SEMINARARBEIT

Bounds on the Jensen gap and implications for mean-concentrated distributions

ausgeführt am

Institut für
Finanz- und Versicherungsmathematik
TU Wien

unter der Anleitung von

Assoc.Prof. Dipl.-Ing. Dr.techn. Stefan Gerhold

durch

Markus Feyertag
Matrikelnummer: 11778904

Wien im Februar 2020

Inhaltsverzeichnis

1 Einleitung

Diese Seminararbeit befassst sich hauptsächlich mit oberen bzw. unteren Schranken der sogenannten *Jensen-Differenz* $\mathcal{J}(f, X) = \mathbb{E}[f(X)] - f(\mathbb{E}[X])$. Die vorliegende Arbeit basiert in weiten Teilen auf dem Paper *Bounds on the Jensen gap, and implications for mean-concentrated distributions* (Erstveröffentlichung 2017) von X. Gao, M. Sitharam und A. Roitberg sowie auf dem Lehrbuch *Maß- und Wahrscheinlichkeitstheorie* von N. Kusolitsch.

Nach einer Vorstellung der Jensen-Ungleichung und der darauf aufbauenden Jensen-Differenz werden erste Schranken für diese präsentiert und bewiesen. Das erste Hauptresultat zeigt eine obere Schranke, während im zweiten Hauptresultat eine untere Schranke untersucht wird. Für Details dazu sei auf die Kapitel 4 und 5 verwiesen. Im weiteren Verlauf der Arbeit soll näher auf Beispiele und praktische Anwendungsmöglichkeiten eingegangen werden. Diese sind, abgesehen von der Mathematik, in der Mechanik, im Machine Learning Bereich sowie in der Physik vorzufinden. Teile dieses abschließenden Kapitels basieren auf dem Buch *Mathematische Statistik* von L. Rüschendorf.

Diese Arbeit setzt ein grundlegendes Verständnis der Maß- und Wahrscheinlichkeitstheorie voraus, allerdings werden zentrale und relevante Resultate wie beispielsweise die Jensen-Ungleichung eingangs wiederholt.

2 Jensen-Ungleichung und Jensen-Differenz

Um die *Jensen-Differenz* (engl. *Jensen-gap* oder *Jensen-difference*) definieren zu können bzw. um zu verstehen, weshalb sie Jensen–Differenz heißt, sei zunächst an die Jensen-Ungleichung aus der Maß- und Wahrscheinlichkeitstheorie erinnert.

2.1 Jensen-Ungleichung

Satz 2.1 (Jensen-Ungleichung). *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$, $X : \Omega \to (a, b)$ eine $\mathbb{P}$-integrierbare Zufallsvariable und $f : (a, b) \to \mathbb{R}$ eine konvexe Funktion. Dann existiert der Erwartungswert von $f \circ X$ und es gilt*

$$f(\mathbb{E}X) \leq \mathbb{E}[f \circ X].$$

Bemerkung 2.2. Unter Integrierbarkeit verstehen wir Folgendes: Eine messbare Funktion f auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt integrierbar bezüglich μ wenn $\max\{\int f^+ d\mu, \int f^- d\mu\} < \infty$. Dabei ist $f^+ = \max(f, 0)$ und $f^- = \max(-f, 0)$. Für Zufallsvariablen auf Wahrscheinlichkeitsräumen wird das Integral, wenn es existiert, Erwartungswert genannt.

Bevor wir Satz 2.1 beweisen können, benötigen wir zuerst zwei Hilfsresultate.

Lemma 2.3. *Sei $f : (a, b) \to \mathbb{R}$ eine konvexe Funktion. Dann gilt für alle $\boldsymbol{x} = (x_1, x_2)$, $\boldsymbol{y} = (y_1, y_2)$ mit $x_1 < x_2$, $y_1 < y_2$ und $\boldsymbol{x} \leq \boldsymbol{y}$ (d.h. $x_1 \leq y_1$ und $x_2 \leq y_2$)*

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(y_2) - f(y_1)}{y_2 - y_1}.$$

Beweis. Aus $x_1 < x_2 \leq y_2$ folgt mit $\alpha := (y_2 - x_2)/(y_2 - x_1) \in [0, 1]$

$$x_2 = \frac{y_2 - x_2}{y_2 - x_1} x_1 + \frac{x_2 - x_1}{y_2 - x_1} y_2 = \alpha x_1 + (1 - \alpha) y_2.$$

Also gilt wegen der vorausgesetzten Konvexität von f

$$f(x_2) \leq \alpha f(x_1) + (1 - \alpha) f(y_2). \tag{2.1}$$

Analog folgt aus $x_1 \leq y_1 < y_2$ mit $\beta := (y_2 - y_1)/(y_2 - x_1) \in [0, 1]$

$$y_1 = \frac{y_2 - y_1}{y_2 - x_1} x_1 + \frac{y_1 - x_1}{y_2 - x_1} y_2 = \beta x_1 + (1 - \beta) y_2.$$

und somit

$$f(y_1) \leq \beta f(x_1) + (1 - \beta) f(y_2). \tag{2.2}$$

Aus der Ungleichung (2.1) erhält man durch Subtraktion von $f(x_1)$ die Ungleichung $f(x_2) - f(x_1) \leq (1-\alpha)(f(y_2) - f(x_1))$, also nach Division durch $x_2 - x_1$

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(y_2) - f(x_1)}{y_2 - x_1}. \tag{2.3}$$

Aus der Ungleichung (2.2) erhält man durch Multiplikation von (-1) und Addition von $f(y_2)$ die Ungleichung $f(y_2) - f(y_1) \geq \beta(f(y_2) - f(x_1))$, also nach Division durch $y_2 - y_1$

$$\frac{f(y_2) - f(y_1)}{y_2 - y_1} \geq \frac{f(y_2) - f(x_1)}{y_2 - x_1}. \tag{2.4}$$

(2.3) und (2.4) ergeben somit gemeinsam die gewünschte Ungleichung

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(y_2) - f(x_1)}{y_2 - x_1} \leq \frac{f(y_2) - f(y_1)}{y_2 - y_1}.$$

□

Lemma 2.4. *Sei wieder $f : (a,b) \to \mathbb{R}$ konvex. Dann sind die linksseitigen (rechtsseitigen) Differenzenquotienten von f in jedem Punkt $x \in (a,b)$ monoton fallend (wachsend) und es gilt*

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(x)}{z - x}, \ \forall y < x < z$$

Beweis. Die Aussage folgt sofort aus Lemma (2.3). □

Bemerkung 2.5. Sei $f : (a,b) \to \mathbb{R}$ abermals konvex. Lemma (2.4) impliziert, dass in jedem Punkt $x \in (a,b)$ die linksseitige Ableitung $f'_-(x)$ von f existiert, wobei $f'_-(x) := \lim_{y \nearrow x} \frac{f(y)-f(x)}{y-x}$. Ebenso existiert in jedem Punkt $x \in (a,b)$ die rechtsseitige Ableitung $f'_+(x)$ von f, wobei $f'_+(x) := \lim_{y \searrow x} \frac{f(y)-f(x)}{y-x}$. Weiters folgt aus Lemma (2.4), dass $f'_-(x) \leq f'_+(x)\ \forall x \in (a,b)$.

Lemma 2.6. *Sei $f : (a,b) \to \mathbb{R}$ konvex. Dann gilt für alle $x \in (a,b)$*

$$f(y) \geq f(x) + k(y-x) \quad \forall\, y \in (a,b),\ k \in [f'_-(x), f'_+(x)]. \tag{2.5}$$

Beweis. Aus den bisherigen Resultaten folgt, dass für $y < x < z$ gilt

$$\frac{f(x) - f(y)}{x - y} \leq f'_-(x) \leq f'_+(x) \leq \frac{f(z) - f(x)}{z - x}.$$

Aus der rechten Ungleichung erhalten wir durch Multiplikation von $(z - x)$ und Addition von $f(x)$

$$f(z) \geq f(x) + f'_+(z - x) \geq f(x) + k(z - x),\ \forall\, z > x,\ k \leq f'_+(x).$$

Analog erhalten wir durch Multiplikation von $(-1(x-y))$ und Addition von $f(x)$ aus der linken Ungleichung

$$f(y) \geq f(x) + f'_-(x)(y - x) \geq f(x) + k(y - x),\ \forall\, y < x,\ k \geq f'_-(x).$$

Also gilt für jedes $k \in [f'_-(x), f'_+(x)]$

$$f(y) \geq f(x) + k(y - x), \forall\ y \in (a, b)$$

Für $y = x$ ist die Ungleichung dabei trivialerweise erfüllt. $\square$

Mit diesen Hilfsresultaten können wir jetzt den Beweis von Satz (2.1), also der Jensen-Ungleichung, nachholen.

