Zufällige Graphen

Christoph Schumacher

04. Mai 2020

RGCN I, Kap. 2.4: Ungleichungen für Wahrscheinlichkeiten

Zusammenfassung

- Markov-Ungleichung bzw. first moment method,
- Chebychev-Ungleichung bzw. second moment method,
- Große Abweichungen

Inhaltsverzeichnis

Fragen

1. —

1 Markov-Ungleichung

Zum Beweis von (2.4.2):

$$\mathbb{P}(X \neq 0) = \mathbb{P}(X \geq 1) \overset{(2.4.1)}{\leq} \mathbb{E}[X] \leq m \tag{1}$$

2 Chebychev-Ungleichung

Für die second moment method gibt es eine etwas bessere Schranke mit Hilfe der Cauchy-Schwarz-Ungleichung:

Lemma. *Sei X eine ZV mit $\mathbb{E}[X^2] > 0$. Dann gilt*

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X^2]}. \tag{2}$$

Beweis. Mit der Cauchy-Schwarz-Ungleichung folgt

$$\begin{aligned}
(\mathbb{E}[X])^2 &= (\mathbb{E}[(\mathbf{1}_{\{X\neq 0\}}X])^2 && (3)\\
&\leq \mathbb{E}[\mathbf{1}^2_{\{X\neq 0\}}]\mathbb{E}[X^2] && (4)\\
&= \mathbb{P}(X \neq 0)\mathbb{E}[X^2] && (5)\\
&= \mathbb{E}[X^2] - \mathbb{P}(X = 0)\mathbb{E}[X^2]. && (6)
\end{aligned}$$

Auflösen ergibt (2). □

3 Große Abweichungen

3.1 Kontext zur Ratenfunktion

Die Terme in der Ratenfunktion haben Namen:

- $M_X\colon (-\varepsilon, \varepsilon) \to \mathbb{R}$, $M_X(t) := \mathbb{E}[\mathrm{e}^{tX}]$ ist die *momentenerzeugende Funktion* der ZV X, (wenn es ein $\varepsilon > 0$ gibt, so dass der Erwartungswert endlich ist). Der Name ist ganz passend, weil die Ableitungen in $t = 0$ die Momente sind:

$$\begin{aligned}
M_X(0) &= \mathbb{E}[\mathrm{e}^{tX}]|_{t=0} &&= 1 && (7)\\
M'_X(0) &= \mathbb{E}[X\mathrm{e}^{tX}]|_{t=0} &&= \mathbb{E}[X] && (8)\\
M''_X(0) &= \mathbb{E}[X^2\mathrm{e}^{tX}]|_{t=0} &&= \mathbb{E}[X^2] && (9)\\
M^{(r)}_X(0) &= \mathbb{E}[X^r\mathrm{e}^{tX}]|_{t=0} &&= \mathbb{E}[X^r] && (10)
\end{aligned}$$

Doch Vorsicht! Wenn man so rechnet, dann vertauscht man den Erwartungswert und die Ableitung, und das geht nicht immmer, sondern muss gerechtfertigt werden. Es genügt, wenn die Ableitung (lokal) uniform in t integrierbar ist, das heißt, wenn es eine ZV Y_r gibt so dass $|X^r \mathrm{e}^{tX}| \leq Y$ fast sicher und $\mathbb{E}[Y_r] < \infty$. Das ist zum Beispiel für alle r der Fall, wenn X fast sicher beschränkt ist, also es ein $C > 0$ gibt mit $X \leq C$ fast sicher, nämlich mit der Wahl $Y_r := C^r \mathrm{e}^{\varepsilon C}$. Für fast sicher nicht-negative Zufallsvariablen wie beispielsweise $X \sim \mathrm{Poi}(\lambda)$ kann man $Y_r := X^r \mathrm{e}^{\varepsilon X}$ nehmen, vorausgesetzt, das ist integrierbar:

$$\mathbb{E}[X^r \mathrm{e}^{\varepsilon X}] = \sum_{k=0}^{\infty} k^r \mathrm{e}^{\varepsilon k} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \sum_{k=0}^{\infty} k^r \frac{\mathrm{e}^{\varepsilon}\lambda}{k!} \mathrm{e}^{-\mathrm{e}^{\varepsilon}\lambda} \mathrm{e}^{\mathrm{e}^{\varepsilon}\lambda} \mathrm{e}^{-\lambda} = \mathrm{e}^{-(1-\mathrm{e}^{\varepsilon})\lambda} \mathbb{E}[Z^r] < \infty \tag{11}$$

mit $Z \sim \mathrm{Poi}(\mathrm{e}^{\mathrm{e}^{\varepsilon}\lambda})$. Frage: Ist die Bedingung (2.1.8) hinreichend, um Ableitung und Erwartungswert zu vertauschen?

