

Technische Universität Dortmund
Fakultät für Mathematik
Sommersemester 2021

Skript:

Konzentrationsungleichungen

Dozent: Prof. Dr. Ivan Veselić
verfasst von: Dennis Andreas Malcherczyk

Dieses Skript ist aus einer vierstündigen Vorlesung entstanden, dass ich 2017/2018 an der TU Dortmund hielt. Es orientierte sich im Wesentlichen an dem Buch: *Concentration Inequalities: A Nonasymptotic Theory of Independence* von Stéphane Boucheron, Gábor Lugosi und Pascal Massart.

Herr Malcherczyk war einer der Hörer und hat im Anschluss ein sehr detailliert ausgearbeitete Skript erstellt. Der vorliegende Text stellt einer leichte Bearbeitung meinerseits dar und verwendet einige von Christoph Schumacher erzeugte Graphiken.

Dortmund, März 2021

Ivan Veselić

Die Stoffauswahl steht noch nicht endgültig im Detail fest, das angezeigte Inhaltsverzeichnis soll Ihnen aber schon einen Einblick in den Stoff der Vorlesung bieten.

INHALTSVERZEICHNIS

1. Motivation	1
2. Grundlegende Ungleichungen	4
2.1. Markov-Ungleichung und Co.	4
2.2. Cramér-Chernoff-Methode	6
Die kumulantenerzeugende Funktion und die Cramér-Transformierte	8
Bestimmung des Supremums mittels der Ableitung	10
Cramér-Transformierte verschiedener Verteilungsklassen	11
2.3. Sub-Gaußsche Zufallsvariablen	16
2.4. Sub-Gamma-Zufallsvariablen	22
2.5. Eine Maximal-Ungleichung	25
Resultat für sub- Γ -Zufallsvariablen	26
2.6. Hoeffding-Ungleichung	28
2.7. Bennett-Ungleichung	29
2.8. Bernstein-Ungleichung	31
2.9. Johnson-Lindenstrauss-Lemma	34
2.10. Assoziations- und Korrelationsungleichungen*	37
2.11. Minkowski-Ungleichung*	40
3. Schranken an die Varianz	43
3.1. Efron-Stein-Ungleichung	43

3.2.	Funktionen mit beschränkter Differenz	47
3.3.	Selbstbeschränkende Funktionen	50
3.4.	Anwendungen: VC-Dimension und Perkolationen	54
3.5.	Eine konvexe Poincaré-Ungleichung	58
3.6.	Anwendung der Efron-Stein-Ungleichung auf Tail-Events	59
3.7.	Gaußsche Poincaré-Ungleichung	65
3.8.	Alternativer Beweis für die ES-Ungleichung	66
4.	Entropie	67
4.1.	Shannon-Entropie und relative Entropie	67
4.2.	Entropie von Produkten und Kettenregel	69
4.3.	Han-Ungleichung	70
4.4.	Isoperimetrische Ungleichung auf BHC	71
4.5.	Kombinatorische Entropien	75
4.6.	Han-Ungleichung für relative Entropien	77
4.7.	Sub-Additivität der Entropie	79
4.8.	Entropie für allgemeine Zufallsvariablen	82
4.9.	Dualität und Variationsformel	84
4.10.	Transportkosten-Abschätzung	87
4.11.	Pinsker-Ungleichung	87
4.12.	Birgé-Ungleichung	87
4.13.	Subadditivität der Entropie für allgemeine Zufallsvariablen	87
4.14.	Brunn-Minkowski-Ungleichung	87
5.	Logarithmische Sobolev-Ungleichung (LSU)	87
5.1.	LSU für symmetrische Bernoulli-Verteilungen	88
5.2.	Herbst-Argument	90
5.3.	Gaußsche LSU	92
5.4.	TSI-Konzentrationsungleichungen für Gauß-Zufallsvariablen	92
5.5.	Konzentrationsungleichung für Suprema von Gauß-Prozessen	92
5.6.	Zufällige Gaußsche Projektionen	92
5.7.	Hyperkontraktivität	92
6.	(evtl.) Ausblick	93
	Literatur	94

1. MOTIVATION

Zunächst sollen klassische Situationen als motivierendes Fundament vorgestellt werden, in denen man Konzentrationsungleichungen begegnen kann.

(A) Gesetz der großen Zahlen:

Für unabhängig, identisch verteilte Zufallsvariablen X_1, \dots, X_n in $\mathcal{L}^1(\Omega, P)$, äquivalent durch $E|X_1| < \infty$ ausgedrückt, gilt das *Gesetz der großen Zahlen*:

$$(1.1) \quad \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} E(X_1).$$

Eine andere nützliche Formulierung ist folgende

$$\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \xrightarrow[n \rightarrow \infty]{} 0.$$

Letztere Schreibweise kann vor allem in Fällen nützlich sein, in denen wir keine identisch verteilten Zufallsvariablen vorliegen haben. Man beachte aber, dass in solchen Fällen nicht allgemein die Konvergenz (1.1) gelten muss.

Bisher haben wir offen gelassen, welche Art von Konvergenz hier vorliegt.

Typischerweise formuliert man das Gesetz der großen Zahlen z.B. in

- fast sicherer Konvergenz (starkes Gesetz der großen Zahlen),
- stochastischer Konvergenz (schwaches Gesetz der großen Zahlen),
- der \mathcal{L}^2 -Norm.

Ein wichtiger Aspekt bei der Anwendung von Grenzwertsätzen in der Stochastik ist die Konvergenzgeschwindigkeit. Typischerweise fragt man sich, wie groß der Approximationsfehler für endliche n ist. Solche *nicht asymptotische* Fragestellungen sind in der Anwendung wichtig, um die Güte einer Approximation für den Erwartungswert von echten gegebenen Daten x_1, \dots, x_n abzuschätzen.

Dabei werden Abschätzungen der folgenden Bauart angestrebt:

$$\left\| \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \right\| \leq f(n, P_{X_1}),$$

wobei $f(n, P_{X_1})$ eine obere Schranke ist, die von dem Stichprobenumfang n und von der Verteilung P_{X_1} selbst abhängt. Wünschenswert wäre es, wenn $f(n, P_{X_1})$ wenige Informationen über die Verteilung benötigte, um möglichst allgemeine Aussagen zu erhalten. In der \mathcal{L}^2 -Norm für unabhängige, aber nicht notwendigerweise identisch verteilte, quadrat-integrierbare Zufallsvariablen

können wir folgende Abschätzung angeben

$$\left\| \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \right\|_{\mathcal{L}^2(P)} \leq \frac{\max_{i=1, \dots, n} \{\sigma_i\}}{\sqrt{n}}.$$

Dabei sind σ_i die Standardabweichungen der ZVen X_i für $i = 1, \dots, n$.

Die hier vorliegende Abschätzung bildet im Fall von identisch verteilten Zufallsvariablen sogar Gleichheit.

In Abschnitt 2 wird u.a. die Frage untersucht, welche Schranken sich durch Verwendung von höheren Momenten finden lassen.

(B) Rekonstruktion von Verteilungen in statistischer Lerntheorie

Wir betrachten den Datensatz $x_1, \dots, x_n \in \mathbb{R}$ als Realisationen von unabhängigen, identisch verteilten ZVen X_1, \dots, X_n mit unbekannter Verteilung $P_X = \mu$.

Frage: Wie lässt sich aus geg. Daten die wahre Verteilung rekonstruieren?

Wir verwenden das sogenannte *empirische Maß*

$$\mu_n = \mu_n^{x_1, \dots, x_n} = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

mit δ_{x_i} als das Punktmaß in x_i . Das liefert eine intuitive Möglichkeit für eine Schätzung von μ . Die empirische Verteilungsfunktion aus den Datensatz x_1, \dots, x_n ist dabei zum empirischen Maß assoziiert.

Auch hier stellt man sich die Frage nach der Güte der Approximation von μ_n zu μ . In welchem Sinne können wir hier überhaupt eine Konvergenz formulieren? Ein elementarer Ansatz ist durch den *Fundamentalsatz der Statistik von Glivenko-Cantelli* gegeben, durch den die Konvergenz der empirischen Verteilungsfunktion gegen die wahren Verteilungsfunktion in der Supremumsnorm geliefert wird.

