Die Azuma-Hoeffding-Ungleichung und verwandte Konzentrationsungleichungen

Bachelorarbeit
im Studiengang Wirtschaftsmathematik
am Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

vorgelegt von
Philipp Hermann
Marburg, September 2014
betreut durch
Prof. Dr. H. Holzmann

Inhaltsverzeichnis

Einleitung

Die Arbeit beschäftigt sich mit Konzentrationsungleichungen, welche ein Teilgebiet der Wahrscheinlichkeitstheorie darstellen. Dieses Fachgebiet hat in den letzten Jahrzehnten einen Wandel vollzogen und an Bedeutung gewonnen. Anwendungen finden sich heutzutage in sehr vielen, verschiedenen Gebieten, z.B. in der Statistik, in der diskreten Mathematik, in der mehrdimensionalen Geometrie, im Operations Research und beim Maschinellen Lernen.

Diese Ausarbeitung basiert dabei weitgehend auf der Arbeit „Concentration Inequalities" von Steven P. Lalley [1].

Konzentrationsungleichungen geben generell Grenzen für die Wahrscheinlichkeit, dass die Abweichung einer Zufallsvariablen von einem Wert, in den meisten Fällen ist dies der Erwartungswert, größer als eine bestimmte Konstante ist, an. Diese Schranke fällt für größer werdende Konstanten und damit auch Abweichungen in allen Fällen, die in dieser Arbeit betrachtetet werden, exponentiell.

Die Gesetze der großen Zahlen besagen, dass das arithmetische Mittel von unabhängigen Zufallsvariablen unter gewissen Bedingungen stochastisch oder fast sicher gegen den Erwartungswert konvergiert. Dies ist ein erstes, grundlegendes Beispiel für Zufallsvariablen, welche sich konzentriert um den Erwartungswert herum befinden.

So ein Verhalten lässt sich auch bei einer großen Klasse von Funktionen von unabhängigen Zufallsvariablen beobachten. Diese Erkenntnisse werden in den Konzentrationsungleichungen dargestellt.

Im ersten Kapitel wird die Azuma-Hoeffding-Ungleichung betrachtet, welche ein Resultat für Martingale mit beschränkten Abständen zwischen den Zufallsvariablen liefert. Dazu werden zu Beginn einige Definitionen aus der Martingaltheorie wiederholt. Desweiteren wird eine Abschätzung für eine bestimmte Klasse von bedingten Erwartungswerten hergeleitet, welche für den Beweis der Azuma-Hoeffding-Ungleichung wichtig ist. Zum Schluss wird diese Ungleichung mit Hilfe des Zentralen Grenzwertsatzes anhand eines Beispiels überprüft.

Ungleichungen für bestimmte Funktionen von unabhängigen Zufallsvariablen werden im zweiten Kapitel untersucht. Dabei werden aus denselben Voraussetzungen zwei unterschiedlich starke Konzentrationsungleichungen hergeleitet, da einmal der Beweis mit Hilfe der Azuma-Hoeffding-Ungleichung geführt wird, und das andere Mal mit dem Hoeffding-Lemma.

Aus der McDiarmid-Ungleichung aus Kapitel 2 wird im letzten Kapitel eine Abschätzung für Funktionen, welche Lipschitz in jeder Variable sind, und unabhängige, standardnormalverteilte Zufallsvariablen entwickelt. Diese Ungleichung wird in einem folgenden Beispiel für zwei konkrete Funktionen überprüft und anschaulich dargestellt. Als letztes wird eine Konzentrationsungleichung für allgemein Lipschitz-stetige Funktionen hergeleitet.

1 Azuma-Hoeffding-Ungleichung

Folgende Definitionen sind eine Wiederholung aus der Wahrscheinlichkeitstheorie zum Thema Martingal, da sich die Azuma-Hoeffding-Ungleichung auf diese bezieht.

Definition 1.1.
Es sei $(\Omega, \mathcal{A})$ ein messbarer Raum. Eine Folge $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ von aufsteigenden Teil-σ-Algebren von $\mathcal{A}$ (mit $\mathcal{F}_0 = \{\emptyset, \Omega\}$) heißt Filtration. Ein Tupel $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\in\mathbb{N}_0}, P)$ nennt man auch einen filtrierten Wahrscheinlichkeitsraum.
Ein stochastischer Prozess $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ mit Werten in $(E, \mathcal{E})$ heißt adaptiert an $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$, falls $\forall\ n \geq 0\ X_n\ \mathcal{F}_n$-$\mathcal{E}$-messbar ist.

Definition 1.2.
Sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ gegeben und außerdem sei $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ ein an $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ adaptierter stochastischer Prozess. $\mathbb{X}$ heißt Martingal bzgl. $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$, falls $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$ P-fast sicher $\forall\ m, n \geq 0$ mit $n \geq m$ gilt.
Gilt $\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m$, so heißt $\mathbb{X}$ Submartingal bzgl. $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$.
Gilt $\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m$, so heißt $\mathbb{X}$ Supermartingal bzgl. $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$.

Im Folgenden führen wir einige Resultate ein, welche man später für den Beweis der Azuma-Hoeffding-Ungleichung braucht.

Satz 1.3.
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, eine Teil-σ-Algebra $\mathcal{B} \subset \mathcal{A}$ und eine Zufallsvariable X mit $X \in [0, 1]$ und bedingtem Erwartungswert $p = \mathbb{E}[X|\mathcal{B}]$. Für jede konvexe Funktion $\varphi\colon [0, 1] \to \mathbb{R}$ gilt:

$$\mathbb{E}[\varphi(X)|\mathcal{B}] \leq p \cdot \varphi(1) + (1 - p) \cdot \varphi(0). \tag{1.1}$$

Beweis. Da $X \in [0,1]$ ist, gilt:

$$\begin{aligned}
& X = X \cdot 1 + (1-X) \cdot 0 \\
\overset{\varphi \text{ konvex}}{\Rightarrow} \quad & \varphi(X) \leq X \cdot \varphi(1) + (1-X) \cdot \varphi(0) \\
\underset{\text{bedingten Erwartungswertes}}{\overset{\text{Monotonie des}}{\Rightarrow}} \quad & \mathbb{E}[\varphi(X)|\mathcal{B}] \leq \mathbb{E}[X \cdot \varphi(1) + (1-X) \cdot \varphi(0)|\mathcal{B}] \\
\underset{\text{bedingten Erwartungswertes}}{\overset{\text{Linearität des}}{\Leftrightarrow}} \quad & \mathbb{E}[\varphi(X)|\mathcal{B}] \leq \mathbb{E}[X|\mathcal{B}] \cdot \varphi(1) + (1 - \mathbb{E}[X|\mathcal{B}]) \cdot \varphi(0) \\
\Leftrightarrow \quad & \mathbb{E}[\varphi(X)|\mathcal{B}] \leq p \cdot \varphi(1) + (1-p) \cdot \varphi(0).
\end{aligned}$$

□

Im anschließenden Korollar übertragen wir diese Eigenschaft auf ein beliebiges Intervall [-A,B].

Korollar 1.4.
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Teil-σ-Algebra $\mathcal{B} \subset \mathcal{A}$. Für jede Zufallsvariable X mit $X \in [-A, B]$ $(A, B \in \mathbb{R}^+)$ und bedingtem Erwartungswert $\mathbb{E}[X|\mathcal{B}] = 0$ und jede konvexe Funktion $\varphi\colon [-A, B] \to \mathbb{R}$ gilt:

$$\mathbb{E}[\varphi(X)|\mathcal{B}] \leq \frac{B}{A+B} \cdot \varphi(-A) + \frac{A}{A+B} \cdot \varphi(B). \tag{1.2}$$

Insbesondere, falls $A = B > 0$ ist, gilt für alle $\theta > 0$:

$$\mathbb{E}[e^{\theta X}|\mathcal{B}] \leq \cosh(\theta A). \tag{1.3}$$

Beweis. Setze

$$Y = \frac{X+A}{A+B} \quad \in [0,1] \quad \text{und} \quad \tilde{\varphi}\colon [0,1] \to \mathbb{R}$$
$$x \mapsto \varphi(x \cdot (B+A) - A).$$

Dann folgt

$$\mathbb{E}[Y|\mathcal{B}] = \mathbb{E}\Big[\frac{X+A}{A+B}\Big|\mathcal{B}\Big] \overset{\text{Linearität}}{=} \frac{1}{A+B} \cdot (\mathbb{E}[X|\mathcal{B}] + A) \overset{\text{Voraussetzung}}{=} \frac{A}{A+B}$$

und

$$\tilde{\varphi}(Y) = \varphi(Y \cdot (B+A) - A) = \varphi\Big(\frac{X+A}{A+B} \cdot (B+A) - A\Big) = \varphi(X).$$

Mit Satz 1.3 gilt nun:

$$
\begin{aligned}
&\mathbb{E}[\tilde{\varphi}(Y)|\mathcal{B}] \leq \mathbb{E}[Y|\mathcal{B}] \cdot \tilde{\varphi}(1) + (1 - \mathbb{E}[Y|\mathcal{B}]) \cdot \tilde{\varphi}(0) \\
\Leftrightarrow \quad &\mathbb{E}[\varphi(X)|\mathcal{B}] \leq \frac{A}{A+B} \cdot \varphi(1 \cdot (B+A) - A) + \left(1 - \frac{A}{A+B}\right) \cdot \varphi(0 \cdot (B+A) - A) \\
\Leftrightarrow \quad &\mathbb{E}[\varphi(X)|\mathcal{B}] \leq \frac{A}{A+B} \cdot \varphi(B) + \frac{B}{A+B} \cdot \varphi(-A).
\end{aligned}
$$

Die Ungleichung (1.3) ist der Spezialfall $\varphi(x) = e^{\theta x}$. Diese Funktion ist für alle $\theta > 0$ konvex. Unter der Berücksichtigung, dass $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ ist, gilt:

$$
\begin{aligned}
\mathbb{E}[\varphi(X)|\mathcal{B}] = \mathbb{E}[e^{\theta X}|\mathcal{B}] &\leq \frac{B}{A+B} \cdot e^{\theta(-A)} + \frac{A}{A+B} \cdot e^{\theta B} \\
&= \frac{A}{A+A} \cdot e^{\theta(-A)} + \frac{A}{A+A} \cdot e^{\theta A} \\
&= \frac{1}{2} \cdot \left(e^{\theta(-A)} + e^{\theta A}\right) \\
&= \cosh(\theta A).
\end{aligned}
$$

□

Nun beweisen wir noch eine Abschätzung für den Kosinus Hyperbolicus.

Lemma 1.5.

$$\cosh(x) \leq e^{x^2/2} \tag{1.4}$$

Beweis. Die Potenzreihe für 2cosh(x) kann man durch die Potenzreihen von e^x und e^{-x} darstellen. Da sich für ungerade n die Terme gegeneinander wegaddieren, erhält man:

$$
\begin{aligned}
\cosh(x) = \frac{1}{2} \cdot (e^x + e^{-x}) &= \frac{1}{2} \cdot \left(\sum_{n=1}^{\infty} \frac{x^n}{n!} + \sum_{n=1}^{\infty} \frac{-x^n}{n!}\right) \\
&= \frac{1}{2} \cdot \left(2 \sum_{n=1}^{\infty} \frac{x^{2n}}{2n!}\right) \\
&\overset{(*)}{\leq} \sum_{n=1}^{\infty} \frac{x^{2n}}{2^n (n!)} = e^{x^2/2}.
\end{aligned}
$$

Die Ungleichung (*) gilt wegen:

$$
\begin{aligned}
(2n)! &= \prod_{k=1}^{n} (2k) \cdot \prod_{k=1}^{n} (2k-1) \\
&\geq 2^n \cdot n! \; .
\end{aligned}
$$

□

Die bisherigen Resultate kann man im nachstehenden Korollar zusammenfassen.

Korollar 1.6.
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, eine Teil-σ-Algebra $\mathcal{B} \subset \mathcal{A}$ und eine Zufallsvariable X mit $-A \leq X \leq A$ $(A \in \mathbb{R}^+)$ und $\mathbb{E}[X|\mathcal{B}] = 0$. Dann gilt für jedes $\theta > 0$:

$$\mathbb{E}[e^{\theta X}|\mathcal{B}] \leq \cosh(\theta A) \leq e^{\frac{\theta^2 A^2}{2}}. \tag{1.5}$$

Beweis. Die erste Ungleichung folgt aus (1.3) und die zweite Ungleichung aus Lemma 1.5. $\square$

Mit Hilfe dieser Abschätzung können wir nun die Azuma-Hoeffding-Ungleichung beweisen.