- $g_X \colon t \mapsto \log \mathbb{E}[\mathrm{e}^{tX}]$ heißt *kumulantenerzeugende Funktion* der ZV X. Die Ableitungen definieren die *Kumulanten* von X:

$$\kappa_0(X) = g_X(0) = \log \mathbb{E}[\mathrm{e}^{tX}]|_{t=0} = 0 \tag{12}$$

$$\kappa_1(X) = g_X'(0) = \left.\frac{\mathbb{E}[X\mathrm{e}^{tX}]}{\mathbb{E}[\mathrm{e}^{tX}]}\right|_{t=0} = \mathbb{E}[X] \tag{13}$$

$$\kappa_2(X) = g_X''(0) = \left.\frac{\mathbb{E}[X^2\mathrm{e}^{tX}]\mathbb{E}[\mathrm{e}^{tX}] - (\mathbb{E}[X\mathrm{e}^{tX}])^2}{(\mathbb{E}[e^{tX}])^2}\right|_{t=0} \tag{14}$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[\mathrm{e}^{tX}])^2 = \mathrm{Var}(X) \tag{15}$$

Wikipedia listet interessante Eigenschaften der Kumulanten auf. Wir brauchen folgende Eigenschaft der kumulantenerzeugenden Funktion.

Lemma. *Die kumulantenerzeugende Funktion von X ist konvex. Wenn X nicht fast sicher konstant ist (d. h. $\mathrm{Var}(X) > 0$), dann ist die kumulantenerzeugende Funktion von X sogar strikt konvex.*

Beweis. Wir verwenden die zweite Ableitung von oben in der Form

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \log \mathbb{E}[e^{tX}] = \frac{\mathbb{E}[X^2\mathrm{e}^{tX}]\mathbb{E}[\mathrm{e}^{tX}] - (\mathbb{E}[X\mathrm{e}^{tX}])^2}{(\mathbb{E}[e^{tX}])^2} \tag{16}$$

$$= \mathbb{E}\Big[X^2 \frac{\mathrm{e}^{tX}}{\mathbb{E}[\mathrm{e}^{tX}]}\Big] - \Big(\mathbb{E}\Big[X \frac{\mathrm{e}^{tX}}{\mathbb{E}[\mathrm{e}^{tX}]}\Big]\Big)^2. \tag{17}$$

Das ist eine Varianz bezüglich dem mit der Dichte $f_t := \frac{\mathrm{e}^{tX}}{\mathbb{E}[\mathrm{e}^{tX}]}$ versehenen Wahrscheinlichkeitsmaß $Q_t := f_t \mathbb{P}$, definiert durch

$$Q_t(A) := \int_A f_t(\omega) \mathbb{P}(\mathrm{d}\omega), \tag{18}$$

denn damit ist beispielsweise

$$\mathbb{E}\Big[X\frac{\mathrm{e}^{tX}}{\mathbb{E}[\mathrm{e}^{tX}]}\Big] = \int X(\omega)\frac{\mathrm{e}^{tX(\omega)}}{\mathbb{E}[\mathrm{e}^{tX}]}\mathbb{P}(\mathrm{d}\omega) = \int X\,\mathrm{d}Q_t \tag{19}$$

Tatsächlich ist f_t eine Wahrscheinlichkeitsdichte, denn $f_t \geq 0$ und

$$\mathbb{E}[f_t] = \mathbb{E}\Big[\frac{\mathrm{e}^{tX}}{\mathbb{E}[\mathrm{e}^{tX}]}\Big] = \frac{\mathbb{E}[\mathrm{e}^{tX}]}{\mathbb{E}[\mathrm{e}^{tX}]} = 1. \tag{20}$$

Dass (17) positiv ist, können wir auch ohne das Maß Q_t direkt prüfen:

$$\begin{aligned}
0 &\leq \mathbb{E}[(X - \mathbb{E}[Xf_t])^2 f_t] && (21)\\
&= \mathbb{E}[(X^2 - 2X\mathbb{E}[Xf_t] + (\mathbb{E}[Xf_t])^2)f_t] && (22)\\
&= \mathbb{E}[X^2 f_t] - 2\mathbb{E}[Xf_t]\mathbb{E}[Xf_t] + \mathbb{E}[f_t](\mathbb{E}[Xf_t])^2 && (23)\\
&= \mathbb{E}[X^2 f_t] - (\mathbb{E}[Xf_t])^2. && (24)
\end{aligned}$$