Auf Ebene der Maßtheorie liegt eine schwache Konvergenz des empirischen Maßes gegen das wahre Wahrscheinlichkeitsmaß für Mengen der Bauart

$$A = (-\infty, x] \text{ für } x \in \mathbb{R}$$

vor. (Hier evtl. noch Arten der schwachen Konvergenz in der Sprache Funktionalanalysis diskutieren.)

Lässt sich die schwache Konvergenz für andere Klassen von Mengen formulieren? Diese Frage wird in einem Exkurs zur statistischen Lerntheorie in Abschnitt 3 untersucht.

(C) Irrfahrten auf \mathbb{Z}^2

Wir betrachten unabhängig, identisch verteilte ZVen der Form

$$X_1, \dots, X_n : \Omega \rightarrow \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}.$$

Die Werte der ZVen beschreiben dabei Bewegungsrichtungen im \mathbb{Z}^2 -Gitter. Die Wahrscheinlichkeit für alle Richtungen soll $\frac{1}{4}$ betragen. Nun summieren wir die ersten n Bewegungsschritte auf und erhalten eine neue ZVe

$$Z_n := \sum_{i=1}^n X_i.$$

Sie beschreibt für verschiedene Zeiten $i = 1, \dots, n$ einen Pfad auf \mathbb{Z}^2 . (Hier evtl. noch eine Abbildung einfügen.) Man nennt solche Prozesse *Irrfahrten*.

Wir fragen uns in diesem Kontext beispielsweise, wie sich der euklidische Abstand $\|Z_n\|$ im Verlauf eines Pfades typischerweise verhält. Es kann z.B. nach einer oberen Schranke für $\max_{i=1, \dots, n} \|Z_i\|$ gefragt werden. Eine sehr einfache Antwort wäre

$$\max_{i=1, \dots, n} \|Z_i\| \leq n,$$

da jeder Schritt maximal Länge 1 hat. Das ist allerdings eine grobe Abschätzung.

Mit dem *Zentralen Grenzwertsatz* können wir oft Aussagen der folgenden Bauart

$$\max_{i=1, \dots, n} \|Z_i\| \leq C \cdot \sqrt{n}$$

gewinnen. Die Frage, wann solche Abschätzungen gelten und wovon die Konstante C abhängt, bleibt noch offen. Unklar bleibt auch, wie die Verteilungen P_{X_i} in die Abschätzung eingehen.

Optional (D) Brownsche Bewegung

Hier interessieren wir uns für Suprema von stochastischen Prozessen. Ein prominentes Beispiel dafür ist die Brownsche Bewegung mit Zeithorizont T

$$B : \Omega \times [0, T] \rightarrow \mathbb{R}.$$

Mit einer funktionalen Version des zentralen Grenzwertsatzes gilt:

$$\text{Irrfahrt} \xrightarrow{\text{Donsker}} \text{Brownsche Bewegung}$$

$t \mapsto B_\omega(t)$ ist f.s. nicht differenzierbare Trajektorie.

Ferner: Für jede Zeit $t > 0$ gilt $\sup_{\omega \in \Omega} \|B_t(\omega)\| = \infty$. Aber es gilt wiederum:

$$\sup_{t \in [0, T]} \|B_t\| \leq C\sqrt{T}, \quad C > 0.$$

Ziel: Präzisiere die Konstante C und die Wahrscheinlichkeit!

2. GRUNDLEGENDE UNGLEICHUNGEN

In diesem Kapitel wollen wir erste Beispiele von Konzentrationsungleichungen kennenlernen.

2.1. Markov-Ungleichung und Co.

Sei X \mathbb{R} -wertige ZVe auf Wahrscheinlichkeitsraum (Ω, P) mit endlichem Erwartungswert $E(X)$.

Frage: Um wie viel weicht X von seinem Erwartungswert $E(X)$ ab?

Wir wollen obere Schranken für $t > 0$ der folgenden Form finden:

$$P(X - E(X) \geq t) \leq \dots$$

$$P(X - E(X) \leq -t) \leq \dots$$

einfachste Antwort: *Markov-Ungleichung*

Sei $Y \geq 0$ eine ZVe mit $E(Y) < \infty$. Dann gilt für alle $t \geq 0$:

$$Y \cdot \mathbf{1}_{\{Y \geq t\}} \geq t \cdot \mathbf{1}_{\{Y \geq t\}} \text{ auf } \Omega$$

Integration liefert:

$$E(Y \cdot \mathbf{1}_{\{Y \geq t\}}) \geq t \cdot E(\mathbf{1}_{\{Y \geq t\}}) = t \cdot P(Y \geq t).$$

Weitere Abschätzung der linken Seite:

$$E(Y \cdot \mathbf{1}_{\{Y \geq t\}}) \leq E(Y).$$

Falls die Dichtefunktion f von Y gegeben ist, ist es natürlich, den Wertebereich von Y auf die horizontale Achse aufzutragen.

Zusammenfassend: Für jede ZVe $Y : \Omega \rightarrow [0, \infty)$ in $\mathcal{L}^1(\Omega, P)$ gilt nach obigen Ausführungen für $t > 0$ die Markov-Ungleichung:

$$(2.1) \quad P(Y \geq t) \leq \frac{1}{t} E(Y \cdot \mathbf{1}_{\{Y \geq t\}}) \leq \frac{1}{t} E(Y).$$

Setze $Y = |E(X) - X|$. Das liefert eine erste Antwort:

$$P(X - E(X) \geq t) + P(X - E(X) \leq -t) = P(Y \geq t) \leq \frac{E(Y)}{t} = \frac{E(|X - E(X)|)}{t}.$$

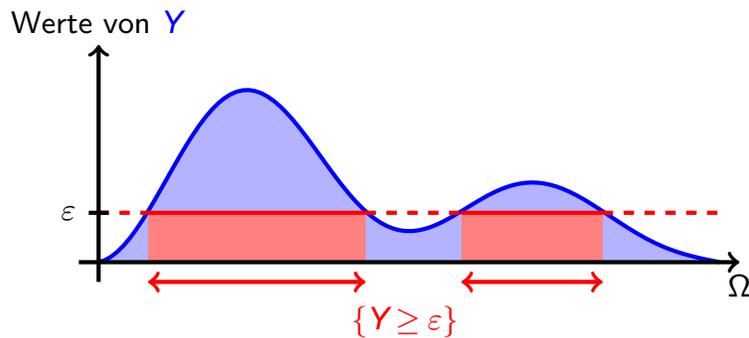


ABBILDUNG 1. Ω auf der horizontalen Achse.

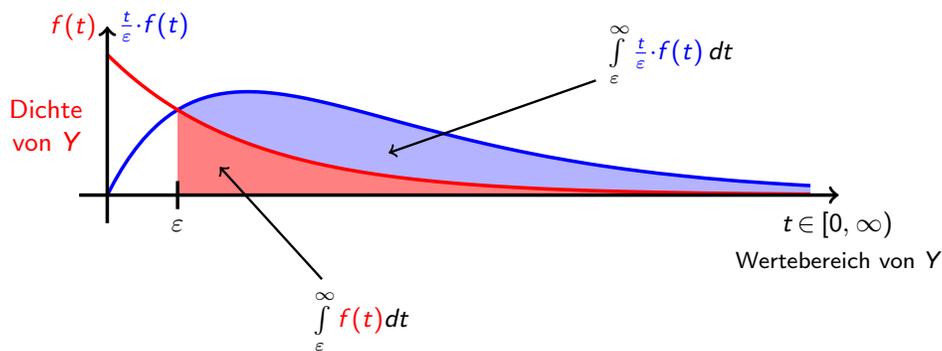


ABBILDUNG 2. Wertebereich auf der horizontalen Achse.

Frage: Gibt es eine bessere Wahl von Y in (2.1)?