Beweis von Satz (2.1). Für den Fall, dass $a = -\infty$, gilt wegen der vorausgesetzten Integrierbarkeit von X, dass $\mathbb{E}[X] > a$. Falls $a \in \mathbb{R}$, gilt wegen $X(\Omega) \subseteq (a, b)$, dass $X - a > 0$ $\mathbb{P}$-f.s., also $\mathbb{E}[X] > a$. Analog dazu sieht man $\mathbb{E}[X] < b$ und daher $\mathbb{E}[X] \in (a, b)$. Somit gilt $f(\mathbb{E}[X]) \in \mathbb{R}$. Wählt man in der Ungleichung (2.5) $x = \mathbb{E}[X]$ und $y = X(\omega)$, so folgt mit $k = f'_+(x)$

$$f(X(\omega)) \geq f(\mathbb{E}[X]) + f'_+(\mathbb{E}[X])(X(\omega) - \mathbb{E}[X]),\ \forall \omega \in \Omega. \tag{2.6}$$

Nun sind $f(\mathbb{E}[X])$, $f'_+(\mathbb{E}[X])$ und $\mathbb{E}[X]$ Konstanten. Weiters ist laut Voraussetzung $X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. Daher ist die rechte Seite dieser Ungleichung integrierbar. Somit existiert auch das Integral von $f \circ X$ bzgl. $\mathbb{P}$ und Integration von Ungleichung (2.6) liefert

$$\begin{aligned}
\mathbb{E}[f \circ X] &= \int (f \circ X) d\mathbb{P} \\
&\geq \underbrace{\int f(\mathbb{E}[X]) d\mathbb{P}}_{=f(\mathbb{E}[X])} + f'_+(\mathbb{E}[X]) \underbrace{\left(\int X d\mathbb{P} - \int \mathbb{E}[X] d\mathbb{P} \right)}_{=\mathbb{E}[X] - \mathbb{E}[X] = 0} \\
&= f(\mathbb{E}[X])
\end{aligned}$$

$\square$

Bemerkung 2.7. Es kann vorkommen, dass $\int (f \circ X) d\mathbb{P} = \infty$. Die Integrierbarkeit von $f \circ X$ bzgl. $\mathbb{P}$ ist im Kontext von Bemerkung (2.2) zu verstehen.

Jetzt können wir definieren, was man unter der Jensen-Differenz versteht und wieso sie ebenfalls nach dem dänischen Mathematiker Johan Jensen (1859 - 1925) benannt ist.

2.2 Jensen-Differenz

Definition 2.8 (Jensen-Differenz). Sei X eine Zufallsvariable mit Verteilung $\mathcal{P}$ und f eine konvexe oder nicht-konvexe Funktion. Die Differenz der beiden Seiten der Jensen-Ungleichung

$$\mathcal{J}(f, X \sim \mathcal{P}) := \mathbb{E}[f(X)] - f(\mathbb{E}[X])$$

bezeichnen wir als *Jensen-Differenz* der Funktion f bezüglich der Zufallsvariable X.

3 Erste Schranken der Jensen-Differenz

Bemerkung 3.1 (Notation)**.** Wir werden im Folgenden stets $\sigma_p := \sqrt[p]{\mathbb{E}[|X-\mu|^p]}$ als Kurzschreibweise für die p-te Wurzel des p-ten absoluten zentralen Moments sowie $\mu := \mathbb{E}[X]$ als Kurzschreibweise für den Erwartungswert der Zufallsvariable X verwenden.

Einen ersten Zusammenhang zwischen dem Jensen-Fehler $\mathcal{J}(f, X \sim \mathcal{P})$ und den Momenten σ_p der Zufallsvariable X erhalten wir, indem wir uns auf Hölder-stetige Funktionen $f : \mathbb{R} \to \mathbb{R}$ beschränken.

3.1 Hölder-Stetigkeit und erste Schranken

Definition 3.2 (**Hölder-Stetigkeit**)**.** Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt Hölder-stetig zum Exponenten $\alpha \in (0, 1]$, wenn es eine Konstante $M > 0$ gibt, sodass für alle $x, y \in \mathbb{R}$ gilt:

$$|f(x) - f(y)| < M|x - y|^\alpha$$

Bemerkung 3.3. Im Spezialfall $\alpha = 1$ erhält man Lipschitz-Stetigkeit. Außerdem ist jede Hölder-stetige Funktion gleichmäßig stetig: Setzt man für gegebenes $\epsilon > 0$ nämlich $\delta := (\frac{\epsilon}{M})^{1/\alpha}$, so folgt aus $|x - y| \leq \delta$ die Ungleichung $|f(x) - f(y)| \leq \epsilon$.

Lemma 3.4. *Sei $f : \mathbb{R} \to \mathbb{R}$ Hölder-stetig. Dann gilt*

$$|\mathcal{J}(f, X \sim P)| \leq M\sigma_\alpha^\alpha$$

Beweis. Aus der Dreiecksungleichung für Integrale und der Definition der Hölder-Stetigkeit folgt

$$\begin{aligned}|\mathcal{J}(f, X \sim P)| &= \left|\mathbb{E}[f(X)] - f(\mu)\right| \leq \int |f(X) - f(\mu)| d\mathcal{P}(X) \leq M \int |X - \mu|^\alpha d\mathcal{P}(X) \\ &= M\sigma_\alpha^\alpha\end{aligned}$$

$\square$

Bemerkung 3.5. Falls umgekehrt $f(x) - f(\mu) \geq M|x - \mu|^\alpha$ oder $f(\mu) - f(x) \geq M|x - \mu|^\alpha$ gilt, erhalten wir

$$|\mathcal{J}(f, X \sim P)| \geq M\sigma_\alpha^\alpha$$

3.2 Verallgemeinerung der bisherigen Schranken

Im Folgenden werden diese ersten beiden Schranken verallgemeinert. Wir zeigen eine obere und eine untere Schranke der Jensen-Differenz. Dabei bedeutet obere Schranke, dass

$$|\mathcal{J}(f, X \sim P)| \leq A$$

für ein $A \geq 0$, d.h.

$$-A \leq \mathcal{J}(f, X \sim P) \leq A$$

Unter einer unteren Schranke verstehen wir hingegen, dass

$$\mathcal{J}(f, X \sim P) \geq A \text{ oder } -\mathcal{J}(f, X \sim P) \geq A$$

Bevor wir die beiden Hauptresultate dieser Seminararbeit angeben und im Anschluss daran beweisen, fassen wir diese zunächst kurz zusammen.