Wenn X nicht fast sicher konstant (bzgl. $\mathbb{P}$ und damit auch bzgl. Q_t) ist, ist diese Varianz sogar strikt positiv und die kumulantenerzeugende Funktion strikt konvex. □

- Die Funktion

$$f^*\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}, \quad f^*(a) := \sup\{at - f(t) \mid t \in \mathbb{R}\}$$

heißt *Legendre-Transformierte* der konvexen Funktion $f\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$. Es gilt für alle $a, t \in \mathbb{R}$

$$f^*(a) \geq ta - f(t) \quad \text{bzw.} \quad f^*(a) + f(t) \geq ta \quad \text{bzw.} \quad f(a) \geq ta - f^*(t), \tag{25}$$

also ist (wenn $f^*(a) < \infty$) $a \mapsto ta - f^*(a)$ eine Stützgerade an f mit Steigung t. Interpretation der Legendre-Transformierten: $-f^*(t)$ ist der y-Achsenabschnitt der Stützgerade mit Steigung t.

Lemma. *Die Legendre-Transformierte f^* ist konvex.*

Beweis. Wir rechnen die Definition von Konvexität mit $a, a' \in \mathbb{R}$ und $\lambda \in [0, 1]$ nach:

$$\begin{aligned}
&f^*\big((1-\lambda)a + \lambda a'\big) && (26)\\
&= \sup\{\big((1-\lambda)a + \lambda a'\big)t - f(t) \mid t \in \mathbb{R}\} && (27)\\
&= \sup\{(1-\lambda)(at - f(t)) + \lambda(a't - f(t)) \mid t \in \mathbb{R}\} && (28)\\
&\leq \sup\{(1-\lambda)(at - f(t)) \mid t \in \mathbb{R}\} + \sup\{\lambda(a't - f(t)) \mid t \in \mathbb{R}\} && (29)\\
&= (1-\lambda)\sup\{at - f(t) \mid t \in \mathbb{R}\} + \lambda \sup\{a't - f(t) \mid t \in \mathbb{R}\} && (30)\\
&= (1-\lambda)f^*(a) + \lambda f^*(a'). && (31)
\end{aligned}$$

□

Bemerkung. Die Konvexität von f wurde in diesem Beweis nicht verwendet.

Die folgende Teilaussage des Satzes von Fenchel–Moreau besagt, dass die Ratenfunktion dieselbe Information wie die kumulantenerzeugende Funktion und wie die momentenerzeugende Funktion über die Verteilung von X enthält. Der Beweis steht der Vollständigkeit halber unten in Abschnitt 5.

Lemma. *Die Legendre-Transformation ist eine Involution auf unterhalbstetigen konvexen Funktionen, d. h. für jede unterhalbstetige konvexe Funktion* $f\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ *gilt* $(f^*)^* = f$.

3.2 Eigenschaften der Ratenfunktion

Eigenschaften der Ratenfunktion I (2.4.12):

- I ist nicht negativ:

$$I(a) = \sup_{t\in\mathbb{R}}\big(ta - \log \mathbb{E}[\mathrm{e}^{tX}]\big) \overset{t=0}{\geq} 0 - \log \mathbb{E}[1] = 0 \tag{32}$$

- I hat eine Nullstelle bei $a = \mathbb{E}[X]$: Die Funktionen $\mathbb{R} \ni x \mapsto \mathrm{e}^{tx}$ sind für alle $t \in \mathbb{R}$ konvex, denn $\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathrm{e}^{tx} = t^2\mathrm{e}^{tx} \geq 0$. Die Jensen-Ungleichung[1] besagt also

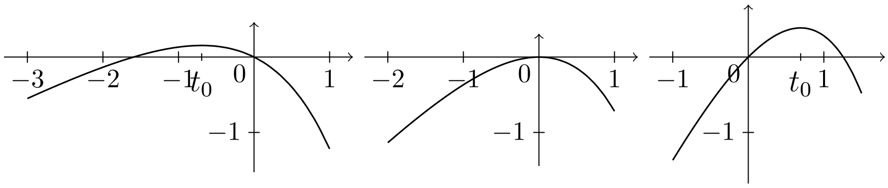
$$\mathbb{E}[\mathrm{e}^{tX}] \geq \mathrm{e}^{t\mathbb{E}[X]}. \tag{33}$$