Satz 1.7. *(Azuma-Hoeffding-Ungleichung)*
Sei $(S_n)_{n\in\mathbb{N}_0}$ ein Martingal bzgl. einer Filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$, dessen Abstände $\xi_k = S_k - S_{k-1}$ für alle $k \in \mathbb{N}$ im Absolutbetrag durch 1 beschränkt sind. Dann gilt für ein beliebiges $\alpha > 0$ und $n \geq 1$:

$$P(S_n - S_0 \geq \alpha) \leq \exp(-\alpha^2/2n). \tag{1.6}$$

Im Allgemeinen gilt unter der Voraussetzung $|\xi_k| \leq \sigma_k$:

$$P(S_n - S_0 \geq \alpha) \leq \exp\Big(-\alpha^2/2\sum_{j=1}^{n} \sigma_j^2\Big). \tag{1.7}$$

Beweis. Es genügt die zweite Ungleichung (1.7) zu zeigen, weil es sich bei der ersten Ungleichung (1.6) nur um einen Spezialfall handelt.
Setze $\tilde{S}_n = S_n - S_0$ für alle $n \in \mathbb{N}_0$. Dann gilt $\forall\ k \in \mathbb{N}$:

$$\tilde{S}_k - \tilde{S}_{k-1} = S_k - S_0 - (S_{k-1} - S_0) = S_k - S_{k-1}$$

und

$$\tilde{S}_0 = S_0 - S_0 = 0.$$

Da $\tilde{S}_n$ eine Zufallsvariable ist und die Funktion $h(x) = e^{\theta x} \geq 0$ für ein beliebiges $\theta > 0$ monoton wachsend ist, gilt für alle $\alpha > 0$ und $n \geq 1$ aufgrund der Markow-Ungleichung:

$$P(\tilde{S}_n \geq \alpha) \leq \frac{\mathbb{E}[h(\tilde{S}_n)]}{h(\alpha)} = \frac{\mathbb{E}[e^{\theta \tilde{S}_n}]}{e^{\theta\alpha}}. \tag{1.8}$$

Da es sich laut Voraussetzung bei $(S_n)_{n\in\mathbb{N}_0}$ um ein Martingal handelt, gilt außerdem $\forall\ n \geq 1$:

$$\begin{aligned}\mathbb{E}[\xi_n|\mathcal{F}_{n-1}] = \mathbb{E}[S_n - S_{n-1}|\mathcal{F}_{n-1}] &= \mathbb{E}[S_n|\mathcal{F}_{n-1}] - \mathbb{E}[S_{n-1}|\mathcal{F}_{n-1}] \\ &= \mathbb{E}[S_n|\mathcal{F}_{n-1}] - S_{n-1} \\ &= S_{n-1} - S_{n-1} = 0.\end{aligned}$$

Aufgrund dieses Resultats und der Voraussetzung, dass ξ_n im Intervall $[-\sigma_n, \sigma_n]$ liegt, kann man nun Korollar 1.6 auf den Zähler der rechten Seite der Ungleichung (1.8) anwenden:

$$\begin{aligned}\mathbb{E}[e^{\theta\tilde{S}_n}] &\overset{totaler\ EW}{=} \mathbb{E}[\mathbb{E}[e^{\theta\tilde{S}_n}|\mathcal{F}_{n-1}]] \\ &= \mathbb{E}[\mathbb{E}[e^{\theta(\tilde{S}_{n-1}+\xi_n)}|\mathcal{F}_{n-1}]] \\ &= \mathbb{E}[\mathbb{E}[e^{\theta\tilde{S}_{n-1}} \cdot e^{\theta\xi_n}|\mathcal{F}_{n-1}]] \\ &= \mathbb{E}[e^{\theta\tilde{S}_{n-1}} \cdot \mathbb{E}[e^{\theta\xi_n}|\mathcal{F}_{n-1}]] \\ &\overset{Korollar\ 1.6}{\leq} \mathbb{E}\big[e^{\theta\tilde{S}_{n-1}} \cdot e^{\frac{\theta^2\sigma_n^2}{2}}\big] \\ &= e^{\frac{\theta^2\sigma_n^2}{2}} \cdot \mathbb{E}[e^{\theta\tilde{S}_{n-1}}].\end{aligned}$$

Man kann nun induktiv dieses Vorgehen benutzen, um auch die Erwartungswerte $\mathbb{E}[e^{\theta\tilde{S}_{n-i}}]$ $\forall\ i = 1, \ldots, n-1$ abzuschätzen. Für den letzten Erwartungswert $\mathbb{E}[e^{\theta\tilde{S}_0}]$ erhält man 1, da $\mathbb{E}[e^{\theta\tilde{S}_0}] = \mathbb{E}[e^0] = \mathbb{E}[1] = 1$.
Am Ende erhält man die Ungleichung:

$$\mathbb{E}[e^{\theta\tilde{S}_n}] \leq \prod_{i=1}^{n} e^{\frac{\theta^2\sigma_i^2}{2}}.$$

Insgesamt ist also für jedes $\theta > 0$:

$$P(S_n - S_0 \geq \alpha) = P(\tilde{S}_n \geq \alpha) \leq e^{-\theta\alpha} \cdot \prod_{i=1}^{n} e^{\frac{\theta^2\sigma_i^2}{2}}. \tag{1.9}$$

Die bestmögliche Abschätzung für $P(S_n - S_0 \geq \alpha)$ erhält man nun, wenn man θ so wählt, dass die rechte Seite minimiert wird.
Die rechte Seite der obigen Ungleichung (1.9) lässt sich auch wie folgt schreiben:

$$e^{-\theta\alpha} \cdot \prod_{i=1}^{n} e^{\frac{\theta^2\sigma_i^2}{2}} = \exp\Big(-\theta\alpha + \sum_{i=1}^{n} \frac{\theta^2\sigma_i^2}{2}\Big).$$

Da die Exponentialfunktion monoton wachsend ist, liegt ihr Minimum genau bei dem θ, bei dem auch die Funktion

$$g(\theta) = -\theta\alpha + \sum_{i=1}^{n} \frac{\theta^2\sigma_i^2}{2}$$

ihr Minimum hat.
Dies ist durch Ableiten von $g(\theta)$ und anschließendem Nullsetzen leicht zu bestimmen:

$$\begin{aligned} && g'(\theta) &= 0 \\ \Leftrightarrow & -\alpha + \sum_{i=1}^{n} \theta\sigma_i^2 &&= 0 \\ \Leftrightarrow && \theta \cdot \sum_{i=1}^{n} \sigma_i^2 &= \alpha \\ \Leftrightarrow && \theta &= \frac{\alpha}{\sum_{i=1}^{n} \sigma_i^2}. \end{aligned}$$

Es handelt sich auch wirklich um ein Minimum, da

$$g''(\theta) = \sum_{i=1}^{n} \sigma_i^2 > 0.$$

Setzt man nun $\theta = \alpha / \sum_{i=1}^{n} \sigma_i^2$ in die rechte Seite von (1.9) ein, erhält man für alle $n \geq 1$:

$$\begin{aligned} P(S_n - S_0 \geq \alpha) &\leq \exp(-\theta\alpha) \cdot \prod_{i=1}^{n} \exp\Big(\frac{\theta^2\sigma_i^2}{2}\Big) \\ &= \exp\Big(\frac{-\alpha^2}{\sum_{i=1}^{n} \sigma_i^2}\Big) \cdot \prod_{i=1}^{n} \exp\Big(\frac{\alpha^2\sigma_i^2}{2(\sum_{i=1}^{n} \sigma_i^2)^2}\Big) \\ &= \exp\Big(\frac{-\alpha^2}{\sum_{i=1}^{n} \sigma_i^2} + \sum_{i=1}^{n} \frac{\alpha^2\sigma_i^2}{2(\sum_{i=1}^{n} \sigma_i^2)^2}\Big) \\ &= \exp\Big(\frac{-\alpha^2}{\sum_{i=1}^{n} \sigma_i^2} + \frac{\alpha^2(\sum_{i=1}^{n} \sigma_i^2)}{2(\sum_{i=1}^{n} \sigma_i^2)^2}\Big) \\ &= \exp\Big(\frac{2(-\alpha^2)}{2\sum_{i=1}^{n} \sigma_i^2} + \frac{\alpha^2}{2\sum_{i=1}^{n} \sigma_i^2}\Big) \\ &= \exp\Big(\frac{-\alpha^2}{2\sum_{i=1}^{n} \sigma_i^2}\Big). \end{aligned}$$

□

Bemerkung 1.8.
Im letzten Satz wurde nur eine Abschätzung für sehr starke Ausreißer nach „oben“ getroffen. Es existiert natürlich auch eine Eingrenzung der Wahrscheinlichkeit für extreme Ausreißer nach „unten“.

Angenommen, $(S_n)_{n\in\mathbb{N}_0}$ erfüllt die Voraussetzungen aus Satz 1.7, dann gilt dies auch für das Martingal $(-S_n)_{n\in\mathbb{N}_0}$, da $\forall\ k \in \mathbb{N}$

$$|-S_k - (-S_{k-1})| = |-S_k + S_{k-1}| = |S_k - S_{k-1}| \leq \sigma_k.$$

Es folgt also aus Satz 1.7:

$$P(-S_n - (-S_0) \geq \alpha) = P(-S_n + S_0 \geq \alpha) \leq \exp\Big(-\alpha^2/2\sum_{j=1}^{n}\sigma_j^2\Big). \tag{1.10}$$

Insgesamt gilt also:

$$P(|S_n - S_0| \geq \alpha) = P(S_n - S_0 \geq \alpha) + P(-S_n + S_0 \geq \alpha) \leq 2\exp\Big(-\alpha^2/2\sum_{j=1}^{n}\sigma_j^2\Big). \tag{1.11}$$

Im Folgenden wird die Aussage der Azuma-Hoeffding-Ungleichung anhand eines Beispiels überprüft und verdeutlicht.
Dazu werden einige Vorkenntnisse zum Zentralen Grenzwertsatz gebraucht, welche im nachstehenden Satz und Korollar noch einmal kurz wiederholt werden.

Satz 1.9. *(Zentraler Grenzwertsatz)*
Sei $(X_{nk})_{n\in\mathbb{N}}^{k=1,\ldots,n}$ ein Dreieckschema mit $\mathbb{E}[X_{nk}] = 0$ und $Var(X_{nk}) < \infty$ für alle n, k. Setze $S_n = \sum_{k=1}^{n} X_{nk}$. Es sei $s_n^2 = \sum_{k=1}^{n} Var(X_{nk}) > 0$ und es gelte für alle $\epsilon > 0$:

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int\limits_{\{|X_{nk}|\geq\epsilon s_n\}} X_{nk}^2 dP_n \to 0. \tag{1.12}$$

Dann folgt:

$$\lim_{n\to\infty} P_n\Big(\frac{S_n}{s_n} \leq y\Big) = \Phi(y), \tag{1.13}$$

wobei $\Phi(y)$ die Verteilungsfunktion der Standardnormalverteilung ist.

Korollar 1.10.
Es sei $X_{nk} = X_k$ für alle $n \in \mathbb{N}, k = 1, \ldots, n$ und die X_k seien unabhängig und identisch verteilt mit endlicher Varianz. Dann gilt:

$$\lim_{n\to\infty} P\Big(\frac{\sum_{k=1}^{n}(X_k - \mathbb{E}[X_1])}{\sqrt{nVar(X_1)}} \leq y\Big) = \Phi(y). \tag{1.14}$$

Beispiel 1.11.
Seien $(Y_n)_{n\in\mathbb{N}_0}$ unabhängige und identisch verteilte Zufallsvariablen mit $P(Y_i = d) = P(Y_i = -d) = \frac{1}{2}$ für alle $i \in \mathbb{N}_0$ und eine Konstante $d > 0$.
Für den Erwartungswert der Y_i gilt: $\mathbb{E}[Y_i] = \frac{1}{2} \cdot d + \frac{1}{2} \cdot (-d) = 0$ und für die Varianz: $Var(Y_i) = \frac{1}{2} \cdot d^2 + \frac{1}{2} \cdot (-d)^2 = d^2$.
$\forall\ k \in \mathbb{N}_0$ setze nun $X_k = \sum_{i=0}^{k} Y_i$ und sei $\mathcal{F}_k = \sigma(X_0, \ldots, X_k) = \sigma(Y_0, \ldots, Y_k)$ die natürliche bzw. die von $(Y_n)_{n\in\mathbb{N}_0}$ erzeugte Filtration.
Der stochastische Prozess $\mathbb{X} = (X_k)_{k\in\mathbb{N}_0}$ ist dann ein Martingal bzgl. der Filtration $(\mathcal{F}_k)_{k\in\mathbb{N}_0}$ und es gilt: $|X_k - X_{k-1}| = |Y_k| = d$ für alle $k \in \mathbb{N}$.
Wendet man nun die Ungleichung (1.11) an, erhält man für ein beliebiges $\alpha > 0$ und $n \in \mathbb{N}$:

$$P(|X_n - X_0| \geq \alpha\sqrt{n}) \leq 2 \cdot \exp\Big(\frac{-(\alpha\sqrt{n})^2}{2\sum_{j=1}^{n} d^2}\Big) = 2 \cdot \exp\Big(\frac{-n\alpha^2}{2nd^2}\Big) = 2 \cdot \exp\Big(\frac{-\alpha^2}{2d^2}\Big).$$