- Für Funktionen f, die sich $f(\mu)$ für $x \to \mu$ nicht langsamer nähern als $|x-\mu|^\alpha,\ \alpha > 0$ und die für $x \to \pm\infty$ nicht schneller wachsen als $\pm|x|^n,\ n \geq \alpha$, werden wir zeigen, dass

  $$|\mathcal{J}(f, X \sim P)| \leq M(1 + \sigma_n^{n-\alpha})\sigma_n^\alpha$$

  Dabei ist

  $$M := \sup_{x \neq \mu} \frac{|f(x) - f(\mu)|}{|x-\mu|^\alpha + |x-\mu|^n}$$

  Das bedeutet, dass sich $\mathcal{J}(f, X \sim P)$ nicht schneller 0 nähert als σ_n^α für $\sigma_n \to 0$

- Für Funktionen f, die sich $f(\mu)$ für $x \to \mu$ von beiden Richtungen kommend nicht schneller von oben oder von unten nähern als $|x - \mu|^\alpha$ und die für $x \to \infty$ nicht langsamer wachsen als $|x|^\beta,\ 0 \leq \beta \leq \alpha$, werden wir zeigen, dass

  $$|\mathcal{J}(f, X \sim P)| \geq M \frac{\sigma_{\alpha/2}^\alpha}{1 + \sigma_{\alpha-\beta}^{\alpha-\beta}}$$

  Dabei ist

  $$M := \inf_{x \neq \mu} \Big(f(x) - f(\mu)\Big)\left(\frac{1}{|X-\mu|^\beta} + \frac{1}{|X-\mu|^\alpha}\right)$$

  Das bedeutet, dass sich $\mathcal{J}(f, X \sim P)$ nicht schneller 0 nähert als $\sigma_{\alpha/2}^\alpha$ für $\sigma_{\alpha/2} \to 0$, sofern $\sigma_{\alpha-\beta}$ nicht gleichzeitig gegen ∞ geht.

4 Erstes Hauptresultat: Obere Schranke der Jensen-Differenz

4.1 Obere Schranke der Jensen-Differenz

Satz 4.1. *Sei $I \subseteq \mathbb{R}$ ein abgeschlossenes Intervall, X eine Zufallsvariable mit Verteilung $\mathcal{P}$ und Erwartungswert $\mathbb{E}[X] =: \mu \in I$, sowie $f : I \to \mathbb{R}$ eine Funktion, die die folgenden Bedingung erfüllt:*

1. *f ist auf jeder kompakten Teilmenge von I beschränkt*
2. $|f(x) - f(\mu)| = \mathcal{O}(|x-\mu|^\alpha)$ *für* $x \to \mu,\ \alpha > 0$
3. $|f(x)| = \mathcal{O}(|x|^n)$ *für* $x \to \infty,\ n \geq \alpha$

Dann gilt

$$\underbrace{|\mathbb{E}[f(X)] - f(\mu)|}_{=|\mathcal{J}(f,X\sim\mathcal{P})|} \leq M(\sigma_\alpha^\alpha + \sigma_n^n) \leq M(1+\sigma_n^{n-\alpha})\sigma_n^\alpha$$

Dabei ist

$$M := \sup_{x\in I\setminus\{\mu\}} \frac{|f(x)-f(\mu)|}{|x-\mu|^\alpha + |x-\mu|^n}$$

unabhängig von der Verteilung $\mathcal{P}$ von X.

Beweis. Wir definieren zunächst die Funktion g durch $g(x) := \frac{|f(x)-f(\mu)|}{|x-\mu|^\alpha+|x-\mu|^n}$ und zeigen, dass g auf $I \setminus \{\mu\}$ beschränkt ist. Voraussetzungsgemäß gilt $|f(x)| = \mathcal{O}(|x|^n)$ für $x \to \infty$, sowie $|x-\mu|^\alpha + |x-\mu|^n = \Theta(|x|^n)$ für $x \to \infty$, da wegen $n \geq \alpha > 0$ gilt $0 < \liminf_{x\to\infty} |x|^{-n}(|x|^\alpha + |x|^n) \leq \limsup_{x\to\infty} |x|^{-n}(|x|^\alpha+|x|^n) < \infty$. Also ist g für hinreichend große Wert von x beschränkt. Mit anderen Worten: Es existiert ein d_1, sodass g auf der Menge $|x-\mu| \geq d_1$ beschränkt ist. Andererseits gilt $|f(x) - f(\mu)| = \mathcal{O}(|x-\mu|^\alpha)$ und $|x-\mu|^\alpha + |x-\mu|^n = \Theta(|x-\mu|^\alpha)$ für $x \to \mu$. Also ist g in einer hinreichend kleinen Umgebung von μ beschränkt. Mit anderen Worten: Es existiert ein $d_2 < d_1$, sodass g auf der Menge $|x-\mu| \leq d_2$ beschränkt ist. Zu zeigen bleibt die Beschränktheit auf der Menge $(d_2 \leq |x-\mu| \leq d_1) =: K$. Da K kompakt ist, ist der Zähler von g darauf laut Voraussetzung beschränkt. Weiters ist der Nenner von g auf K von unten beschränkt durch $d_2^\alpha + d_2^n$. Somit ist g auf K beschränkt. Insgesamt haben wir also gezeigt, dass g auf ganz $I \setminus \{\mu\}$ beschränkt ist. Sei nun M so wie in der Voraussetzung:

$$M := \sup_{x\in I\setminus\{\mu\}} \frac{|f(x)-f(\mu)|}{|x-\mu|^\alpha + |x-\mu|^n} = \sup_{x\in I\setminus\{\mu\}} g(x)$$

Damit und aus der Definition von g folgt

$$|f(x) - f(\mu)| = (|x-\mu|^\alpha + |x-\mu|^n) \cdot g(x) \leq M(|x-\mu|^\alpha + |x-\mu|^n)$$

Also gilt für die Jensen-Differenz

$$\begin{aligned}
|\mathcal{J}(f, X\sim\mathcal{P})| &= |\mathbb{E}[f(X)] - f(\mu)| \\
&\leq \int_{\mathbb{R}} |f(X) - f(\mu)| d\mathcal{P}(X) \\
&\leq M \int_{\mathbb{R}} |X-\mu|^\alpha + |X-\mu|^n d\mathcal{P}(X) \\
&= M(\underbrace{\sigma_\alpha^\alpha}_{\leq \sigma_n^\alpha} + \sigma_n^n) \\
&\leq M(1+\sigma_n^{n-\alpha})\sigma_n^\alpha
\end{aligned}$$

□

Bemerkung 4.2. Wenn wir uns auf Verteilungen von Zufallsvariablen beschränken, die um den Erwartungswert μ konzentriert sind, können wir obige Ungleichung vereinfachen zu

$$|\mathbb{E}[f(X)] - f(\mu)| \leq M'\sigma_n^\alpha \text{ für hinreichend kleine } \sigma_n \tag{4.1}$$

Dabei ist $M' > 0$ unabhängig von der Verteilung der Zufallsvariable X.

4.2 Schärfe der Abschätzung

Wir zeigen nun, dass die Ungleichung (4.1) scharf ist (modulo der Konstanten M').

Proposition 4.3. *Sei $I = \mathbb{R}$ und f eine Funktion so wie in Satz (4.1) gefordert, die zusätzlich $|f(x) - f(\mu)| \geq M|x - \mu|^\alpha$ für ein $M > 0$ und alle $x \in \mathbb{R}$ erfüllt. Dann existiert für alle $\sigma_n > 0$ eine Wahrscheinlichkeitsverteilung $\mathcal{P}$, sodass*

$$|\mathbb{E}[f(X)] - f(\mu)| \geq M\sigma_n^\alpha$$

gilt.

Beweis. Sei $\mathcal{P}$ eine diskrete Wahrscheinlichkeitsverteilung definiert durch

$$\begin{cases} \mathcal{P}(\{\mu + \sigma_n\}) = \mathcal{P}(\{\mu - \sigma_n\} = 1/2 \\ \mathcal{P}(\mathbb{R} \setminus \{\mu + \sigma_n, \mu - \sigma_n\}) = 0 \end{cases}$$

Für den Jensen-Fehler gilt dann:

$$\begin{aligned} |\mathcal{J}(f, X \sim \mathcal{P})| &= |\mathbb{E}[f(X)] - f(\mu)| = |\mathbb{E}[f(X) - f(\mu)]| \geq M\mathbb{E}[|X - \mu|^\alpha] \\ &= M\big((1/2) \cdot |\sigma_n|^\alpha + (1/2) \cdot |-\sigma_n|^\alpha\big) = M\sigma_n^\alpha \end{aligned}$$

□

Wir zeigen nun, dass das σ_n in (4.1) durch kein σ_m mit $m < n$ ersetzt werden kann