Damit können wir $I(\mathbb{E}[X])$ nach oben abschätzen:

$$0 \leq I(\mathbb{E}[X]) = \sup_{t\in\mathbb{R}}\big(t\mathbb{E}[X] - \log \mathbb{E}[\mathrm{e}^{tX}]\big) \tag{34}$$

$$\leq \sup_{t\in\mathbb{R}}\big(t\mathbb{E}[X] - \log \mathrm{e}^{t\mathbb{E}[X]}\big) = 0. \tag{35}$$

- Warum ist $t \mapsto ta - \log \mathbb{E}[\mathrm{e}^{tX}]$ konkav? — Weil die kumulantenerzeugende Funktion konvex ist.
- Konkave Funktionen mit Ableitung in Null negativ, gleich Null und positiv:



[1] Die Jensen-Ungleichung besagt: Für jede konvexe Funktion ψ gilt $\mathbb{E}[\psi(X)] \geq \psi(\mathbb{E}[X])$. Moralische Begründung: Der Erwartungswert bildet eine (verallgemeinerte) Konvexkombination der Werte von X plus Definition von konvex. Eselsbrücke: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$

Die Maximalstelle t_0 der Funktion ist entsprechend negativ, gleich Null bzw. positiv.

• Weil die kumulantenerzeugende Funktion immer immer im Ursprung eine Nullstelle hat, siehe (12), erkennt man auch, dass das Supremum von $ta - \log \mathbb{E}[\mathrm{e}^{tX_1}]$ bei $a \neq \mathbb{E}[X_1]$ strikt positiv ist. Das bedeutet $I(a) > 0$ für alle $a \neq \mathbb{E}[X]$.

• Als Legendre-Transformierte der kumulantenerzeugenden Funktion ist die Ratenfunktion konvex, also ist $a = \mathbb{E}[X]$ die einzige Nullstelle von I.

Zum Beweis von Theorem 2.19: Zu (2.4.16):

$$\mathbb{P}\Big(\sum_{j=1}^n X_j \geq na\Big) \overset{(2.4.15)}{\leq} \inf_{t\geq 0}\big(\mathrm{e}^{-ta}\mathrm{e}^{\log \mathbb{E}[\mathrm{e}^{tX_1}]}\big)^n \tag{36}$$

$$= \big(\inf_{t\geq 0} \mathrm{e}^{-(ta-\log \mathbb{E}[\mathrm{e}^{tX_1}])}\big)^n \tag{37}$$

$$= \big(\mathrm{e}^{\inf_{t\geq 0}(-(ta-\log \mathbb{E}[\mathrm{e}^{tX_1}]))}\big)^n \tag{38}$$

$$= \big(\mathrm{e}^{-\sup_{t\geq 0}(ta-\log \mathbb{E}[\mathrm{e}^{tX_1}])}\big)^n \tag{39}$$

$$= \mathrm{e}^{-n\sup_{t\geq 0}(ta-\log \mathbb{E}[\mathrm{e}^{tX_1}])} \tag{40}$$

Für (2.4.9) wiederholt man am einfachsten den Beweis. Weil man dabei $t \leq 0$ verwendet, steht gleich am Anfang

$$\mathbb{P}\Big(\sum_{j=1}^n X_j \leq na\Big) = \mathbb{P}\Big(\mathrm{e}^{t\sum_{j=1}^n X_j} \geq \mathrm{e}^{tna}\Big), \tag{41}$$

und ab da ist der Beweis ganz genau der aus dem Buch, nur dass man immer $t \leq 0$ verwendet.

4 Anwendung bei der Binomialverteilung

Frage: Warum vergleicht man nach Korollar 2.20 die Bernoulli-Verteilung $\mathrm{Bin}(1,p)$ mit der Poisson-Verteilung $\mathrm{Poi}(p)$? — Die Summe von n unabhängigen $\mathrm{Poi}(p)$-verteilten ZVn ist $\mathrm{Poi}(np)$-verteilt, und das vergleichen wir mit $\mathrm{Bin}(n,p)$.