Da die $(Y_n)_{n\in\mathbb{N}_0}$ unabhängig und identisch verteilt sind, kann man den Zentralen Grenzwertsatz, insbesondere Korollar 1.10, anwenden. Wenn $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung ist, gilt also:

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n - X_0| \geq \alpha\sqrt{n}) &= \lim_{n\to\infty} P\Big(\frac{X_n - X_0}{\sqrt{n}} \geq \alpha\Big) + \lim_{n\to\infty} P\Big(\frac{X_n - X_0}{\sqrt{n}} \leq -\alpha\Big) \\
&= \lim_{n\to\infty} P\Big(\frac{X_n - X_0}{\sqrt{n}d} \geq \frac{\alpha}{d}\Big) + \lim_{n\to\infty} P\Big(\frac{X_n - X_0}{\sqrt{n}d} \leq \frac{-\alpha}{d}\Big) \\
&= \lim_{n\to\infty} P\Big(\frac{X_n - X_0}{\sqrt{nd^2}} \geq \frac{\alpha}{d}\Big) + \lim_{n\to\infty} P\Big(\frac{X_n - X_0}{\sqrt{nd^2}} \leq \frac{-\alpha}{d}\Big) \\
&= \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^{n} Y_i}{\sqrt{nd^2}} \geq \frac{\alpha}{d}\Big) + \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^{n} Y_i}{\sqrt{nd^2}} \leq \frac{-\alpha}{d}\Big) \\
&= \Big(1 - \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^{n} Y_i}{\sqrt{nd^2}} \leq \frac{\alpha}{d}\Big)\Big) + \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^{n} Y_i}{\sqrt{nd^2}} \leq \frac{-\alpha}{d}\Big) \\
&= \Big(1 - \Phi\Big(\frac{\alpha}{d}\Big)\Big) + \Phi\Big(\frac{-\alpha}{d}\Big) \\
&= \Big(1 - \Phi(\frac{\alpha}{d})\Big) + \Big(1 - \Phi\Big(\frac{\alpha}{d}\Big)\Big) \\
&= 2 \cdot \Big(1 - \Phi\Big(\frac{\alpha}{d}\Big)\Big).
\end{aligned}$$

Es gilt weiter, dass $1 - \Phi(x) = Q(x)$, wobei $Q(x)\ \forall\ x \in \mathbb{R}$ wie folgt definiert ist:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_x^{\infty} \exp\Big(\frac{-t^2}{2}\Big) dt.$$

Mit Hilfe der Substitution $v = t^2/2$ lässt sich für $Q(x)$ eine obere Schranke $\forall\ x > 0$

finden:

$$
\begin{aligned}
Q(x) &= \frac{1}{\sqrt{2\pi}} \cdot \int\limits_{x}^{\infty} \exp\Big(\frac{-t^2}{2}\Big) dt < \frac{1}{\sqrt{2\pi}} \cdot \int\limits_{x}^{\infty} \frac{t}{x} \cdot \exp\Big(\frac{-t^2}{2}\Big) dt \\
&= \frac{1}{\sqrt{2\pi}} \cdot \int\limits_{\frac{x^2}{2}}^{\infty} \frac{\exp(-v)}{x} dv = \frac{1}{\sqrt{2\pi}} \cdot \Big(-\frac{\exp(-v)}{x}\Big)\Big|_{\frac{x^2}{2}}^{\infty} \\
&= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{x} \cdot \exp\Big(-\frac{x^2}{2}\Big).
\end{aligned}
$$

Ebenso kann mit

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big) \\
f'(x) &= -x \cdot \Big(\frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big)\Big) = -x \cdot f(x)
\end{aligned}
$$

und der Quotientenregel eine untere Grenze $\forall\ x > 0$ bestimmt werden:

$$
\begin{aligned}
\Big(1 + \frac{1}{x^2}\Big) \cdot Q(x) &= \int\limits_{x}^{\infty} \Big(1 + \frac{1}{x^2}\Big) \cdot f(t) dt > \int\limits_{x}^{\infty} \Big(1 + \frac{1}{t^2}\Big) \cdot f(t) dt \\
&= \int\limits_{x}^{\infty} \Big(\frac{f(t) \cdot t^2 + f(t)}{t^2}\Big) dt = \frac{-f(t)}{t}\Big|_{x}^{\infty} \\
&= \frac{-\frac{1}{\sqrt{2\pi}} \cdot \exp(\frac{-t^2}{2})}{t}\Bigg|_{x}^{\infty} = \frac{1}{x} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow Q(x) > \frac{1}{x} \cdot \frac{1}{1 + \frac{1}{x^2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big) &= \frac{x}{x^2} \cdot \frac{1}{1 + \frac{1}{x^2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big) \\
&= \frac{x}{1 + x^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big).
\end{aligned}
$$

Da $\alpha/d > 0$ ist, gilt nun also insgesamt für $\lim_{n\to\infty} P(|X_n - X_0| \geq \alpha\sqrt{n})$:

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n - X_0| \geq \alpha\sqrt{n}) &= 2 \cdot \left(1 - \Phi\left(\frac{\alpha}{d}\right)\right) = 2 \cdot Q\left(\frac{\alpha}{d}\right) < \frac{2}{\sqrt{2\pi}(\frac{\alpha}{d})} \cdot \exp\left(\frac{-(\frac{\alpha}{d})^2}{2}\right) \\
&= \frac{1}{(\frac{\alpha}{d})} \cdot \frac{1}{\sqrt{2\pi}} \cdot 2 \cdot \exp\left(\frac{-\alpha^2}{2d^2}\right) \\
\lim_{n\to\infty} P(|X_n - X_0| \geq \alpha\sqrt{n}) &= 2 \cdot Q\left(\frac{\alpha}{d}\right) > 2 \cdot \frac{(\frac{\alpha}{d})}{1 + (\frac{\alpha}{d})^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-(\frac{\alpha}{d})^2}{2}\right) \\
&= \frac{(\frac{\alpha}{d})}{1 + (\frac{\alpha}{d})^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot 2 \cdot \exp\left(\frac{-\alpha^2}{2d^2}\right).
\end{aligned}$$

Der Grenzwert kann dementsprechend eingegrenzt werden:

$$\begin{aligned}
\frac{(\frac{\alpha}{d})^2}{(\frac{\alpha}{d}) + (\frac{\alpha}{d})^3} \cdot \frac{1}{\sqrt{2\pi}} \cdot 2 \cdot \exp\left(\frac{-\alpha^2}{2d^2}\right) &< \lim_{n\to\infty} P(|X_n - X_0| \geq \alpha\sqrt{n}) \\
&< \frac{(\frac{\alpha}{d})^2 + 1}{(\frac{\alpha}{d}) + (\frac{\alpha}{d})^3} \cdot \frac{1}{\sqrt{2\pi}} \cdot 2 \cdot \exp\left(\frac{-\alpha^2}{2d^2}\right).
\end{aligned}$$

Umso größer $(\frac{\alpha}{d})$ ist, umso kleiner ist das Intervall, welches durch die beiden Grenzen festgelegt wird. Man kann also gut erkennen, dass der Exponent, den man durch die Azuma-Hoeffding-Ungleichung erhält, der exakte Exponent in diesem Beispiel ist.
Es kann aber kein Fazit gezogen werden im Hinblick darauf, ob die mit Hilfe des Zentralen Grenzwertsatzes ermittelte Ungleichung oder die Azuma-Hoeffding-Ungleichung genauer ist, da dies von der Wahl von d und α abhängt.

2 McDiarmid-Ungleichung

In diesem Kapitel beschäftigen wir uns mit einer Verallgemeinerung der Azuma-Hoeffding-Ungleichung. Wir betrachten nun nicht-lineare Funktionen von Datenmengen $f(X_1, \dots, X_n)$.
Um eine erste Konzentrationsungleichung für diesen neuen Fall zu beweisen, werden wieder einige Vorbetrachtungen benötigt.
Im Folgenden werden kurz einige Erkenntnisse aus der Wahrscheinlichkeitstheorie zu bedingten Erwartungswerten wiederholt.

Lemma 2.1. *(Faktorisierungslemma)*
Sei $Y\colon \Omega \to \Omega'$ eine Abbildung von $(\Omega, \mathcal{A})$ nach $(\Omega', \mathcal{A}')$ und $Z\colon \Omega \to \bar{\mathbb{R}}$, wobei $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ ist, eine messbare Funktion. Genau dann ist Z $\sigma(Y)$-$\mathcal{B}(\bar{\mathbb{R}})$-messbar, wenn eine messbare Funktion $g\colon \Omega' \to \bar{\mathbb{R}}$ existiert, mit $Z = g \circ Y$.

Satz 2.2.
Jede $\mathcal{A}'$-messbare Funktion g mit $\mathbb{E}[X|Y] = g(Y)$ ist P^Y-integrierbar und erfüllt

$$\int_{A'} g dP^Y = \int_{\{Y \in A'\}} X dP \quad \text{für } A' \in \mathcal{A}'. \tag{2.1}$$

Sie ist fast sicher eindeutig bestimmt.
Umgekehrt, ist g $\mathcal{A}'$-messbar, P^Y-integrierbar und erfüllt (2.1), so ist $g \circ Y$ eine Version von $\mathbb{E}[X|Y]$.

Definition 2.3.
Für ein integrierbares X sei g eine der Bedingung (2.1) genügende, $\mathcal{A}'$-messbare, P^Y-integrierbare, reelle Funktion. Dann heißt $g(y)$ für jedes $y \in \Omega'$ die bedingte Erwartung von X gegeben $Y = y$. Notation: $\mathbb{E}[X|Y = y]$.

Nun können wir einen Hilfssatz beweisen, welchen wir für den Beweis der ersten Ungleichung brauchen.

Satz 2.4.
Seien $X\colon (\Omega, \mathcal{A}, P) \to (\Omega_1, \mathcal{A}_1)$ *und* $Y\colon (\Omega, \mathcal{A}, P) \to (\Omega_2, \mathcal{A}_2)$ *stochastisch unabhängige Zufallsvariablen und* $h\colon (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ *derart, dass* $h(X,Y)$ *quasi-integrierbar ist. Dann gilt:*

$$\mathbb{E}[h(X,Y)|Y=y] = \mathbb{E}[h(X,y)] = \int h(X(\omega), y)P(d\omega) \quad P^Y\text{-f.s.} \tag{2.2}$$

bzw. für den nicht faktorisierten Erwartungswert:

$$\mathbb{E}[h(X,Y)|Y](\omega) = \int h(X(\omega'), Y(\omega))P(d\omega') \quad P\text{-f.s. .} \tag{2.3}$$

Beweis. Setze $Z = h(X,Y)$. Für jedes $A \in \mathcal{A}_2$ folgt mit dem Satz von Fubini:

$$\begin{aligned}\int\limits_{\{Y\in A\}} Z dP = \int\limits_{\{Y\in A\}} h(X,Y)dP = \int\limits_{\Omega} \mathbb{1}_A(Y) \circ h(X,Y)dP &= \int\limits_{A}\int\limits_{\Omega_1} h(x,y)P^X(dx)P^Y(dy) \\ &= \int\limits_{A} \mathbb{E}[h(X,y)]P^Y(dy) \\ &:= \int\limits_{A} g(y)P^Y(dy).\end{aligned}$$

Mit Satz 2.2 folgt nun Aussage (2.2):

$$\mathbb{E}[h(X,Y)|Y=y] = \mathbb{E}[Z|Y=y] = g(y) = \mathbb{E}[h(X,y)] \overset{Definition}{\underset{EW}{=}} \int h(X(\omega), y)P(d\omega) \quad P^Y\text{-f.s. .}$$

Die Aussage (2.3) folgt direkt aus (2.2). $\square$

Wir betrachten nun die erste Konzentrationsungleichung für nicht-lineare Funktionen.