Proposition 4.4. *Es existiert eine Funktion f so wie in Satz (4.1) gefordert, sodass für jedes m mit $0 < m < n$ und $\sigma_n > 0$ eine Wahrscheinlichkeitsverteilung $\mathcal{P}$ existiert, sodass $\frac{|\mathcal{J}(f, X \sim P)|}{\sigma_m^\alpha}$ beliebig groß wird.*

Beweis. Sei $\mathcal{P}$ eine diskrete Wahrscheinlichkeitsverteilung definiert durch

$$\begin{cases} \mathcal{P}(\{\mu + a\}) = \mathcal{P}(\{\mu - a\}) = p/2 \\ \mathcal{P}(\{\mu\}) = 1 - p \\ \mathcal{P}(\mathbb{R} \setminus \{\mu, \mu + a, \mu - a\}) = 0 \end{cases}$$

Damit gilt

$$\sigma_m = \sqrt[m]{\mathbb{E}[|X - \mu|^m]} = \sqrt[m]{(p/2)|a|^m + (p/2)|-a|^m} = \sqrt[m]{p} \cdot a$$

Weiters definieren wir f durch $f(x) := |x-\mu|^\alpha + |x-\mu|^n$. Damit folgt

$$\mathbb{E}[X] = \frac{p}{2}(\mu+a) + \frac{p}{2}(\mu-a) + (1-p)\mu = \mu \Rightarrow f(\mathbb{E}[X]) = 0$$

und

$$\mathbb{E}[f(X)] = \frac{p}{2}(|a|^\alpha + |a|^n) + \frac{p}{2}(|-a|^\alpha + |-a|^n) = p(a^\alpha + a^n)$$

Also

$$|\mathcal{J}(f, X \sim \mathcal{P})| = p(a^\alpha + a^n)$$

und daher

$$\frac{|\mathcal{J}(f, X \sim \mathcal{P}|)}{\sigma_m^\alpha} = \frac{p(a^\alpha + a^n)}{\underbrace{(\sqrt[m]{p}a)^\alpha}_{Nenner = p^{(\alpha/m)}a^\alpha}} = p^{1-\alpha/m}(1 + a^{n-\alpha})$$

Wegen $\sigma_n = \sqrt[n]{p}a \Leftrightarrow a = \frac{\sigma_n}{\sqrt[n]{p}}$ folgt daraus durch Einsetzen und Ausmultiplizieren

$$\frac{|\mathcal{J}(f, X \sim \mathcal{P})|}{\sigma_m^\alpha} = p^{1-\alpha/m}\Big(1 + \frac{\sigma_n^{n-\alpha}}{p^{1-\alpha/n}}\Big) = p^{1-\alpha/m} + p^{\alpha(1/n-1/m)}\sigma_n^{n-\alpha}$$

Wegen $1/n - 1/m < 0$ und da p jeden Wert in $(0,1)$ annehmen kann, ist es möglich, dass dieser Ausdruck beliebig groß wird. □

4.3 Linearer shift

Für Zufallsvariablen, deren Verteilung um den Erwartungswert konzentriert ist, d.h. Zufallsvariablen mit kleinem σ_n, bedeutet ein großes α in der Ungleichung 4.1, dass die obere Schranke der Jensen-Differenz eng ist. Es gibt jedoch bestimmte Funktionen, bei denen es unmöglich ist, ein $\alpha > 1$ zu finden. Dazu zählen Funktionen mit $f'(\mu) \neq 0$ und konvexe Funktionen, die bei $x = \mu$ streng monoton wachsend sind. Dies verdeutlicht das folgende Lemma.

Lemma 4.5. *Sei f eine konvexe Funktion, die bei $x = \mu$ streng monoton wachsend ist. Für jedes $\alpha > 1$ gilt dann*

$$\lim_{x\to\mu} \frac{|x-\mu|^\alpha}{f(x) - f(\mu)} = 0$$

Bemerkung 4.6. Das bedeutet, dass die zweite Bedingung des ersten Hauptresultates für $\alpha > 1$ nicht erfüllt ist.

Beweis. Da f voraussetzungsgemäß eine konvexe Funktion ist, die zusätzlich bei $x = \mu$ streng monoton wachsend, gilt $f'_+(\mu) > 0$ und $f'_-(\mu) > 0$. Daraus folgt

$$\lim_{x\to\mu+} \frac{|x-\mu|^\alpha}{f(x)-f(\mu)} = \lim_{x\to\mu+}(x-\mu)^{\alpha-1}\frac{x-\mu}{f(x)-f(\mu)} = 0 \cdot \frac{1}{\underbrace{f'_+(\mu)}_{>0}} = 0$$

und analog dazu

$$\lim_{x\to\mu-} \frac{|x-\mu|^\alpha}{f(x)-f(\mu)} = 0$$

□

Das Problem, dass für solche Funktionen kein $\alpha > 1$ gefunden werden kann, lässt sich beheben, indem wir die ursprüngliche Funktion durch eine lineare Funktion verschieben. Dieser shift ändert nichts an der Konvexität und am Jensen-Fehler der ursprünglichen Funktion. Für Funktionen f, deren Ableitung an der Stelle $x = \mu$ existiert, gilt nach dem Satz von Taylor mit der Restglieddarstellung von Peano

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + o(x - \mu)$$

Also können wir statt f die Funktion g definiert durch $g(x) := f(x) - f'(\mu)(x - \mu)$ untersuchen. Aus der Definition von g erhalten wir

$$g(x) - g(\mu) = f(x) - f'(\mu)(x - \mu) - f(\mu) = o(x - \mu)$$

welches einen α-Wert hat, der mindestens so groß ist wie jener von f. Wenn f darüber hinaus eine wohldefinierte 2. Ableitung bei $x = \mu$ hat, folgt mit dem Mittelwertsatz, dass

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi_L)}{2}(x - \mu)^2$$

mit einem ξ_L zwischen x und μ. Also

$$g(x) - g(\mu) = \frac{f''(\xi_L)}{2}(x - \mu)^2$$

Daraus folgt $\alpha = 2$.

5 Zweites Hauptresultat: Untere Schranke der Jensen-Differenz

5.1 Untere Schranke der Jensen-Differenz

Satz 5.1. *Sei $I \subseteq \mathbb{R}$ ein abgeschlossenes Intervall, X eine Zufallsvariable mit Verteilung $\mathcal{P}$ und Erwartungswert $\mathbb{E}[X] =: \mu \in I$, sowie $f : I \to \mathbb{R}$ eine Funktion, die die folgenden Bedingung erfüllt:*

1. $f(x) - f(\mu) > 0$ *für* $x \neq \mu$
2. $f(x) - f(\mu) = \Omega(|x - \mu|^\alpha)$ *für* $x \to \mu, \ \alpha > 0$
3. $f(x) - f(\mu) = \Omega(|x - \mu|^\beta)$ *für* $x \to \infty, \ 0 \leq \beta \leq \alpha$

Dann gilt

$$\mathcal{J}(f, X \sim \mathcal{P}) \geq M \frac{\sigma_{\alpha/2}^{\alpha}}{1 + \sigma_{\alpha-\beta}^{\alpha-\beta}}$$

Dabei ist

$$M := \inf_{x \in I \setminus \{\mu\}} (f(x) - f(\mu)) \left(\frac{1}{|x-\mu|^\beta} + \frac{1}{|x-\mu|^\alpha} \right) > 0$$

unabhängig von der Verteilung $\mathcal{P}$ von X.