Nebenrechnungen:

• Zu (2.4.21): Der Erwartungswert ist

$$\mathbb{E}[\mathrm{e}^{tX_1}] = \sum_{k=0}^1 \mathrm{e}^{tk}\mathbb{P}(X_1 = k) = (1-p) + \mathrm{e}^t p. \tag{42}$$

Wir bestimmen das Supremum durch Ableiten und Nullsetzen. Ich nenne die Nullstelle der Ableitung t_0:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=t_0}\big(ta - \log(p\mathrm{e}^t + (1-p))\big) = a - \frac{p\mathrm{e}^{t_0}}{p\mathrm{e}^{t_0} + (1-p)} \overset{!}{=} 0 \tag{43}$$

$$\iff \quad a = \frac{pe^{t_0}}{pe^{t_0} + (1-p)} \tag{44}$$
$$\iff \quad a(pe^{t_0} + (1-p)) = pe^{t_0} \tag{45}$$
$$\iff \quad a(1-p) = pe^{t_0} - ape^{t_0} \tag{46}$$
$$\iff \quad a(1-p) = p(1-a)e^{t_0} \tag{47}$$
$$\iff \quad \frac{a(1-p)}{p(1-a)} = e^{t_0} \tag{48}$$
$$\iff \quad \log\frac{a(1-p)}{p(1-a)} = t_0 \tag{49}$$

Jetzt brauchen wir den Wert an der Stelle t_0:

$$t_0 a - \log(pe^{t_0} + (1-p)) = a\log\frac{a(1-p)}{p(1-a)} - \log\underbrace{\Big(p\frac{a(1-p)}{p(1-a)} + (1-p)\Big)}_{=(\frac{a}{1-a}+1)(1-p)=\frac{1-p}{1-a}} \tag{50}$$
$$= a\log\frac{a}{p} - a\log\frac{1-a}{1-p} + \log\frac{1-p}{1-a} \tag{51}$$
$$= a\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p}. \tag{52}$$

Der Beweis von Theorem 2.21 ist lang und rechenlastig. Grobe Struktur:

A) Bis inklusive (2.4.28): Markov-Ungleichung anwenden und die Problemstellung mit unterschiedlichen p_i auf das einheitliche $p = \lambda/n$ zurückführen.

B) Bis inklusive (2.4.31): Bisherige Schranke umschreiben und mit Hilfe von φ ausdrücken

C) Bis zum Schluss: φ hat eine quadratische Nullstelle ($\varphi(0) = \varphi'(0) = 0 \neq \varphi''(0)$) und wird möglichst gut mit einer quadratischen Funktion abgeschätzt und der Ausdruck so vereinfacht und die wesentliche Form herausarbeitet.

Details:

A) Das Ziel sind (2.4.24) und (2.4.25), das heißt, die Verteilung von X muss sich um $\mathbb{E}[X]$ konzentrieren. Dafür ist der Fall $p_j = p = \lambda/n$ der worst case unter allen p_j, $j \in \{1, \ldots, n\}$. Als Maß dafür, wie gut sich die Verteilung konzentriert bzw. umgekehrt sie sehr die Verteilung streut, können wir die Varianz betrachten:

$$\mathrm{Var}(X) \overset{(X_j)_j \text{ unabhängig}}{=} \sum_{j=1}^{n} \mathrm{Var}(X_j) = \sum_{j=1}^{n} p_j(1-p_j). \tag{53}$$

Optimierung über alle $p_j \in \mathbb{R}$ unter der Nebenbedingung $\sum_{j=1}^{n} p_j = \lambda$ mit dem Lagrange-Multiplikator μ:

$$\frac{\partial}{\partial p_j}\Big(\mathrm{Var}(X) + \mu\Big(\sum_{j=1}^{n} p_j - \lambda\Big)\Big) = 1 - 2p_j + \mu \overset{!}{=} 0 \tag{54}$$

$$\implies p_j = \frac{\mu+1}{2} \implies \lambda = \sum_{j=1}^{n} p_j = n\frac{\mu+1}{2} \tag{55}$$

$$\implies \mu = \frac{2\lambda}{n} - 1 \implies p_j = \frac{\frac{2\lambda}{n} - 1 + 1}{2} = \frac{\lambda}{n} \tag{56}$$

Der gefundene kritische Punkt ist ein Maximum, denn $\sum_{j=1}^{n} p_j(1-p_j) \to -\infty$ für $\|(p_j)_j\| \to \pm\infty$. Also ist $\mathrm{Var}(X) \le \mathrm{Var}(Y) = \lambda(1-\frac{\lambda}{n})$.
(Übrigens: $\mathrm{Var}(Y) \nearrow \lambda = \mathrm{Var}(Z)$ für $n \to \infty$ mit $Z \sim \mathrm{Poi}(\lambda)$!)