Satz 2.5.
Seien $X_1, \dots, X_n$ *unabhängige Zufallsvariablen mit* $X_i \in \mathcal{X}_i \ \forall\ i = 1, \dots, n$ *für beliebige messbare Räume* $\mathcal{X}_i$. *Angenommen,* $f\colon \prod_{i=1}^n \mathcal{X}_i \to \mathbb{R}$ *ist „Lipschitz" im folgenden Sinn: für alle* $1 \le k \le n$ *und alle* $x, x' \in \prod_{i=1}^n \mathcal{X}_i$, *welche sich nur in der* k*-ten Koordinate unterscheiden, gilt:* $|f(x) - f(x')| \le \sigma_k$. *Sei* $Y = f(X_1, \dots, X_n)$. *Dann gilt für jedes* $\alpha > 0$*:*

$$P(|Y - \mathbb{E}[Y]| \ge \alpha) \le 2\exp\Big(-\alpha^2/2\sum_{k=1}^n \sigma_k^2\Big). \tag{2.4}$$

Beweis. Da immer auf die ersten k Zufallsvariablen X_i bedingt werden soll, benötigen wir die von diesen Zufallsvariablen erzeugten σ-Algebren $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$ für $k = 1, \dots, n$ und $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
Setzte für $k = 1, \dots, n$:

$$Y_k = \mathbb{E}[Y|\mathcal{F}_k] = \mathbb{E}[Y|X_1, \dots, X_k]$$

und

$$Y_0 = \mathbb{E}[Y|\mathcal{F}_0] = \mathbb{E}[Y|\{\emptyset, \Omega\}] = \mathbb{E}[Y].$$

Die Folge $(Y_k)_{k=0,\dots,n}$ ist ein Martingal bezüglich $(\mathcal{F}_k)_{k=0,\dots,n}$, da $\forall\ k = 0, \dots, n-1$ aufgrund der Turmeigenschaft für bedingte Erwartungswerte gilt:

$$\mathbb{E}[Y_{k+1}|F_k] = \mathbb{E}[\mathbb{E}[Y|F_{k+1}]|F_k] = \mathbb{E}[Y|F_k] = Y_k.$$

Weiter sei nun $Y' = f(X')$, wobei X' von $X = (X_1, \dots, X_n)$ abgeleitet wurde, indem die k-te Koordinate X_k durch eine unabhängige Kopie X'_k ersetzt wurde. Es gilt $\forall\ k = 1, \dots, n$:

$$\begin{aligned}\mathbb{E}[Y'|F_k] &= \mathbb{E}[f(X_1, \dots, X_{k-1}, X'_k, X_{k+1}, \dots, X_n)|X_1, \dots, X_k] \\ &= \mathbb{E}[f(X_1, \dots, X_{k-1}, X'_k, X_{k+1}, \dots, X_n)|X_1, \dots, X_{k-1}].\end{aligned}$$

Da f quasi-integrierbar sein muss, gilt mit Satz 2.4:

$$= \int f(X_1, \dots, X_{k-1}, x'_k, x_{k+1}, \dots, x_n)\ P^{X'_k}(dx'_k) \prod_{j=k+1}^{n} P^{X_j}(dx_j).$$

Da X_k und X'_k die gleiche Verteilung haben, gilt:

$$\begin{aligned}&= \int f(X_1, \dots, X_{k-1}, x_k, x_{k+1}, \dots, x_n)\ \prod_{j=k}^{n} P^{X_j}(dx_j) \\ &= \mathbb{E}[f(X_1, \dots, X_{k-1}, X_k, X_{k+1}, \dots, X_n)|X_1, \dots, X_{k-1}] \\ &= \mathbb{E}[Y|\mathcal{F}_{k-1}] = Y_{k-1}.\end{aligned}$$

Da sich X und X' nur in der k-ten Koordinate unterscheiden, gilt laut Voraussetzung:

$$|Y - Y'| = |f(X) - f(X')| \le \sigma_k.$$

Und daraus folgt insgesamt $\forall\ k = 1, \dots, n$:

$$|Y_k - Y_{k-1}| = |\mathbb{E}[Y|\mathcal{F}_k] - \mathbb{E}[Y'|\mathcal{F}_k]| = |\mathbb{E}[Y - Y'|\mathcal{F}_k]| \le \mathbb{E}[|Y - Y'||\mathcal{F}_k] \le \sigma_k.$$

Die Ungleichung (2.4) folgt jetzt unmittelbar aus der Azuma-Hoeffding-Ungleichung und der Tatsache, dass $Y_n = \mathbb{E}[Y|\mathcal{F}_n] = Y$ (alle Information sind vorhanden) ist. Für beliebiges $\alpha > 0$ gilt:

$$\begin{aligned} & P(|Y_n - Y_0| \geq \alpha) \leq 2\exp\Big(-\alpha^2/2\sum_{k=1}^{n}\sigma_k^2\Big) \\ \Leftrightarrow \quad & P(|\mathbb{E}[Y|\mathcal{F}_n] - \mathbb{E}[Y|\mathcal{F}_0]| \geq \alpha) \leq 2\exp\Big(-\alpha^2/2\sum_{k=1}^{n}\sigma_k^2\Big) \\ \Leftrightarrow \quad & P(|Y - \mathbb{E}[Y]| \geq \alpha) \leq 2\exp\Big(-\alpha^2/2\sum_{k=1}^{n}\sigma_k^2\Big). \end{aligned}$$

□

Man hat nun sehr gut gesehen, wie man mit Hilfe der Azuma-Hoeffding-Ungleichung auf ein sehr ähnliches Resultat für nicht-lineare Funktionen schließen kann. Es existiert aber eine noch kleinere obere Schranke für die Wahrscheinlichkeit $P(|Y - \mathbb{E}[Y]| \geq \alpha)$. Diese Abschätzung wird als McDiarmid-Ungleichung bezeichnet. Bevor wir zu der Betrachtung des Satzes kommen, beweisen wir noch ein Lemma, welches wir für den späteren Beweis benötigen.

Lemma 2.6. *(Hoeffding-Lemma)*
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, eine Teil-σ-Algebra $\mathcal{B} \subset \mathcal{A}$ und eine reellwertige Zufallsvariable X mit $a \leq X \leq b$ P-fast sicher $(a, b \in \mathbb{R})$ und bedingtem Erwartungswert $\mathbb{E}[X|\mathcal{B}] = 0$. Dann gilt für jedes $\theta > 0$:

$$\mathbb{E}[e^{\theta X}|\mathcal{B}] \leq \exp\Big(\frac{\theta^2 \cdot (b-a)^2}{8}\Big). \tag{2.5}$$

Beweis. Da $\varphi(x) = e^{\theta x}$ eine konvexe Funktion ist, folgt mit $A = -a$ und $B = b$ aus Korollar 1.4:

$$\begin{aligned} \mathbb{E}[e^{\theta X}|\mathcal{B}] &\leq \frac{B}{A+B} \cdot \varphi(-A) + \frac{A}{A+B} \cdot \varphi(B) \\ &= \frac{b}{b-a} \cdot e^{\theta a} - \frac{a}{b-a} \cdot e^{\theta b}. \end{aligned}$$

Mit $p = -\frac{a}{b-a}$ (Beachte: $a < 0$, da $\mathbb{E}[X|\mathcal{B}] = 0$, und deshalb $p \in [0, 1]$) und $1 - p = \frac{b}{b-a}$ erhält man:

$$\begin{aligned} &= (1-p)e^{\theta a} + pe^{\theta b} \\ &= (1 - p + pe^{\theta b - \theta a})e^{\theta a}. \end{aligned}$$

Benutzt man die Äquivalenz $p = -\frac{a}{b-a} \Leftrightarrow a = -p(b-a)$ und setzt dies für a ein, folgt:

$$= (1 - p + pe^{\theta(b-a)})e^{-\theta p(b-a)}.$$

Definiere nun $u = \theta(b-a)$ und $\phi(x) = -px + \ln(1 - p + pe^x)$. Benutze nun $\phi(u) = -\theta p(b-a) + \ln(1 - p + pe^{\theta(b-a)})$:

$$\begin{aligned} &= e^{-p\theta(b-a)+\ln(1-p+pe^{\theta(b-a)})} \\ &= e^{\phi(u)}. \end{aligned}$$

Im Folgenden versuchen wir eine obere Schranke für $\phi(u)$ zu finden. Wir betrachten dazu das Taylor-Polynom 1. Grades von $\phi(x)$ mit Entwicklungsstelle 0. Für ein $z \in [0, u]$ gilt:

$$\phi(u) = \phi(0) + \phi'(0)u + \frac{1}{2}\phi''(z)u^2$$

und als obere Grenze:

$$\phi(u) \leq \phi(0) + \phi'(0)u + \sup_{z\in[0,u]} \frac{1}{2}\phi''(z)u^2.$$

Die erste und zweite Ableitung von $\phi(x)$ sind:

$$\phi'(x) = -p + \frac{pe^x}{1 - p + pe^x}$$

$$\begin{aligned} \phi''(x) = \frac{pe^x \cdot (1 - p + pe^x) - (pe^x)^2}{(1 - p + pe^x)^2} &= \frac{pe^x - p^2e^x + p^2e^{2x} - p^2e^{2x}}{(1 - p + pe^x)^2} \\ &= \frac{p(1-p)e^x}{(1 - p + pe^x)^2}. \end{aligned}$$

Es ist leicht zu erkennen, dass $\phi(0) = \ln(1) = 0$ und $\phi'(0) = -p + p = 0$. Nun muss noch $\phi''(x)$ maximiert werden. Mit der Substitution $v = e^x$ ergibt sich:

$$\begin{aligned} \frac{d}{dv}\ \frac{p(1-p)v}{(1-p+pv)^2} &= \frac{p(1-p)(1-p+pv)^2 - 2p(1-p+pv)p(1-p)v}{(1-p+pv)^4} \\ &= \frac{p(1-p)(1-p+pv) - 2p^2(1-p)v}{(1-p+pv)^3} \\ &= \frac{p(1-p) - p^2(1-p) + p^2(1-p)v - 2p^2(1-p)v}{(1-p+pv)^3} \\ &= \frac{p(1-p) - p^2(1-p) - p^2(1-p)v}{(1-p+pv)^3} \\ &= \frac{p(1-p)(1-p-pv)}{(1-p+pv)^3}. \end{aligned}$$

Setzt man den ermittelten Ausdruck $\frac{d}{dv}\frac{p(1-p)v}{(1-p+pv)^2} = 0$, erhält man für $v = \frac{1-p}{p}$ (und somit nach Definition $e^x = \frac{1-p}{p}$) ein Extrempunkt. Aufgrund der Tatsache, dass $\phi''(x)$ konkav ist (da für $v > 0$ eine lineare durch eine quadratische Funktion geteilt wird), handelt es sich um ein Maximum.
Für $\phi''(x)$ folgt daher:

$$\phi''(x) \leq \phi''\Big(\ln\Big(\frac{1-p}{p}\Big)\Big) = \frac{p(1-p)\frac{1-p}{p}}{(1-p+p\ \frac{1-p}{p})^2} = \frac{(1-p)^2}{4(1-p)^2} = \frac{1}{4}.$$

Es ergibt sich also nun für $\phi(u)$ aufgrund der Taylorentwicklung und wegen $u = \theta(b-a)$ die obere Schranke:

$$\phi(u) \leq 0 + 0u + \frac{1}{2}u^2\frac{1}{4} = \frac{1}{8}u^2 = \frac{1}{8}\theta^2(b-a)^2.$$

Und insgesamt für die Ungleichung:

$$\mathbb{E}[e^{\theta X}|\mathcal{B}] \leq \exp(\phi(u)) \leq \exp\Big(\frac{\theta^2 \cdot (b-a)^2}{8}\Big).$$

$\square$

Bemerkung 2.7.
Betrachtet man den Fall, dass $b = -a$ ist, folgt aus dem Hoeffding-Lemma 2.6 das gleiche Resultat, wie in Korollar 1.6.
Begründung:

$$\mathbb{E}[e^{\theta X}|\mathcal{B}] \leq \exp\Big(\frac{\theta^2 \cdot (-a-a)^2}{8}\Big) = \exp\Big(\frac{\theta^2 \cdot 4a^2}{8}\Big) = \exp\Big(\frac{\theta^2 \cdot a^2}{2}\Big).$$

Im Folgenden werden wir nun die McDiarmid-Ungleichung betrachten, welche eine genauere Abschätzung als Satz 2.5 liefert.