Beweis. Sei

$$g(x) := \left(\frac{1}{|x-\mu|^\beta} + \frac{1}{|x-\mu|^\alpha} \right)^{-1}$$

Zunächst zeigen wir, dass M positiv ist. Klarerweise gilt $g(x) > 0$ für $x \neq \mu$. Aus der Definition von M folgt, dass $f(x) - f(\mu) \geq Mg(x)$. Wie im Beweis des ersten Hauptresultates sehen wir, dass $g(x) = \Theta(|x-\mu|^\alpha)$ für $x \to \mu$. Also existiert eine Konstante $M_1 > 0$ und eine Zahl $d_1 > 0$, sodass $f(x) - f(\mu) \geq M_1 \cdot g(x)$ für $|x-\mu| \leq d_1$. Ebenso sehen wir, dass $g(x) = \Theta(|x-\mu|^\beta)$ für $x \to \infty$. Also existiert eine Konstante $M_2 > 0$ und eine Zahl $d_2 \geq d_1$, sodass $f(x) - f(\mu) \geq M_2 \cdot g(x)$ für $|x-\mu| \geq d_2$. Da die Menge $d_1 \leq |x-\mu| \leq d_2$ kompakt ist, und sowohl $f(x) - f(\mu)$ als auch $g(x)$ darauf positiv sind, existiert eine Konstante $M_3 > 0$, sodass $f(x) - f(\mu) \geq M_3 \cdot g(x)$. Bezeichne nun $M' > 0$ das Minimum von $\{M_1, M_2, M_3\}$. Damit gilt $f(x) - f(\mu) \geq M'g(x)$. Das bedeutet, dass $\frac{f(x)-f(\mu)}{g(x)}$ von unten durch eine positive Zahl beschränkt ist.

$$\Rightarrow M = \inf_{x \in I \setminus \{\mu\}} \frac{f(x) - f(\mu)}{g(x)} > 0.$$

Da, wie eingangs erwähnt, $f(x) - f(\mu) \geq M \cdot g(x)$ gilt, folgt also für die Jensen-Differenz:

$$\begin{aligned} \mathcal{J}(f, X \sim P) = \mathbb{E}[f(X) - f(\mu)] &\geq M \int \Big(\frac{1}{|X-\mu|^\beta} + \frac{1}{|X-\mu|^\alpha} \Big)^{-1} d\mathcal{P}(X) \\ &= M \int \frac{|X-\mu|^\alpha}{|X-\mu|^{\alpha-\beta} + 1} d\mathcal{P}(X) \geq M \frac{\sigma_{\alpha/2}^\alpha}{1 + \sigma_{\alpha-\beta}^{\alpha-\beta}} \end{aligned}$$

Dabei folgt die letzte Ungleichung aus der Ungleichung von Cauchy-Schwarz. Siehe dazu folgendes Lemma. □

Lemma 5.2. *Mithilfe der Ungleichung von Cauchy-Schwarz erhält man*

$$M \int \frac{|X-\mu|^\alpha}{|X-\mu|^{\alpha-\beta} + 1} d\mathcal{P}(X) \geq M \frac{\sigma_{\alpha/2}^\alpha}{1 + \sigma_{\alpha-\beta}^{\alpha-\beta}}$$

Beweis. Sei

$$g_1(X) := \sqrt{\frac{|X-\mu|^\alpha}{|X-\mu|^{\alpha-\beta} + 1}}$$

und

$$g_2(X) := \sqrt{|X-\mu|^{\alpha-\beta} + 1}$$

Wegen der Ungleichung von Cauchy-Schwarz gilt

$$\mathbb{E}[g_1^2(X)] \cdot \mathbb{E}[g_2^2(X)] \geq \mathbb{E}[g_1(X)g_2(X)]^2$$

Das ist äquivalent zu

$$\mathbb{E}[g_1^2(X)] \geq \frac{\mathbb{E}[g_1(X)g_2(X)]^2}{\mathbb{E}[g_2^2(X)]} \tag{5.1}$$

Dabei ist (indem wir den Erwartungswert als Integral nach dem Wahrscheinlichkeitsmaß schreiben)

$$\mathbb{E}[g_1^2(X)] = \int \frac{|X-\mu|^\alpha}{|X-\mu|^{\alpha-\beta}+1} d\mathcal{P}(X)$$

$$\mathbb{E}[g_2^2(X)] = \int |X-\mu|^{\alpha-\beta} + 1 d\mathcal{P}(X) = 1 + \sigma_{\alpha-\beta}^{\alpha-\beta}$$

$$\mathbb{E}[g_1(X)g_2(X)]^2 = \underbrace{\Big(\int |X-\mu|^{\alpha/2} d\mathcal{P}(X)\Big)}_{=\sigma_{\alpha/2}^{\alpha/2}}^2 = \sigma_{\alpha/2}^{\alpha}$$

Somit erhalten wir aus Ungleichung 5.1

$$\int \frac{|X-\mu|^\alpha}{|X-\mu|^{\alpha-\beta}+1} d\mathcal{P}(X) \geq \frac{\sigma_{\alpha/2}^{\alpha}}{1+\sigma_{\alpha-\beta}^{\alpha-\beta}}$$

Also gilt in der Tat

$$M \int \frac{|X-\mu|^\alpha}{|X-\mu|^{\alpha}-\beta+1} d\mathcal{P}(X) \geq M \frac{\sigma_{\alpha/2}^{\alpha}}{1+\sigma_{\alpha-\beta}^{\alpha-\beta}}$$

□

5.2 Untere Schranke und konvexe Funktionen

Bemerkung 5.3. Obwohl die Bedingungen aus Satz (5.1) für allgemeine Funktionen in der Regel nicht erfüllt sind, werden sie von konvexen Funktionen fast automatisch erfüllt. „Fast“ deshalb, weil eine konvexe Funktion f zusätzlich nicht-wachsend auf dem Intervall $(-\infty, \mu]$ und nicht-fallend auf dem Intervall $[\mu, +\infty)$ sein muss, wofür ggf. ein shift der Funktion f durch eine lineare Funktion notwendig ist:

Proposition 5.4. *Sei f eine beliebige konvexe Funktion und $a \in \mathbb{R}$, sodass $f'_-(\mu) \leq a \leq f'_+(\mu)$. Dann ist der lineare shift $g(x) := f(x) - a(x-\mu)$ nicht-wachsend auf $(-\infty, \mu]$ und nicht-fallend auf $[\mu, +\infty)$.*

Beweis. Sei $x \geq \mu$ und $x' > x$. Dann folgt

$$\frac{g(x') - g(x)}{x' - x} = \frac{1}{x'-x}\Big(f(x') - a(x'-\mu) - f(x) + a(x-\mu)\Big) = \frac{f(x')-f(x)}{x'-x} - a \geq f'_+(\mu) - a \geq 0$$

Dabei geht beim vorletzten Ungleichheitszeichen die Konvexität von f ein. Für $x \leq \mu$ und $x' < x$ verläuft der Beweis analog:

$$\frac{g(x) - g(x')}{x - x'} = \frac{1}{x-x'}\Big(f(x) - a(x-\mu) - f(x') + a(x'-\mu)\Big) = \frac{f(x)-f(x')}{x-x'} - a \leq f'_-(\mu) - a \leq 0$$

□

Bemerkung 5.5. Die folgende Proposition zeigt, dass für konvexe Funktionen α nur Werte in $[1, +\infty)$ annehmen kann.

Proposition 5.6. *Es existiert keine konvexe Funktion f, sodass $f(x) - f(\mu) = \Omega(|x-\mu|^\alpha)$ für $x \to \mu$, $\alpha < 1$*

Beweis. Wir führen den Beweis durch Widerspruch. Angenommen $f(x)-f(\mu) = \Omega(|x-\mu|^\alpha)$ für $x \to \mu$, wobei $\alpha < 1$. Dann existieren positive Konstanten d und M, sodass $M|x-\mu|^\alpha \leq f(x) - f(\mu)$ auf der Menge $|x - \mu| \leq d$. Dadurch erhalten wir für alle x mit $\mu < x < \mu + d$

$$\frac{f(x) - f(\mu)}{x - \mu} \geq \frac{M \cdot |x - \mu|^\alpha}{x - \mu} = M(x - \mu)^{\alpha - 1} \tag{5.2}$$

Weil f konvex ist, ist der Quotient $\frac{f(x)-f(\mu)}{x-\mu}$ nicht-fallend in x. Weiters wird wegen $\alpha < 1$ der Term $(x - \mu)^{\alpha-1}$ beliebig groß für $x \to \mu$. Deshalb erhalten wir für $x \to \mu$ einen Widerspruch zur Ungleichung (5.2) □

