Technische Universität Dortmund
Fakultät für Mathematik
Sommersemester 2021

Vorlesungsskript:

Konzentrationsungleichungen

Dozent: Prof. Dr. Ivan Veselić
erstellt 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 ausgearbeitetes Skript erstellt. Der vorliegende Text stellt einer 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

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

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

Eine andere nützliche Formulierung ist folgende

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

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 $x_1, ..., x_n$ abzuschätzen.
Dabei werden Abschätzungen der folgenden Bauart angestrebt:

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

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

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

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 unabhängig, 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

$$X_1, ..., X_n \colon \Omega \to \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}.$$

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

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

Sie beschreibt für verschiedene Zeiten $i = 1, ..., n$ 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 typischeweise verhält. Es kann z.B. nach einer oberen Schranke für $\max\limits_{i=1,...,n} \|Z_i\|$ gefragt werden. Eine sehr einfache Antwort wäre

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

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

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

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

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

$$B \colon \Omega \times [0, T] \to \mathbb{R}.$$

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 $\sup\limits_{\omega\in\Omega} \|B_t(\omega)\| = \infty$. Aber es gilt wiederum:

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

<u>Ziel:</u> 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:

$$P(X - E(X) \ge t) \le \ldots$$
$$P(X - E(X) \le -t) \le \ldots$$

<u>einfachste Antwort:</u> *Markov-Ungleichung*
Sei $Y \ge 0$ eine ZVe mit $E(Y) < \infty$. Dann gilt für alle $t \ge 0$:

$$Y \cdot \mathbb{1}_{\{Y\ge t\}} \ge t \cdot \mathbb{1}_{\{Y\ge t\}} \text{ auf } \Omega$$

Integration liefert:

$$E\left(Y \cdot \mathbb{1}_{\{Y\ge t\}}\right) \ge t \cdot E\left(\mathbb{1}_{\{Y\ge t\}}\right) = t \cdot P(Y \ge t).$$

Weitere Abschätzung der linken Seite:

$$E\left(Y \cdot \mathbb{1}_{\{Y\ge t\}}\right) \le E(Y).$$

Falls die Dichtefunktion f von Y gegeben ist, ist es natürlich, den Wertebereich von Y auf die horizontale Achse aufzutragen.
<u>Zusammenfassend:</u> Für jede ZVe $Y\colon \Omega \to [0,\infty)$ in $\mathcal{L}^1(\Omega, P)$ gilt nach obigen Ausführungen für $t > 0$ die Markov-Ungleichung:

$$P(Y \ge t) \le \frac{1}{t} E\left(Y \cdot \mathbb{1}_{\{Y\ge t\}}\right) \le \frac{1}{t} E(Y). \tag{2.1}$$

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

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

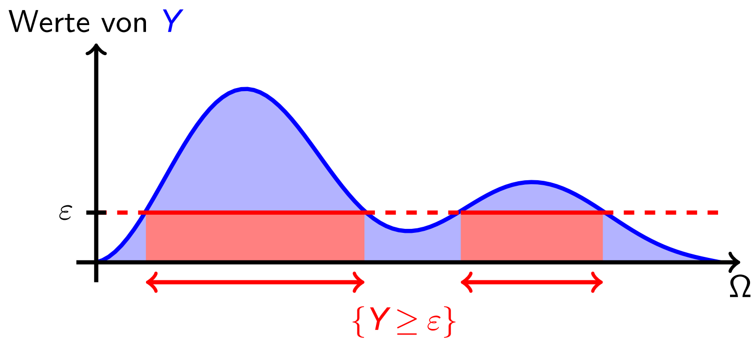


ABBILDUNG 1. Ω auf der horizontalen Achse.

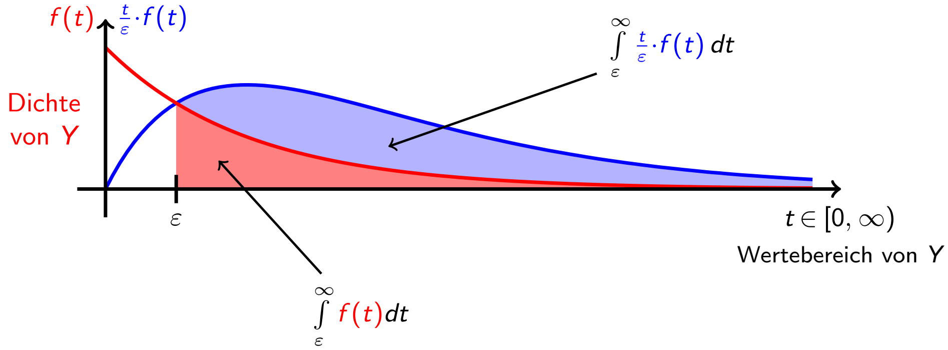


ABBILDUNG 2. Wertebereich auf der horizontalen Achse.

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

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

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

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

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

und erhalte damit die *Markov-Čebyšev-Ungleichung*. Die Ungleichung (2.2) gilt für sämtliche isotone (monoton wachsende) Funktionen $\Phi\colon I \to [0,\infty)$ auf einem Intervall $I \subset \mathbb{R}$, sodass $\Phi(t) > 0$ ist. [1] Für die Anwendbarkeit der

[1] In dieser Vorlesung benutzen wir der Kürze halber den Begriff *isoton* für monoton wachsend und *antiton* für monoton fallend.

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 $\forall\, q, t \geq 0$:

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

$\to$ schöne Form, da linke Seite unabhängig von q ist (Optimierungsaspekt)

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

<u>Erstes Fazit</u>: 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 $X_1, ..., X_n$ unabhängige ZVen mit endlichen Varianzen. Für $Z = \sum_{i=1}^n X_i$ gilt nach dem Additionssatz (Bienaymé):

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

$\to$ Formulierung einer Konzentrationsungleichung für gemittelte ZVen mit Hilfe von Ungleichung (2.2):

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

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

<u>Bemerkung:</u> 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 Kapiteln 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:

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

$\rightarrow$ Schranke mit exponentiellem Abfall gewonnen.
Studiere nun die Schranken genauer. Die folgende Abbildung

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

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 $X_1, ..., X_n$. 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 $X_1, ..., X_n$. 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:

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

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

2.2.1. *Die kumulantenerzeugenden Funktion und die Cramér-Transformierte.* Sei Z ZVe mit MEF $M_Z(\lambda)$, d.h. (äquivalent zu (2.4)):

$$P(Z \geq t) \leq e^{-\lambda t} M(\lambda) = e^{-\lambda t + \ln(M(\lambda))}. \tag{2.5}$$

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*:

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

Die Tansformierte ψ^* einer konvexen Funktion ψ wird abstrakt in Lemma 2.14 untersucht. Bedeutung der Cramér-Transformierte:
Minimiere die Schranke von (2.5) über λ und betrachte nur noch den Exponenten. Beachte: Vorzeichenwechsel liefert Maximierungsproblem:

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

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.

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

$$(2.6)\quad \tilde{\psi}_Z(t) := \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) = \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda)) = \psi_Z^*(t) \text{ für } t \geq E(Z).$$

Wir nennen die Funktion $\tilde{\psi}_Z$ in (2.6) 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

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

Um diese Situation zu vermeiden, nehmen wir in diesem Kapitel an, dass ein $\lambda_0 > 0$ existiert, sodass $E\left(e^{\lambda_0 Z}\right) < \infty$ 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) $\psi \colon I \to \mathbb{R}$ ist $\mathcal{C}^\infty$

Für zentrierte Zufallsvariablen Z gilt darüber hinaus:

(d) $\psi_Z : [0, b) \to \mathbb{R}$ 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:

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

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

2.2.2. *Bestimmung des Supremums mittels der Ableitung.* Ansatz:

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

$$\begin{aligned} 0 &= \frac{d}{d\lambda}(\lambda t - \psi(\lambda)) = t - \psi'(\lambda) \\ \Leftrightarrow \quad t &= \psi'(\lambda) \end{aligned}$$

Sei λ_t eine Lösung dieser Gleichung. Falls man den trivialen Fall einer fast sicher konstanten Zufallsvariable ausschließt, ist ψ strikt konvex.
$\Rightarrow \quad \lambda_t$ ist eindeutig.

Definition: Sei $B := \psi'_Z(b)$. Dann ist $\psi'_Z : I \to (0, B)$ 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.

2.2.3. *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.

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

$\to$ 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 Bennett-Ungleichung 2.20 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:

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

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

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

<u>Ansatz:</u> $t = \psi'_Z(\lambda)$, 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:

$$h(x) := (1 + x) \cdot \ln (1 + x) - x.$$

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 $Y : \Omega \to \{0,1\}$ 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 $h_p \colon (0,1) \to \mathbb{R}$, $h_p(a) = (1-a) \cdot \ln\left(\frac{1-a}{1-p}\right) + a \cdot \ln\left(\frac{a}{p}\right)$ 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 $X_1, ..., X_n$. 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 $X_1, ..., X_n$.

Für λ mit $\psi_{X_1}(\lambda) < \infty$ bestimmen wir die KEF von Z:

$$
\begin{aligned}
\psi_Z(\lambda) &= \ln\left(E\left(e^{\lambda Z}\right)\right) \;=\; \ln\left(E\left(e^{\lambda \sum_{i=1}^n X_i}\right)\right) \\
&= \ln\left(\prod_{i=1}^n \left(E\left(e^{\lambda X_i}\right)\right)\right) \text{ (wegen Unabhängigkeit)} \\
&= \ln\left(E\left(e^{\lambda X_1}\right)^n\right) \text{ (wegen identischer Verteilung)} \\
&= n \cdot \ln\left(E\left(e^{\lambda X_1}\right)\right) \\
&= n \cdot \psi_{X_1}(\lambda).
\end{aligned} \tag{2.7}
$$

Falls die X_i nur unabhängig sind, erhalten wir mit denselben Schritten zumindest

$$
\psi_Z(\lambda) = \sum_{i=1}^n \psi_{X_i}(\lambda). \tag{2.8}
$$

Bemerkung 2.3. Wir nehmen in diesem Abschnitt durchgängig an, dass die betrachtete ZV X integrierbar ist und ein $\lambda > 0$ existiert mit $M(\lambda) < \infty$, was sich dann auch auf die KEF überträgt. Insbesondere ist $b > 0$ in der Definition des Intervalls $I = (0, b)$. Im Spezialfall einer f.s. konstanten ZV gilt

$$
\psi^*(t) = \begin{cases} 0, & \text{falls } t = EX \\ +\infty, & \text{falls } t > EX, \end{cases}
$$

anderenfalls

$$
\psi^*(t) = \begin{cases} 0, & \text{falls } t = EX \\ \in (0, \infty), & \text{falls } t > EX. \end{cases}
$$

Auch die Cramér-Trafo hat ein einfaches verhalten bei einer i.i.d.-Summe:

Lemma 2.4. *Seien $X, X_1, ..., X_n$ unabhängige, identisch verteilten ZVen und $Z := \sum_{i=1}^n X_i$. Dann 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:

$$
\begin{aligned}
\psi_Z'(\lambda) &\overset{(2.7)}{=} n \cdot \psi_X'(\lambda) \\
&= \mathcal{M}_n(\psi_X'(\lambda)) = (\mathcal{M}_n \circ \psi_X')(\lambda)
\end{aligned}
$$

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

$$\Rightarrow \quad \lambda_t = (\psi_Z')^{-1}(t) = (\psi_X')^{-1}((\mathcal{M}_n)^{-1}(t)) = (\psi_X')^{-1}\left(\frac{t}{n}\right).$$

Für die Cramér-Trafo von Z folgt schließlich

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

□

(e) Cramér-Transformierte für zentrierte Binomialverteilungen::
Jetzt sind alle Vorbereitungen getroffen, die Cramér-Trafo der Binomialverteilung zu berechnen. Seien $X_1, ..., X_n \sim Ber(p)$ 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 $n \to \infty$, aber wie schnell und wohin konvergiert h_p für $n \to \infty$? $\rightarrow$ Noch zu untersuchen!