Satz 2.8. *(McDiarmid-Ungleichung)*
Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit $X_i \in \mathcal{X}_i\ \forall\ i = 1, \dots, n$ für beliebige messbare Räume $\mathcal{X}_i$. Angenommen, $f\colon \prod_{i=1}^n \mathcal{X}_i \to \mathbb{R}$ ist „Lipschitz“ im folgenden Sinn: für alle $1 \leq k \leq n$ und alle $x, x' \in \prod_{i=1}^n \mathcal{X}_i$, welche sich nur in der k-ten Koordinate unterscheiden, gilt: $|f(x) - f(x')| \leq \sigma_k$. Sei $Y = f(X_1, \dots, X_n)$. Dann gilt für jedes $\alpha > 0$:

$$P(|Y - \mathbb{E}[Y]| \geq \alpha) \leq 2\exp\Big(-2\alpha^2 / \sum_{k=1}^n \sigma_k^2\Big). \tag{2.6}$$

Beweis. Sei für $k = 1, \dots, n$ wieder $\mathcal{F}_k$ die von den ersten k Zufallsvariablen erzeugte σ-Algebra $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$ und $Y_k = \mathbb{E}[Y|\mathcal{F}_k] = \mathbb{E}[Y|X_1, \dots, X_k]$, sowie $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $Y_0 = \mathbb{E}[Y]$ und $Y_n = Y$.
Wir betrachten zu Beginn die Zufallsvariablen $Y_k - Y_{k-1}$ für $k = 1, \dots, n$. Für ihre bedingten Erwartungswerte gilt aufgrund der Turmeigenschaft:

$$\begin{aligned}\mathbb{E}[Y_k - Y_{k-1}|\mathcal{F}_{k-1}] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_k]|\mathcal{F}_{k-1}] - \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_{k-1}]|\mathcal{F}_{k-1}] &= \mathbb{E}[Y|\mathcal{F}_{k-1}] - \mathbb{E}[Y|\mathcal{F}_{k-1}] \\ &= Y_{k-1} - Y_{k-1} = 0.\end{aligned}$$

Definiert man:

$$\begin{aligned}U_k &= \sup_u \{\mathbb{E}[Y|X_1, \dots, X_{k-1}, u] - \mathbb{E}[Y|X_1, \dots, X_{k-1}]\} \\ L_k &= \inf_l \{\mathbb{E}[Y|X_1, \dots, X_{k-1}, l] - \mathbb{E}[Y|X_1, \dots, X_{k-1}]\}\end{aligned}$$

so gilt:

$$\begin{aligned}& L_k \leq \mathbb{E}[Y|X_1, \dots, X_k] - \mathbb{E}[Y|X_1, \dots, X_{k-1}] \leq U_k \\ \Leftrightarrow \quad & \qquad\qquad L_k \leq Y_k - Y_{k-1} \leq U_k.\end{aligned}$$

Da $X_1, \dots, X_n$ unabhängig sind, folgt mit Satz 2.4:

$$\begin{aligned}U_k - L_k &= \sup_u \{\mathbb{E}[Y|X_1, \dots, X_{k-1}, u]\} - \inf_l \{\mathbb{E}[Y|X_1, \dots, X_{k-1}, l]\} \\ &= \sup_{u,l} \{\mathbb{E}[Y|X_1, \dots, X_{k-1}, u] - \mathbb{E}[Y|X_1, \dots, X_{k-1}, l]\} \\ &= \sup_{u,l} \{\mathbb{E}[f(X_1, \dots, X_n)|X_1, \dots, X_{k-1}, u] - \mathbb{E}[f(X_1, \dots, X_n)|X_1, \dots, X_{k-1}, l]\} \\ &\leq \sup_{u,l} \Big\{ \int f(X_1, \dots, X_{k-1}, u, x_{k+1}, \dots, x_n) \prod_{j=k+1}^n P^{X_j}(dx_j) \\ &\qquad\qquad - \int f(X_1, \dots, X_{k-1}, l, x_{k+1}, \dots, x_n) \prod_{j=k+1}^n P^{X_j}(dx_j) \Big\} \\ &\leq \sup_{u,l} \Big\{ \int f(X_1, \dots, X_{k-1}, u, x_{k+1}, \dots, x_n) \\ &\qquad\qquad - f(X_1, \dots, X_{k-1}, l, x_{k+1}, \dots, x_n) \prod_{j=k+1}^n P^{X_j}(dx_j) \Big\}.\end{aligned}$$

Mit der Voraussetzung, dass $|f(x) - f(x')| \leq \sigma_k$ für beliebige $x, x' \in \prod_{i=1}^n \mathcal{X}_i$, welche sich nur in der k-ten Koordinate unterscheiden, gilt, folgt:

$$\leq \int \sigma_k \prod_{j=k+1}^n P^{X_j}(dx_j) = \sigma_k.$$

Insgesamt folgt also für $Y_k - Y_{k-1}$ und beliebiges $\theta > 0$ mit Lemma 2.6:

$$\mathbb{E}[e^{\theta(Y_k - Y_{k-1})}|\mathcal{F}_{k-1}] \leq \exp\Big(\frac{\theta^2 \cdot (U_k - L_k)^2}{8}\Big) \leq \exp\Big(\frac{\theta^2 \cdot \sigma_k^2}{8}\Big).$$

Mit Hilfe dieser Abschätzung können wir die gewünschte Ungleichung relativ leicht beweisen. Dazu benutzen wir zuerst die Markow-Ungleichung mit $h(x) = e^{\theta x}$:

$$\begin{aligned} P(Y - \mathbb{E}[Y] \geq \alpha) &\leq \frac{\mathbb{E}[e^{\theta(Y - \mathbb{E}[Y])}]}{e^{\theta\alpha}} \\ &= e^{-\theta\alpha} \cdot \mathbb{E}[e^{\theta(Y_n - Y_0)}]. \end{aligned}$$

Unter Verwendung der Teleskopsumme und der Turmeigenschaft für bedingte Erwartungswerte gilt weiter:

$$\begin{aligned} &= e^{-\theta\alpha} \cdot \mathbb{E}[e^{\theta \sum_{j=1}^{n}(Y_j - Y_{j-1})}] \\ &= e^{-\theta\alpha} \cdot \mathbb{E}[\mathbb{E}[e^{\theta \sum_{j=1}^{n}(Y_j - Y_{j-1})}|\mathcal{F}_{n-1}]] \\ &= e^{-\theta\alpha} \cdot \mathbb{E}[e^{\theta \sum_{j=1}^{n-1}(Y_j - Y_{j-1})} \cdot \mathbb{E}[e^{\theta(Y_n - Y_{n-1})}|\mathcal{F}_{n-1}]]. \end{aligned}$$

Und nun mit der Abschätzung von oben:

$$\begin{aligned} &\leq e^{-\theta\alpha} \cdot \mathbb{E}[e^{\theta \sum_{j=1}^{n-1}(Y_j - Y_{j-1})} \cdot e^{\frac{\theta^2\sigma_n^2}{8}}] \\ &\leq e^{-\theta\alpha} \cdot e^{\frac{\theta^2\sigma_n^2}{8}} \cdot \mathbb{E}[e^{\theta \sum_{j=1}^{n-1}(Y_j - Y_{j-1})}]. \end{aligned}$$

Führt man dieses Verfahren induktiv weiter, erhält man am Ende:

$$\begin{aligned} P(Y - \mathbb{E}[Y] \geq \alpha) &\leq \exp(-\theta\alpha) \cdot \prod_{j=1}^{n} \exp\Big(\frac{\theta^2\sigma_j^2}{8}\Big) \\ &= \exp\Big(-\theta\alpha + \sum_{j=1}^{n} \frac{\theta^2\sigma_j^2}{8}\Big). \end{aligned}$$

Im letzten Schritt bestimmt man den Punkt θ' der Funktion $g(\theta) = -\theta\alpha + \sum_{j=1}^{n} \frac{\theta^2\sigma_j^2}{8}$, bei dem sie ihr Minimum hat. Da die Exponentialfunktion streng monoton wachsend ist, hat $e^{g(\theta)}$ sein Minimum auch an der Stelle $e^{\theta'}$. θ' erhält man durch:

$$\begin{aligned} g'(\theta) &= 0 \\ \Leftrightarrow -\alpha + 2\theta \cdot \sum_{j=1}^{n} \frac{\sigma_j^2}{8} &= 0 \\ \Leftrightarrow \qquad \theta \cdot \sum_{j=1}^{n} \frac{\sigma_j^2}{4} &= \alpha \\ \Leftrightarrow \qquad \theta &= \frac{4\alpha}{\sum_{j=1}^{n} \sigma_j^2}. \end{aligned}$$

Da

$$g''(\theta) = \sum_{j=1}^{n} \frac{\sigma_j^2}{4} > 0$$

gilt, ist an der Stelle $\theta' = 4\alpha / \sum_{j=1}^{n} \sigma_j^2$ ein Minimum.
Setzt man dieses θ' nun in die obige Ungleichung ein, erhält man:

$$\begin{aligned}
P(Y - \mathbb{E}[Y] \geq \alpha) &\leq \exp\Big(\frac{-4\alpha^2}{\sum_{j=1}^{n} \sigma_j^2} + \sum_{j=1}^{n} \frac{(4\alpha)^2 \sigma_j^2}{8(\sum_{j=1}^{n} \sigma_j^2)^2} \Big) \\
&= \exp\Big(\frac{-4\alpha^2}{\sum_{j=1}^{n} \sigma_j^2} + \sum_{j=1}^{n} \frac{16\alpha^2 \sigma_j^2}{8(\sum_{j=1}^{n} \sigma_j^2)^2} \Big) \\
&= \exp\Big(\frac{-4\alpha^2}{\sum_{j=1}^{n} \sigma_j^2} + \frac{2\alpha^2}{\sum_{j=1}^{n} \sigma_j^2} \Big) \\
&= \exp\Big(\frac{-2\alpha^2}{\sum_{j=1}^{n} \sigma_j^2} \Big).
\end{aligned}$$

Wenn f die Voraussetzungen erfüllt, so genügt auch $-f$ diesen Bedingungen, da für beliebige Folgen $x, x' \in \prod_{i=1}^{n} \mathcal{X}_i$, welche sich nur in der k-ten Koordinate unterscheiden, gilt:

$$|-f(x) - (-f(x'))| = |-f(x) + f(x')| = |f(x) - f(x')| \leq \sigma_k.$$

Es folgt also für $-Y = -f(X_1, \ldots, X_n)$ und beliebiges $\alpha > 0$:

$$P(-Y - \mathbb{E}[-Y] \geq \alpha) = P(-Y + \mathbb{E}[Y] \geq \alpha) \leq \exp\Big(-2\alpha^2 / \sum_{j=1}^{n} \sigma_j^2 \Big).$$

Und insgesamt:

$$P(|Y - \mathbb{E}[Y]| \geq \alpha) = P(Y - \mathbb{E}[Y] \geq \alpha) + P(-Y + \mathbb{E}[Y] \geq \alpha) \leq 2\exp\Big(-2\alpha^2 / \sum_{j=1}^{n} \sigma_j^2 \Big).$$

$\square$

Bemerkung 2.9.
Sowohl in Satz 2.5 als auch in Satz 2.8 werden dieselben Voraussetzungen gestellt, aber man erhält unterschiedliche Ungleichungen. Dies geschieht aufgrund der unterschiedlichen Beweisstruktur.

In beiden Fällen wird zu Beginn das Martingal $(Y_i)_{i=0,...n}$ konstruiert und dann der Abstand $\xi_k = Y_k - Y_{k-1} \; \forall \; k = 1, \ldots, n$ untersucht.
In Satz 2.5 erhält man im Beweis eine Einschränkung des Betrages von ξ_k durch σ_k, also $|\xi_k| \leq \sigma_k$ bzw. $\sigma_k \leq \xi_k \leq \sigma_k$. Dagegen werden in Satz 2.8 Grenzen L_k und U_k gefunden, sodass $L_k \leq \xi_k \leq U_k$ und $U_k - L_k \leq \sigma_k$ gilt. Das Intervall, welches den Abstand ξ_k beschränkt, ist im Beweis von Satz 2.5 demzufolge doppelt so groß, wie in dem von Satz 2.8. Daher kommt am Ende der Unterschied in den Ungleichungen.
Eine genauere Eingrenzung des Abstandes ξ_k bewirkt also eine genauere Abschätzung der Wahrscheinlickeit $P(|Y - \mathbb{E}[Y]| \geq \alpha)$ für ein beliebiges $\alpha > 0$, was auch logisch erscheint.

Bemerkung 2.10.
In dem zugrunde liegenden Paper von Steven P. Lalley wurde nur die McDiarmid-Ungleichung aus Satz 2.8 betrachtet. Der Beweis dazu hat aber nicht gepasst, da der aus Satz 2.5 verwendet wurde. Ich habe mich deswegen dazu entschieden, zu Beginn eine schwächere Ungleichung mit Hilfe der Azuma-Hoeffding-Ungleichung zu beweisen (Satz 2.5) und dann den Beweis von Satz 2.8 präzise mit Hilfe des Hoeffding-Lemmas zu führen.

3 Gauß-Konzentration

Im folgenden Korollar wollen wir eine Ungleichung für standardnormalverteilte Zufallsvariablen betrachten, welche sich aus der McDiarmid-Ungleichung herleiten lässt.