6 Beispiele

Beispiel 6.1. Als erstes Beispiel wollen wir die Jensen-Differenz der Funktion $x \mapsto \sin(x)$ und einer Zufallsvariable $X \sim P$ mit Erwartungswert $\mathbb{E}[X] = \mu = 0$ untersuchen. Die Taylorreihendarstellung von $\sin(x)$ an der Entwicklungsstelle $x_0 = 0$ lautet $\sin(x) = x - \frac{x^3}{6} + \frac{x^5}{120} + \mathcal{O}(x^6)$. Mit $\alpha = n = 1$ sind die Bedingungen des ersten Hauptresultates erfüllt, wobei I die Menge der reellen Zahlen sei: In der Tat ist $x \mapsto \sin(x)$ als stetige Funktion auf jeder kompakten Teilmenge von $\mathbb{R}$ beschränkt. Weiters gilt $|\sin(x) - \sin(\mu)| = |\sin(x)| = \mathcal{O}(|x - \mu|^1) = \mathcal{O}(|x|^1)$ für $x \to \mu$ und $|\sin(x)| = \mathcal{O}(|x|^1)$ für $x \to \infty$. Wir erhalten also mit $\mu = 0$ und $\alpha = n = 1$ für M:

$$M = \sup_{x \in \mathbb{R}\setminus\{\mu\}} \frac{|\sin(x) - \sin(\mu)|}{|x - \mu|^\alpha + |x - \mu|^n} = \sup_{x \in \mathbb{R}\setminus\{0\}} \frac{|\sin(x) - \sin(0)|}{|x - 0|^1 + |x - 0|^1} = \sup_{x \in \mathbb{R}\setminus\{0\}} \frac{|\sin(x)|}{2|x|} = \frac{1}{2}$$

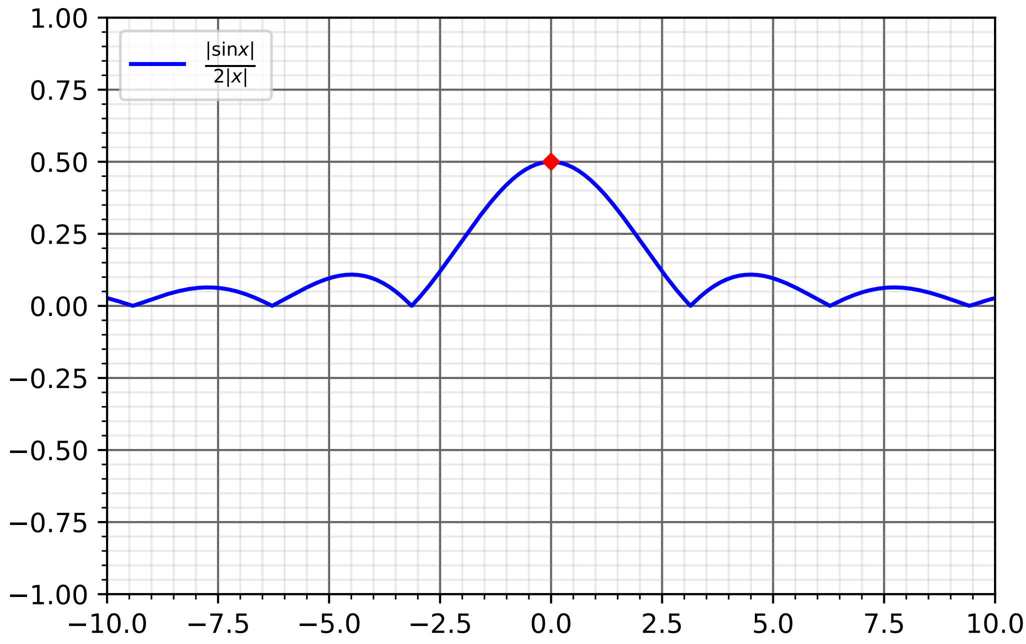


Abbildung 6.1: Die Funktion $x \mapsto \frac{|\sin(x)|}{2|x|}$ mit eingezeichnetem Supremum

und

$$(1+\sigma_n^{n-\alpha})\sigma_n^\alpha = (1+\sigma_1^{1-1})\sigma_1^1 = 2\sigma_1$$

Insgesamt erhalten wir somit

$$|\mathcal{J}(\sin, X \sim \mathcal{P})| \leq M \cdot 2\sigma_1 = \sigma_1$$

Für Verteilungen, die um den Erwartungswert konzentriert sind, bedeutet dies, dass die Jensen-Differenz $|\mathcal{J}(\sin, X \sim \mathcal{P})|$ nicht schneller gegen 0 geht als $\sim \sigma_1$.

Beispiel 6.2. Als nächstes Beispiel wollen wir die Funktion $x \mapsto \ln(x)$ und eine Zufallsvariable $X \in [a, +\infty),\ a > 0$ mit Erwartungswert $\mu = 1$ betrachten. Da $\ln'(x)\Big|_{x=\mu} = 1 \neq 0$, können wir stattdessen die verschobene Funktion

$$g(x) := f(x) - f'(\mu)(x-\mu) = \ln(x) - (x-1)$$

betrachten. Die Taylorreihendarstellung von $\ln(x)$ um die Entwicklungsstelle $x_0 = 1$ lautet: $\ln(x) = (x-1) - \frac{1}{2}(x-1)^2 \pm \ldots$. Mit $\alpha = n = 2$ sind die Bedingungen des ersten Hauptresultates erfüllt:

$$\limsup_{x\to\mu} \frac{|g(x)-g(\mu)|}{|x-\mu|^\alpha} = \limsup_{x\to 1} \frac{x-\ln(x)-1}{(x-1)^2} = \frac{1}{2} < \infty$$

und

$$\limsup_{x\to\infty} \frac{|g(x)|}{|x|^n} = \limsup_{x\to\infty} \frac{x-\ln(x)-1}{x^2} = 0 < \infty$$

Siehe dazu auch im Anhang. Weiters ist g als Zusammensetzung stetiger Funktionen stetig, also auf jeder kompakten Menge beschränkt. Um M zu ermitteln, setzen wir in die Definition ein. Man beachtet, dass $\frac{|g(x)|}{2(x-1)^2}$ streng monoton fallend ist.

$$M = \sup_{x\in[a,+\infty)} \frac{|g(x)-g(1)|}{2|x-1|^2} = \sup_{x\in[a,+\infty)} \frac{|g(x)|}{2(x-1)^2} = \frac{a-1-\ln(a)}{2(1-a)^2}$$

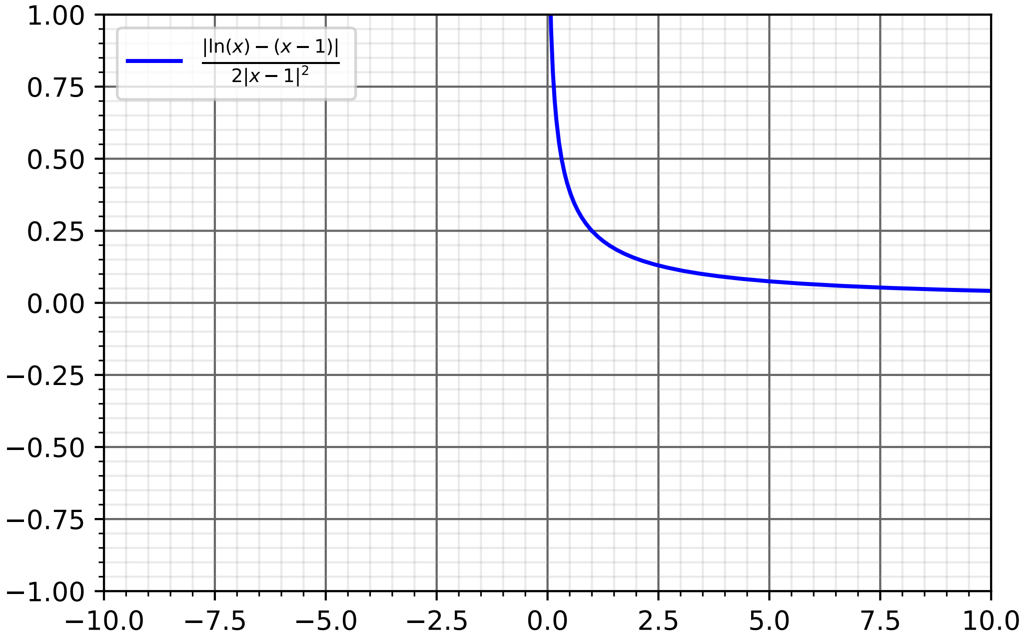


Abbildung 6.2: Die Funktion $x \mapsto \frac{|\ln(x)-(x-1)|}{2(x-1)^2}$

Mit $(1+\sigma_2^{2-2})\sigma_2^2 = 2\sigma_2^2$ erhalten wir somit insgesamt

$$|\mathcal{J}(\ln, X\sim\mathcal{P})| \leq \frac{a-1-\ln(a)}{(1-a^2)}\sigma_2^2$$