Hat X z.B. eine endliche Varianz $\text{Var}(X)$, so gilt für $\Phi(Y)$ mit $\Phi(y) = y^2$

$$E(\Phi(Y)) = E(Y^2) = E(|X - E(X)|^2) = \text{Var}(X) < \infty \Rightarrow \Phi(Y) \in \mathcal{L}^1(\Omega, P).$$

Wende nun für $t \geq 0$ die Markov-Ungleichung an:

$$(2.2) \quad P(|X - E(X)| \geq t) = P(\Phi(Y) \geq \Phi(t)) \leq \frac{E(\Phi(Y))}{\Phi(t)} = \frac{\text{Var}(X)}{t^2}$$

und erhalte damit die *Markov-Čebyšev-Ungleichung*. Die Ungleichung (2.2) gilt für sämtliche monoton wachsende Funktionen $\Phi : I \rightarrow [0, \infty)$ auf einem Intervall $I \subset \mathbb{R}$, sodass $\Phi(t) > 0$ ist. Für die Anwendbarkeit der Methode braucht man

$$\Phi(Y) = \Phi(|X - E(X)|) \in \mathcal{L}^1(\Omega, P).$$

Gilt für eine ZVe X : $E|X^q| < \infty \forall q \in \mathbb{N}$, erhalten wir $\forall q, t \geq 0$:

$$P(|X - E(X)| \geq t) \leq \frac{E(|X - E(X)|^q)}{t^q}.$$

→ schöne Form, da linke Seite unabhängig von q ist (Optimierungsaspekt)

$$(2.3) \quad \Rightarrow \quad P(|X - E(X)| \geq t) \leq \inf_{q>0} \frac{E(|X - E(X)|^q)}{t^q}.$$

Erstes Fazit: Je mehr Informationen über eine ZVe vorliegen, desto breiter das Spektrum an potentiellen Abschätzungen und Methoden.

Warum spielt der Fall $q = 2$ eine zentrale Rolle?

Seien X_1, \dots, X_n unabhängige ZVen mit endlichen Varianzen. Für $Z = \sum_{i=1}^n X_i$ gilt nach dem Additionssatz (Bienaymé):

$$\text{Var}(Z) = \sum_{i=1}^n \text{Var}(X_i).$$

→ Formulierung einer Konzentrationsungleichung für gemittelte ZVen mit Hilfe von Ungleichung (2.2):

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i))\right| > t\right) &= P\left(\left|\sum_{i=1}^n (X_i - E(X_i))\right| > t \cdot n\right) \\ &\leq \frac{\text{Var}(Z)}{t^2 \cdot n^2} = \frac{\sigma^2}{t^2 \cdot n}, \end{aligned}$$

wobei $\sigma^2 := \frac{1}{n} \sum_{i=1}^n \text{Var}(X_i)$ die gemittelte Varianz ist.

Bemerkung: In dieser Vorlesung spielt die Eigenschaft der *Identischen Verteilung* eine weniger zentrale Rolle als die Unabhängigkeit, da uns nicht der explizite Wert des Limes interessiert, sondern gute Abschätzungen für *endliches* n (bzw. die Konvergenzordnung). Natürlich vereinfachen sich viele Aussagen, falls die ZVen identisch verteilt sind.

2.2. Cramér-Chernoff-Methode.

Die Methoden dieses Kapitels werden in den nachfolgenden Kapitel 2.3 - 2.9 benötigt. Zusammenhänge zur Entropie werden in Kapitel 4.9 dargestellt. In Kapitel 5.2, 5.4, 5.5 taucht sie ebenso auf.

Idee: Wähle statt $\Phi(t) = t^2$ nun $\Phi(t) = e^{\lambda t}$ für $\lambda > 0$.

Analog zu (2.2) mit $\Phi(y) = e^{\lambda y}$ nach Anwendung der Markov-Ungleichung:

$$(2.4) \quad P(X \geq t) = P\left(e^{\lambda X} \geq e^{\lambda t}\right) \leq \frac{E\left(e^{\lambda X}\right)}{e^{\lambda t}}.$$

→ Schranke mit exponentiellem Abfall gewonnen.

Studiere nun die Schranken genauer. Die folgende Abbildung

$$M: \mathbb{R} \rightarrow [0, \infty], M(\lambda) := E\left(e^{\lambda X}\right)$$

nennen wir die *momentenerzeugenden Funktion* von X (kurz: MEF von X).

Sie ist gegebenenfalls unendlich. Betrachte $Z = \sum_{i=1}^n X_i$ mit unabhängigen ZVen X_1, \dots, X_n . Für die MEF von $Z - E(Z)$ gilt dann

$$E\left(e^{\lambda \sum_{i=1}^n (X_i - E(X_i))}\right) = \prod_{i=1}^n E\left(e^{\lambda (X_i - E(X_i))}\right),$$

wegen der Unabhängigkeit der X_1, \dots, X_n . Sind $(X_i - E(X_i))$ zusätzlich identisch verteilt mit MEF $M(\lambda) := M_{X_1 - E(X_1)}(\lambda)$, so gilt mit (2.4):

$$P\left(\frac{1}{n}(Z - E(Z)) \geq t\right) \leq \frac{\prod_{i=1}^n M(\lambda)}{e^{\lambda t n}} = \frac{(M(\lambda))^n}{e^{\lambda t n}}.$$

Vorgehensweise bisher:

- Gewinne Klassen von Ungleichungen für verschiedene λ
- Schranke über den Parameter λ optimieren (Minimierungsaufgabe)
- Lösen der Optimierungsaufgabe liefert eine (hoffentlich) gute Abschätzung

Bemerkungen:

- (a) Im Allg. liefern polynomielle Transformationen $\Phi(t) = t^q$ aus (2.3) bessere Schranken als exponentielle Transformationen $\Phi(t) = e^{\lambda t}$ aus (2.4). D.h.: für beliebige $t > 0$ und ZVe $X \geq 0$:

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

Beweisidee: Taylorentwicklung der Exponentialfunktion (Übung).

- (b) Wir interessieren uns für die Wahrscheinlichkeit der Abweichungen vom Mittelwert mit folgender Bauart:

$$P(|Z - E(Z)| \geq t) = P(Z - E(Z) \geq t) + P(Z - E(Z) \leq -t) \quad \text{für } t > 0.$$

Wegen (b) genügt es o.B.d.A. die zentrierte Version der ZVe $\tilde{Z} := (Z - E(Z))$ zu betrachten und Abschätzungen für $P(|\tilde{Z}| \geq t)$ herzuleiten.

Die kumulantenerzeugende Funktion und die Cramér-Transformierte.

Sei Z ZVe mit MEF $M_Z(\lambda)$, d.h. (äquivalent zu (2.4)):

$$(2.4^*) \quad P(Z \geq t) \leq e^{-\lambda t} M(\lambda) = e^{-\lambda t + \ln(M(\lambda))}.$$

Fasse die obere Schranke als Klasse von Funktionen auf, um sie dann zu minimieren. Dazu definieren wir die *kumulantenerzeugende Funktion* $\psi_Z(\lambda)$ (kurz: KEF):

$$\psi_Z(\lambda) := \ln(M(\lambda)) = \ln(E(e^{\lambda Z})).$$

Betrachte dazu die sogenannte *Cramér-Transformierte*:

$$\psi_Z^*(t) := \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda)) \quad \text{für } \lambda \geq 0.$$

Bedeutung der Cramér-Transformierte:

Minimiere die Schranke von (2.4*) über λ und betrachte nur noch den Exponenten. Beachte: Vorzeichenwechsel liefert Maximierungsproblem:

$$P(Z \geq t) \leq \inf_{\lambda \geq 0} e^{-(\lambda t - \ln(M(\lambda)))} = e^{-\psi_Z^*(t)}.$$

Untersuche nun den Definitions- und Wertebereich der Cramér-Trafo ψ_Z^* . Zunächst bemerkt man eine Eigenschaft der KEF für beliebige ZVen Z :

$$\psi_Z(0) = \ln(1) = 0.$$

Dieser Zusammenhang ist nützlich für Randwertuntersuchungen. Auf die Cramér-Trafo überträgt sich dies, wenn man für $\lambda = 0$ setzt, wie folgt:

$$\psi_Z^*(t) = \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda)) \geq 0 - 0 = 0.$$

Insbesondere wissen wir nun, dass der Wertebereich von ψ_Z^* nichtnegativ ist.

Warum betrachten wir nicht $\lambda \in \mathbb{R}$, sondern nur $\lambda \geq 0$ im Supremum?