B) Beim Nachrechnen von (2.4.29) ersetzt man überall $p = \lambda/n$.
Um (2.4.30) in (2.4.31) umzuformen: $X' := n - X = \sum_{j=1}^{n}(1-X_j)$ ist von der gleichen Form wie X, nur mit $\lambda' := \sum_{j=1}^{n}(1-p_j) = n-\lambda$. Für X' liest sich (2.4.30) so: Für alle $t \in [0, n-\lambda'] = [0,\lambda]$:

$$\mathbb{P}(X' \ge \lambda' + t) \le \exp\Big(-\lambda'\varphi\Big(\frac{t}{\lambda'}\Big) - (n-\lambda')\varphi\Big(-\frac{t}{n-\lambda'}\Big)\Big) \tag{57}$$

$$\iff \mathbb{P}(n - X \ge n - \lambda + t) \le \exp\Big(-(n-\lambda)\varphi\Big(\frac{t}{n-\lambda}\Big) - \lambda\varphi\Big(-\frac{t}{\lambda}\Big)\Big), \tag{58}$$

und das ist (2.4.31).

C) Zur Abschätzung von φ:

- der „zweite Term“ in (2.4.30) bzw. (2.4.31) ist zumindest für große n vernachlässigbar, denn
$$(n-\lambda)\varphi\Big(\pm\frac{t}{n-\lambda}\Big) \xrightarrow{n\to\infty} 0, \tag{59}$$
da φ eine Nullstelle zweiter Ordnung in 0 hat. In (2.4.31) passiert dasselbe. (Frage: Wir werfen hier die n-Abhängigkeit weg. Sind die Abschätzungen (2.4.24) und (2.4.25) dann auch für Poisson-verteiltes X richtig? Siehe Anhang.)
- Die Ungleichung $\varphi'(x) = \log(x+1) \le x$ gilt für alle $x \in [-1,\infty)$, denn der Logarithmus ist konvex und die Winkelhalbierende ist Tangente/Stützgerade an φ im Ursprung. Die Abschätzung für $x \in [-1,0]$ ist etwas verwirrend, weil sowohl x also auch $\log(x+1)$ negativ sind:

$$\varphi(x) = -\varphi(0) - (-\varphi(x)) = -\int_x^0 \varphi'(\xi)\,\mathrm{d}\xi \tag{60}$$

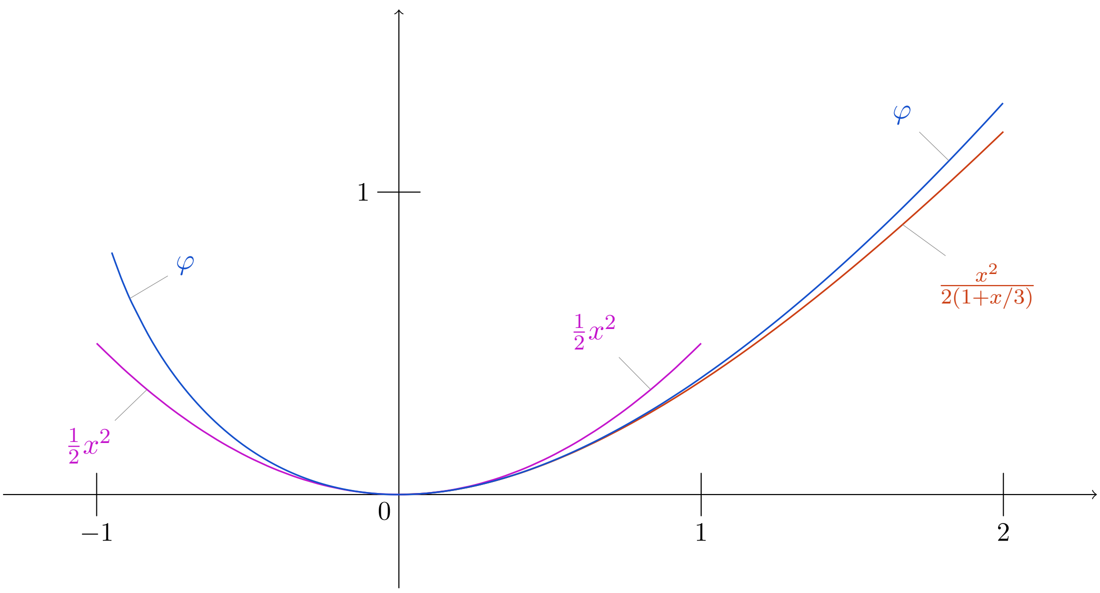
$$\ge -\int_x^0 \xi\,\mathrm{d}\xi = -\frac{\xi^2}{2}\Big|_x^0 = -0 - \Big(-\frac{x^2}{2}\Big) = \frac{x^2}{2}. \tag{61}$$

Für $x > 0$ geht die Abschätzung in die andere Richtung.