2.3. **Sub-Gaußsche Zufallsvariablen.**
Gauß-ZVen $\widehat{=}$ gut umgängliche Klasse von ZVen
Klasse von Verteilungen, die von Gauß-ZVen dominiert werden, sind ähnlich gut umgänglich.

Definition 2.5. Eine reellwertige, zentrierte ZVe X heißt *sub-gaußsch* mit Varianzfaktor ν, in Zeichen: $P_X \in \mathcal{G}(\nu)$, falls

$$\forall \lambda \in \mathbb{R} : \psi_X(\lambda) \leq \frac{\lambda^2 \nu}{2}. \tag{2.9}$$

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 2.6 (Eigenschaften sub-gaußscher Zufallsvariablen). (a) Varianz: Es folgt $\mathrm{Var}(X) \leq \nu$, im Allg. gilt aber $\mathrm{Var}(X) \neq \nu$. 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

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

(c) In der Tat, Bedingung (2.9) kann sogar als Vergleich mit Gauß-ZV verstanden werden:

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

(d) Verträglichkeit mit der Faltung:
Sind $X_1, ..., X_n$ unabhängige ZVen mit $X_i \in \mathcal{G}(\nu_i)$ für $i = 1, ..., n$, 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:

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

(e) Inklusionen

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

Definition 2.7. Ein messbares $X\colon \Omega \to \{-1, 1\}$ 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 2.8 (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 2.9.
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 $\forall\, q \in \mathbb{N}$:

$$E(X^{2q}) \leq 2q!(2\nu)^q \leq q!(4\nu)^q. \tag{2.10}$$

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

$$\forall\; q \in \mathbb{N} : E(X^{2q}) \leq q! C^q, \tag{2.11}$$

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

$$\max\{P(X > s), P(X < -s)\} \leq e^{-\frac{s^2}{8C}}. \tag{2.12}$$

Theorem 2.9 kann als Hilfsmittel dienen, um die Voraussetzung der Bernstein-Ungleichung 2.22 nachzurechnen, wie es im Beweis vom Johnson-Lindenstrauß-Lemma 2.26 in 2.9 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 (2.10):

$$E\left(Y^{2q}\right) = \int_0^\infty P\left(|Y|^{2q} > x\right)\, \mathrm{d}x.$$

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

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

Anwendung der Voraussetzung ermöglicht folgende Abschätzung:

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

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

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

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

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

Für allgemeinere Varianzen ν folgt:

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

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

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

<u>Frage/Übung</u>: Existieren überhaupt die ungeraden Momente? Sind sie summierbar?
Wegen Konvexität von $x \mapsto x^{2q} = x^m$ für $m \in 2\mathbb{N}$ folgt (vgl. auch Abb. ??):

$$(a - b)^m \leq 2^{m-1}(a^m - b^m) \leq 2^{m-1}(a^m + b^m)$$

$$\Rightarrow \quad E\left((X - \tilde{X})^{2q}\right) \leq 2^{2q-1}\left(E\left(X^{2q}\right) + E(\tilde{X}^{2q})\right) \overset{\text{id. vert.}}{=} 2^{2q} E\left(X^{2q}\right)$$

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

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

$$\Rightarrow \quad E\left(e^{\lambda X}\right) \cdot E\left(e^{-\lambda X}\right) \overset{\text{s.o.}}{\leq} \sum_{q \in \mathbb{N}_0} 2^{2q} \lambda^{2q} C^q \frac{q!}{(2q)!}$$
$$\overset{\text{N.R.}}{\leq} \sum_{q \in \mathbb{N}_0} 2^{2q} \lambda^{2q} C^q \frac{1}{2^q q!}$$
$$= \sum_{q \in \mathbb{N}_0} \frac{2^q \lambda^{2q} C^q}{q!} \quad = \quad e^{2C\lambda^2}.$$

Da X zentriert, folgt mit der Jensen-Ungleichung:

$$\Rightarrow \quad M_X(\lambda) \leq E\left(e^{\lambda X}\right) \cdot \underbrace{E\left(e^{-\lambda X}\right)}_{\geq 1} \leq e^{2C\lambda^2}$$

$$\Rightarrow \quad \text{Also: } \psi_X(\lambda) \leq \frac{4C}{2}\lambda^2 \quad \Rightarrow \quad X \in \mathcal{G}(4C)$$

□

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

Lemma 2.10. *Die Momentenbedingung* (2.11) *ist äquivalent zu folgender Bedingung :*

(2.13) $$\exists \alpha > 0, \text{ sodass } E\left(e^{\alpha X^2}\right) \leq 2.$$

Interpretation: Da exp(.) schnell wächst, muss αX^2 schnell abfallen, falls (2.13) gilt.

Beweis. Nach Voraussetzung und Konvergenzsatz

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

Alle Summanden nichtnegativ $\Rightarrow$ $\forall\, k \in \mathbb{N}$ gilt: $E(X^{2k}) \leq \alpha^{-k} k!$.
Satz 2.9 Teil b) $\Rightarrow$ $X \in \mathcal{G}\left(\frac{4}{\alpha}\right)$.
Gegenrichtung: mit Satz 2.9 Teil a)

$$
X \in \mathcal{G}(\nu) \quad \overset{\text{Satz 2.9}}{\Rightarrow} \quad E\left(X^{2q}\right) \leq C^q q! \text{ mit } C = 4\nu.
$$

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

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

□

Quantitative Variante der Charakterisierung (2.13)? Es sei $\alpha > 0$ und X zentriert. Dann:

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

Beschränkte zentrierte ZVen sind sub-gaußsch:

Lemma 2.11 (Hoeffding-Lemma)**.**
Sei Y eine $[a,b]$-wertige (schreibe zukünftig $Y \in [a,b]$) zentrierte ZVe. Sei $\psi_Y(\lambda) = \ln(E(e^{\lambda Y}))$.

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

Das Hoeffding-Lemma 2.11 wird beim Beweis der Hoeffding-Ungleichung 2.18 in 2.6 benötigt.

Beweis.

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

Sei P_Y Verteilung von Y und $P_\lambda(\mathrm{dx})=e^{-\psi_Y(\lambda)}e^{\lambda x}P_Y(\mathrm{dx})$ modifiziertes absolutstetiges Maß. Sei $Z\in\mathbb{R}$ mit Verteilung $P_Z=P_\lambda$.
Frage/Übung: Ist überhaupt P_λ ein W-Mass?
Da $Z\in[a,b]$ gilt ebenso:

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

Direkte Rechnung ergibt:

$$\begin{aligned}\psi_Y''(\lambda)&=\frac{\mathrm{d}^2}{\mathrm{d}^2\lambda}\ln(M_Y(\lambda))\\&=e^{-\psi_Y(\lambda)}E\left(Y^2e^{\lambda Y}\right)-e^{-2\psi_Y(\lambda)}\left(E\left(Ye^{\lambda Y}\right)\right)^2\\&=e^{-\psi_y(\lambda)}\int_{\mathbb{R}}y^2e^{\lambda y}\,\mathrm{d}P_Y(y)-\left(e^{-\psi_Y(\lambda)}\int_{\mathbb{R}}ye^{\lambda y}\,\mathrm{d}P_Y(y)\right)^2\\&=\int_{\mathbb{R}}y^2\,\mathrm{d}P_\lambda(y)-\left(\int_{\mathbb{R}}y\,\mathrm{d}P_\lambda(y)\right)^2\\&=\operatorname{Var}(Z)\le\frac{(b-a)^2}{4}\;\forall\,\lambda\in\mathbb{R}.\end{aligned}$$

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

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

□

Beispiel: Ungleichung des Lemmas ist scharf: Führe sogenannte *Rademacher-Zufallsvariable* $X\colon\Omega\to\{-1,1\}$ 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 2.11 zeigen wir:

$$\Rightarrow\quad\psi_X''(\lambda)\le\frac{(b-a)^2}{4}=1\,\forall\lambda\ge0$$

Nutzen nun die charakteristische Eigenschaft der KEF. $\lambda=0$ einsetzen liefert: $\operatorname{Var}(X)=\psi_X''(0)\le1$.

Nun berechnen wir die Varianz exakt:

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

$\rightarrow$ 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 2.12. Sei $X \in \mathcal{L}^1$ mit $E(X) = 0$. X heißt *sub-*Γ*-verteilt von rechts* mit Varianzfaktor $\nu > 0$ und Skalenparameter $c \geq 0$, in Symbolen

$$X \in \Gamma_+(\nu, c) :\Leftrightarrow \begin{cases} \forall \lambda \in \left(0, \frac{1}{c}\right) \text{ gilt } \psi_X(\lambda) \leq \dfrac{\lambda^2 \nu}{2(1 - c\lambda)}, & \text{falls } c > 0 \\ \forall \lambda \in (0, \infty) \text{ gilt } \psi_X(\lambda) \leq \dfrac{\lambda^2 \nu}{2}, & \text{falls } c = 0. \end{cases}$$

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 $c \Leftrightarrow X \in \Gamma(\nu, c) := \Gamma_+(\nu, c) \cap \Gamma_-(\nu, c)$.

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.

Einige Bemerkungen:

(a) Insbesondere $\Gamma(\nu, 0) = \mathcal{G}(\nu)$

(b) Sei ZVe Y gammaverteilt mit Parameter $a, b \geq 0$, $X := Y - E(Y)$ zentrierte Version. Y hat die Dichte

$$\forall x \geq 0 : f(x) = \frac{x^{a-1} e^{-\frac{x}{b}}}{\Gamma(a) b^a}$$

$$\Rightarrow \quad E(Y) = ab, \operatorname{Var}(Y) = ab^2 \text{ und}$$

$$E\left(e^{\lambda X}\right) = \int_0^\infty e^{\lambda(y-ab)} f(y) dy = e^{-\lambda ab - a\ln(1-\lambda b)}$$

$$\Rightarrow \quad \forall \lambda \in \left(0, \frac{1}{c}\right) : \psi_X(\lambda) = a(-\lambda b - \ln(1-\lambda b))$$

$$\overset{\text{NR.}}{\leq} \frac{\lambda^2 \nu}{2(1-c\lambda)}, \text{ mit } \nu = ab^2, c = b$$

Nebenrechung: Für $x := \lambda b \in (0,1)$ gilt mit der Reihenentwicklung des Logarithmus:

$$-\ln(1-x) - x = \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots = \frac{x^2}{2}\left(1 + \frac{2}{3}x + \frac{2}{4}x^2 + \ldots\right)$$

$$\leq \frac{x^2}{2}\left(1 + x + x^2 + \ldots\right) \leq \frac{x^2}{2}\frac{1}{1-x}$$

Also sind Γ-ZVen sub-$\Gamma(ab^2, b)$. Allerdings ist X nicht symmetrisch verteilt. Die Verteilung von $-X$ fällt sogar schneller ab, da ja

$$\psi_{-X}(\lambda) = \ln E\left(e^{-\lambda X}\right) = \ln E\left(e^{(-\lambda)X}\right) = a\left(\lambda b - \ln(1+\lambda b)\right) \leq \frac{\lambda^2}{2} ab^2 = \frac{\lambda^2}{2}\nu$$

wegen

$$y - \ln(1-y) = y - y + \frac{y^2}{2} - \frac{y^3}{3} + \frac{y^4}{4} - \ldots$$

$$= \frac{y^2}{2} - \left(\frac{y^3}{3} - \frac{y^4}{4}\right) - \left(\frac{y^5}{5} - \frac{y^6}{6}\right) - \ldots \leq \frac{x^2}{2} \quad \text{für } y \in (0,1).$$

Also: $X \in \Gamma_-(ab^2, 0) \subset \Gamma_-(ab^2, b)$ und damit $X \in \Gamma(ab^2, b)$.