Korollar 3.1.
Seien $X_1, \ldots, X_n$ unabhängige, standardnormalverteilte Zufallsvariablen und sei $g \colon \mathbb{R}^n \to \mathbb{R}$ stetig und Lipschitz in jeder Variable mit Lipschitz-Konstante 1, d.h. für jedes $j = 1, \ldots, n$ und $\forall\ x_1, \ldots, x_n, x'_j \in \mathbb{R}$ gilt:

$$|g(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n) - g(x_1, \ldots, x_{j-1}, x'_j, x_{j+1}, \ldots, x_n)| \leq |x_j - x'_j|.$$

Setze $Y = g(X_1, \ldots, X_n)$. Dann gilt für beliebiges $\alpha > 0$:

$$P(|Y - \mathbb{E}[Y]| \geq \alpha) \leq 2 \exp(-\alpha^2/2n). \tag{3.1}$$

Beweis. Wir betrachten zu Beginn unabhängige, identisch verteilte Zufallsvariablen $Y_1, \ldots, Y_{m \cdot n}$ mit $P(Y_i = 1) = P(Y_i = -1) = 0,5$ für alle $i = 1, \ldots, m \cdot n$ und $m, n \in \mathbb{N}$. Weiter sei die Funktion

$$\begin{aligned} x_m \colon \mathbb{R}^{m \cdot n} &\to \mathbb{R}^n \\ (y_1, \ldots, y_{m \cdot n}) &\mapsto \Big(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i, \frac{1}{\sqrt{m}} \sum_{i=m+1}^{2m} y_i, \ldots, \frac{1}{\sqrt{m}} \sum_{i=(n-1)m+1}^{nm} y_i \Big) \end{aligned}$$

gegeben. Definiere nun $f_m \colon \{-1, 1\}^{m \cdot n} \to \mathbb{R}$ mit $f_m(y) = g(x_m(y))$. Für $y, y' \in \{-1, 1\}^{m \cdot n}$, welche sich nur in einer Koordinate k unterscheiden, wobei $jm + 1 \leq k \leq (j + 1)m$ für ein $j \in \{0, 1, \ldots, n-1\}$ gelten muss, gilt aufgrund der Lipschitz-Eigenschaft von g:

$$\begin{aligned}
|f_m(y) - f_m(y')| &= |g(x_m(y)) - g(x_m(y'))| \\
&= \Big| g\Big(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i, \ldots, \frac{1}{\sqrt{m}} \sum_{i=jm+1}^{(j+1)m} y_i, \ldots, \frac{1}{\sqrt{m}} \sum_{i=(n-1)m+1}^{nm} y_i\Big) \\
&\quad - g\Big(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i, \ldots, \frac{1}{\sqrt{m}} \sum_{\substack{i=jm+1,\\ i\neq k}}^{(j+1)m} y_i + y_k', \ldots, \frac{1}{\sqrt{m}} \sum_{i=(n-1)m+1}^{nm} y_i\Big)\Big| \\
&\leq \Big| \frac{1}{\sqrt{m}} \sum_{i=jm+1}^{(j+1)m} y_i - \Big(\frac{1}{\sqrt{m}} \sum_{\substack{i=jm+1,\\ i\neq k}}^{(j+1)m} y_i + y_k'\Big)\Big| \\
&= \Big| \frac{1}{\sqrt{m}}(y_k - y_k')\Big| = \frac{2}{\sqrt{m}}.
\end{aligned}$$

Aus der McDiarmid-Ungleichung folgt jetzt für die unabhängigen, identisch verteilten Zufallsvariablen $Y_1, \ldots, Y_{m \cdot n}$, das für beliebiges $\alpha > 0$ gilt:

$$\begin{aligned}
P(|f_m(Y_1, \ldots, Y_{m\cdot n}) - \mathbb{E}[f_m(Y_1, \ldots, Y_{m\cdot n})]| \geq \alpha) &\leq 2\exp\Big(-2\alpha^2/nm\Big(\frac{2}{\sqrt{m}}\Big)^2\Big) \\
&= 2\exp(-2\alpha^2/4n) = 2\exp(-\alpha^2/2n).
\end{aligned}$$

Zum Schluss stellt man nun den Zusammenhang zu standardnormalverteilten Zufallsvariablen her. Dazu benutzen wir noch einmal Korollar 1.10. Die Zufallsvariablen $Y_1, \ldots, Y_{m\cdot n}$ haben den Erwartungswert 0 und die Varianz 1, da $\forall\ i \in \{1, \ldots, m \cdot n\}$ $\mathbb{E}[Y_i] = 0,5 \cdot 1 + 0,5 \cdot (-1) = 0$ und $Var(Y_i) = 0,5 \cdot 1^2 + 0,5 \cdot (-1)^2 = 1$. Das genannte Korollar impliziert nun:

$$\lim_{m\to\infty} P\Big(\frac{\sum_{k=1}^{m}(Y_k - \mathbb{E}[Y_1])}{\sqrt{mVar(Y_1)}} \leq y\Big) = \lim_{m\to\infty} P\Big(\frac{1}{\sqrt{m}} \cdot \sum_{k=1}^{m} Y_k \leq y\Big) = \Phi(y)\ ,$$

wobei Φ die Verteilungsfunktion einer standardnormalverteilten Zufallsvariable ist. Also konvergiert für $m \to \infty$ $\quad x_m(Y_1, \ldots, Y_{m\cdot n})$ in Verteilung gegen $(X_1, \ldots, X_n)$, wobei die X_i standardnormalverteilt sind, und somit auch $f_m(Y_1, \ldots, Y_{m\cdot n})$ gegen $g(X_1, \ldots, X_n)$. Insgesamt gilt also:

$$\begin{aligned}
&P(|g(X_1, \ldots, X_n) - \mathbb{E}[g(X_1, \ldots, X_n)]| \geq \alpha) \\
&= \lim_{m\to\infty} P(|f_m(Y_1, \ldots, Y_{m\cdot n}) - \mathbb{E}[f_m(Y_1, \ldots, Y_{m\cdot n})]| \geq \alpha) \\
&\leq 2\exp(-\alpha^2/2n).
\end{aligned}$$

□

Beispiel 3.2.
In diesem Beispiel wird für zwei konkrete Funktionen g_1 und g_2 die Ungleichung aus Korollar 3.1 veranschaulicht. Wir betrachten dazu einmal die Summe der Zufallsvariablen und einmal die Summe der Beträge. Deshalb werden g_1 und g_2 wie folgt definiert:

$$\begin{aligned} g_1 \colon \mathbb{R}^n &\to \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto \sum_{i=1}^{n} x_i \\ g_2 \colon \mathbb{R}^n &\to \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto \sum_{i=1}^{n} |x_i|. \end{aligned}$$

Beide Funktionen sind offensichtlich stetig und erfüllen die Lipschitz-Bedingung in jeder Variable, da $\forall\ j = 1, \dots, n$ und $x'_j \in \mathbb{R}$ gilt:

$$\begin{aligned} |g_1(x_1, \dots, x_n) - g_1(x_1, \dots, x_{j-1}, x'_j, x_{j+1}, \dots, x_n)| &= \Big| \sum_{i=1}^{n} x_i - \Big(\sum_{\substack{i=1,\\ i \neq j}}^{n} x_i + x'_j \Big) \Big| \\ &= |x_j - x'_j| \\ |g_2(x_1, \dots, x_n) - g_2(x_1, \dots, x_{j-1}, x'_j, x_{j+1}, \dots, x_n)| &= \Big| \sum_{i=1}^{n} |x_i| - \Big(\sum_{\substack{i=1,\\ i \neq j}}^{n} |x_i| + |x'_j| \Big) \Big| \\ &= \big| |x_j| - |x'_j| \big| \leq |x_j - x'_j|. \end{aligned}$$

Seien nun n unabhängige, standardnormalverteilte Zufallsvariablen $X_1, \dots, X_n$ gegeben. Setzt man $Y_n^1 = g_1(X_1, \dots, X_n)$ und $Y_n^2 = g_2(X_1, \dots, X_n)$, so gilt für die Erwartungswerte dieser Zufallsvariablen:

$$\begin{aligned} \mathbb{E}[Y_n^1] = \mathbb{E}[g_1(X_1, \dots, X_n)] &= \mathbb{E}\Big[\sum_{i=1}^{n} X_i \Big] \\ &= \sum_{i=1}^{n} \mathbb{E}[X_i] = 0 \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[Y_n^2] &= \mathbb{E}[g_2(X_1,\dots,X_n)] = \mathbb{E}\Big[\sum_{i=1}^{n} |X_i|\Big] \\
&= \sum_{i=1}^{n} \mathbb{E}[|X_i|] = n \cdot \mathbb{E}[|X_1|] = n \cdot \int\limits_{-\infty}^{\infty} |x| \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big) dx \\
&= n \cdot \int\limits_{0}^{\infty} 2x \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(\frac{-x^2}{2}\Big) dx = n \cdot \frac{2}{\sqrt{2\pi}} \cdot \int\limits_{0}^{\infty} x \cdot \exp\Big(\frac{-x^2}{2}\Big) dx \\
&= n \cdot \frac{2}{\sqrt{2\pi}} \cdot \Big(-\exp\Big(\frac{-x^2}{2}\Big)\Big)\Big|_0^{\infty} = n \cdot \frac{2}{\sqrt{2\pi}} \cdot (0-(-1)) \\
&= n \cdot \frac{2}{\sqrt{2\pi}}.
\end{aligned}
$$

Wir simulieren nun in einer R-Umgebung n standardnormalverteilte Zufallsvariablen und bestimmen dann $Y_n^1 - \mathbb{E}[Y_n^1]$ bzw. $Y_n^2 - \mathbb{E}[Y_n^2]$. Anschließend bestimmen wir für ein beliebiges $t > 0$ die Wahrscheinlichkeit $P(|Y_n^1 - \mathbb{E}[Y_n^1]| \geq t)$ bzw. $P(|Y_n^2 - \mathbb{E}[Y_n^2]| \geq t)$. Damit man eine aussagekräftige Wahrscheinlichkeit erhält, wiederholen wir diesen Vorgang und bilden dann den Mittelwert. Anschließend kann man diese empirisch ermittelten Wahrscheinlichkeiten mit der oberen Schranke $2\exp(-t^2/2n)$ vergleichen. Dieses gesamte Vorgehen wiederholen wir wiederum mehrmals, um wirklich ein aussagekräftiges Schaubild zu erhalten.
Als erstes betrachten wir nun Y_{10}^1 und $t = 6$ bzw. $t = 4$. Man kann erkennen, dass für ein größeres t die obere Schranke kleiner ist, was auch aufgrund der Definition logisch ist. Alle Wahrscheinlichkeiten liegen klar unterhalb der Grenze, welche immer rot in den Diagrammen gekennzeichnet ist. Umso größer t wird, umso kleiner werden selbstverständlich die Wahrscheinlichkeiten, da größere Y_n^1 bzw. größere Abstände zwischen Y_n^2 und $\mathbb{E}[Y_n^2]$ seltener auftreten.

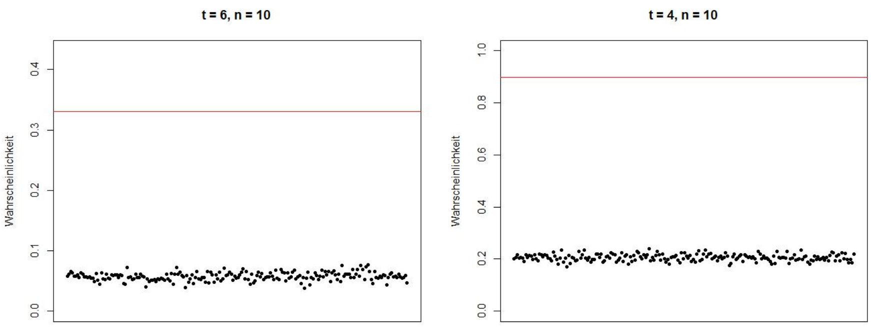

Abbildung 3.1

Als nächstes betrachten wir in Abbildung 3.2 Y_{100}^1 und $t = 25$ bzw. $t = 20$. Es ergeben sich eigentlich die selben Resultate wie in Abbildung 3.1. Die Wahrscheinlichkeiten sind insgesamt kleiner geworden, da sich Y_n^1 durch eine sehr große Anzahl von Zufallsvariablen $X_1, \ldots, X_n$ immer besser dem Erwartungswert annähert.

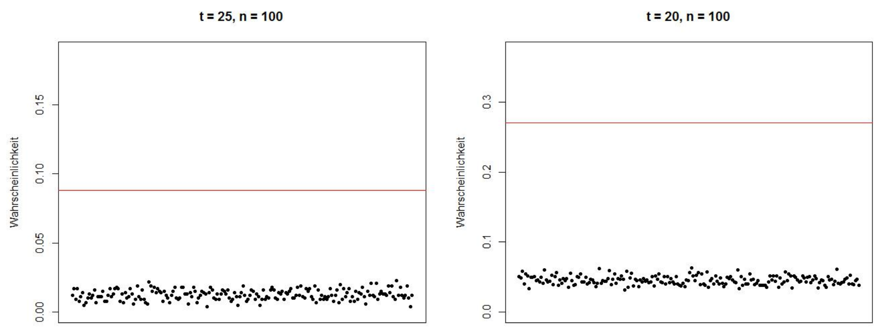

Abbildung 3.2

Zum Schluss schauen wir uns die Zufallsvariablen, welche aus der Summe der Beträge der standardnormalverteilten Zufallsvariablen entstehen, an. Dazu werfen wir einen Blick auf Y_{100}^2 und $t = 13$, sowie Y_{10}^2 und $t = 4$. Es ist auffällig, dass die Wahrscheinlichkeiten in beiden Fällen sehr nah bei 0 liegen. Dies ist dadurch begründet, dass für die Varianz $Var(|X|)$ einer standardnormalverteilten Zufallsvariable X gilt:

$$Var(|X|) = \mathbb{E}[|X|^2] - \mathbb{E}[|X|]^2 = \mathbb{E}[X^2] - \left(\frac{2}{\sqrt{2\pi}}\right)^2 = 1 - \frac{4}{2\pi} \approx 0,3634$$

und somit Y_n^2 auch schon für kleinere n sehr nah an $\mathbb{E}[Y_n^2]$ liegt.