Falls für die ZV $Z \in \mathcal{L}^1$ gilt, so gilt nach der Jensen-Ungleichung

$$e^{\lambda E(Z)} \leq E(e^{\lambda Z}) = M(\lambda),$$

wobei der letzte Ausdruck auch unendlich sein könnte. Logarithmieren dieser Ungleichung ergibt

$$\lambda \cdot E(Z) \leq \ln(M(\lambda)) = \psi_Z(\lambda).$$

Betrachte $\lambda < 0$ und $t \geq E(Z)$ und schätze die linke Seite ab

$$\lambda \cdot E(Z) \geq \lambda \cdot t$$

Insgesamt folgt aus beiden Ungleichungen für $\lambda < 0$ und $t \geq E(Z)$:

$$\lambda \cdot t - \psi_Z(\lambda) \leq 0$$

Die Annahme $t \geq E(Z)$ lässt sich allgemein dadurch motivieren, dass wir zentrierte ZVe Z mit $E(Z) = 0$ betrachten wollen.

Fazit: Obige Randbetrachtung für $\lambda = 0$ das liefert, dass das Supremum über λ nicht im negativen Bereich angenommen wird, d.h.

$$(2.5) \quad \tilde{\psi}_Z(t) := \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) = \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda)) = \psi_Z^*(t) \text{ für } t \geq E(Z).$$

Wir nennen die Funktion $\tilde{\psi}_Z$ in (2.5) die *Fenchel-Legendre-Transformierte* oder auch *Fenchel-Legendre-Duale* von ψ_Z .

Nicht jedes $t \geq 0$ liefert brauchbare Chernoff-Schranken. Falls $\psi_Z^*(t) = 0$ ist, ergibt sich eine triviale Schranke $e^{-\psi_Z^*(t)} = 1$. In welchen Fällen tritt dies noch ein?

Einerseits ist der Fall $\psi_Z(\lambda) \equiv \infty$ für $\lambda > 0$ problematisch. Dann folgt

$$\psi_Z^*(t) = \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda)) = 0.$$

Andererseits ist der Fall $t \leq E(Z)$ problematisch wegen

$$\begin{aligned} \lambda t &\leq \lambda E(Z) \leq \psi_Z(\lambda) \quad \text{für } \lambda \geq 0 \\ \Rightarrow \lambda t - \psi_Z(\lambda) &\leq 0 \text{ und } = 0 \text{ für } \lambda = 0. \end{aligned}$$

Um diese Situation zu vermeiden, nehmen wir in diesem Kapitel an, dass ein $\lambda_0 > 0$ existiert, sodass $E(e^{\lambda_0 Z}) < \infty$ gilt. Mit der Hölder-Ungleichung lässt sich zeigen, dass dann auch das exponentielle Moment für $\lambda \leq \lambda_0$ existiert (Übung). Es gilt dann also:

$$\text{für alle } \lambda \in [0, \lambda_0] : E(e^{\lambda Z}) < \infty.$$

Setze dazu die Zahl $b := \sup\{\lambda \geq 0 \mid E(e^{\lambda Z}) < \infty\} \in [0, \infty]$. Da die Voraussetzung $E(e^{\lambda Z}) < \infty$ für $\lambda = 0$ immer erfüllt ist, genügt es auch hier statt $\lambda \in \mathbb{R}$ nur den Bereich $\lambda \geq 0$ zu untersuchen. In den meisten Fällen ist $b \in \{0, \infty\}$. Dagegen ist für exponentialverteilte ZVen der Wert b gerade der Parameter der Exponentialverteilung.

Bemerkung 2.1 (Eigenschaften der kumulantenerzeugenden Funktion). Folgende Eigenschaften werden sich für Optimierungsaufgaben als nützlich erweisen:

- (a) $\psi = \psi_Z$ ist konvex auf $I := (0, b)$ (auch gültig, falls $b = \infty$)
- (b) ψ ist strikt konvex auf I , falls Z nicht fast sicher konstant
- (c) $\psi : I \rightarrow \mathbb{R}$ ist \mathcal{C}^∞

Für zentrierte Zufallsvariablen Z gilt darüber hinaus:

- (d) $\psi_Z : [0, b) \rightarrow \mathbb{R}$ ist \mathcal{C}^1 . Beachte: 0 ist *zusätzlich* im Definitionsbereich.
- (e) $\psi'_Z(0) = 0$ (zusätzlich zu $\psi_Z(0) = 0$)
- (f) Es genügt die Cramér-Trafo auf dem Intervall I zu bestimmen:

$$\psi_Z^*(t) = \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda)) = \sup_{\lambda \in I} (\lambda t - \psi_Z(\lambda))$$

Die Beweise von (a) - (f) werden als Übungsaufgaben gestellt.

Bestimmung des Supremums mittels der Ableitung. Ansatz:

- ψ_Z ist $\mathcal{C}^1 \rightarrow$ mittels Ableitung stationäre Punkte berechnen
- strikte Konvexität liefert Eindeutigkeit des Optimums auf I

$$0 = \frac{d}{d\lambda} (\lambda t - \psi(\lambda)) = t - \psi'(\lambda)$$

$$\Leftrightarrow t = \psi'(\lambda)$$

Sei λ_t eine Lösung dieser Gleichung. Falls man den trivialen Fall einer fast sicher konstanten Zufallsvariable ausschließt, ist ψ strikt konvex.

$\Rightarrow \lambda_t$ ist eindeutig.

Definition: Sei $B := \psi'_Z(b)$. Dann ist $\psi'_Z : I \rightarrow (0, B)$ wegen strikter Monotonie bijektiv mit strikt monotoner Inversen $(\psi'_Z)^{-1}$.

$$\text{Daher gilt für alle } t \in (0, B) : \lambda_t = (\psi'_Z)^{-1}(t).$$

Diese Formel können wir nun nutzen, um für konkrete Verteilungen die Cramér-Trafo auszurechnen.

Cramér-Transformierte verschiedener Verteilungsklassen.

(a) Cramér-Transformierte für zentrierte Normalverteilungen:

Die Kenntnis der Cramér-Transformierten der zentrierten Normalverteilung ist beim Verständnis von sub-gaußschen Zufallsvariablen in Kapitel 2.2 relevant. Sei $Z \sim \mathcal{N}(0, \sigma^2)$ mit Varianz σ^2 . Beachte, dass wir Zentriertheit brauchen.

- zunächst berechne MEF: $M(\lambda) = e^{\frac{\lambda^2 \sigma^2}{2}}$ (Übung)
- MEF ergibt sofort die KEF $\psi_Z(\lambda) = \frac{\lambda^2 \sigma^2}{2}$.
- erste Ableitung ist $\psi'_Z(\lambda) = \lambda \sigma^2$
- obige Formel ergibt $\lambda_t = (\psi')^{-1}(t) = \frac{t}{\sigma^2}$ als Lösung der Optimierungsaufgabe

$\forall t > 0$ besitzt die Cramér-Trafo also die folgende Gestalt

$$\begin{aligned}\psi_Z^*(t) &= \lambda_t t - \psi_Z(\lambda_t) \\ &= \frac{t^2}{\sigma^2} - \frac{\sigma^2 t^2}{2\sigma^4} \\ &= \frac{t^2}{2\sigma^2}.\end{aligned}$$

Das liefert die Chernoff-Schranke für $\forall t > 0 : P(Z \geq t) \leq e^{-\frac{t^2}{2\sigma^2}}$.

Wie gut ist diese Schranke? Kann man sie noch verbessern? Zur Beantwortung dieser Fragen formuliere die Chernoff-Abschätzung um

$$P(Z \geq t) \cdot e^{\frac{t^2}{2\sigma^2}} \leq 1.$$

Diese Abschätzung lässt sich aber verbessern.

Für alle $t > 0$ gilt $P(Z \geq t) \cdot e^{\frac{t^2}{2\sigma^2}} \leq \frac{1}{2}$ (Übung).

→ globale Vorfaktor $\frac{1}{2}$ wird verschenkt, Größenordnung passt zumindest.