Um das tail-Verhalten von sub-gamma ZV zu verstehen, untersuchen wir die Fenchel-Legendre-Duale von $\psi(\lambda) = \frac{\lambda^2 \nu}{2(1-c\lambda)}$. Setze dazu $h_1(u) = 1 + u - \sqrt{1+2u}$ für $u > 0$. In der Übung wird gezeigt, dass

$$\psi^*(t) = \sup_{\lambda \in \left(0, \frac{1}{c}\right)} \left(t\lambda - \frac{\lambda^2 \nu}{2(1-c\lambda)}\right) = \frac{\nu}{c^2} h_1\left(\frac{ct}{\nu}\right), \tag{2.14}$$

dass die Funktion $h_1 : (0, \infty) \to (0, \infty)$ strikt monoton wachsend, und bijektiv mit $h_1^{-1}(u) = u + \sqrt{2u}$ ist. Damit ergibt sich:

$$
\begin{aligned}
(\psi^*)^{-1}(u) &= \left(\mathcal{M}_{c/\nu}\right)^{-1} \circ h_1^{-1} \circ \left(\mathcal{M}_{\nu/c^2}\right)^{-1}(u) \\
&= \frac{\nu}{c} h_1^{-1}\left(\frac{c^2 u}{\nu}\right) = \frac{\nu}{c}\left(\frac{c^2 u}{\nu} + \sqrt{2c^2 u/\nu}\right) = \sqrt{2\nu u} + cu.
\end{aligned} \tag{2.15}
$$

Die Chernoff-Schranken lauten:

(i) $X \in \Gamma_+(\nu, c) \Rightarrow \forall t > 0$ gilt $P(X > t) \leq \exp\left(-\frac{\nu}{c^2} h_1\left(\frac{ct}{\nu}\right)\right)$.
Mit der Substitution $s := \psi^*(t) = \frac{\nu}{c^2} h_1\left(\frac{ct}{\nu}\right)$ und der berechteten Formel für die Inverse erhalten wir die oftmals praktischere äquivalente Darstellung

$$\forall s > 0 \text{ gilt } P(X > \sqrt{2\nu s} + cs) \leq e^{-s}.$$

(ii) $X \in \Gamma(\nu, c) \Rightarrow \forall t > 0:$

$$\max\{P(X > \sqrt{2\nu t} + ct), P(-X > \sqrt{2\nu t} + ct)\} \leq e^{-t}.$$

Es gilt wieder (fast) eine Äquivalenz zu einer Momentenbedingung:

Satz 2.13.
Sei $X \in \mathcal{L}^1$ eine zentrierte Zufallsvariable.
(a) Gilt für ein $\nu > 0$ für jedes $t > 0$:

$$\max\left\{P(X > \sqrt{2\nu t} + ct), P(X < -(\sqrt{2\nu t} + ct))\right\} \leq e^{-t}, \tag{2.16}$$

so folgt für jedes $q \in \mathbb{N}$

$$E(X^{2q}) \leq q!(8\nu)^q + (2q)!(4C)^{2q}.$$

(b) Gilt umgekehrt für zwei Parameter $A, B \geq 0$ und jedes $q \in \mathbb{N}$

$$E(X^{2q}) \leq q! A^q + (2q)! B^{2q}, \tag{2.17}$$

so ist bereits $X \in \Gamma(4(A + B^2), 2B)$.

Man beachte die Analogien zu Kapitel 2.3, u.a. Satz 2.9.

2.5. **Eine Maximal-Ungleichung.**

In Teil (C) der Motivation wollten wir das Supremum einer Familie von ZVen abschätzen. Konkrete Schranke wird in diesem Kapitel für den Erwartungswert des Supremums hergeleitet.
Seien ZVen $Z_1, ..., Z_N \in \mathbb{R}$ mit $Z_i \in \mathcal{G}(\nu)$ für $i = 1, ..., N$ und $\nu > 0$.

Ziel: Schätze $E\left(\max_{i=1,\dots,N} Z_i\right)$ nach oben ab.
Idee: Nutze sub-gaußsche Eigenschaft der ZVen wie folgt aus.

$$
\begin{aligned}
\exp\left(\lambda E\left(\max_{i=1,\dots,N} Z_i\right)\right) &\overset{\text{Jensen}}{\leq} E\left(\exp\left(\lambda \max_{i=1,\dots,N} Z_i\right)\right) \\
&\leq E\left(\max_{i=1,\dots,N} \exp\left(\lambda Z_i\right)\right) \\
&\leq E\left(\sum_{i=1}^{N} \exp\left(\lambda Z_i\right)\right) \\
&= \sum_{i=1}^{N} M_{Z_i}(\lambda) \leq N \cdot \exp\left(\frac{\lambda^2 \nu}{2}\right).
\end{aligned}
\tag{2.18}
$$

Durch Logarithmieren erhalten wir eine Abschätzung in gewünschter Form:

$$
E\left(\max_{i=1,\dots,N} Z_i\right) \leq \frac{\ln(N) + \frac{\lambda^2 \nu}{2}}{\lambda}.
$$

Wähle nun

$$
\lambda = \sqrt{2(\ln N)/\nu}. \tag{2.19}
$$

Wir erhalten:

$$
E\left(\max_{i=1,\dots,N} Z_i\right) \leq \frac{\sqrt{\ln(N)}}{\sqrt{2/\nu}} + \frac{\sqrt{\nu \ln(N)}}{\sqrt{2}} = \sqrt{2\nu \ln(N)}. \tag{2.20}
$$

Fragen:

- Kann man λ geschickter wählen als in (2.19)?
- Ist die Schranke in (2.20) optimal?

Zur zweiten Frage: Sind $Z_1, \dots, Z_N \sim \mathcal{N}(0,1)$ unabhängig, folgt aus dem obigen:

$$
\frac{E\left(\max_{i=1,\dots,N} Z_i\right)}{\sqrt{2\nu \ln(N)}} \leq 1.
$$

Optimalität würde bedeuten, dass sogar:

$$
\frac{E\left(\max_{i=1,\dots,N} Z_i\right)}{\sqrt{2\nu \ln(N)}} = 1.
$$

Für dieses Beispiel gilt zumindest (Übung):

$$\lim_{N\to\infty} \frac{E\left(\max\limits_{i=1,\dots,N} Z_i\right)}{\sqrt{2\nu\ln(N)}} = 1.$$

2.5.1. *Ähnliche Resultate gelten auch für sub-*Γ*-Zufallsvariablen.*
Dazu holen wir etwas aus und entwickeln allgemeinere Resultate. Die folgenden Resultate dieses Kapitels sind anwendbar auf Klassen von Verteilungen, die geeignet dominiert werden. Die sub-gauß oder sub-Γ-ZVen sind lediglich Beispiele für solche Klassen, sodass auch ohne Kapitel 2.4 nachfolgende Ausführungen (bis auf Korollar 2.16) verstanden werden können.
Untersuchung von Eigenschaftern der abstrakten Legendre-Fenchel-Dualen.

Lemma 2.14.
Seien $b \in [0,\infty)$ *und* $\psi\colon [0,b) \to \mathbb{R}$ *konvex und* $\mathcal{C}^1([0,\infty))$ *mit* $\psi(0) = \psi'(0) = 0$*. Für* $t \geq 0$ *setzen wir:*

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

Dann ist ψ^* *auf* $[0,\infty)$ *nichtnegativ, monoton wachsend, konvex und unbeschränkt. Die* verallgemeinerte Inverse (Pseudo-Inverse):

$$(\psi^*)^{-1}(y) := \inf\{t \geq 0 \mid \psi^*(t) > y\}$$

erfüllt die Gleichung: $(\psi^*)^{-1}(y) = \inf\limits_{\lambda\in(0,b)} \left(\dfrac{y+\psi(\lambda)}{\lambda}\right).$

Die Voraussetzungen sind für die KEF ψ_X einer zentrierten ZV X erfüllt.
Man beachte inhaltliche Analogien zum Kapitel 2.2 .

Beweis. $t \mapsto \lambda t - \psi(\lambda)$ ist affin-linear, konvex und isoton für $\lambda \geq 0$.
Übung $\Rightarrow$ ψ^* auch konvex und isoton als Supremum solcher Funktionen.

$$\psi^*(0) = \sup_{0<\lambda<b} \big(0 - \psi(\lambda)\big) = -\inf_{\lambda\in(0,b)} \psi(\lambda) = 0,$$

Denn: ψ konvex $\Rightarrow$ ψ' isoton, mit $\psi'(0) = 0$. Also ist ψ' nichtnegativ.
$\Rightarrow$ ψ isoton, mit $\psi(0) = 0$ und somit ist ψ nichtnegativ.
Damit ist auch ψ^* nichtnegativ.

Sei nun $\lambda_0 \in (0,b)$ fix. Da für jedes $t \geq 0$

$$\psi^*(t) \geq \sup_{\lambda \in [0,b)} (\lambda t - \psi(\lambda)) \geq \lambda_0 t - \psi(\lambda_0)$$

vorliegt, ist ψ^* unbeschränkt und $\{t \geq 0 \mid \psi^*(t) > y\} \neq \emptyset$, $\forall\, y \geq 0$.
Setze $u := \inf_{\lambda \in (0,b)} \left(\frac{y + \psi(\lambda)}{\lambda} \right)$, so gilt $\forall\, t \geq 0$:

$$\begin{aligned} u \geq t \quad &\Leftrightarrow \quad \forall\, \lambda \in (0,b) : \frac{y+\psi(\lambda)}{\lambda} \geq t \\ &\Leftrightarrow \quad \forall\, \lambda \in (0,b) : y \geq \lambda t - \psi(\lambda) \\ &\Leftrightarrow \quad y \geq \sup_{\lambda \in (0,b)} (\lambda t - \psi(\lambda)) = \psi^*(t) \end{aligned}$$

Komplementbildung liefert äquivalente Aussage dazu:

$$u < t \quad \Leftrightarrow \quad \psi^*(t) > y.$$

Also: $(u, \infty) = \{t \geq 0 \mid \psi^*(t) > y\}$
$\Rightarrow \quad u = \inf\big((u,\infty)\big) = \inf\{t \geq 0 \mid \psi^*(t) > y\} = \underbrace{(\psi^*)^{-1}(t).}_{\text{verallg. Inverse}}$ $\square$

Vergleichen Sie die obige Vorgehensweise mit der aus der Stochastik bekannten Definition/Konstruktion der Quantil-Transformierten.
Anwendung von Lemma 2.14:

Satz 2.15 (Maximal-Ungleichung)**.**
Seien $Z_1, ..., Z_N \in \mathbb{R}$ ZVen, $b \in (0,\infty)$, $\psi \in C^1\big([0,b)\big)$ konvex mit $\psi(0) = \psi'(0) = 0$, sodass

$$\psi_{Z_i}(\lambda) = \ln(M_{Z_i}(\lambda)) \leq \psi(\lambda) \qquad \textit{für } \lambda \in [0,b) \textit{ und } i = 1, ..., N.$$

Dann folgt

$$\textit{für } i = 1,...,N : Z_i \in \mathcal{L}^1 \textit{ und } E\left(\max_{i=1,\dots,N} Z_i\right) \leq (\psi^*)^{-1}(\ln(N)).$$

Es wird keine Unabhängigkeit verlangt!

