigen Sie

$$|\mathbb{E}(X) - \mathrm{Me}(X)| \leq \sqrt{\mathrm{Var}(X)}.$$

Beweis. i) Sei $t \geq \mathrm{Me}(X)$. Dann gilt

$$\begin{aligned}
&|X - t| - |X - \mathrm{Me}(X)| \\
&= (\mathrm{Me}(X) - t)\mathbb{1}_{\{X \geq t\}} + (\mathrm{Me}(X) + t - 2X)\mathbb{1}_{\{\mathrm{Me}(X) < X < t\}} \\
&\quad + (t - \mathrm{Me}(X))\mathbb{1}_{\{X \leq \mathrm{Me}(X)\}} \\
&= (t - \mathrm{Me}(X))(2\mathbb{1}_{\{X \leq \mathrm{Me}(X)\}} - 1) + 2(t - X)\mathbb{1}_{\{\mathrm{Me}(X) < X < t\}}.
\end{aligned}$$

Somit folgt

$$\begin{aligned}
&\mathbb{E}\left(|X - t| - |X - \mathrm{Me}(X)|\right) \\
&= (t - \mathrm{Me}(X))(2\mathbb{P}(X \leq \mathrm{Me}(X)) - 1) + 2\mathbb{E}[(t - X)\mathbb{1}_{\{\mathrm{Me}(X) < X < t\}}] \geq 0 + 0
\end{aligned}$$

Der Fall $t \leq \mathrm{Me}(X)$ funktioniert analog.

ii) Der Betrag $|\cdot|$ ist konvex, daher können wir die Jensensche Ungleichung nutzen und erhalten

$$|\mathbb{E}(X) - \mathrm{Me}(X)| = |\mathbb{E}(X - \mathrm{Me}(X))| \leq \mathbb{E}|X - \mathrm{Me}(X)|.$$

Der Median $\mathrm{Me}(X)$ minimiert $t \mapsto \mathbb{E}|X - t|$, daher können wir weiter abschätzen

$$... \leq \mathbb{E}|X - \mathbb{E}(X)| \leq \mathbb{E}((X - \mathbb{E}(X))^2)^{1/2},$$

wobei wir im letzten Schritt wieder Jensen genutzt haben.

□

Übung 4 (4 Punkte)**.** *Zeigen sie für $n \in \mathbb{N}$, $D \in \mathbb{N}$ mit $1 \leq D < \frac{n}{2}$*

$$\sum_{j=0}^{D} \binom{n}{j} \leq \left(\frac{en}{D}\right)^D$$

mithilfe der Cramér-Chernoff-Methode für binomialverteilte Zufallsvariablen.

Beweis. Anmerkung: Da $\left(\frac{D}{D+1}\right)^D \geq \frac{1}{\mathrm{e}}$ ist die rechte Seite der zu zeigenden Gleichung monoton wachsend in D und somit größer als $\left(\sqrt{2\mathrm{e}}\right)^n > 2^n$ falls $D \geq n/2$. Die linke Seite wiederum ist immer kleiner als $\sum_{j=0}^n \binom{n}{j} = 2^n$, somit gilt die Ungleichung auch für $n/2 \leq D \leq n$.
Den Ausgangspunkt der Lösung bildet die Beobachtung, dass für eine mit $Bin(n, \frac{1}{2})$ verteilte Zufallsvariable S_n gilt, dass

$$\mathbb{P}(S_n \geq n - D) = \mathbb{P}(S_n \leq D) = \sum_{j=0}^{D} \binom{n}{j} \left(\frac{1}{2}\right)^n.$$

Wie in der Vorlesung gesehen lässt sich die Cramér-Transformierte einer binomialverteilten ZV Z über

$$\psi_Z^*(t) = n \cdot \psi_X^*\left(\frac{t}{n}\right)$$

aus der Cramér-Transformierten einer Bernoulli-verteilten ZV berechnen.
Dazu berechnen wir zunächst die Kumulantenerzeugende Funktion einer Bernoulli-ZV X mit $p = 1/2$:

$$\psi_X(\lambda) = \ln\left(\frac{1}{2}\mathrm{e}^\lambda + \frac{1}{2}\right)$$

sowie deren Ableitung

$$\psi_X'(\lambda) = \frac{1}{1 + \mathrm{e}^{-\lambda}}.$$

Der minimierende Parameter λ_t ergibt sich dann für ein $t \in \left(\frac{1}{2}, 1\right)$ über

$$\psi_X'(\lambda_t) \stackrel{!}{=} t$$

und ist $\lambda_t = \ln\left(\frac{t}{1-t}\right)$.
Die Cramér-Transformierte ist somit

$$\begin{aligned}
\psi_X^*(t) &= t\lambda_t - \psi_X(\lambda_t) \\
&= t\ln\left(\frac{t}{1-t}\right) - \ln\left(\frac{1}{2}\mathrm{e}^{\ln\left(\frac{t}{1-t}\right)} + \frac{1}{2}\right) \\
&= \ln(2) + t\ln(t) + (1-t)\ln(1-t) =: h(t)
\end{aligned}$$

womit wir zu S_n zurückkehren können. Mit der Cramér-Chernoff-Methode ergibt sich die Abschätzung

$$\mathbb{P}(S_n \geq n - D) \leq \exp\left(-nh\left(\frac{n-D}{n}\right)\right) = \exp\left(-nh\left(1 - \frac{D}{n}\right)\right).$$

Mit $y := \frac{D}{n}$ ergibt sich

$$h\left(1 - \frac{D}{n}\right) = h(1-y) = \ln(2) + y\ln(y) + (1-y)\ln(1-y).$$

Über differenzieren lässt sich zeigen, dass

$$\forall y < 1: \ (1-y)\ln(1-y) + y \geq 0$$

sodass

$$h(1-y) \geq \ln(2) - y + y\ln(y)$$

folgt.
Daraus ergibt sich dann

$$\begin{aligned}\sum_{j=0}^{D}\binom{n}{j} &= 2^n\mathbb{P}(S_n \geq n-D) \leq 2^n \exp\left(-n(\ln(2) - y + y\ln(y))\right) \\ &= \exp\left(D - D\ln\left(\frac{D}{n}\right)\right) = \left(\frac{\mathrm{e}n}{D}\right)^D\end{aligned}$$

$\square$

Abgabe bis zum 26.04.2021