Das bedeutet, dass $|\mathcal{J}(\ln, X\sim\mathcal{P})|$ nicht langsamer fällt als $\approx \sigma_2^2$, also die Varianz der Verteilung für $\sigma_2 \to 0$.
Um eine untere Schranke der Jensen-Differenz zu ermitteln, zeigt man zunächst analog zu den Ausführungen oben, dass mit $\alpha = 2$ und $\beta = 1$ die Bedingungen des zweiten Hauptresultates erfüllt sind. Damit erhält man

$$\inf_{x\in[a,+\infty)\setminus\{1\}} -(g(x)-g(1))\left(\frac{1}{|x-1|^1} + \frac{1}{|x-1|^2}\right) = \frac{1}{2} \tag{6.1}$$

Gemeinsam mit

$$\frac{\sigma_{\alpha/2}^{\alpha}}{1+\sigma_{\alpha-\beta}^{\alpha-\beta}} = \frac{\sigma_1^2}{1+\sigma_1}$$

erhalten wir somit als untere Schranke der Jensen-Differenz

$$-\mathcal{J}(\ln, X \sim \mathcal{P}) \geq \frac{1}{2}\frac{\sigma_1^2}{1+\sigma_1}$$

Das „−“ Zeichen stammt dabei aus Gleichung 6.1.

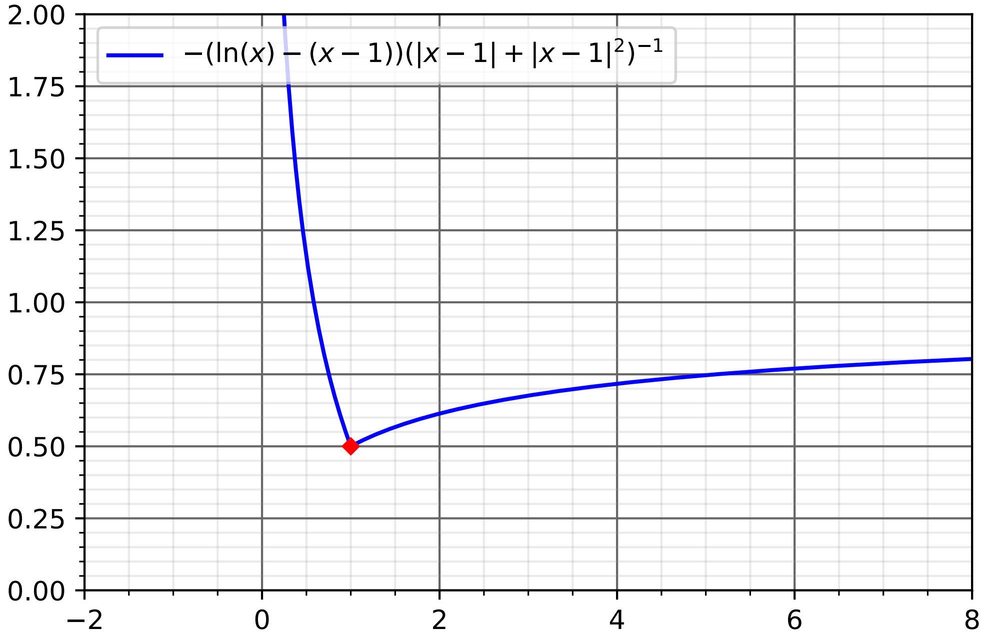


Abbildung 6.3: Die Funktion $x \mapsto -(\ln(x) - (x-1))\left(\frac{1}{|x-1|} + \frac{1}{|x-1|^2}\right)$ mit eingezeichnetem Infimum

Beispiel 6.3. Für die Jensen-Differenz der Funktion $x \mapsto \cos(x)$, $\cos(x) = 1 - \frac{x^2}{2} \pm \ldots$ und eine Zufallsvariable X mit Erwartungswert $\mu = 0$, können wir im ersten Hauptresultat $\alpha = n = 2$ wählen und erhalten so analog zum ersten Beispiel mit $M = \frac{1}{4}$

$$|\mathcal{J}(\cos, X \sim \mathcal{P})| \leq \frac{\sigma_2^2}{2}$$

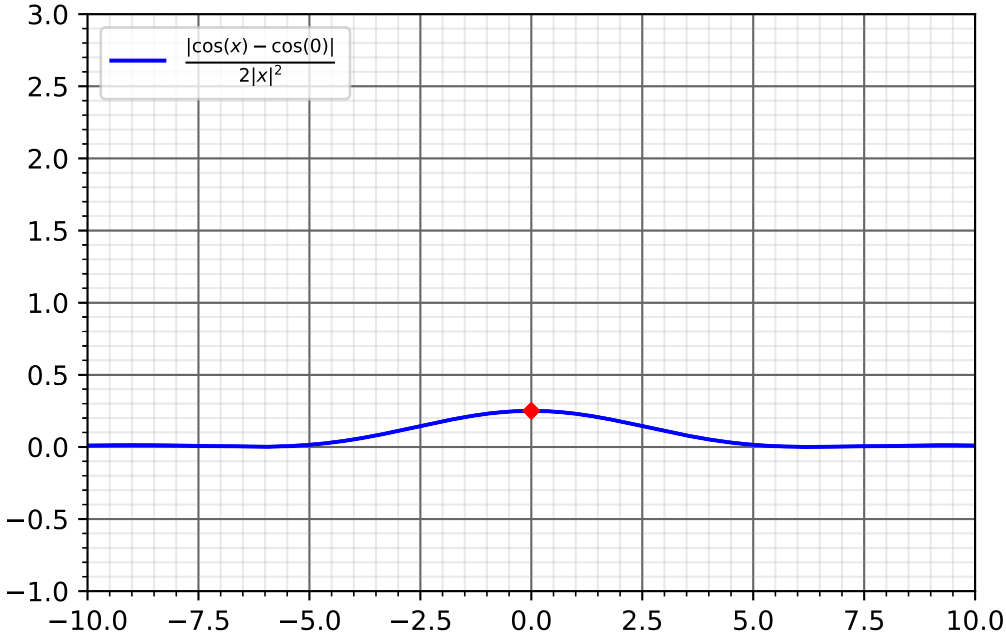


Abbildung 6.4: Die Funktion $x \mapsto \frac{|\cos(x)-\cos(0)|}{2|x|^2}$ mit eingezeichnetem Supremum

Also fällt $|\mathcal{J}(\cos, X \sim \mathcal{P})|$ nicht langsamer als $\approx \sigma_2^2$, d.h. die Varianz der Verteilung.