Beweis. Mit Jensen-Ungleichung ist bekannt für $\lambda \in (0,b)$:

$$\begin{aligned} \exp\left(\lambda E\left(\max_{i=1,\dots,N} Z_i\right)\right) &\leq E\left(\exp\left(\lambda \max_{i=1,\dots,N} Z_i\right)\right) \\ &\leq E\left(\left(\max_{i=1,\dots,N} e^{\lambda Z_i}\right)\right) \leq \sum_{i=1}^N E\left(e^{\lambda Z_i}\right) \overset{\text{Vorr.}}{\leq} N e^{\psi(\lambda)} \end{aligned}$$

$$\Rightarrow\ \forall\,\lambda \in (0,b):\ \lambda E\left(\max_{i=1,...,N} Z_i\right) \leq \psi(\lambda) + \ln(N)$$

$$\Rightarrow\ \forall\,\lambda \in (0,b):\ E\left(\max_{i=1,...,N} Z_i\right) \leq \frac{\psi(\lambda) + \ln(N)}{\lambda}$$

$$\Leftrightarrow\ E\left(\max_{i=1,...,N} Z_i\right) \leq \inf_{\lambda\in(0,b)} \frac{\psi(\lambda) + \ln(N)}{\lambda} \overset{2.14}{=} (\psi^*)^{-1}(\ln(N))$$

□

Korollar 2.16.
Seien $Z_1, ..., Z_n \in \mathcal{L}^1$ zentrierte ZVen. Es gelten:

(a) *Falls $Z_i \in \mathcal{G}(\nu)$ für $i = 1, ..., N$, so:*

$$E\left(\max_{i=1,...,N} Z_i\right) \leq \sqrt{2\nu\ln(N)}.$$

(b) *Falls $Z_i \in \Gamma_+(\nu, c)$ für $i = 1, ..., N$, so:*

$$E\left(\max_{i=1,...,N} Z_i\right) \leq \sqrt{2\nu\ln(N)} + c\ln(N).$$

Beweis. zu (a): siehe Idee/Rechnung (2.18) am Anfang von Kapitel 2.5.
zu (b): Einsetzen von (2.15) in $(\psi^*)^{-1}(\ln(N))$ liefert die angegebene obere Schranke. □

Beispiel 2.17 (χ^2-Verteilung).
Sind $Y_1, ..., Y_k \sim \mathcal{N}(0,1)$ unabhängige ZVen, so ist $X := \sum_{i=1}^k Y_i^2$ χ^2-verteilt mit k Freiheitsgraden. Die Dichte von X ist

$$f(x) = \frac{x^{k/2-1}e^{-x/2}}{\Gamma(k/2)\sqrt[k]{2}},$$

d.h. Dichte der Γ-Verteilung mit $a = k/2, b = 2$. Insbesondere $X - E(X) = X - k \in \Gamma_+(2k, 2) \cap \Gamma_-(2k, 0)$. Für $X_1, ..., X_N \sim \chi^2$ mit k Freiheitsgraden impliziert Korollar 2.16

$$E\left(\max_{i=1,...,N} X_i - k\right) \leq 2\sqrt{k\ln(N)} + 2\ln(N)\square$$

2.6. **Hoeffding-Ungleichung.**
Nächstes Ziel: Schranken für Wahrscheinlichkeiten für große Werte von Summen von unabhängigen ZVen. Seien $X_1, ..., X_n \in \mathbb{R}$ unabhängige $\mathcal{L}^1$-ZVen, sodass ein echtes Intervall $I \subset \mathbb{R}$ existiert mit $M_{X_i}(\lambda) = E(e^{\lambda X_i}) < \infty$ für $i = 1, ..., n, \lambda \in I$. Für $S := \sum_{i=1}^n (X_i - E(X_i))$ gilt:

$$\forall\,\lambda \in I:\ \psi(\lambda) = \sum_{i=1}^n \ln(E(e^{\lambda(X_i - E(X_i))})).$$

Sind die X_i sogar beschränkt, genauer: $X_i \in [a_i, b_i]$ für $i = 1, ..., n$, so gilt nach Hoeffding-Lemma 2.11:

$$\begin{aligned}
\psi''_{X_i - E(X_i)}(\lambda) \leq \frac{(b_i - a_i)^2}{4} \quad &\Rightarrow \quad \psi_{X_i - E(X_i)}(\lambda) \leq \frac{\lambda^2 (b_i - a_i)^2}{8} \\
(2.8) \Rightarrow \quad &\psi_S(\lambda) \leq \frac{\lambda^2}{8} \sum_{i=1}^{n} (b_i - a_i)^2 \\
\Rightarrow \quad &S \in \mathcal{G}\left(\frac{1}{4} \sum_{i=1}^{n} (b_i - a_i)^2 \right)
\end{aligned}$$

und somit

Satz 2.18 (Hoeffding-Ungleichung)**.**
Seien $X_1 \in [a_1, b_1], \ldots, X_n \in [a_n, b_n]$ unabhängige ZVen mit zentrierter Summe S. Dann gilt $\forall\, t \geq 0$:

$$P(S \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right)$$

Beweis. s.o. und Bemerkung 2.8. $\square$

<u>Bemerkung/Beispiel:</u> Wähle $X_i = \alpha_i \varepsilon_i$, wobei $\varepsilon_1, \ldots, \varepsilon_n$ unabh. Rademacher-ZVen:

$$\overset{\text{Satz 2.18}}{\Rightarrow} \quad P(S \geq t) \leq \exp\left(-\frac{2t^2}{4 \sum_{i=1}^{n} \alpha_i^2} \right)$$

Da $\mathrm{Var}(S) = \mathrm{Var}(\alpha_1 \epsilon_1 + \ldots + \alpha_n \epsilon_n) = \sum_i \alpha_i^2 \,\mathrm{Var}(\epsilon_i) = \sum_i \alpha_i^2$ identifizieren wir

$$\exp\left(-\frac{2t^2}{4 \sum_{i=1}^{n} \alpha_i^2} \right) = \exp\left(-\frac{t^2}{2\,\mathrm{Var}(S)} \right)$$

$$\underline{\text{Im Allgemeinen gilt aber:}}\ \mathrm{Var}(S) < \frac{1}{4} \sum_{i=1}^{n} (b_i - a_i)^2.$$

In diesen Fällen gibt es feinere Ungleichungen: die *Bennett-* und *Bernsteinungleichung.*

Bemerkung 2.19 (Extremalitätseigenschjaft der Rademacher-ZVen)*.* Wie lässt sich die Varianz beschränkter ZVen maximieren?
→ schiebe Werte so weit wie möglich nach außen, an die Ränder des Wertebereiches
→ Rademacher-ZVen sind gerade diejenigen ZVen, die auf einem Intervall $[-\alpha_i, \alpha_i]$ die Varianz maximieren (zum Wert $\sum_{i=1}^{n} \alpha_i^2$).

2.7. **Bennett-Ungleichung.**
Wie zuvor: $X_1, ..., X_n$ unabhängige ZVen. $S := \sum_{i=1}^n (X_i - E(X_i))$.

$$\psi_S(\lambda) = \sum_{i=1}^n \Big(\ln(E(e^{\lambda X_i})) - \lambda E(X_i)\Big) \tag{2.21}$$

$$\leq \sum_{i=1}^n E(e^{\lambda X_i} - \lambda X_i - 1) \tag{2.22}$$

da ja $\ln u \leq u - 1$ für $u > 0$.

(2.21) und (2.22) sind Startpunkte für Bennett- bzw. Bernstein-Ungleichung.

Satz 2.20 (Bennett-Ungleichung)**.**
Sei $b > 0$. Seien $X_1, ..., X_n \in \mathcal{L}^2$ unabhängige ZVen, sodass $X_i \leq b$ fast sicher für $i = 1, ..., n$ (einseitige Beschränktheit).
Sei $\nu = \sum_{i=1}^n E(X_i^2)$. Dann gilt:

$$\psi_S(\lambda) \leq n \ln\Big(1 + \frac{\nu}{nb^2}\phi(b\lambda)\Big) \leq \frac{\nu}{b^2}\phi(b\lambda) \tag{2.23}$$

mit $\phi(u) := e^u - u - 1$ für $u \in \mathbb{R}$. Ferner gilt:

$$P(S \geq t) \leq \exp\left(-\frac{\nu}{b^2} h\left(\frac{bt}{\nu}\right)\right),$$

wobei $h(u) := (1 + u)\ln(1 + u) - u$ für $u \geq 0$.

Die Bennett-Ungleichung wird mit der Bernstein-Ungleichung in Kapitel 2.8 verglichen. Ansonsten wird die Bennett-Ungleichung im weiteren Verlauf des Skripts nicht verwendet.

Beweis. O.E. sei $b = 1$ (skaliere sonst nach). Die Abbildung

$\mathbb{R} \setminus \{0\} \ni u \mapsto \phi(u)/u^2$ ist isoton und stetig auf $\mathbb{R}$ fortsetzbar (Übung).

Sei $\lambda > 0$. Nach Voraussetzung ist $\lambda X_i \leq \lambda b = \lambda$. Also gilt $\forall i = 1, ..., n$

$$\begin{aligned} & e^{\lambda X_i} - \lambda X_i - 1 \leq X_i^2(e^\lambda - \lambda - 1), \\ \Rightarrow \quad & E(e^{\lambda X_i}) \leq 1 + \lambda E(X_i) + E(X_i^2)\phi(\lambda). \end{aligned}$$

Mit Aufsummieren und (2.21) folgt:

$$\psi_S(\lambda) \overset{(2.21)}{=} \sum_{i=1}^{n} \left(\ln(E(e^{\lambda X_i})) - \lambda E(X_i) \right)$$

$$\leq \frac{n}{n} \sum_{i=1}^{n} \left(\ln(1 + \lambda E(X_i) + E(X_i^2)\phi(\lambda)) - \lambda E(X_i) \right)$$

(2.24)

$$\overset{konkav}{\leq} n \left(\ln \left(1 + \lambda \frac{\sum_{i=1}^{n} E(X_i)}{n} + \underbrace{\frac{\sum_{i=1}^{n} E(X_i^2)}{n}}_{=\nu/n} \phi(\lambda) \right) - \lambda \frac{\sum_{i=1}^{n} E(X_i)}{n} \right)$$

$$\overset{\ln(x+1)\leq x}{\leq} \nu\phi(\lambda).$$

Um die andere in (2.23) behauptete Schranke zu erhalten, schätzen wir (2.24) ab mit Hilfe von:

$$\text{Für } x \geq 1, a > 0 : 1 + \frac{a}{x} \leq 1 + a \leq e^a$$

$$\Rightarrow \text{ Für } a, b > 0 : \ \ln(1 + a + b) - a \leq \ln(1 + b).$$

In Kapitel 2.2 wurde bereits gesehen:

$$\nu\phi(\lambda) \text{ ist KEF von } Y = X - E(X) \text{ für } X \sim Poi(\nu)$$

Aus der Ungleichung zwischen zwei KEFs ergibt sich eine zwischen den entsprechnden Cramer-Transformierten:

$$\psi_S^*(t) \geq \psi_Y^*(t) = \nu h\left(\frac{t}{\nu}\right) \overset{\text{Chernoff}}{\Rightarrow} P(S \geq t) \leq \exp\left(-\psi_S^*(t)\right) \leq \exp\left(-\nu h\left(\frac{t}{\nu}\right)\right)$$

□

Bemerkung 2.21. In einer Übungsaufgabe wird gezeigt

$$h(u) = (1 + u)\ln(1 + u) - u \geq \frac{u^2}{2\left(1 + \frac{u}{3}\right)}$$

$$\overset{\text{Bennett}}{\Rightarrow} \underbrace{P(S \geq t) \leq \exp\left(-\frac{t^2}{2\left(\nu + \frac{bt}{3}\right)}\right)}_{\text{Bernstein-Ungleichung}}$$

Das ist die *Bernstein-Ungleichung*. Für $\nu \gg tb$ sind Bennett und Bernstein im Wesentlichen äquivalent. Für $t \gg \frac{\nu}{b}$ verliert Bernstein gegenüber Bennett einen $\ln(t)$-Faktor im Exponenten.
→ Bernstein lässt sich aber allgemeiner beweisen (für schwächere Annahmen an ZVen)!