- Die Ungleichung in (2.4.32) prüft man mit der Bernoulli-Ungleichung oder dem binomischen Lehrsatz:

$$(1+\tfrac{x}{3})^3 = 1 + 3\tfrac{x}{3} + 3(\tfrac{x}{3})^2 + (\tfrac{x}{3})^3 \geq 1 + x. \tag{62}$$

  (Da ist wieder eine Tangente.)
- Die Skizze soll verdeutlichen,
  - dass $\varphi(x) \geq 0$,
  - $\varphi(0) = \varphi'(0) = 0$,
  - wie gut die Abschätzung von φ ist und
  - dass $\frac{x^2}{2}$ für $x > 0$ nicht funktioniert.



5 Zum Beweis des Satzes von Fenchel–Moreau

Lemma. *Die Legendre-Transformation ist eine Involution auf unterhalbstetigen konvexen Funktionen, d. h. für jede unterhalbstetige konvexe Funktion $f\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ gilt $(f^*)^* = f$.*

Beweis. Sei $f\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ konvex und $s \in \mathbb{R}$ mit $f(s) < \infty$. Dann gilt

$$\begin{aligned}(f^*)^*(s) &= \sup_{a\in\mathbb{R}} (as - f^*(a)) && (63)\\ &= \sup_{a\in\mathbb{R}} \big(as - \sup_{t\in\mathbb{R}} (at - f(t))\big) && (64)\\ &= \sup_{a\in\mathbb{R}} \inf_{t\in\mathbb{R}} \big(a(s-t) + f(t)\big) && (65)\end{aligned}$$

$$\overset{t:=s}{\leq} \sup_{a\in\mathbb{R}} f(s) = f(s). \tag{66}$$

Für die andere Ungleichung nutzen wir die Konvexität von f in s und wählen $a_0 \in \mathbb{R}$ als die Steigung einer Stützgerade an f in s, das heißt, es soll für alle $t \in \mathbb{R}$

$$f(t) \geq a_0(t-s) + f(s) \tag{67}$$

gelten. Damit schätzen wir ab:

$$
\begin{aligned}
(f^*)^*(s) &= \sup_{a\in\mathbb{R}}(as - f^*(a)) && (68)\\
&\geq a_0 s - f^*(a_0) && (69)\\
&= a_0 s - \sup_{t\in\mathbb{R}}(a_0 t - f(t)) && (70)\\
&= a_0 s + \inf_{t\in\mathbb{R}}(f(t) - a_0 t). && (71)
\end{aligned}
$$

Sei nun $\varepsilon > 0$. Wir finden ein $t_\varepsilon \in \mathbb{R}$, so dass

$$
\begin{aligned}
a_0 s + \inf_{t\in\mathbb{R}}(f(t) - a_0 t) &\geq a_0 s + f(t_\varepsilon) - a_0 t_\varepsilon - \varepsilon && (72)\\
&= f(t_\varepsilon) - a_0(t_\varepsilon - s) - \varepsilon && (73)\\
&\overset{(67)}{\geq} f(s) - \varepsilon. && (74)
\end{aligned}
$$

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, erhalten wir die gewünschte Ungleichung $(f^*)^*(s) \geq f(s)$.

Falls das Intervall $I_f := \{t \in \mathbb{R} \mid f(t) < \infty\}$ nur einen einzigen Punkt umfasst, sind wir fertig. Ansonsten umfasst die Menge der Punkte, in denen eine Stützgerade an f gewählt werden kann, das Innere von I_f und ist damit dicht in I_f. Mit Hilfe der Unterhalbstetigkeit folgt die Gleichheit von f und $(f^*)^*$ auch im Rand I_f und damit auf ganz $\mathbb{R}$. □

Bemerkung. Der Beweis zeigt noch mehr. Die Ungleichung $(f^*)^*(s) \leq f(s)$ gilt auch, wenn f nicht konvex ist, und an allen Stellen s mit Stützgerade gilt die Gleichheit $(f^*)^*(s) = f(s)$. Damit ist $(f^*)^*$ die konvexe Einhüllende von f, also die größte unterhalbstetige konvexe Funktion kleiner gleich f.