Zudem kann man zeigen, dass letztere Abschätzung scharf ist:

$$\sup_{t>0} \left(P(Z \geq t) \cdot e^{\frac{t^2}{2\sigma^2}} \right) = \frac{1}{2} \text{ (ebenfalls Übung).}$$

Unter Normalität sind Gewinnungen von Abschätzungen mittels anderer Techniken noch einfach, wodurch wir einen Vergleich zwischen ihnen und der Chernoff-Methode erhalten. In anderen Fällen ist dies nicht mehr so einfach möglich.

(b) Cramér-Transformierte für zentrierte Poisson-Verteilungen:

Die Kenntnis der KEF der zentrierten Poisson-Verteilung wird im Beweis der Bennet-Ungleichung 2.9 in Kapitel 2.7 benutzt. Sei $Y \sim Poi(\nu)$ für ein $\nu > 0$. Nach Definition der Poisson-Verteilung gilt

$$\forall k \in \mathbb{N}_0 : P(Y = k) = \frac{\nu^k}{k!} \cdot e^{-\nu}.$$

$\Rightarrow E(Y) = \nu$. Wir arbeiten mit zentrierten ZVen und definieren daher $Z := Y - E(Y)$ mit $E(Z) = 0$. Berechne im ersten Schritt die MEF:

$$\begin{aligned} M_Z(\lambda) &= E(e^{\lambda Z}) = e^{-\lambda\nu} \sum_{k \in \mathbb{N}_0} \left(e^{\lambda k} \cdot \frac{\nu^k}{k!} \right) \cdot e^{-\nu} \\ &= e^{-\lambda\nu - \nu} \sum_{k \in \mathbb{N}_0} \frac{(\nu e^\lambda)^k}{k!} = e^{-(\lambda+1)\nu} \cdot e^{\nu e^\lambda}. \end{aligned}$$

Logarithmieren liefert für $\lambda > 0$ die KEF ψ_Z . Berechne außerdem ψ'_Z

$$\psi_Z(\lambda) = \nu \left(e^\lambda - \lambda - 1 \right), \quad \psi'_Z(\lambda) = \nu \left(e^\lambda - 1 \right).$$

Ansatz: $t = \psi'_Z(\lambda)$, um λ_t als Lösung des Optimierungsproblems herzuleiten:

$$\begin{aligned} t &= \psi'_Z(\lambda) = \nu \left(e^{\lambda_t} - 1 \right) \\ \Leftrightarrow e^{\lambda_t} &= \frac{t}{\nu} + 1 \\ \Leftrightarrow \lambda_t &= \ln \left(\frac{t}{\nu} + 1 \right). \end{aligned}$$

Der optimierende Parameter λ_t liefert nun die Gestalt der Cramér-Trafo:

$$\begin{aligned} \psi_Z^*(t) &= t\lambda_t - \psi_Z(\lambda_t) \\ &= t \ln \left(\frac{t}{\nu} + 1 \right) - \nu \left(\frac{t}{\nu} + 1 - \ln \left(\frac{t}{\nu} + 1 \right) - 1 \right) \\ &= (t + \nu) \cdot \ln \left(\frac{t}{\nu} + 1 \right) - t \\ &= \nu \cdot h \left(\frac{t}{\nu} \right), \end{aligned}$$

wobei wir für $x \geq -1$ die Funktion h einführen:

$$h(x) := (1+x) \cdot \ln(1+x) - x. \quad x = \frac{t}{\nu}$$

Die ZVe $-Z$ ist ebenfalls zentriert und analog folgt:

$$\psi_{-Z}^*(t) = \nu \cdot h\left(-\frac{t}{\nu}\right), \text{ sofern } t \leq \nu \text{ gilt.}$$

(c) Cramér-Transformierte für zentrierte Bernoulli-Verteilungen:

Eigentliches Ziel: Cramér-Trafo von binomialverteilten ZVen.

Betrachte zunächst in (c) Bernoulli-ZVen und in (d) Summen von unabhängigen ZVen, um in Teil (e) die Binomialverteilung zu untersuchen.

Sei $Y : \Omega \rightarrow \{0, 1\}$ eine ZVe mit $P(Y = 1) = p = 1 - P(Y = 0)$ für ein $p \in [0, 1]$. Der Erwartungswert ist dann $E(Y) = p$. Die ZVe $Z := Y - p$ ist dann die zentrierte Version. Die MEF von Z ergibt sich aus der Definition

$$M_Z(\lambda) = \left(p \cdot e^\lambda + (1-p)\right) e^{-\lambda p},$$

mit KEF

$$\psi_Z(\lambda) = -\lambda p + \ln\left(pe^\lambda + 1 - p\right).$$

Schließlich folgt mit gleicher Strategie (Übung) für $t \in (0, 1-p)$:

$$\begin{aligned} \psi_Z^*(t) &= (1-p-t) \cdot \ln\left(\frac{1-p-t}{1-p}\right) + (p+t) \cdot \ln\left(\frac{p+t}{p}\right) \\ &= (1-a) \cdot \ln\left(\frac{1-a}{1-p}\right) + a \cdot \ln\left(\frac{a}{p}\right) \\ &=: h_p(a). \end{aligned}$$

Hierbei haben wir $a := t + p$ gesetzt, so dass $a \in (p, 1)$.

Definition 2.2. Die Funktion $h_p : (0, 1) \rightarrow \mathbb{R}$, $h_p(a) = (1-a) \cdot \ln\left(\frac{1-a}{1-p}\right) + a \cdot \ln\left(\frac{a}{p}\right)$ nennt man die *Kulback-Leibler-Divergenz* $D(P_a \| P_p)$ zwischen zwei Bernoulli-Verteilungen mit Parameter a bzw. p .

Wir werden im Kapitel 4 den Begriff der *Entropie* einführen und auch die Funktion h_p diesem Begriff zuordnen können (vgl. insbesondere die *Dualitätsformel 4.14 aus Kapitel 4.9*). Die Schranken einiger Konzentrationsungleichungen entsprechen solchen Entropien.

(d) Cramér-Transformierte für Summen unabhängiger Zufallsvariablen:

Die Cramér-Chernoff-Methode erlaubt einfachen Umgang mit Summen von

unabhängig, identisch verteilten ZVen X_1, \dots, X_n . Dazu betrachten wir die ZVe $Z := \sum_{i=1}^n X_i$ und die KEF ψ_{X_1} , sowie die Cramér-Trafo $\psi_{X_1}^*$ für X_1 .

Ziel: Darstellung der KEF von Z in Abhängigkeit der KEF der X_1, \dots, X_n .
Für λ mit $\psi_{X_1}(\lambda) < \infty$ bestimmen wir die KEF von Z :

$$\begin{aligned}
 \psi_Z(\lambda) &= \ln \left(E \left(e^{\lambda Z} \right) \right) = \ln \left(E \left(e^{\lambda \sum_{i=1}^n X_i} \right) \right) \\
 &= \ln \left(\prod_{i=1}^n \left(E \left(e^{\lambda X_i} \right) \right) \right) \quad (\text{wegen Unabhängigkeit}) \\
 (2.1) \quad &= \ln \left(E \left(e^{\lambda X_1} \right)^n \right) \quad (\text{wegen identischer Verteilung}) \\
 &= n \cdot \ln \left(E \left(e^{\lambda X_1} \right) \right) \\
 &= n \cdot \psi_{X_1}(\lambda).
 \end{aligned}$$

Falls nur unabh. vorliegt

$$\psi_Z(\lambda) = \sum_{i=1}^n \ln E(e^{\lambda X_i}) = \sum_{i=1}^n \psi_{X_i}(\lambda)$$

Ein ähnlicher Zusammenhang gilt auch für die Cramér-Trafo:

Lemma 2.3. Seien X, X_1, \dots, X_n unabhängige, identisch verteilten ZVen und $Z := \sum_{i=1}^n X_i$. Für $\lambda \in \mathbb{R}$ mit $\psi_{X_1}(\lambda) < \infty$ gilt:

$$\psi_Z^*(t) = n \cdot \psi_{X_1}^* \left(\frac{t}{n} \right).$$

Beweis. Es sei an die Gestalt des optimierenden Parameters $\lambda_t = (\psi')^{-1}(t)$ der Cramér-Trafo erinnert. Also untersuchen wir:

$$\begin{aligned}
 \psi'_Z(\lambda) &\stackrel{(2.1)}{=} n \cdot \psi'_X(\lambda) \\
 &= \mathcal{M}_n(\psi'_X(\lambda)) = (\mathcal{M}_n \circ \psi'_X)(\lambda)
 \end{aligned}$$

mit $\mathcal{M}_n(x) = n \cdot x$. Außerdem ist $(\mathcal{M}_n)^{-1}(x) = \frac{x}{n}$.