2.8. **Bernstein-Ungleichung.**

Satz 2.22 (Bernstein-Ungleichung)**.** *Seien $X_1, ..., X_n$ unabhängige ZVen mit $\nu, c > 0$, sodass $\sum_{i=1}^n E(X_i^2) \leq \nu$ und $\sum_{i=1}^n E((X_i)_+)^q \leq \frac{q!}{2}\nu c^{q-2}$ für alle $q \in \mathbb{N}_{\geq 3}$. Dann gilt $\forall\, \lambda \in \left(0, \frac{1}{c}\right)$, $\forall\, t > 0$:*

$$\psi_S(\lambda) \leq \frac{\nu\lambda^2}{2(1-c\lambda)}, \quad \psi_S^*(t) \geq \frac{\nu}{c^2} h_1\left(\frac{ct}{\nu}\right),$$

wobei $h_1(u) = 1 + u - \sqrt{1+2u}$, $u > 0$. Insbesondere gilt $\forall\, t > 0$:

$$P(S \geq \sqrt{2\nu t} + ct) \leq e^{-t}.$$

Die Bernstein-Ungleichung wird in Kapitel 2.9 beim Johnson-Lindenstrauß-Lemma verwendet.

Beweis. Im Beweis der Bennett-Ungleichung wurde $\phi(u) := e^u - u - 1$ definiert. Es gilt $\phi(u) \leq \frac{u^2}{2}$ für $u \leq 0$ (Übung). Also folgt für $\lambda > 0, i = 1, ..., n$:

$$\begin{aligned}
\phi(\lambda X_i) &\leq \begin{cases} \sum_{q=3}^\infty \dfrac{\lambda^q (X_i)^q}{q!}, & \text{falls } X \geq 0 \\ \dfrac{\lambda^2 (X_i)^2}{2}, & \text{falls } X \leq 0. \end{cases} \\
&\leq \frac{\lambda^2 (X_i)^2}{2} + \sum_{q=3}^\infty \frac{\lambda^q (X_i)_+^q}{q!}
\end{aligned}$$

$$\Rightarrow \quad E(\phi(\lambda X_i)) \leq \frac{\lambda^2 E(X_i^2)}{2} + \sum_{q=3}^\infty \frac{\lambda^q E((X_i)_+^q)}{q!}$$

$$\Rightarrow \quad \sum_{i=1}^n E(\phi(\lambda X_i)) \leq \frac{\nu}{2} \sum_{q=2}^\infty \lambda^q c^{q-2}. \qquad \text{nach Vorauss.}$$

Letzte Summe ist endlich, falls $\lambda \in \left(0, \frac{1}{c}\right)$. Mit (2.22) ergibt sich

$$\psi_S(\lambda) \leq \sum_{i=1}^n E(\phi(\lambda X_i)) \leq \frac{\nu}{2} \sum_{q=2}^\infty \lambda^q c^{q-2} = \frac{\nu\lambda^2}{2(1-c\lambda)}.$$

(Der letzte Schritt beleuchtet, warum der Bruch ein Bezugswert ist.) Damit folgt mit (2.15) insgesamt

$$\psi_S^*(t) \geq \sup_{\lambda \in \left(0, \frac{1}{c}\right)} \left(t\lambda - \frac{\lambda^2 \nu}{2(1-c\lambda)}\right) = \frac{\nu}{c^2} h_1\left(\frac{ct}{\nu}\right)$$

Die Abschätzung an $P(S \geq t)$ folgt mit der Bemerkung über h_1^* vor Satz 2.13 in Kapitel 2.4, vgl. Übung. $\square$

$$-\cancel{\nu}\frac{t^2}{2(\nu^{\cancel{2}}+ct\cancel{\nu})}$$

Korollar 2.23. Bernstein

Seien $X_1, ..., X_n$ unabhängige ZVen wie in Theorem 2.22. Dann gilt $\forall\, t > 0$:

$$P(S \geq t) \leq \exp\left(-\frac{t^2}{2(\nu + ct)}\right)$$

Beweis. Direkte Folgerung aus $h_1(u) \geq \frac{u^2}{2(1+u)}$ (Übungsaufgabe). $\square$

$$P(S \geq t) \leq e^{-\psi_S^*(t)} = \exp\left(-\frac{\nu}{c^2} h_1\left(\frac{ct}{\nu}\right)\right) \leq e^{-\frac{\nu}{\cancel{c^2}}\frac{\cancel{c^2}t^2}{\nu^2 2(1+\frac{ct}{\nu})}}$$

Man kann zeigen, dass dies für $X_i \leq b$ die gleiche Schranke liefert, wie am Ende von Kapitel 2.7 aus Bennett hergeleitet wurde (Übungsaufgabe).

Beispiel 2.24 (Gaußsches Chaos der Ordnung Zwei).
Sei $X = (X_1, ..., X_n)$ mit unabhängigen Komponenten $X_j \sim \mathcal{N}(0, 1)$, also $X \sim \mathcal{N}(0, Id_n)$. Sei $A = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ symmetrisch mit $a_{jj} = 0$ für $j = 1, ..., n$, insbesondere $\text{Spur}(A) = \sum_{j=1}^n a_{jj} = 0$.
Definiere ZVe Z als quadratische Form: $Z = X^\top AX = \sum_{i,j=1}^n X_i a_{ij} X_j$.

$$\Rightarrow \quad E(Z) = \sum_{i,j=1, i\neq j}^n a_{ij} \underbrace{E(X_iX_j)}_{E(X_i)E(X_j)=0} + \sum_{i=1}^n \underbrace{a_{ii}}_{=0} E(X_iX_i) = 0.$$

<u>Frage:</u>
Wie stark schwankt die ZVe Z um den Mittelwert 0? ($\hat{=}$ Konzentration).

Da A symmetrisch, existiert $B \in \mathbb{R}^{n\times n}$ orthogonal mit $A = B^\top DB$, wobei

$$D = \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mu_4 \end{pmatrix}$$ Diagonalmatrix mit Eigenwerten $\mu_1, ..., \mu_n$ von A.

Sei $B = (b_{ij})_{i,j=1}^n$. Setze $Y_i = \sum_{j=1}^n b_{ij} X_j$ für $i = 1, ..., n$.

$$\Rightarrow \quad \sum_{i=1}^n \mu_i Y_i^2 = \sum_{i=1}^n \mu_i (\sum_{j=1}^n b_{ij} X_j)^2 = \sum_{i=1}^n \mu_i (BX)_i^2$$
$$= \langle X, B^\top DBX\rangle = X^\top AX = Z.$$

Ist diese andere Darstellung von Z über Y_i besser?
Da $X \sim \mathcal{N}(0, Id_n)$ und B orthogonal, ist auch $Y = (Y_1, ..., Y_n) \sim \mathcal{N}(0, Id_n)$

wegen Rotationsinvarianz.

$$\Rightarrow \quad P_X = P_Y,\ P_{(X_1^2,\dots,X_n^2)} = P_{(Y_1^2,\dots,Y_n^2)}.$$
$$\Rightarrow \quad P_Z = P_{\sum_{i=1}^n \mu_i Y_i^2} = P_{\sum_{i=1}^n \mu_i X_i^2}.$$

Da $0 = \operatorname{Spur}(A) = \sum_{i=1}^n \mu_i$ gilt:

$$\sum_{i=1}^n \mu_i X_i^2 = \sum_{i=1}^n \mu_i (X_i^2 - 1).$$

Eigenschaft von X_i^2 : $E(X_i^2) = \operatorname{Var}(X_i) = 1 \quad \Rightarrow \quad X_i^2 - 1$ zentriert
$X_i^2 \sim \chi^2$-verteilt mit einem Freiheitsgrad, also insbesondere Γ-verteilt mit Parameter $a = \frac{1}{2}$ und $b = 2$. Nach Bsp. 2.17in Kapitel 2.4 gilt:

$$\begin{aligned}\psi_{X_i^2-1}(\lambda) &= \frac{1}{2}\left[-\ln(1-2\lambda) - 2\lambda\right] \\ &\le 2\frac{\lambda^2}{2(1-2\lambda)} = \frac{\lambda^2}{1-2\lambda}.\end{aligned}$$

Also folgt für die KEF ψ_Z von Z ähnlich wie (2.8) (in Kapitel 2.6) wegen Unabhängigkeit:

$$\begin{aligned}\psi_Z(\lambda) = \psi_{\sum_{i=1}^n \mu_i(X_i^2-1)}(\lambda) &= \sum_{i=1}^n \psi_{\mu_i(X_i^2-1)}(\lambda) \\ &= \sum_{i=1}^n \psi_{X_i^2-1}(\mu_i\lambda) = \frac{1}{2}\sum_{i=1}^n(-\ln(1-2\mu_i\lambda) - 2\mu_i\lambda) \\ &\le \sum_{i=1}^n \frac{\mu_i^2\lambda^2}{1-2\mu_i\lambda}, \quad \text{sofern } \lambda \in (0, (2\max_{i=1,\dots,n}\mu_i)^{-1}) =: J\end{aligned}$$

Nun gilt $\forall\, j = 1, \dots, n$

Länge des abg. Vektors

$$\mu_i \le |\mu_i| \le \|A\| = \underbrace{\sup_{\|x\|=1} \|Ax\|}.$$

längste Streckungsf.

Sei $\|A\|_{\mathrm{HS}} := \sqrt{\sum_{i=1}^n \mu_i^2}$ die *Hilbert-Schmidt-* oder *Frobenius-Norm* von A. Dann folgt für $\lambda \in J$:

$$\psi_Z(\lambda) \le \sum_{i=1}^n \frac{\mu_i^2\lambda^2}{1-2\lambda\|A\|} = \frac{\lambda^2\|A\|_{\mathrm{HS}}^2}{1-2\lambda\|A\|}.$$

In der Notation der sub-Γ-Verteilungen ist $\nu = 2\|A\|_{\text{HS}}^2$, $c = 2\|A\|$ und nach Kapitel 2.4 gilt:

$$P(Z > 2\|A\|_{\text{HS}}\sqrt{t} + 2\|A\|t) \le e^{-t} \text{ (vgl. Satz 2.13), oder}$$

$$P(Z > t) < \exp\left(-\frac{\|A\|_{\text{HS}}^2}{2\|A\|} h_1\left(\frac{\|A\|t}{\|A\|_{\text{HS}}^2}\right)\right) \text{ oder}$$

$$P(Z > t) \le \exp\left(-\frac{t^2}{4(\|A\|_{\text{HS}}^2 + t\|A\|)}\right) \text{ Übung wie in Kor. 2.23,}$$

wobei h_1 die schon bekannte Entropie-Funktion ist.

2.9. **Johnson-Lindenstrauss-Lemma.**

Folgende Ausführungen werden in einem viel allgemeineren Rahmen in Kapitel 5.6 ausgeführt.

Definition 2.25. Sei $\mathcal{H}$ ein separabler Hilbertraum, z.B. $\mathcal{H} = \mathbb{R}^D$ mit Dimension $D \in \mathbb{N}$. Für gegebenes $\varepsilon \in (0,1)$ und $A \subset \mathcal{H}$ heißt $f\colon \mathcal{H} \to \mathbb{R}^d$ *ε-Isometrie auf A*, falls für alle $a, b \in A$ gilt:

$$(1-\varepsilon)\|a-b\|_{\mathcal{H}} \le \|f(a) - f(b)\|_{\mathbb{R}^d} \le (1+\varepsilon)\|a-b\|_{\mathcal{H}} \tag{2.25}$$

Falls $\mathcal{H}$ hochdimensional, A große Kardinalität hat und $\varepsilon \in (0,1)$ und $d \in \mathbb{N}$ klein sind, ist völlig unklar, ob ein solches f existiert.