6 Anhang

Zum Beweis von Theorem 2.21: Man kann die n-abhängigen Terme schon früher loswerden, indem man nicht mit der $\mathrm{Bin}(n, \lambda/n)$-verteilten ZV Y sondern mit einer

Poi(λ)-verteilten ZV Z vergleicht. Dazu berechnen wir

$$\mathbb{E}[\mathrm{e}^{uZ}] = \sum_{k=0}^{\infty} \mathrm{e}^{uk} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{(\mathrm{e}^u \lambda)^k}{k!} = \mathrm{e}^{-\lambda} \mathrm{e}^{\mathrm{e}^u \lambda} = \mathrm{e}^{(\mathrm{e}^u - 1)\lambda} \tag{75}$$

und stellen fest, dass für alle $u \in \mathbb{R}$

$$\mathbb{E}[\mathrm{e}^{uY}] = \big(1 + (\mathrm{e}^u - 1)\lambda/n\big)^n \overset{\star}{\leq} \big(\exp((\mathrm{e}^u - 1)\lambda/n)\big)^n = \mathrm{e}^{(\mathrm{e}^u - 1)\lambda} = \mathbb{E}[\mathrm{e}^{uZ}]. \tag{76}$$

In der Ungleichung $\star$ verwenden wir $1 + x \leq \mathrm{e}^x$ und die Monotonie von $[0, \infty) \ni x \mapsto x^n$. Dafür brauchen wir $(\mathrm{e}^u - 1)\lambda \geq -n$, und das gilt wegen $\lambda = \sum_{j=1}^n p_j \leq n$. Damit können wir in (2.4.28) gleich so abschätzen:

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}[X] + t) &= \mathbb{P}(\mathrm{e}^{uX} \geq \mathrm{e}^{u(\mathbb{E}[X]+t)}) & (77)\\ &\leq \mathrm{e}^{-u(\lambda+t)} \mathbb{E}[\mathrm{e}^{uX}] \leq \mathrm{e}^{-u(\lambda+t)} \mathbb{E}[\mathrm{e}^{uZ}] & (78)\\ &= \mathrm{e}^{-u(\lambda+t)} \mathrm{e}^{(\mathrm{e}^u-1)\lambda} = \mathrm{e}^{(\mathrm{e}^u-1)\lambda - u(\lambda+t)}. & (79) \end{aligned}$$

Dieser Ausdruck führt schneller auf φ. Das optimale u erfüllt

$$\frac{\mathrm{d}}{\mathrm{d}u}\big((\mathrm{e}^u - 1)\lambda - u(\lambda + t)\big) = \lambda \mathrm{e}^u - (\lambda + t) \overset{!}{=} 0 \implies \mathrm{e}^u = \frac{\lambda + t}{\lambda}. \tag{80}$$

Bemerke, dass $\mathrm{e}^u = \frac{\lambda+t}{\lambda} \geq 1$ wegen $t \geq 0$, und das bedeutet $u \geq 0$ und u ist zugelassen. Also ist

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}[X] + t) &\leq \exp\Big(\Big(\frac{\lambda + t}{\lambda} - 1\Big)\lambda - (\lambda + t)\log\Big(\frac{\lambda + t}{\lambda}\Big)\Big) & (81)\\ &= \cdots = \exp(-\lambda\varphi(t/\lambda)). & (82) \end{aligned}$$

Der Nachteil bei dieser Abkürzung ist, dass man für (2.4.25) nochmal neu rechnen muss. Dazu nehmen wir $u \leq 0$:

$$\mathbb{P}(Y \leq \mathbb{E}[Y] - t) = \mathbb{P}(\mathrm{e}^{uY} \geq \mathrm{e}^{u(\mathbb{E}[Y]-t)}) \leq \cdots \leq \mathrm{e}^{(\mathrm{e}^u-1)\lambda - u(\lambda-t)}. \tag{83}$$

Diesmal erfüllt das optimale u

$$\frac{\mathrm{d}}{\mathrm{d}u}((\mathrm{e}^u - 1)\lambda - u(\lambda - t)) = \lambda \mathrm{e}^u - (\lambda - t) \overset{!}{=} 0 \implies \mathrm{e}^u = \frac{\lambda - t}{\lambda}. \tag{84}$$

Bemerke, dass $\mathrm{e}^u = \frac{\lambda-t}{\lambda} \leq 1$ wegen $t \leq 0$, und das bedeutet $u \leq 0$, wie gefordert. Wir erhalten

$$\begin{aligned} \mathbb{P}(Y \leq \mathbb{E}[Y] - t) &\leq \exp\Big(\Big(\frac{\lambda - t}{\lambda} - 1\Big)\lambda - (\lambda - t)\log\Big(\frac{\lambda - t}{\lambda}\Big)\Big) & (85)\\ &= \exp(-\lambda\varphi(-t/\lambda)). & (86) \end{aligned}$$