$$\Rightarrow \lambda_t = (\psi'_Z)^{-1}(t) = (\psi'_X)^{-1}((\mathcal{M}_n)^{-1}(t)) = (\psi'_X)^{-1} \left(\frac{t}{n} \right).$$

Für die Cramér-Trafo von Z folgt schließlich

$$\begin{aligned}
 \psi_Z^*(t) &= t\lambda_t - \psi_Z(\lambda_t) \\
 \lambda_t^X &= (\psi_X')^{-1}(t) = t(\psi_Z')^{-1}(t) - \psi_Z((\psi_Z')^{-1}(t)) \\
 &= t(\psi_X')^{-1}\left(\frac{t}{n}\right) - n\psi_X\left((\psi_X')^{-1}\left(\frac{t}{n}\right)\right) \\
 \lambda_{t/m}^X &= (\psi_X')^{-1}\left(\frac{t}{m}\right) = n\psi_X^*\left(\frac{t}{n}\right) \\
 &= \cancel{n} \lambda_{t/m} \cdot \cancel{\frac{t}{m}} - n \psi_X(\lambda_{t/m}) \\
 &= t(\psi_X')^{-1}\left(\frac{t}{m}\right) - n\psi_X\left((\psi_X')^{-1}\left(\frac{t}{m}\right)\right)
 \end{aligned}$$

□

(e) Cramér-Transformierte für zentrierte Binomialverteilungen:

Jetzt sind alle Vorbereitungen getroffen, die Cramér-Trafo der Binomialverteilung zu berechnen. Seien $X_1, \dots, X_n \sim Ber(p)$ unabhängig.

$\Rightarrow Y := \sum_{i=1}^n X_i \sim Bin(n, p)$. Betrachte zentrierte Version $Z := Y - np = \sum_{i=1}^n (X_i - p)$, da ja $E(X_i) = p$. Für $t \in (0, n(1-p))$ gilt nach (d) und (c) mit $a = p + \frac{t}{n}$:

$$\psi_Z^*(t) = n \cdot \psi_{X-p}^* \left(\frac{t}{n} \right) = n \cdot h_p \left(\frac{t}{n} + p \right),$$

wobei h_p die in (c) bereits eingeführte Kulback-Leibler-Divergenz zwischen zwei Bernoulli-Verteilungen ist.

Wir erhalten insbesondere die Chernoff-Schranke:

$$P(Z \geq t) \leq \exp \left(-n \cdot h_p \left(\frac{t}{n} + p \right) \right).$$

\rightarrow Guter Vorfaktor n für $n \rightarrow \infty$, aber wie schnell und wohin konvergiert h_p für $n \rightarrow \infty$? \rightarrow Noch zu untersuchen!

2.3. Sub-Gaußsche Zufallsvariablen.

Gauß-ZVen $\hat{=}$ gut umgängliche Klasse von ZVen

Klasse von Verteilungen, die von Gauß-ZVen dominiert werden, sind ähnlich gut umgänglich.

Definition 2.4. Eine reellwertige, zentrierte ZVe X heißt *sub-gaußsch* mit Varianzfaktor ν , in Zeichen: $P_X \in \mathcal{G}(\nu)$, falls

$$(2.2) \quad \forall \lambda \in \mathbb{R} : \psi_X(\lambda) \leq \frac{\lambda^2 \nu}{2}.$$

Die Klasse aller solchen ZVen, bezeichnen wir (auch) mit $\mathcal{G}(\nu)$. Die Schranke entspricht der KEF einer zentrierten Normalverteilung mit Varianz ν . Wir betrachten also die Klasse von Verteilungen, die durch gaußsche ZVen im Sinne der KEF dominiert werden.

Die Konzept von sub-gaußsch findet in den nachfolgenden Kapiteln 2.5 und 2.9 Anwendung.

Bemerkung 2.5 (Eigenschaften sub-gaußscher Zufallsvariablen). (a) Varianz:

Es folgt $\text{Var}(X) \leq \nu$, im Allg. gilt aber $\text{Var}(X) \neq \nu$. Die Ungleichung kann man durch eine Taylorentwicklung zeigen (Übung). Sie ist insbesondere scharf. Betrachte dazu Rademacher-verteilte ZVen.

- (b) Für normalverteilte ZVen $X \sim \mathcal{N}(m, \nu)$ ist die momentenerzeugende Funktion $M_X(\lambda) = \exp\left(m\lambda + \frac{\lambda^2}{2}\nu\right)$, wie man durch direkte Rechnung sieht. Also gilt

$$X \sim \mathcal{N}(0, \nu) \implies X \in \mathcal{G}(\nu)$$

- (c) In der Tat, Bedingung (2.2) kann sogar als Vergleich mit Gauß-ZV verstanden werden:

$$\cancel{X} = Y - E(Y) \in \mathcal{G}(\nu) \iff \psi_{\cancel{X}} \leq \psi_{\mathcal{N}(0, \nu)} \text{ auf } \mathbb{R}$$

- (d) Verträglichkeit mit der Faltung: / $\int \cup m me$

Sind X_1, \dots, X_n unabhängige ZVen mit $X_i \in \mathcal{G}(\nu_i)$ für $i = 1, \dots, n$, so gilt folgende Stabilitätseigenschaft für $Z = \sum_{i=1}^n X_i \in \mathcal{G}(\sum_{i=1}^n \nu_i)$. Nachweis erfolgt z.B. mit dem Additionssatz der Varianz und der Stabilität der KEF:

siehe oben
$$\psi_Z \ominus \sum_{i=1}^n \psi_{X_i} \leq \sum_{i=1}^n \psi_{\mathcal{N}(0, \nu_i)} \ominus \psi_{\mathcal{N}(0, \sum_{i=1}^n \nu_i)}.$$

- (e) Inklusionen

$$\lambda \leq \tilde{\lambda} \implies \mathcal{G}(\lambda) \subset \mathcal{G}(\tilde{\lambda})$$

Definition 2.6. Ein messbares $X: \Omega \rightarrow \{-1, 1\}$ mit $P(X = 1) = P(X = -1) = \frac{1}{2}$ heißt *Rademacher-Zufallsvariable*.

Statt über Ungleichungen für KEF ψ_X kann man die Eigenschaft sub-gaußsch auch über Ungleichungen für Momente oder über Ungleichungen von Abfall bei ∞ (engl. *tail-probability*) charakterisiert werden.

Bemerkung 2.7 (Äquivalente *tail*-Charakterisierungen von sub-gaußsch). Sei $X \in \mathcal{G}(\nu)$. Chernoff liefert für $t > 0$

$$\max\{P(X > t), P(X < -t)\} \leq e^{-\frac{t^2}{2\nu}}.$$

Für $X \in \mathcal{G}(\nu)$ gilt nämlich $\psi_X^*(t) \geq \psi_{\mathcal{N}(0, \nu)}^*(t)$ (Übung) und damit

$$P(X > t) \leq e^{-\psi_X^*(t)} \leq e^{-\psi_{\mathcal{N}(0, \nu)}^*(t)} = e^{-t^2/2\nu}$$

Analoge Rechnung für $P(-X > t)$ wegen $-X \in \mathcal{G}(\nu)$ liefert Behauptung.

Satz 2.8.

Sei $X \in \mathcal{L}^1$ eine zentrierte ZVe.