Das Johnson-Lindenstrauss-Lemma besagt, dass es eine universelle Konstante κ gibt, sodass für jedes A mit Kardinalität $\text{card}(A) = n < \infty$ und

$$d \ge \frac{\kappa}{\varepsilon^2}\ln(n)$$

tatsächlich ein $f\colon \mathcal{H} \to \mathbb{R}^d$ mit Eigenschaft (2.25) existiert.

Allerdings kann man dieses f nicht explizit angeben, da es mit einem *Zufallsmechanismus* konstruiert wird. (Vergleiche die *probabilistic method* von Paul Erdős in der Kombinatorik und Graphentheorie.) Dafür ist es möglich f, linear zu wählen. Wir zeigen sogar, dass aus einem vorgegebenen Ebsemble linearer Funktionen $f\colon \mathcal{H} \to \mathbb{R}^d$, die meisten (2.25) erfüllen.
Einfachheitshalber nehmen wir $\mathcal{H} = \mathbb{R}^D$ und typischerweise $D \gg d$ an.
<u>Ansatz:</u> Sei $W : \mathbb{R}^D \to \mathbb{R}^d$ linear und zufällig gewählt mit der Eigenschaft

$$\forall\, \alpha \in \mathbb{R}^D : \; E\left(\|W\alpha\|^2\right) = \|\alpha\|^2,$$

d.h. im Mittel haben wir eine exakte Isometrie. Gewünscht ist die Eigenschaft, dass die Zufallsvariable $\|W\alpha\|^2$ wenig streut.

Konstruktion von W

Seien $X_{ij}, i = 1, ..., d$ und $j = 1, ..., D$ unabhängig identisch verteilte Zufallsvariablen. Ferner seien $E(X_{ij}) = 0$ und $\mathrm{Var}(X_{ij}) = 1$. Für $\alpha = (\alpha_1, ..., \alpha_D) \in \mathbb{R}^D$ und $i = 1, ..., d$ setze

$$\tilde{W}_i(\alpha) = \sum_{j=1}^{D} \alpha_j X_{ij} \text{ und } W_\alpha = \left(\frac{1}{\sqrt{d}} \tilde{W}_i(\alpha) \right)_{i=1}^{d} . \tag{2.26}$$

Wegen der Unabhängigkeit gilt für für jedes $i = 1, ..., d$:

$$E\left(\tilde{W}_i(\alpha)^2\right) = E\left(\left(\sum_{j=1}^{D} \alpha_j X_{ij}\right)^2\right) = \sum_{i,j=1}^{D} \alpha_i \alpha_j E(X_{ij}^2)\delta_{ij}, = \sum_{i=1}^{D} \alpha_i^2 \cdot 1$$

wobei δ_{ij} das Kronecker-Delta ist. Daraus folgt für alle $\alpha \in \mathbb{R}^D$

$$E\left(\|W_\alpha\|^2\right) = E\left(\frac{1}{d} \sum_{i=1}^{d} (\tilde{W}_i(\alpha))^2\right) = \frac{1}{d} \sum_{i=1}^{d} \|\alpha\|^2 = \|\alpha\|^2.$$

Wir wollen solche Zufallsvariablen X_{ij} wählen, dass ihr Verhalten meistens nahe am Mittelwert ist. Das ist eine typische Konzentrationsbedingung, wie sie z.B. für sub-gaußsche Zufallsvariablen erfüllt ist.

Satz 2.26 (Johnson-Lindenstrauss-Lemma)**.**
Seien $A \subset \mathbb{R}^D$ mit $card(A) = n \in \mathbb{N}$, sowie $\varepsilon, \delta \in (0,1)$. Für $\nu \geq 1$ seien $X_{ij} \in \mathcal{G}(\nu), i = 1, ..., d$ und $j = 1, ..., D$ unabhängig identisch verteilt und W wie in (2.26).

$$\text{Sobald } d \geq 100 \cdot \frac{\nu^2}{\varepsilon^2} \ln\left(\frac{n}{\sqrt{\delta}}\right) \text{ gilt:}$$
$$P(W \text{ ist eine } \varepsilon\text{-Isometrie auf } A) \geq 1 - \delta.$$

Beweis. Sei $S \subset \mathbb{R}^d$ die Einheitssphäre und $T \subset S$ definiert durch

$$T = \left\{ \frac{a-b}{\|a-b\|} \mid a, b \in A, a \neq b \right\}$$

Wir wollen beweisen, dass mit hoher Wahrscheinlichkeit gilt:

$$\max_{\alpha \in T} \left| \|W\alpha\|^2 - 1 \right| \leq \varepsilon.$$

Denn dann folgt wegen Linearität $\forall\, a, b \in A$:

$$\left| \|Wa - Wb\|^2 - \|a-b\|^2 \right| \leq \varepsilon \|a-b\|^2$$
$$\Rightarrow \ (1-\varepsilon)\|a-b\|^2 \leq \|Wa - Wb\|^2 \leq (1+\varepsilon)\|a-b\|^2.$$

Für jedes $\alpha \in S$ und $i \leq d$ gilt:

$$
\begin{aligned}
E(\exp(\lambda \tilde{W}_i(\alpha))) &= E(\exp(\lambda \sum_{j=1}^{D} \alpha_j X_{ij})) \\
&= \prod_{j=1}^{D} E(\exp(\lambda \alpha_j X_{ij})) \leq \prod_{j=1}^{D} \exp\left(\frac{\lambda^2 \alpha_j^2 \nu}{2}\right) \\
&= \exp\left(\frac{\lambda^2}{2} \nu \sum_{j=1}^{D} \alpha_j^2\right) = \exp\left(\frac{\lambda^2 \nu \|\alpha\|^2}{2}\right) = \exp\left(\frac{\lambda^2 \nu}{2}\right),
\end{aligned}
$$

also $\tilde{W}_i(\alpha) \in \mathcal{G}(\nu)$.
Insbesondere impliziert Satz 2.9 $\forall\, k \in \mathbb{N}_{\geq 2}$: $E(\tilde{W}_i(\alpha)^{2k}) \leq \frac{k!}{2}(4\nu)^k$.
Weiterhin sind alle Komponenten $\tilde{W}_1(\alpha), ..., \tilde{W}_d(\alpha)$ unabhängig $\forall\, \alpha \in S$.
Damit können wir die Bernstein-Ungleichung 2.22 anwenden:

$$
\Rightarrow \forall \alpha \in T, t > 0 : P\left(\left|\sum_{i=1}^{d} (\tilde{W}_i(\alpha)^2 - 1)\right| \geq 4\nu\sqrt{2dt} + 4\nu t\right) \leq 2e^{-t}.
$$

$\delta = n^2 e^{-t}$

subaddit.

$\frac{e^t}{n^2} = \frac{1}{\delta}$

$$
\begin{aligned}
\Rightarrow\quad & P\left(\max_{\alpha \in T}\left|\sum_{i=1}^{d} (\tilde{W}_i(\alpha)^2 - 1)\right| \geq 4\nu\sqrt{2dt} + 4\nu t\right) \\
&= P\left(\bigcup_{\alpha \in T}\left\{\left|\sum_{i=1}^{d} (\tilde{W}_i(\alpha)^2 - 1)\right| \geq 4\nu\sqrt{2dt} + 4\nu t\right\}\right) \\
&\leq \sum_{\alpha \in T} P\left(\left|\sum_{i=1}^{d} (\tilde{W}_i(\alpha)^2 - 1)\right| \geq 4\nu\sqrt{2dt} + 4\nu t\right) \\
&\leq 2e^{-t}|T| \leq 2e^{-t} n^2
\end{aligned}
$$

Wähle $\boxed{t = \ln\left(\frac{n^2}{\delta}\right)}$, dann: $= 2 \ln\left(n / \sqrt{\delta}\right)$

$\frac{1}{d}$

$$
P\left(\max_{\alpha \in T}\left|\sum_{i=1}^{d} (\tilde{W}_i(\alpha)^2 - \underset{d}{1})\right| \geq 8\nu\sqrt{d \ln\left(\frac{n}{\sqrt{8}}\right)} + 8\nu \ln\left(\frac{n}{\sqrt{\delta}}\right)\right) \leq \delta \qquad 2
$$

$\sqrt{d}$

$$
\Leftrightarrow P\left(\max_{\alpha \in T}\left|\|W\alpha\|^2 - 1\right| \geq 8\nu\left(\sqrt{\frac{\ln(n/\sqrt{\delta})}{d}} + \frac{\ln(n/\delta)}{d}\right)\right) \leq \delta \qquad 2
$$

Falls $d \geq 100\frac{\nu^2}{\varepsilon^2}\ln\left(\frac{n}{\sqrt{\delta}}\right)$ gewählt wird, folgt

$$8\nu\left[\sqrt{\frac{\ln(n/\sqrt{\delta})}{d}}+\frac{\ln(n/\sqrt{\delta})}{d}\right] \leq \frac{4\varepsilon}{5}+\frac{2\varepsilon^2}{25\nu} \leq \frac{20\varepsilon}{25}+\frac{2\varepsilon}{25} \leq \varepsilon,$$

da nach Annahme $\nu \geq 1$. Also haben wir tatsächlich gezeigt:

$$P\left(\sup_{\alpha\in T}\left|\|W\alpha\|^2-1\right| \leq \varepsilon\right) \geq 1-\delta \, 2$$

$\square$

Übung: Welche Konstanten verbessern sich, falls $X_{ij} \sim \mathcal{N}(0,1)$?

2.10. **Assoziations- und Korrelationsungleichungen.**
Dieser Abschnitt wird bei der Janson-Ungleichung und der Perkolationstheorie (optional) angewendet. Für das sonstige weitere Verständnis kann dieses Kapitel übersprungen werden. Since X, Y unabh. ZV, gelten praktische Rechenregeln für Produkte von X und Y oder von Funktionen davon. Unter welchen Annahmen kann man etwas von diesen Rechenregeln ‚retten', falls X und Y nicht unabh. sind?

Satz 2.27 (Čebyšev-Assoziationsungleichung)**.**
Seien $f, g : \mathbb{R} \to \mathbb{R}$ isoton und ZVen $X \in \mathbb{R}$ und $Y \in \mathbb{R}_+$. Dann gilt auf $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ $f(X), g(X) \in \mathcal{L}^2(P) \Leftrightarrow f, g \in \mathcal{L}^2(P_X)$

$$E(Y)E(Yf(X)g(X)) \geq E(Yf(X))E(Yg(X))$$

Bemerkung: Sei $Y \equiv 1$. Dann impliziert Satz 2.27:

$$\underbrace{E(f(X)g(X)) \geq E(f(X))E(g(X))}_{=} \Leftrightarrow E(f(X)g(X)) - E(f(X))E(g(X)) \geq 0$$

$$\overset{f,g\in\mathcal{L}^2}{\Leftrightarrow} \mathrm{Cov}(f(X), g(X)) \geq 0.$$

$f(X)$, $g(X)$ unabh

Ist f antiton und g weiterhin isoton, gilt
~~Verändern wir die Annahme an f zu monoton fallend, gilt~~

$$E(Y)E(Yf(X)g(X)) \leq E(Yf(X))E(Yg(X)).$$

Beweis. Sei der Vektor $(X', Y') : \Omega \to \mathbb{R} \times \mathbb{R}_+$ unabhängige Kopie von (X, Y) mit $P_{(X,Y)} = P_{(X',Y')}$. Sind f, g monoton wachsend, so gilt:

$$
\begin{aligned}
& (f(X) - f(X'))(g(X) - g(X')) \geq 0 \\
\Rightarrow\ & YY'(f(X) - f(X'))(g(X) - g(X')) \geq 0 \\
\Rightarrow\ & E(YY'(f(X) - f(X'))(g(X) - g(X'))) \geq 0 \\
\overset{\text{Linearität}}{\Leftrightarrow}\ & E(YY'f(X)g(X) + YY'f(X')g(X')) \\
& \quad \geq E(YY'f(X)g(X') + YY'f(X')g(X)) \\
\overset{\text{unabh.}}{\Leftrightarrow}\ & E(Y')E(Yf(X)g(X)) + E(Y)E(Y'f(X')g(X')) \\
& \quad \geq E(Yf(X))E(Y'g(X')) + E(Yg(X))E(Y'f(X')) \\
\overset{\text{id. vert.}}{\Leftrightarrow}\ & 2E(Y)E(Yf(X)g(X)) \geq 2E(Yf(X))E(Yg(X)).
\end{aligned}
$$

Ausmultiplizieren.

