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.

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

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

Dabei sind $\sigma_i$ die Standardabweichungen der ZVen $X_i$ für $i = 1, \dots, n$.
Die hier vorliegende Abschätzung bildet im Fall von identisch verteilten Zufallsvariablen sogar Gleichheit.

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

$$\left( \frac{1}{n} \big( Z_n - E(Z_n) \big)^2 \right)^{1/2} = \sqrt{\operatorname{Var} \frac{Z_n}{n}} = \frac{1}{n} \sqrt{\operatorname{Var} Z_n}$$

unabh.

$$= \frac{1}{n} \sqrt{\sum_{i=1}^{n} \sigma_i^2} \overset{=}{\le} \frac{1}{n} \sqrt{n\, v} = \frac{\sqrt{v}}{\sqrt{n}}$$

falls $X_i$ identisch verteilt.

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 $x_1, ..., x_n \in \mathbb{R}$ als Realisationen von unabhängig, identisch verteilten ZVen $X_1, ..., X_n$ mit unbekannter Verteilung $P_X = \mu$.
Frage: Wie lässt sich aus geg. Daten die wahre Verteilung rekonstruieren?
Wir verwenden das sogenannte *empirische Maß*

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

$$F_n(x) = \int_0^x \mu_n^{x_1 \dots x_n}$$

mit $\delta_{x_i}$ als das Punktmaß in $x_i$. Das liefert eine intuitive Möglichkeit für eine Schätzung von $\mu$. Die empirische Verteilungsfunktion aus den Datensatz $x_1, ..., x_n$ ist dabei zum empirischen Maß assoziiert.
Auch hier stellt man sich die Frage nach der Güte der Approximation von $\mu_n$ zu $\mu$. 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.

Bei der Konvergenz von Maßen $\mu_n \to \mu$ (auf $(\mathbb{R}, \mathcal{B})$) kann ich mit verschiedenen Klassen von Fkt. oder Mengen testen.

z.B.

$\mu_n \to \mu$ schwach $\Leftrightarrow$ $\forall f \in C_b(\mathbb{R})$ : $\int f \mu_n \to \int f \mu$

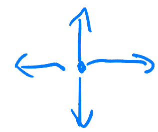
|| vage $\forall f \in C_c(\mathbb{R})$ ||



Hier testen wir mit einer Klasse von Mengen $\mathcal{A}$

$\forall A \in \mathcal{A}$ : $\int_A \mu_n \to \int_A \mu$

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 $\mathbb{Z}^2$**

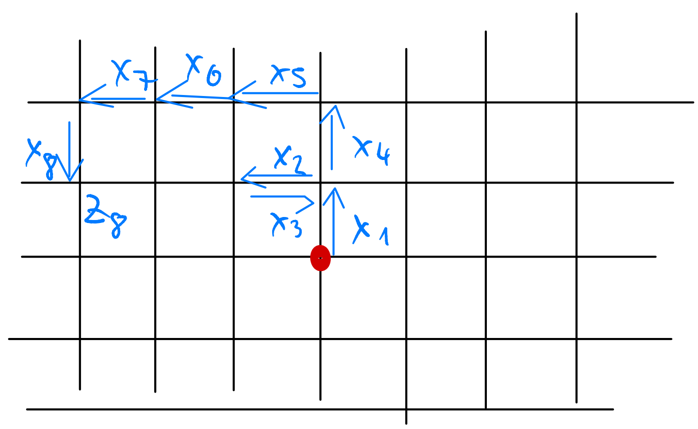
Wir betrachten unabhängig, identisch verteilte ZVen der Form

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

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*.