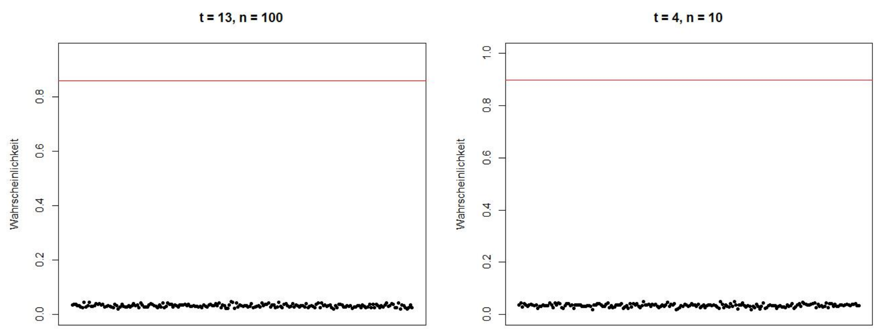

Abbildung 3.3

Wie man gesehen hat, wurde für die vorherige Ungleichung in Korollar 3.1 die Lipschitz-Stetigkeit in jeder einzelnen Variablen vorausgesetzt. Zum Abschluss möchten wir uns nun eine ähnliche Ungleichung für standardnormalverteilte Zufallsvariablen anschauen, bei der aber die Funktion die normale Lipschitz-Bedingung bzgl. der euklidischen Metrik erfüllen muss. Dazu betrachten wir als nächstes erst einmal ein Lemma, welches wir für den späteren Beweis brauchen.

Lemma 3.3.
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine reelle Zufallsvariable Y. Wenn Konstanten $C, A < \infty$ existieren, so dass $\mathbb{E}[e^{\lambda Y}] \leq Ce^{A\lambda^2}$ für alle $\lambda > 0$ ist, so gilt für ein beliebiges $\alpha > 0$:

$$P(Y \geq \alpha) \leq C \cdot \exp\Big(\frac{-\alpha^2}{4A}\Big). \tag{3.2}$$

Beweis. Da $h(x) = e^{\lambda x}$ für $\lambda > 0$ monoton wachsend ist, folgt mit der Markow-Ungleichung:

$$P(Y \geq \alpha) \leq \frac{\mathbb{E}[e^{\lambda Y}]}{e^{\lambda\alpha}} \overset{\textit{Voraussetzung}}{\leq} e^{-\lambda\alpha} \cdot Ce^{A\lambda^2} = C \cdot e^{A\lambda^2 - \lambda\alpha}.$$

Die beste Abschätzung erhält man nun, wenn man den Ausdruck $e^{g(\lambda)}$ mit $g(\lambda) = A\lambda^2 - \lambda\alpha$ minimiert. Der Punkt λ', an dem die Funktion $g(\lambda)$ ihr Optimum hat, ist auch die Stelle, an der $e^{g(\lambda)}$ ein Optimum hat, da e^x ja monoton wachsend ist. λ' erhält man durch:

$$\begin{aligned} & g'(\lambda) = 0 \\ \Leftrightarrow\ & 2A\lambda - \alpha = 0 \\ \Leftrightarrow\ & \qquad \lambda = \frac{\alpha}{2A}. \end{aligned}$$

Setzt man nun $\lambda' = \alpha/2A$ in die Ungleichung ein, so erhält man:

$$\begin{aligned} P(Y \geq \alpha) \leq C \cdot \exp(A\lambda'^2 - \lambda'\alpha) &= C \cdot \exp\Big(A\Big(\frac{\alpha}{2A}\Big)^2 - \frac{\alpha}{2A}\alpha\Big) \\ &= C \cdot \exp\Big(A\frac{\alpha^2}{4A^2} - \frac{\alpha^2}{2A}\Big) \\ &= C \cdot \exp\Big(\frac{\alpha^2}{4A} - \frac{2\alpha^2}{4A}\Big) \\ &= C \cdot \exp\Big(\frac{-\alpha^2}{4A}\Big). \end{aligned}$$

□

Wir betrachten nun eine Konzentrationsungleichung für n-dimensionale, standardnormalverteilte Zufallsvektoren und allgemein Lipschitz-stetige Funktionen.

Satz 3.4. *(Gauß-Konzentration)*
Sei X ein n-dimensionaler, standardnormalverteilter Zufallsvektor und sei $F\colon \mathbb{R}^n \to \mathbb{R}$ Lipschitz-stetig bzgl. der euklidischen Metrik $\|\cdot\|$ auf $\mathbb{R}^n$ mit Lipschitz-Konstante 1, d.h für beliebige $x, y \in \mathbb{R}^n$ gilt: $|F(x) - F(y)| \leq \|x - y\|$. Dann gilt für beliebiges $\alpha > 0$:

$$P(F(X) - \mathbb{E}[F(X)] \geq \alpha) \leq \exp(-2\alpha^2/\pi^2) \tag{3.3}$$

bzw.

$$P(|F(X) - \mathbb{E}[F(X)]| \geq \alpha) \leq 2\exp(-2\alpha^2/\pi^2). \tag{3.4}$$

Bemerkung 3.5.
Es ist zu beachten, dass die Ungleichungen in Satz 3.4 nicht explizit von der Dimension n abhängen. Die vorherigen Konzentrationsungleichungen waren hingegen immer, in einer gewissen Weise, von der Anzahl der Zufallsvariablen abhängig.

Beweis von Satz 3.4. Wir beweisen die Ungleichung (3.3) nur für Funktionen F, welche glatt, also unendlich oft stetig differenzierbar, sind und für deren Gradient ∇F gilt: $\|\nabla F\| \leq 1$.
Aus dem Schrankensatz folgt dann für diese Funktionen die Lipschitz-Bedingung $|F(x) - F(y)| \leq \|x - y\|$ für beliebige $x, y \in \mathbb{R}^n$.
Der Beweis beruht auf einem bestimmten Vorgehen, welches Dublikat-Trick genannt wird. Dieses ist im Zusammenhang mit Konzentrationsungleichungen oft sehr nützlich. Die Grundidee dabei ist, eine unabhängige Kopie der Zufallsvariablen oder des Zufallsvektors, welcher in der Ungleichung vorkommt, zu erstellen und dann eine Annahme aufzustellen, welche sowohl das Original als auch die unabhängige Kopie enthält. Sei nun also $X = (X_1, \ldots, X_n)$ ein $N_n(0,\mathrm{I})$-Zufallsvektor. Aus Lemma 3.3 wissen wir, dass es genügt, wenn wir zeigen, dass $\mathbb{E}[e^{\lambda(F(X)-\mathbb{E}[F(X)])}] \leq Ce^{A\lambda^2}$ für $A = \pi^2/8$, $C = 1$ und alle $\lambda > 0$ gilt. Aus dem genannten Lemma folgt dann für ein beliebiges $\alpha > 0$:

$$P(F(X) - \mathbb{E}[F(X)] \geq \alpha) \leq \exp\Big(\frac{-\alpha^2}{4A}\Big) = \exp\Big(\frac{-\alpha^2}{4 \cdot \frac{\pi^2}{8}}\Big) = \exp(-2\alpha^2/\pi^2).$$

Wir betrachten nun im Folgenden also zwei unabhängige, n-dimensionale, standardnormalverteilte Zufallsvektoren X und X'.

Da die Funktion $f(x) = e^{\lambda' x}$ für alle $\lambda' \in \mathbb{R}$ konvex ist, gilt für integrierbare Zufallsvariablen Y gemäß der Jensen-Ungleichung:

$$e^{\lambda' \mathbb{E}[Y]} \leq \mathbb{E}[e^{\lambda' Y}].$$

Da $\mathbb{E}[F(X) - \mathbb{E}[F(X)]] = 0$ ist, gilt also:

$$1 = e^{\lambda' \cdot 0} = e^{\lambda' \mathbb{E}[F(X) - \mathbb{E}[F(X)]]} \leq \mathbb{E}[e^{\lambda'(F(X) - \mathbb{E}[F(X)])}].$$

Für die unabhängige Kopie X' von X gilt diese Ungleichung natürlich gleichermaßen. Weil $\mathbb{E}[F(X)] = \mathbb{E}[F(X')]$ gilt, reicht es nun, $\mathbb{E}[\exp(\lambda F(X) - \lambda F(X'))] \leq Ce^{A\lambda^2}$ für alle $\lambda > 0$ zu zeigen, da

$$\begin{aligned}
& \mathbb{E}[\exp(\lambda F(X) - \lambda F(X'))] \leq Ce^{A\lambda^2} \\
\Leftrightarrow\quad & \mathbb{E}[\exp(\lambda(F(X) - \mathbb{E}[F(X)]) - \lambda(F(X') - \mathbb{E}[F(X')]))] \leq Ce^{A\lambda^2} \\
\Leftrightarrow\quad & \mathbb{E}[\exp(\lambda(F(X) - \mathbb{E}[F(X)])) \cdot \ \exp(-\lambda(F(X') - \mathbb{E}[F(X')]))] \leq Ce^{A\lambda^2} \\
\Leftrightarrow\quad & \mathbb{E}[\exp(\lambda(F(X) - \mathbb{E}[F(X)]))] \cdot \mathbb{E}[\exp(-\lambda(F(X') - \mathbb{E}[F(X')]))] \leq Ce^{A\lambda^2}
\end{aligned}$$

und mit $1 \leq \mathbb{E}[\exp(\lambda'(F(X') - \mathbb{E}[F(X')]))]$ für beliebiges $\lambda' \in \mathbb{R}$ folgt:

$$\Rightarrow\quad \mathbb{E}[\exp(\lambda(F(X) - \mathbb{E}[F(X)]))] \leq Ce^{A\lambda^2}.$$

Im nächsten Schritt versuchen wir die Konstanten $A = \pi^2/8$ und $C = 1$ zu finden. Wir betrachten dazu zu Beginn einen glatten Pfad zwischen $X_0 = X'$ und $X_1 = X$ und die Funkion

$$\begin{aligned}
f \colon \mathbb{R} &\to \mathbb{R} \\
t &\mapsto F(X_t).
\end{aligned}$$

Die Ableitung von f ist demzufolge:

$$f'(t) = \nabla F(X_t)^T \ \frac{dX_t}{dt} \qquad \text{(totales Differential)}.$$

Aufgrund des Fundamentalsatzes der Analysis gilt für $f(t)$ und $f'(t)$:

$$F(X) - F(X') = F(X_1) - F(X_0) = f(1) - f(0) = \int_0^1 f'(t)dt = \int_0^1 \nabla F(X_t)^T \ \frac{dX_t}{dt} \ dt.$$

Der offensichtlichste Pfad X_t wäre der direkte Weg zwischen X' und X, aber es ist besser, wenn wir

$$X_t = \cos(\pi t/2) \cdot X' + \sin(\pi t/2) \cdot X$$

wählen. Für die Ableitung nach t ergibt sich:

$$\begin{aligned} dX_t/dt &= -(\pi/2) \cdot \sin(\pi t/2) \cdot X' + (\pi/2) \cdot \cos(\pi t/2) \cdot X \\ &= (\pi/2) \cdot (-\sin(\pi t/2) \cdot X' + \cos(\pi t/2) \cdot X) =: (\pi/2) \cdot Y_t. \end{aligned}$$

Dieser Pfad ist optimaler, da für jedes $t \in [0,1]$ X_t und Y_t standardnormalverteilte Zufallsvektoren sind und sie zusätzlich noch stochastisch unabhängig sind. Dies wird im Folgenden kurz bewiesen. Beachte dabei, dass wenn X und X' unabhängig sind, auch die einzelnen Koordinaten $X_1, \ldots, X_n, X'_1, \ldots, X'_n$ unabhängig sind.
Für die Erwartungswerte von X_t und Y_t gilt:

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}[\cos(\pi t/2) \cdot X' + \sin(\pi t/2) \cdot X] \\ &= \cos(\pi t/2) \cdot \mathbb{E}[X'] + \sin(\pi t/2) \cdot \mathbb{E}[X] \\ &= \cos(\pi t/2) \cdot 0 + \sin(\pi t/2) \cdot 0 = 0 \\ \mathbb{E}[Y_t] &= \mathbb{E}[-\sin(\pi t/2) \cdot X' + \cos(\pi t/2) \cdot X] \\ &= -\sin(\pi t/2) \cdot \mathbb{E}[X'] + \cos(\pi t/2) \cdot \mathbb{E}[X] \\ &= -\sin(\pi t/2) \cdot 0 + \cos(\pi t/2) \cdot 0 = 0. \end{aligned}$$