□

Der Beweis illustriert die Methode der *unabhängigen Kopie.* Führe dazu unabhängig, identisch verteilte ZVen ein und im zweiten Schritt durch Erwartungswerte auf die ursprünglichen ZVen zurück.
Es gibt auch eine multivariate Variante dieser Ungleichung:
Dazu:

Definition 2.28. $f : \mathbb{R}^n \to \mathbb{R}$ heißt *isoton* (oder n-*isoton*) $:\Leftrightarrow$

$\forall\, i \in \{1, ..., n\}$ und $(y_1, ..., y_{i-1}, y_{i+1}, ..., y_n) \in \mathbb{R}^{n-1}$ ist

$\mathbb{R} \ni x \mapsto f(y_1, ..., y_{i-1}, x, y_{i+1}, ..., y_n)$ monoton wachsend bzw. isoton,

d.h. f isoton in jeder Koordinate
f heißt *antiton* (oder n-*antiton*), falls $-f$ isoton ist.

Satz 2.29 (Harris-Ungleichung)**.**
Seien $f, g : \mathbb{R}^n \to \mathbb{R}$ n-isoton und unabhängige ZVen $X_1, ..., X_n \in \mathbb{R}$. Setze $X = (X_1, ..., X_n) \in \mathbb{R}^n$. Dann gilt

$$E(f(X)g(X)) \geq E(f(X))E(g(X)). \tag{2.27}$$

Bemerkung 2.30 (FKG-Ungleichung)*.* Die Aussage gilt auch im Fall $n = \infty$, d.h. falls $X_k, k \in \mathbb{N}$, eine Folge unabhängiger ZV, $f, g : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ isoton in jedem Argument sind und $f(X), g(X)$ endliche Varianz besitzen. Diese Aussage wird mit Hilfe eines Martingal-KOnverngezsatzes bewiesen.

(2.27) gilt.

Bemerkung 2.31. Da die X_i unabhängig sind, integrieren wir bezüglich eines Produktmaßes. Dennoch ist es in solchen Situationen sinnvoll, bedingte

$$E(f(X)) = \int_{\mathbb{R}^3} f(x_1, x_2, x_3) \; P_X(dx_1, dx_2, dx_3)$$

Erwartungen zu benutzen, um Schreibarbeit zu sparen.

Illustration für $n = 3$: Sei $f(X_1, X_2, X_3) = f(X)$. Nach dem Trafo-Satz gilt:

$$\begin{aligned}
E(f(X) \mid X_1) &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(X_1, t, s) \mathrm{d}P_{X_2}(t) \mathrm{d}P_{X_3}(s) \\
E(f(X) \mid X_1, X_2) &= \int_{\mathbb{R}} f(X_1, X_2, s) \mathrm{d}P_{X_3}(s) \\
E(E(f(X) \mid X_1, X_2)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(r, t, s) \mathrm{d}P_{X_3}(s) \right) \mathrm{d}P_{X_2}(t) \mathrm{d}P_{X_1}(r) \\
&= E(f(X)),
\end{aligned}$$

was sich auch aus der Turmeigenschaft für bedingte Erwartungen herleiten lässt.

(IV)

Beweis. Zeige (2.27) in Satz 2.29 durch vollständige Induktion.
Fall $n = 1$ folgt aus Bemerkung zu Satz 2.27. Nehme an, dass (2.27) bekannt ist für alle $n < k$. Für das 1-dim. Integral bezüglich P_{X_k} gilt:

$$E(f(X)g(X) \mid X_1, ..., X_{k-1}) \geq E(f(X) \mid X_1, ..., X_{k-1}) E(g(X) \mid X_1, ..., X_{k-1}).$$

wegen Satz 2.27, denn bei eingefrorenen $X_1, ..., X_{k-1}$ sind f und g isotone Funktionen im $k$ten Argument. Turmeigenschaft und Monotonie des Integrals ergeben:

$$\begin{aligned}
E(f(X)g(X)) &= E(E(f(X)g(X) \mid X_1, ..., X_{k-1})) \\
&\geq E\left[E(f(X) \mid X_1, ..., X_{k-1}) E(g(X) \mid X_1, ..., X_{k-1})\right]
\end{aligned}$$

Nun ist

$$\begin{aligned}
f_{k-1} \colon (X_1, ..., X_{k-1}) &\mapsto E(f(X) \mid X_1, ..., X_{k-1}) \\
&\overset{\text{unabh.}}{=} \int_{\mathbb{R}} f(X_1, ..., X_{k-1}, t) \mathrm{d}P_{X_k}(t) \quad (k-1)\text{-isoton,}
\end{aligned}$$

ebenso wie

$$g_{k-1} \colon \; (X_1, ..., X_k) \mapsto E(g(X) \mid X_1, ..., X_{k-1}).$$

Wende Induktionsvorraussetzung (IV) an und erhalte:

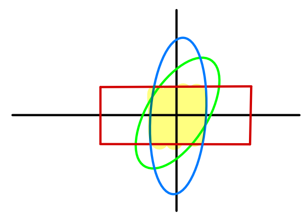
$$
\begin{aligned}
&E\left[\underbrace{E(f(X) \mid X_1, ..., X_{k-1})}_{= f_{k-1}(\dots)} \underbrace{E(g(X) \mid X_1, ..., X_{k-1})}_{g_{k-1}(\dots)}\right] \\
&\overset{\text{(IV)}}{\geq} E[E(f(X) \mid X_1, ..., X_{k-1})]E[E(g(X) \mid X_1, ..., X_{k-1})] \\
&\overset{\text{Turmeig.}}{=} E(f(X))E(g(X)).
\end{aligned}
$$

□

Bemerkung 2.32 (Frage/Überlegung:). Funktionieren solche Aussage auch für andere Klassen von Funktionen? Z.B.:

- sphärisch-symmetrische,
- radial-abfallende

Funktionen? Betrachte dazu folgenden Satz.



Satz 2.33 (Gaußsche-Korrelationsungleichung)**.**
Sei $P = \mathcal{N}(0, C)$, mit C positiv definit, ein zentriertes Gaußmaß auf $\mathbb{R}^d$ und A, B zwei (bzgl. Spiegelung am Ursprung) symmetrische[2]*, konvexe Mengen.*

$x \in A \Rightarrow -x \in A$

$$\Rightarrow \ P(A \cap B) \geq P(A)P(B).$$

Falls $P(A) > 0$, gilt das besser interpretierbare

$$P(B \mid A) \geq P(B)$$

Diesen Satz gibt es als Vermutung seit den 60ern motiviert durch zwei Arbeiten in 1955 und 1959. Beweis und Veröffentlichung von Thomas Royen im Jahr 2014. Seit 2017 gibt es weiteren, strukturierteren Beweis von Latala et. al..

2.11. **Anwendung der Harris-Ungleichung: Janson-Ugleichung.** Wir betrachten folgende kombinatorische Situation, die man in der Theoretischen Informatik, der probabilistischen Methode in der Kombinatorik und dem Studium von zufälligen Graphen antrifft.
Seien $n \in \mathbb{N}$, $[n] := \{1, \dots, n\}$, $X_1, \dots, X_n \in \{0, 1\}$ unabh. ZV mit

$$p_k := P(X_k = 1) = 1 - P(X_k = 0) \text{ für } k \in [n] \text{ und } p_1, \dots, p_n \in [0, 1]$$

[2] *d.h. $x \in A \Rightarrow -x \in A$*

$$1_{M_1} + 1_{M_2} + 1_{M_3} = 1_{M_1 \cup M_2 \cup M_3}$$ disjunkt.

Für $A \subset [n]$ setzte

$$Y_A = \prod_{i \in A} X_i$$

und $\mathcal{I} \subset \mathcal{P}([n])$

$$Z := \sum_{A \in \mathcal{I}} Y_A$$

was ein Polynom in den 0/1-Variablen $X_1, \ldots, X_n$ ist.

Bemerkung 2.34 (Übung). Seien $f, g \colon \mathbb{R}^n \to \mathbb{R}$ zwei n-antitone Funktionen. Leiten Sie aus der Harris-Ungleichung

$$E(f(X)g(X)|Y_A = 1) \geq E(f(X)|Y_A = 1)E(g(X)|Y_A = 1)$$

her. Hierbei ist $X = (X_1, \ldots, X_n)$. Elementar z.z., da $|\Omega| = 2^n$

Offensichtlich gilt für $A, B \in \mathcal{I}$ mit $A \cap B = \emptyset$ Y_A, Y_B unabh.

$$E(Y_A Y_B) = E(\prod_{i \in A} X_i \prod_{k \in B} X_k) = \prod_{i \in A} \prod_{k \in B} E(X_i)E(X_k) = E(Y_A)E(Y_B)$$

Daher reduziert sich die

$$\mathrm{Var}(Z) = E(Z^2) - E(Z)^2 = \sum_{A,B \in \mathcal{I}} \left(E(Y_A Y_B) - \cancel{\sum_{A,B \in \mathcal{I}}} E(Y_A)E(Y_B) \right)$$

(2.28)

$$= \sum_{A,B \in \mathcal{I}, A \cap B \neq \emptyset} [E(Y_A Y_B) - \underbrace{E(Y_A)E(Y_B)]}_{\geq 0} \leq \sum_{A,B \in \mathcal{I}, A \cap B \neq \emptyset} E(Y_A Y_B) \quad =: \Delta$$

Čebyšev liefert die symmetrische Schranke

$$P(|Z - EZ| > t) \leq \frac{\Delta}{t}$$

Auch wenn die Summanden von Z nicht unabhängig sind, gibt es eine (zumindest einseitige) exponentielle *tail*-Schranke.