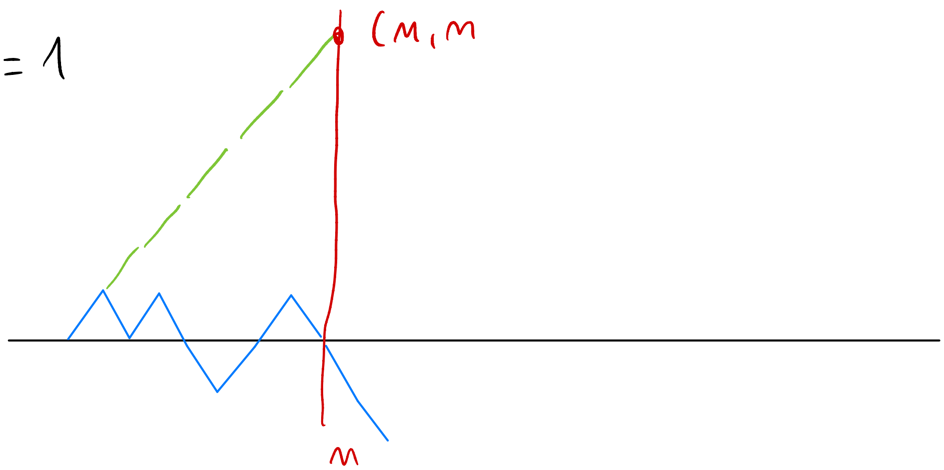


Wir fragen uns in diesem Kontext beispielsweise, wie sich der euklidische Abstand $\|Z_n\|$ im Verlauf eines Pfades typischeweise verhält. Es kann z.B. nach einer oberen Schranke für $\max\limits_{i=1,\ldots,n} \|Z_i\|$ gefragt werden. Eine sehr einfache Antwort wäre

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

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

$d = 1$



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

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

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:

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

<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 $(\Omega, P)$ 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:

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

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

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

Integration liefert:

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

Weitere Abschätzung der linken Seite:

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

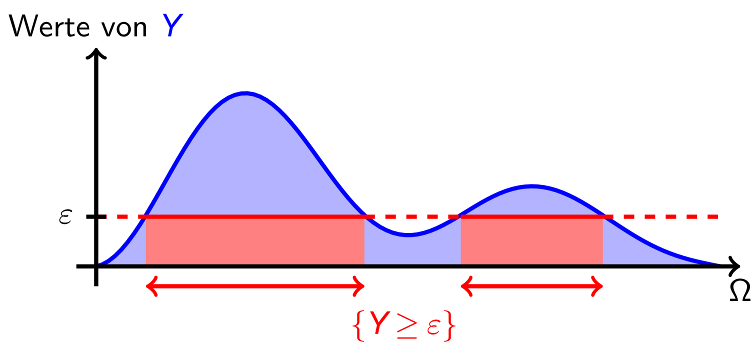



Abbildung 1. $\Omega$ auf der horizontalen Achse.

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

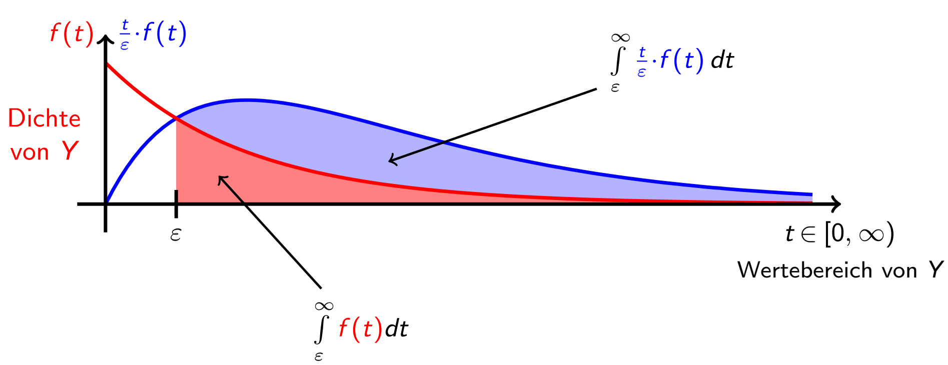



Abbildung 2. Wertebereich auf der horizontalen Achse.

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

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

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

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

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 $\Phi(y) = y^2$

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

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 monoton wachsende Funktionen $\Phi : I \to [0, \infty)$ 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 $\forall \, q, t \geq 0$:

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

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

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

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

**Warum spielt der Fall $\mathbf{q = 2}$ 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):

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

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

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

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:

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

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

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

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 $Z$ 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 $M_X(\lambda)$, so gilt mit (2.4):

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

Vorgehensweise bisher:

- Gewinne Klassen von Ungleichungen für verschiedene $\lambda$
- Schranke über den Parameter $\lambda$ 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:

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

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.