Da außerdem für alle $i, j \in \{1, \ldots, n\}, i \neq j$ gilt:

$$\begin{aligned} Var(\cos(\pi t/2) \cdot X'_i + \sin(\pi t/2) \cdot X_i) &= \cos(\pi t/2)^2 \cdot Var(X'_i) + \sin(\pi t/2)^2 \cdot Var(X_i) \\ &= \cos(\pi t/2)^2 + \sin(\pi t/2)^2 = 1 \end{aligned}$$

$$\begin{aligned} Var(\cos(\pi t/2) \cdot X_i - \sin(\pi t/2) \cdot X'_i) &= \cos(\pi t/2)^2 \cdot Var(X_i) + (-\sin(\pi t/2))^2 \cdot Var(X'_i) \\ &= \cos(\pi t/2)^2 + \sin(\pi t/2)^2 = 1 \end{aligned}$$

$$\begin{aligned} &Cov(\cos(\pi t/2) \cdot X'_i + \sin(\pi t/2) \cdot X_i, \cos(\pi t/2) \cdot X'_j + \sin(\pi t/2) \cdot X_j) \\ &= \mathbb{E}[(\cos(\pi t/2) \cdot X'_i + \sin(\pi t/2) \cdot X_i) \cdot (\cos(\pi t/2) \cdot X'_j + \sin(\pi t/2) \cdot X_j)] \\ &\quad - \mathbb{E}[\cos(\pi t/2) \cdot X'_i + \sin(\pi t/2) \cdot X_i] \cdot \mathbb{E}[\cos(\pi t/2) \cdot X'_j + \sin(\pi t/2) \cdot X_j] \\ &= \mathbb{E}[\cos(\pi t/2)^2 \cdot X'_i X'_j + \cos(\pi t/2) \cdot \sin(\pi t/2) \cdot X'_i X_j \\ &\quad + \cos(\pi t/2) \cdot \sin(\pi t/2) \cdot X_i X'_j + \sin(\pi t/2)^2 \cdot X_i X_j] \\ &= \cos(\pi t/2)^2 \cdot \mathbb{E}[X'_i]\mathbb{E}[X'_j] + \cos(\pi t/2) \cdot \sin(\pi t/2) \cdot \mathbb{E}[X'_i]\mathbb{E}[X_j] \\ &\quad + \cos(\pi t/2) \cdot \sin(\pi t/2) \cdot \mathbb{E}[X_i]\mathbb{E}[X'_j] + \sin(\pi t/2)^2 \cdot \mathbb{E}[X_i]\mathbb{E}[X_j] \\ &= 0 \end{aligned}$$

$$
\begin{aligned}
&Cov(-\sin(\pi t/2)\cdot X_i' + \cos(\pi t/2)\cdot X_i, -\sin(\pi t/2)\cdot X_j' + \cos(\pi t/2)\cdot X_j)\\
&= \mathbb{E}[(-\sin(\pi t/2)\cdot X_i' + \cos(\pi t/2)\cdot X_i)\cdot(-\sin(\pi t/2)\cdot X_j' + \cos(\pi t/2)\cdot X_j)]\\
&\quad - \mathbb{E}[-\sin(\pi t/2)\cdot X_i' + \cos(\pi t/2)\cdot X_i]\cdot\mathbb{E}[-\sin(\pi t/2)\cdot X_j' + \cos(\pi t/2)\cdot X_j]\\
&= \mathbb{E}[\sin(\pi t/2)^2\cdot X_i'X_j' - \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot X_i'X_j\\
&\quad - \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot X_iX_j' + \cos(\pi t/2)^2\cdot X_iX_j]\\
&= \sin(\pi t/2)^2\cdot\mathbb{E}[X_i']\mathbb{E}[X_j'] - \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot\mathbb{E}[X_i']\mathbb{E}[X_j]\\
&\quad - \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot\mathbb{E}[X_i]\mathbb{E}[X_j'] + \cos(\pi t/2)^2\cdot\mathbb{E}[X_i]\mathbb{E}[X_j]\\
&= 0,
\end{aligned}
$$

erhält man für die Kovarianzmatrizen von X_t und Y_t die Einheitsmatrix I.
Die Kovarianz von X_t und Y_t ist die Nullmatrix, da

$$
\begin{aligned}
&Cov(\cos(\pi t/2)\cdot X_i' + \sin(\pi t/2)\cdot X_i, -\sin(\pi t/2)\cdot X_j' + \cos(\pi t/2)\cdot X_j)\\
&= \mathbb{E}[(\cos(\pi t/2)\cdot X_i' + \sin(\pi t/2)\cdot X_i)\cdot(-\sin(\pi t/2)\cdot X_j' + \cos(\pi t/2)\cdot X_j)]\\
&\quad - \mathbb{E}[\cos(\pi t/2)\cdot X_i' + \sin(\pi t/2)\cdot X_i]\cdot\mathbb{E}[-\sin(\pi t/2)\cdot X_j' + \cos(\pi t/2)\cdot X_j]\\
&= \mathbb{E}[-\cos(\pi t/2)\cdot\sin(\pi t/2)\cdot X_i'X_j' + \cos(\pi t/2)^2\cdot X_i'X_j\\
&\quad - \sin(\pi t/2)^2\cdot X_iX_j' + \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot X_iX_j]
\end{aligned}
$$

für $i = j$:

$$
\begin{aligned}
&= -\cos(\pi t/2)\cdot\sin(\pi t/2)\cdot\mathbb{E}[X_i'^2] + \cos(\pi t/2)^2\cdot\mathbb{E}[X_i']\mathbb{E}[X_i]\\
&\quad - \sin(\pi t/2)^2\cdot\mathbb{E}[X_i]\mathbb{E}[X_i'] + \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot\mathbb{E}[X_i{}^2]\\
&= -\cos(\pi t/2)\cdot\sin(\pi t/2)\cdot Var(X_i') + \cos(\pi t/2)^2\cdot\mathbb{E}[X_i']\mathbb{E}[X_i]\\
&\quad - \sin(\pi t/2)^2\cdot\mathbb{E}[X_i]\mathbb{E}[X_i'] + \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot Var(X_i)\\
&= \cos(\pi t/2)^2\cdot\mathbb{E}[X_i']\mathbb{E}[X_i] - \sin(\pi t/2)^2\cdot\mathbb{E}[X_i]\mathbb{E}[X_i']\\
&= 0,
\end{aligned}
$$

für $i \neq j$:

$$
\begin{aligned}
&= -\cos(\pi t/2)\cdot\sin(\pi t/2)\cdot\mathbb{E}[X_i']\mathbb{E}[X_j'] + \cos(\pi t/2)^2\cdot\mathbb{E}[X_i']\mathbb{E}[X_j]\\
&\quad - \sin(\pi t/2)^2\cdot\mathbb{E}[X_i]\mathbb{E}[X_j'] + \cos(\pi t/2)\cdot\sin(\pi t/2)\cdot\mathbb{E}[X_i]\mathbb{E}[X_j]\\
&= 0.
\end{aligned}
$$

Aus der Unkorreliertheit von X_t und Y_t folgt direkt die stochastische Unabhängigkeit, weil beide standardnormalverteilte Zufallsvektoren sind.
Betrachten wir nun im Folgenden wieder den Ausdruck $\mathbb{E}[\exp(\lambda F(X) - \lambda F(X'))]$, für den wir die Konstanten $A = \pi^2/8$ und $C = 1$ nachweisen wollen. Ersetzen wir in dem

Ausdruck $F(X)$ und $F(X')$ durch das Integral, erhalten wir:

$$\mathbb{E}[\exp(\lambda F(X) - \lambda F(X'))] = \mathbb{E}\Big[\exp\Big(\lambda \int_0^1 \nabla F(X_t)^T \ \frac{dX_t}{dt} \ dt\Big)\Big].$$

Nun benutzen wir die stetige Variante der Jensenschen Ungleichung mit der konvexen Funktion $g(x) = e^{\lambda x}$:

$$\begin{aligned}
&= \mathbb{E}\Big[g\Big(\int_0^1 f'(t)dt\Big)\Big] \\
&\leq \mathbb{E}\Big[\int_0^1 g(f'(t))dt\Big] \\
&= \mathbb{E}\Big[\int_0^1 \exp\Big(\lambda \nabla F(X_t)^T \ \frac{dX_t}{dt}\Big)dt\Big].
\end{aligned}$$

Ersetzt man nun dX_t/dt durch $(\pi/2)Y_t$ und benutzt den Satz von Fubini folgt:

$$= \int_0^1 \mathbb{E}\Big[\exp\Big(\frac{\lambda\pi}{2} \nabla F(X_t)^T \ Y_t\Big)\Big]dt.$$

Als nächstes untersuchen wir die Zufallsvariable $\nabla F(X_t)^T \ Y_t$ etwas genauer. Bedingt man auf X_t, so hat $\nabla F(X_t)$ konkrete Werte und es gilt deshalb für den Erwartungswert und die Varianz Folgendes:

$$\begin{aligned}
&\mathbb{E}[\nabla F(X_t)^T \ Y_t | X_t] = \nabla F(X_t)^T \cdot 0_v = 0 \\
&Var(\nabla F(X_t)^T \ Y_t | X_t) = (\nabla F(X_t)^2)^T \cdot 1_v = \|\nabla F(X_t)\|^2 \leq 1.
\end{aligned}$$

Die letzte Ungleichung gilt, da nach Voraussetzung $\|\nabla F\| \leq 1$ ist.
Da für die momenterzeugende Funktion einer normalverteilten Zufallsvariable Z mit Erwartungswert μ und Varianz σ^2 gilt, dass $\mathbb{E}[e^{tZ}] = \exp(\mu t + \frac{1}{2}t^2\sigma^2)$ ist, folgt:

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\frac{\lambda\pi}{2} \nabla F(X_t)^T \ Y_t\Big)\Big| X_t\Big] \\
&\quad = \exp\Big(\mathbb{E}[\nabla F(X_t)^T \ Y_t | X_t] \cdot \frac{\lambda\pi}{2} + \frac{1}{2} \cdot \Big(\frac{\lambda\pi}{2}\Big)^2 \cdot Var(\nabla F(X_t)^T \ Y_t | X_t)\Big) \\
&\quad = \exp\Big(\frac{\lambda^2\pi^2}{8} \cdot Var(\nabla F(X_t)^T \ Y_t | X_t)\Big) \\
&\quad \leq \exp\Big(\frac{\lambda^2\pi^2}{8}\Big).
\end{aligned}$$

Da diese Abschätzung für jeden dieser bedingten Erwartungswerte gilt, ist auch

$$\mathbb{E}\Big[\exp\Big(\frac{\lambda\pi}{2}\nabla F(X_t)^T\ Y_t\Big)\Big]\leq \exp\Big(\frac{\lambda^2\pi^2}{8}\Big).$$

Insgesamt erhalten wir also für alle $\lambda > 0$:

$$\begin{aligned}\mathbb{E}[\exp(\lambda F(X) - \lambda F(X'))] &\leq \int_0^1 \mathbb{E}\Big[\exp\Big(\frac{\lambda\pi}{2}\nabla F(X_t)^T\ Y_t\Big)\Big]dt\\ &\leq \int_0^1 \exp\Big(\frac{\lambda^2\pi^2}{8}\Big)dt\\ &= \exp\Big(\frac{\lambda^2\pi^2}{8}\Big)\end{aligned}$$

und damit aufgrund unserer Vorüberlegungen auch:

$$\mathbb{E}[\exp(\lambda(F(X) - \mathbb{E}[F(X)]))] \leq \exp\Big(\frac{\lambda^2\pi^2}{8}\Big).$$

Wir haben nun also die gesuchten Konstanten $C = 1$ und $A = \pi^2/8$ nachgewiesen.

Weil für beliebige $x, y \in \mathbb{R}^n$

$$|-F(x) - (-F(y))| = |-F(x) + F(y)| = |F(x) - F(y)| \leq \|x - y\|$$

gilt, erfüllt auch $-F$ die Lipschitz-Bedingung. Aufgrund dieser Tatsache erhält man zusätzlich:

$$\begin{aligned} & P(-F(X) - \mathbb{E}[-F(X)] \geq \alpha) \leq \exp(-2\alpha^2/\pi^2)\\ \Leftrightarrow\quad & P(-F(X) + \mathbb{E}[F(X)] \geq \alpha) \leq \exp(-2\alpha^2/\pi^2).\end{aligned}$$

Somit gilt für die zweiseitige Abgrenzung:

$$\begin{aligned}P(|F(X) - \mathbb{E}[F(X)]| \geq \alpha) &= P(F(X) - \mathbb{E}[F(X)] \geq \alpha) + P(-F(X) + \mathbb{E}[F(X)] \geq \alpha)\\ &\leq 2\exp(-2\alpha^2/\pi^2).\end{aligned}$$

$\square$

Literaturverzeichnis

[1] Lallay, S. P.: *Concentration Inequalities.* University of Chicago

[2] Vetter, M.: *Skript zum Modul „Wahrscheinlichkeitstheorie“.* WS 13/14

[3] Klenke, A.: *Wahrscheinlichkeitstheorie.* Springer Spektrum, 3. Auflage, 2013

[4] Boucheron, S., Lugosi, G., Massart, P.: *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press, 2013

[5] Duchi, J.: *Probability Bounds.*

[6] Raginsky, M., Sason, I.: *Concentration of Measure Inequalities in Information Theory, Communications, and Coding.* Foundations and Trends in Communications and Information Theory, volume 10, 2013

Erklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit eigenständig verfasst habe. Ich habe mich keiner anderen als der von mir ausgewiesenen Quellen und Hilfsmittel bedient. Diese Arbeit hat in vorliegender oder ähnlicher Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, den 10. September 2014

(Philipp Hermann, Verfasser)