Satz 2.35. *Seien $\mathcal{I} \subset \mathcal{P}([n])$ sowie Z und Δ wie soeben definiert. Dann gilt für alle $\lambda \leq 0$*

$$\psi_{Z-EZ}(\lambda) \leq \phi\left(\frac{\lambda \Delta}{EZ}\right) \frac{(EZ)^2}{\Delta}, \qquad \textit{mit } \phi(x) = e^x - x - 1$$

Insbesondere gilt für $t \in [0, EZ]$

$$P(Z - EZ < -t) \leq \exp\left(\frac{-t^2}{2\Delta}\right)$$

Beweis. Für die KEF von $Z - EZ$ gilt

$$\psi'(\lambda) = \frac{E(Ze^{\lambda Z})}{E(e^{\lambda Z})} - E(Z) = \sum_{A \in \mathcal{I}} \frac{E(Y_A e^{\lambda Z})}{E(e^{\lambda Z})} - E(Z)$$

Wie wollen jeden A-Summanden auf der rechten Seite einzeln abschätzen. Zu jedem $A \in \mathcal{I}$ setze

$$\begin{aligned} U_A &= \sum_{B \in \mathcal{I}, A \cap B \neq \emptyset} Y_B \\ Z_A &= \sum_{B \in \mathcal{I}, A \cap B = \emptyset} Y_B, \end{aligned}$$

so dass $Z = U_A + Z_A \geq Z_A$ für jedes $A \in \mathcal{I}$. Wegen der Fallunterscheidungsformel

$$E(Y_A e^{\lambda Z}) = E(Y_A e^{\lambda Z} | Y_A = 1) P(Y_A = 1) + 0 = E(e^{\lambda Z} | Y_A = 1) E(Y_A)$$

reicht es, die bedingte Erwartung von oben abzuschätzen. Da λ negativ ist, sind $X \mapsto e^{\lambda U_A}$ und $X \mapsto e^{\lambda Z_A}$ antitone Funktionen. Es folgt

$$\begin{aligned} E(e^{\lambda Z} | Y_A = 1) &= E(e^{\lambda U_A} e^{\lambda Z_A} | Y_A = 1) \\ \text{(Harris)} \quad &\geq E(e^{\lambda U_A} | Y_A = 1) E(e^{\lambda Z_A} | Y_A = 1) \\ (Z_A, Y_A \text{unabhängig}) \quad &= E(e^{\lambda U_A} | Y_A = 1) E(e^{\lambda Z_A}) \\ (Z \geq Z_A) \quad &\geq E(e^{\lambda U_A} | Y_A = 1) E(e^{\lambda Z}) \\ \text{(Jensen)} \quad &\geq e^{\lambda E(U_A | Y_A = 1)} E(e^{\lambda Z}) \end{aligned} \tag{2.29}$$

Damit erhalten wir

$$\begin{aligned} \frac{E(Ze^{\lambda Z})}{E(Z)} &= \sum_{A \in \mathcal{I}} \frac{E(Y_A e^{\lambda Z})}{E(Z)} = \sum_{A \in \mathcal{I}} \frac{E(e^{\lambda Z} | Y_A = 1) E(Y_A)}{E(Z)} \\ &\geq E(e^{\lambda Z}) \underbrace{\sum_{A \in \mathcal{I}} \frac{E(Y_A)}{E(Z)}} e^{\lambda E(U_A | Y_A = 1)} \\ &\geq E(e^{\lambda Z}) \exp\left[\lambda \sum_{A \in \mathcal{I}} \frac{E(Y_A)}{E(Z)} E(U_A | Y_A = 1) \right] \end{aligned}$$

Nun wollen wir die bedingte Erwartung loswerden.

$$\begin{aligned}
\Delta &= \sum_{A,B\in\mathcal{I},A\cap B\neq\emptyset} E(Y_A Y_B) = \sum_{A\in\mathcal{I}} E(Y_A U_A) \\
&= \sum_{A\in\mathcal{I}} E(Y_A U_A|Y_A = 1)E(Y_A) + E(Y_A U_A|Y_A = 0)P(Y_A = 0) \\
&= \sum_{A\in\mathcal{I}} E(U_A|Y_A = 1)E(Y_A)
\end{aligned}$$

Umstellen ergibt

$$\frac{E(Ze^{\lambda Z})}{E(e^{\lambda Z})} \geq E(Z)\exp\left[\lambda\frac{\Delta}{E(Z)}\right]$$

Damit folgt für die Ableitung der KEF

$$\frac{E(Ze^{\lambda Z})}{E(e^{\lambda Z})} - E(Z) \geq E(Z)\left[\exp(\lambda\frac{\Delta}{E(Z)}) - 1\right]$$

und wegen $\psi(0) = 0$ für die KEF selbst :

$$\begin{aligned}
\psi(\lambda) &= \psi(0) - \int_\lambda^0 \psi'(t)\mathrm{d}t \leq -E(Z)\int_\lambda^0 (e^{t\Delta/E(Z)} - 1)\mathrm{d}t \\
&= -E(Z)\int_{\lambda\Delta/E(Z)}^0 (e^s - 1)\frac{E(Z)}{\Delta}\mathrm{d}s = \frac{(EZ)^2}{\Delta}\left[e^s - s\right]_{\lambda\Delta/E(Z)}^0 \\
&= \frac{(EZ)^2}{\Delta}\phi\left(\frac{\lambda\Delta}{E(Z)}\right) \\
&= \frac{\lambda^2}{2}\Delta
\end{aligned}$$

Somit folgt auch $P(Z - EZ \leq -t) \leq e^{-t^2/(2\nu)}$ mit $\nu = \Delta$. $\square$

Optional: Anwendung der Harris-Ungleichung: Perkolation. Gegeben:

- Gitter $\mathbb{Z}^d$
- $p \in [0,1]$
- Kante e wird mit Wahrscheinlichkeit p beibehalten bzw. mit Wahrscheinlichkeit $(1-p)$ entfernt (Perkolationswahrscheinlichkeit)
- Modell entspricht Produktmaß von Bernoulli-Verteilungen mit Parameter p bzgl. $\mathbb{Z}^d$, in Zeichen: $\mathcal{B}_p^{\otimes \mathbb{Z}^d}$

Man interessiert sich für die erzeugten Graphenstrukturen nach Entfernung der Kanten (vgl. Abbildung ??). Es entsteht ein zufälliger Untergraph auf $\mathbb{Z}^d$ (im Bild $d = 2$). Welche Eigenschaften besitzen die Cluster?

<u>Beispiel-Frage:</u> Existiert abhängig von p ein unendlicher Cluster?
$\rightarrow$ Zugang über endliche Gitter in Box $\Lambda_L \subset \mathbb{Z}^d$. Unendliche Cluster sind bei endlichen Unterstrukturen nie möglich.

<u>Strategie:</u> Gibt es einen Pfad ohne Lücken zwischen den Rändern in Λ_L?
$\rightarrow$ Studiere dieses Verhalten in Abhängigkeit der Länge L der Box Λ_L.
$\rightarrow$ Lasse die Größe der Box dann gegen unendlich laufen, um asymptotisches Perkolationsverhalten zu verstehen.

Das *0-1-Gesetz von Kolmogorov* liefert:

(1) Entweder existiert ein ∞-Cluster fast sicher oder
(2) es existiert fast sicher nicht.

$\overset{\text{Monotonie}}{\Rightarrow}$ Es existiert ein kritischer Wert $p_c \in (0,1)$, sodass

$$p > p_c \;\Rightarrow\; (1)$$
$$p < p_c \;\Rightarrow\; (2)$$

Bei $p = p_c$ unklar, was passiert!

Was hat das mit der Harris-Ungleichung zutun?
Bei der Konstruktion von Links-Rechts-Durchquerungen und Oben-Unten-Durchquerungen (vgl. Abbildung ??) spielen Ereignisse der folgenden Bauart

$A \hat{=}$ „Gibt es einen Weg von der linken zur rechten Kante?“ oder
$B \hat{=}$ „Gibt es einen Weg von der unteren zur oberen Kante?“

eine Rolle (vgl. Abbildung ??). Wie stehen die Wahrscheinlichkeiten der Ereignisse A, B zueinander?

$$\text{Klar ist: } P(A \cap B) \neq P(A)P(B), \text{ da keine Unabhängigkeit.}$$

Wegen Monotonie liefert Anwendung der Harris-Ungleichung:

$$P(A \cap B) \geq P(A)P(B).$$

2.12. **Minkowski-Ungleichung***.
Die Minkowski-Ungleichung ist unabhängig von den restlichen Kapiteln und kann bei Bedarf übersprungen werden. Bekannte Minkowski-Ungleichung: $X, Y \in \mathcal{L}^q$, dann gilt

$$E(|X+Y|^q) \leq E(|X|^q)^{\frac{1}{q}} + E(|X|^p)^{\frac{1}{p}}.$$

Satz 2.36 (Minkowski-Ungleichung)**.**
Seien $X\colon \Omega \to E, Y : \Omega \to F$ unabhängige ZVen und

$$f\colon E \times F \to \mathbb{R}, \text{ messbar und } Z = f(X,Y) : \Omega \to \mathbb{R} \quad P_X\text{-f.s. } P_Y\text{-intbar}$$

Für $q \geq 1$ gilt (auch für $q = \infty$):

$$E_X(|E_Y(Z)|^q)^{\frac{1}{q}} \leq E_Y((E_X(|Z|^q))^{\frac{1}{q}}), \text{ wobei}$$
$$E_X(Z) = E(Z \mid Y) = \int_{\mathbb{R}} f(t,Y)\mathrm{d}P_X(t) \text{ die Integration bzgl. } P_X.$$

Also ist

$$\left(\int_{\mathbb{R}}\int_{\mathbb{R}} |f(t,s)|^q \mathrm{d}P_X(t)\mathrm{d}P_Y(s)\right)^{\frac{1}{q}} \leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(t,s)|^q \mathrm{d}P_X(t)\right)^{\frac{1}{q}} \mathrm{d}P_Y(s).$$

Für den Fall $q = 1$ werden lediglich Betragsstriche reingezogen. Gibt es einen Zusammenhang zur klassischen Minkowski-Ungleichung? Sei dazu

- $F = \{1,2\}$, $P_Y(\{1\}) = P_Y(\{2\}) = \frac{1}{2}$
- $X = (X_1, X_2), f(X,1) = X_1, f(X,2) = X_2$

$$\Rightarrow \left(E_X\left(\left|\frac{1}{2}(X_1+X_2)\right|^q\right)\right)^{\frac{1}{q}} \leq \frac{1}{2}\left((E_X(|X_1|^q))^{\frac{1}{q}} + (E_X(|X_2|^q))^{\frac{1}{q}})\right)$$
$$\leq \frac{1}{2}\left(\int_{\mathbb{R}} |X_1(e)|^q \mathrm{d}P_X(e)\right)^{\frac{1}{q}} + \frac{1}{2}\left(\int_{\mathbb{R}} |X_2(e)|^q \mathrm{d}P_X(e)\right)^{\frac{1}{q}}$$

Beweis. Für $q = 1$: Wende Fubini und Dreiecksungleichung an.
Für $q = \infty$: $E_X(\text{ess-sup}(|E_Y(Z)|)) \leq E_Y(E_X(\text{ess-sup}|Z|))$.

Nun $q = (1, \infty)$: o.E. sei $Z \geq 0$, daher $|Z| = Z$. Sei U unabhängige Kopie von Y und unabhängig von X.

$$
\begin{aligned}
E_X(E_Y(Z)^{q-1+1}) &= E_X\left[E_U(f(X,U))^{q-1}E_Y(f(X,Y))\right] \\
\\
&= E_Y\left[E_X((E_U(f(X,U))^{q-1}f(X,Y)))\right] \\
&\leq E_Y\left[(E_X(E_U(f(X,U)))^q)^{\frac{q-1}{q}}(E_X(f(X,Y)^q))^{\frac{1}{q}}\right] \\
&= \left[E_X(E_U(f(X,U))^q)\right]^{\frac{q-1}{q}} E_Y\left[(E_X(f(X,Y)^q))^{\frac{1}{q}}\right] \\
&= \left[E_X((E_Y(Z))^q)\right]^{1-\frac{1}{q}} E_Y\left[(E_X(Z^q))^{\frac{1}{q}}\right]
\end{aligned}
$$

Dividiere durch $(E_X(E_Y(Z))^q)^{1-\frac{1}{q}}$, dann folgt

$$
(E_X(E_Y(Z)^q))^{\frac{1}{q}} \leq E_Y((E_X(Z^q))^{\frac{1}{q}})
$$

$\square$