- (a) Gibt es ein $\nu > 0$, sodass $\forall s > 0$

$$\max\{P(X > s), P(X < -s)\} \leq e^{-\frac{s^2}{2\nu}}$$

gilt, so folgt $\forall q \in \mathbb{N}$:

$$(*) \quad E(X^{2q}) \leq 2q!(2\nu)^q \leq q!(4\nu)^q.$$

(b) Existiert ein $C \in (0, \infty)$ mit

$$(2.3) \quad \forall q \in \mathbb{N} : E(X^{2q}) \leq q!C^q,$$

so gilt $X \in \mathcal{G}(4C)$. Insbesondere folgt daraus

$$(**) \quad \max \{P(X > s), P(X < -s)\} \leq e^{-\frac{s^2}{8C}}.$$

Theorem 2.3 kann als Hilfsmittel dienen, um die Voraussetzung der Bernstein-Ungleichung in 2.10 nachzurechnen, wie es im Johnson-Lindenstrauß-Lemma in 2.11 getan wird.

Beweis. zu (a): Gilt $X \in \mathcal{G}(\nu)$, so ist $Y = \frac{1}{\sqrt{\nu}}X \in \mathcal{G}(1)$, denn

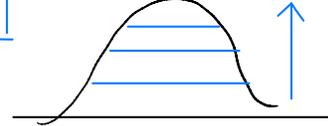
$$\psi_Y(\lambda) = \ln \left(E \left(e^{\lambda Y} \right) \right) = \ln \left(E \left(e^{\frac{\lambda X}{\sqrt{\nu}}} \right) \right) = \psi_X \left(\frac{\lambda}{\sqrt{\nu}} \right) \leq \frac{\frac{\lambda^2}{\nu} \cdot \nu}{2} = \frac{\lambda^2}{2}.$$

Betrachte zunächst den Fall $\nu = 1$ und beginne auf der linken Seite von (*):

$$E(Y^{2q}) = \int_0^\infty P(|Y|^{2q} > x) dx.$$

Erste Substitution mit $y = x^{\frac{1}{2q}}$ ergibt

$$= 2q \int_0^\infty P(|Y| > y) \cdot y^{2q-1} dy.$$



Anwendung der Voraussetzung ermöglicht folgende Abschätzung:

$$\leq 4q \int_0^\infty e^{-\frac{y^2}{2}} \cdot y^{2q-1} dy.$$

Mit zweiter Substitution $t = \frac{y^2}{2}$ erhalten wir

$$= 4q \int_0^\infty e^{-t} \cdot (2t)^{q-\frac{1}{2}} \cdot (2t)^{-\frac{1}{2}} dt.$$

Diese Substitution ist nützlich, da das Integral explizite Darstellung hat (vgl. Bronstein):

$$E(Y^{2q}) \leq 2^{q+1}q!.$$

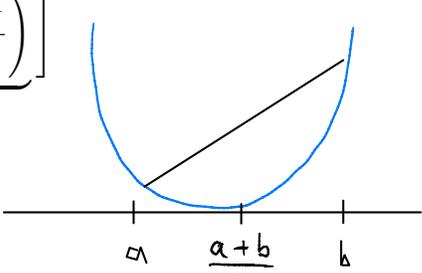
Für allgemeinere Varianzen ν folgt:

$$E(X^{2q}) = E\left((\sqrt{\nu}Y)^{2q}\right) = \nu^q E(Y^{2q}) \leq 2^{q+1}\nu^q q! \leq 2 \cdot 2^q \nu^q q! = (4\nu)^q q!.$$

zu (b): Es gelte $E(X^{2q}) \leq q!C^q$.

Sei \tilde{X} eine unabhängig, identisch verteilte Kopie von X .

$\Rightarrow X - \tilde{X}$ ist symmetrisch verteilt, d.h. $P(X - \tilde{X} > s) = P(\tilde{X} - X > s) \forall s \in \mathbb{R}$. Wir nutzen den Satz zur monotone Konvergenz und dass die ungeraden Momente verschwinden:

$$\begin{aligned}
 & E\left(e^{\lambda X}\right) \cdot E\left(e^{-\lambda X}\right) \stackrel{\text{id. vert.}}{=} E\left(e^{\lambda X}\right) \cdot E\left(e^{-\lambda \tilde{X}}\right) \stackrel{\text{unabh.}}{=} E\left(e^{-\lambda(X-\tilde{X})}\right) \\
 & = E\left(\sum_{q \in \mathbb{N}_0} \left[\frac{\lambda^{2q}(X-\tilde{X})^{2q}}{(2q)!} + \frac{\lambda^{2q+1}(X-\tilde{X})^{2q+1}}{(2q+1)!} \right]\right) \quad \left(\frac{a+b}{2}\right)^m \leq \frac{1}{2}(a^m + b^m) \\
 & = \sum_{q \in \mathbb{N}_0} \left[E\left(\frac{\lambda^{2q}(X-\tilde{X})^{2q}}{(2q)!}\right) + \underbrace{E\left(\frac{\lambda^{2q+1}(X-\tilde{X})^{2q+1}}{(2q+1)!}\right)}_{=0 \text{ wegen Symmetrie}} \right] \\
 & = \sum_{q \in \mathbb{N}_0} \frac{\lambda^{2q} E\left((X-\tilde{X})^{2q}\right)}{(2q)!} \quad \forall \lambda \in \mathbb{R}
 \end{aligned}$$


Frage/Übung: Existieren überhaupt die ungeraden Momente? Sind sie summierbar?

$k \leq m$ Setze $p = \frac{m}{k}$, Hölder $E(X^k \cdot 1) \leq \sqrt[p]{E(X^m)}$

Wegen Konvexität von $x \mapsto x^{2q} = x^m$ für $m \in 2\mathbb{N}$ folgt (vgl. auch Abb. ??):

$$\begin{aligned}
 & (a-b)^m \leq 2^{m-1}(a^m - b^m) \leq 2^{m-1}(a^m + b^m) \\
 \Rightarrow & E\left((X-\tilde{X})^{2q}\right) \leq 2^{2q-1} \left(E(X^{2q}) + E(\tilde{X}^{2q})\right) \stackrel{\text{id. vert.}}{=} 2^{2q} E(X^{2q})
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow & E\left(e^{\lambda X}\right) \cdot E\left(e^{-\lambda X}\right) \stackrel{\text{s.o.}}{=} \sum_{q \in \mathbb{N}_0} \frac{\lambda^{2q} E\left((X-\tilde{X})^{2q}\right)}{(2q)!} \\
 & \leq \sum_{q \in \mathbb{N}_0} \frac{\lambda^{2q}}{(2q)!} 2^{2q} \underbrace{E(X^{2q})}_{\leq q! C^q \text{ nach Vorr.}} \\
 & \leq \sum_{q \in \mathbb{N}_0} 2^{2q} \lambda^{2q} C^q \frac{q!}{(2q)!}
 \end{aligned}$$

Nebenrechnung.: $\frac{(2q)!}{q!} = \prod_{j=1}^q (q+j) \geq \prod_{j=1}^q 2j = 2^q \cdot q!$

$$\begin{aligned}
\Rightarrow E(e^{\lambda X}) \cdot E(e^{-\lambda X}) &\stackrel{\text{s.o.}}{\leq} \sum_{q \in \mathbb{N}_0} 2^{2q} \lambda^{2q} C^q \frac{q!}{(2q)!} \\
&\stackrel{\text{N.R.}}{\leq} \sum_{q \in \mathbb{N}_0} 2^{2q} \lambda^{2q} C^q \frac{1}{2^q q!} \\
&= \sum_{q \in \mathbb{N}_0} \frac{2^q \lambda^{2q} C^q}{q!} = e^{2C\lambda^2}.
\end{aligned}$$

Da X zentriert, folgt mit der Jensen-Ungleichung:

$$\begin{aligned}
\Rightarrow M_X(\lambda) &\leq E(e^{\lambda X}) \cdot \underbrace{E(e^{-\lambda X})}_{\geq 1} \leq e^{2C\lambda^2} \\
\Rightarrow \text{Also: } \psi_X(\lambda) &\leq 2C\lambda^2 \Rightarrow X \in \mathcal{G}(4\nu)
\end{aligned}$$

□

Werfen wir noch einen zweiten Blick auf das tail-Verhalten:

Lemma 2.9. Die Momentenbedingung (2.3) ist äquivalent zu folgender Bedingung :

$$(2.4) \quad \exists \alpha > 0 \text{ sodass } E(e^{\alpha X^2}) \leq 2$$

Interpretation: Da $\exp(\cdot)$ schnell wächst, muss αX^2 schnell abfallen, falls (2.4) gilt.