$\Rightarrow P(|\tilde{Z}| \geq t) = P(|Z - EZ| \geq t)$

**Die kumulantenerzeugenden Funktion und die Cramér-Transformierte**

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

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

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):

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

Momente der ZV $X$: $E(X), E(X^2), E(X^3)$

zentrierte Mom., absolute Mom $E|X|^q$

oder Kumulanten: Mittelwert,

Varianz,

sind Koeffiz in Entwicklung $\psi_Z$ bei $0$

Betrachte dazu die sogenannte *Cramér-Transformierte*:

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

<u>Bedeutung der Cramér-Transformierte:</u>
Minimiere die Schranke von (2.4*) über $\lambda$ und betrachte nur noch den Exponenten. Beachte: Vorzeichenwechsel liefert Maximierungsproblem:

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

Untersuche nun den Definitions- und Wertebereich der Cramér-Trafo $\psi_Z^*$. 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 $\psi_Z^*$ 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 $\lambda$ nicht im negativen Bereich angenommen wird, d.h.

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



$, \geq t) + P(-Z + E(Z) \leq -t)$

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

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.

**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) $\psi$ ist strikt konvex auf $I$, falls $Z$ nicht fast sicher konstant
(c) $\psi : I \to \mathbb{R}$ 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:

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



Falls $Z = c$ f.s. $\Rightarrow M(\lambda) = E e^{\lambda Z} = e^{\lambda c}$

$\psi(\lambda) = \ln M(\lambda) = \lambda c$ | $\lambda \mapsto \lambda c$ ist konvex aber nicht strikt.

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

$E(Z) = 0$ zentriert.

$$\frac{d}{d\lambda} \psi_Z(\lambda) = \frac{d}{d\lambda}\left(\ln M(\lambda)\right)$$

$$= \frac{1}{M(\lambda)} M'(\lambda) \stackrel{\lambda=0}{=\!=} 1 \cdot 0$$

Denn allg. gilt (falls $M : [0, \delta) \to \mathbb{R}$ wohldef.)

$$M^{(n)}(0) = E(Z^n)$$

insbes. $M'(0) = E(Z) = 0$

**Bestimmung des Supremums mittels der Ableitung**

Ansatz:

- $\psi_Z$ ist $\mathcal{C}^1$ $\rightarrow$ 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 $\lambda_t$ eine Lösung dieser Gleichung. Falls man den trivialen Fall einer fast sicher konstanten Zufallsvariable ausschließt, ist $\psi$ strikt konvex.

$\Rightarrow \quad \lambda_t$ ist eindeutig. $\psi' : I = (0,b) \to (0,B)$ isoton

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.

**Cramér-Transformierte verschiedener Verteilungsklassen**

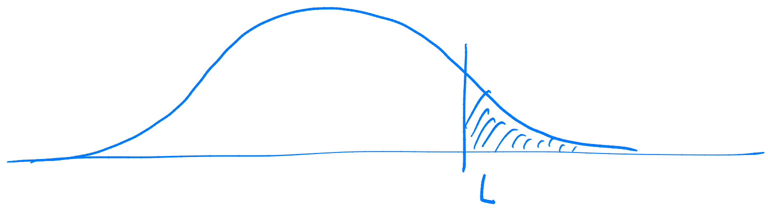
(a) Cramér-Transformierte für zentrierte Normalverteilungen:

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 $\sigma^2$. 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

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




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.

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

$\rightarrow$ 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:

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

Logarithmieren liefert für $\lambda > 0$ die KEF $\psi_Z$. Berechne außerdem $\psi_Z'$

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

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

$$\begin{aligned} t &= \nu\left(e^{\lambda_t} - 1\right) \\ \Leftrightarrow \quad e^{\lambda_t} &= \frac{t}{\nu} + 1 \\ \Leftrightarrow \quad \lambda_t &= \ln\left(\frac{t}{\nu} + 1\right). \end{aligned}$$

Der optimierende Parameter $\lambda_t$ 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:

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