Beweis. Nach Voraussetzung und Konvergenzsatz

$$\begin{aligned}
2 &\geq \sum_{k \in \mathbb{N}_0} \frac{\alpha^k E(X^{2k})}{k!} \\
\Leftrightarrow 1 &\geq \sum_{k \in \mathbb{N}} \frac{\alpha^k E(X^{2k})}{k!}
\end{aligned}$$

Alle Summanden nichtnegativ $\Rightarrow \forall k \in \mathbb{N}$ gilt: $E(X^{2k}) \leq \alpha^{-k} k!$.

Satz 2.1 Teil b) $\Rightarrow X \in \mathcal{G}(\frac{4}{\alpha})$.

Gegenrichtung: mit Satz 2.1 Teil a)

$$X \in \mathcal{G}(\nu) \stackrel{\text{Satz 2.1}}{\Rightarrow} E(X^{2q}) \leq C^q q! \text{ mit } C = 4\nu.$$

Setze: $\alpha = \frac{1}{2C} = \frac{1}{8\nu}$

$$\begin{aligned} \Rightarrow E(e^{\alpha X^2}) &= \sum_{q \in \mathbb{N}_0} \frac{\alpha^q E(X^{2q})}{q!} \\ &\leq \sum_{q \in \mathbb{N}_0} \left(\frac{1}{2C}\right)^q \frac{C^q q!}{q!} = \sum_{q \in \mathbb{N}_0} \left(\frac{1}{2}\right)^q = 2. \end{aligned}$$

□

Quantitative Variante der Charakterisierung (2.4)? Es sei $\alpha > 0$ und X zentriert. Dann:

$$X \in \mathcal{G}\left(\frac{1}{8\alpha}\right) \Rightarrow E(e^{\alpha X^2}) \stackrel{\leq 2}{\Rightarrow} \mathcal{G}\left(\frac{4}{\alpha}\right) \ni X$$

Beschränkte zentrierte ZVen sind sub-gaußsch:

Lemma 2.10 (Hoeffding).

Sei Y eine $[a, b]$ -wertige (schreibe zukünftig $Y \in [a, b]$) zentrierte ZVe. Sei $\psi_Y(\lambda) = \ln(E(e^{\lambda Y}))$.

$$\begin{aligned} \Rightarrow \psi_Y''(\lambda) &\leq \frac{(b-a)^2}{4} \text{ und} \\ Y &\in \mathcal{G}\left(\frac{(b-a)^2}{4}\right) \end{aligned}$$

Das Hoeffding-Lemma 2.3 wird beim Beweis der Hoeffding-Ungleichung 2.6 in 2.8 benötigt.

Beweis.

$$\begin{aligned} \left|Y - \frac{b+a}{2}\right| &\leq \frac{b-a}{2} \Rightarrow \text{Var}(Y) = \text{Var}\left(Y - \frac{b+a}{2}\right) \\ &\leq \left(\frac{b-a}{2}\right)^2 = \frac{(b-a)^2}{4}. \end{aligned}$$

Sei P_Y Verteilung von Y und $P_\lambda(dx) = e^{-\psi_Y(\lambda)} e^{\lambda x} P_Y(dx)$ modifiziertes absolutstetiges Maß. Sei $Z \in \mathbb{R}$ mit Verteilung $P_Z = P_\lambda$. *Ist das normiert?*

$$\begin{aligned} \int_{\mathbb{R}} P_\lambda(dx) &= e^{-\psi_Y(\lambda)} \int_{\mathbb{R}} e^{\lambda x} P_Y(dx) \\ &= (e^{\psi_Y(\lambda)})^{-1} E(e^{\lambda Y}) = M(\lambda)^{-1} M(\lambda) = 1 \end{aligned}$$

Da $Z \in [a, b]$ gilt ebenso:

$$\text{Var}(Z) \leq \left(\frac{b-a}{2}\right)^2.$$

Direkte Rechnung ergibt:

$$\begin{aligned} \psi_Y''(\lambda) &= \frac{d^2}{d^2\lambda} \ln(M_Y(\lambda)) \\ &= e^{-\psi_Y(\lambda)} E\left(Y^2 e^{\lambda Y}\right) - e^{-2\psi_Y(\lambda)} \left(E\left(Y e^{\lambda Y}\right)\right)^2 \\ &= e^{-\psi_Y(\lambda)} \int_{\mathbb{R}} y^2 e^{\lambda y} dP_Y(y) - \left(e^{-\psi_Y(\lambda)} \int_{\mathbb{R}} y e^{\lambda y} dP_Y(y)\right)^2 \\ &= \int_{\mathbb{R}} y^2 dP_\lambda(y) - \left(\int_{\mathbb{R}} y dP_\lambda(y)\right)^2 \\ &= \text{Var}(Z) \leq \frac{(b-a)^2}{4} \quad \forall \lambda \in \mathbb{R}. \end{aligned}$$

Da Y zentriert: $\psi_Y(0) = \psi_Y'(0) = 0$ und $\psi \in \mathcal{C}^2(0, \infty) \cap \mathcal{C}^1[0, \infty)$. Taylor-entwicklung mit Lagrange-Restglied liefert:

$$\begin{aligned} \exists \theta \in [0, \lambda] : \psi_Y(\lambda) &= \psi_Y(0) + \lambda \psi_Y'(0) + \frac{\lambda^2}{2} \psi_Y''(\theta) \\ &\leq 0 + 0 + \frac{\lambda^2(b-a)^2}{8} \Rightarrow Y \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right) \end{aligned}$$

□

Beispiel: Ungleichung des Lemmas ist scharf: Führe sogenannte *Rademacher-Zufallsvariable* $X: \Omega \rightarrow \{-1, 1\}$ ein mit $P(X = 1) = P(X = -1) = \frac{1}{2}$. Es gilt $X \in [-1, 1]$ mit $a = -1$ und $b = 1$. Wie in Lemma 2.2 zeigen wir:

$$\Rightarrow \psi_X''(\lambda) \leq \frac{(b-a)^2}{4} = 1 \quad \forall \lambda \geq 0$$

Nutzen nun die charakteristische Eigenschaft der KEF. $\lambda = 0$ einsetzen liefert: $\text{Var}(X) = \psi_X''(0) \leq 1$.

Nun berechnen wir die Varianz exakt:

$$\text{Var}(X) = E(X^2) = \frac{1}{2}(-1)^2 + \frac{1}{2}(-1)^2 = 1.$$

→ Ungleichung lässt sich nicht mehr verbessern.

2.4. Sub-Gamma-Zufallsvariablen.

Einige wichtige Verteilungen haben Dichten, die schnell abfallen bei $\pm\infty$, jedoch nicht ganz so schnell wie $\mathcal{G}(\nu)$. Daher führen wir ein:

Definition 2.11. Sei $X \in \mathcal{L}^1$ mit $E(X) = 0$. X heißt *sub- Γ -verteilt von rechts* mit Varianzfaktor ν und Skalenparameter c (in Symbolen $X \in \Gamma_+(\nu, c)$)

$$:\Leftrightarrow \quad \forall \lambda \in \left(0, \frac{1}{c}\right) \text{ gilt } \psi_X(\lambda) \leq \frac{\lambda^2 \nu}{2(1 - c\lambda)}$$

Die Klasse $\Gamma_-(\nu, c)$ führen wir ein, indem wir setzen:

$$X \in \Gamma_-(\nu, c) \Leftrightarrow -X \in \Gamma_+(\nu, c),$$

d.h.: X ist sub- Γ von links mit Varianzfaktor ν und Skalenparameter c $:\Leftrightarrow -X \in \Gamma_+(\nu, c) \Leftrightarrow X \in \Gamma_-(\nu, c)$.

X heißt sub- Γ -verteilt mit Varianzfaktor ν und Skalenparameter c $\Leftrightarrow X \in \Gamma_+(\nu, c) \cap \Gamma_-(\nu, c)$.

Das Konzept von sub-gamma findet im Kapitel 2.5 Verwendung, kann (dort) aber auch weggelassen werden. Es soll ein alternatives, flexibleres Kriterium zur sub-gauß-Eigenschaft bereitstellen.