Beispiel 6.4. Sei $f(x) := \sqrt{x}$ und X eine nichtnegative Zufallsvariable mit Erwartungswert $\mu = 1$. Da $f'(1) = \frac{1}{2} \neq 0$, betrachten wir stattdessen die geshiftete Funktion $g(x) := f(x) - f'(\mu)(x - \mu) = \sqrt{x} - \frac{x-1}{2}$. Es gilt

$$\sqrt{x} = 1 + \frac{x-1}{2} - \frac{1}{8}(x-1)^2 + \dots$$

Indem wir $\alpha = n = 2$ wählen, erhalten wir für die obere Schranke der Jensen-Differenz

$$|\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P})| \leq M(1 + \sigma_2^{2-2})\sigma_2^2 = \frac{1}{4} \cdot 2 \cdot \sigma_2^2 = \frac{1}{2}\sigma_2^2$$

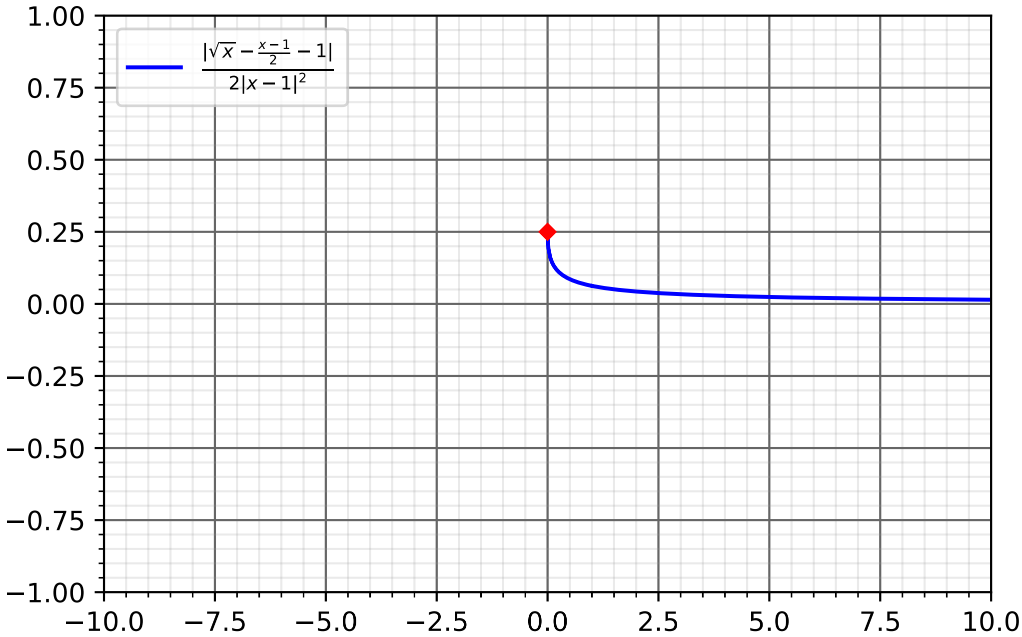


Abbildung 6.5: Die Funktion $x \mapsto \frac{|\sqrt{x} - \frac{x-1}{2} - 1|}{2|x-1|^2}$ mit eingezeichnetem Supremum

Also fällt $|\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P})|$ nicht langsamer als die Varianz der Verteilung. Für die untere Schranke erhalten wir mittels $\alpha = 2, \beta = 1$

$$-\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P}) \geq \frac{1}{8} \frac{\sigma_1^2}{1 + \sigma_1}$$

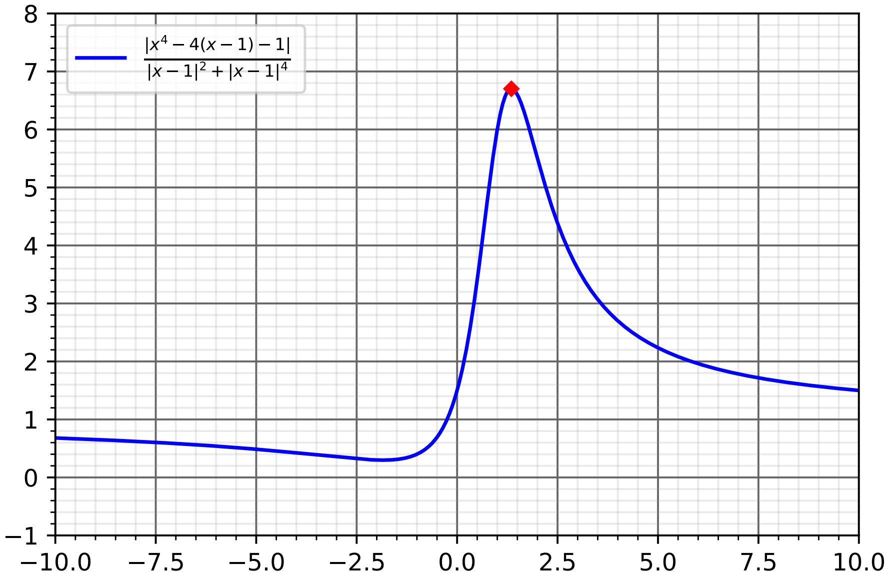
Das heißt die Jensen-Differenz nähert sich 0 nicht schneller als σ_1^2.

Beispiel 6.5. Sei $f(x) := x^4$ und X eine Zufallsvariable mit Erwartungswert $\mu = 1$. Da $f'(1) = 4 \neq 0$, betrachten wir $g(x) := f(x) - f'(\mu)(x - \mu) = x^4 - 4(x - 1)$ statt $f(x)$. Mit $\alpha = 2, n = 4$ erhält man für die Jensen Differenz

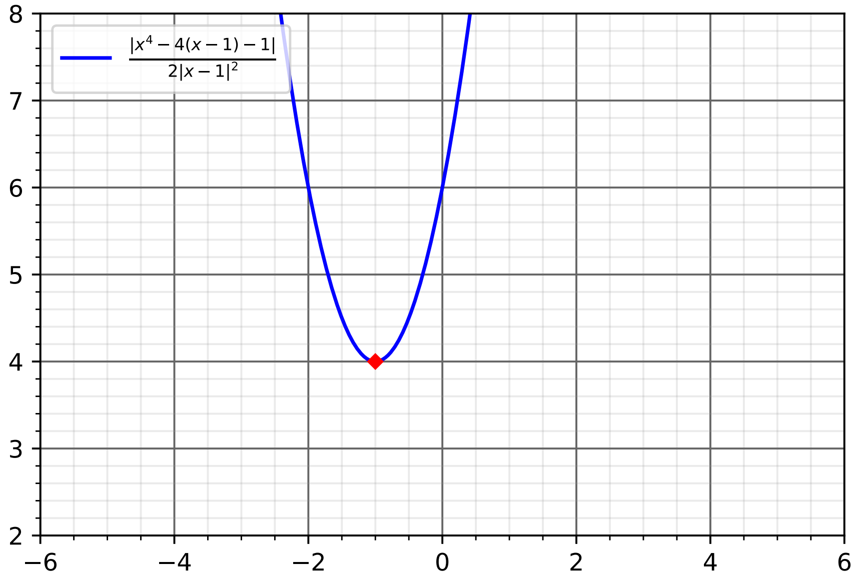
$$|\mathcal{J}(f, X \sim \mathcal{P})| \leq \frac{7 + \sqrt{41}}{2}(1 + \sigma_4^2)\sigma_4^2$$

Also fällt $\mathcal{J}(f, X \sim P)$ nicht schneller als $\approx \sigma_4^2$. Mit $\alpha = \beta = 2$ erhält man für die untere Schranke

$$J(f, X \sim P) \geq 2\sigma_1^2$$



(a) $x \mapsto \frac{|x^4-4(x-1)-1|}{|x-1|^2+|x-1|^4}$ mit eingezeichnetem Supremum



(b) $x \mapsto \frac{|x^4-4(x-1)-1|}{2|x-1|^2}$ mit eingezeichnetem Infimum

Abbildung 6.6

7 Anwendungen

7.1 Mathematik

Die Berechnung der Jensen-Differenz sowie oberer und unterer Schranken dieser, auch mithilfe der in dieser Arbeit vorgestellten Methoden, findet sich vor allem in der reinen Mathematik wieder. So zum Beispiel in der Maß- und Wahrscheinlichkeitstheorie, aber auch in der Statistik. Darüber hinaus ist man auch in Wissenschaften, die die Mathematik nur als Hilfswissenschaft verwenden, an der Jensen-Differenz interessiert, wie beispielsweise in der statistischen Mechanik, in der Physik oder im Bereich des machine learning.

7.2 Statistische Mechanik: Jarzynski-Gleichung

Als erstes soll ein Anwendungsgebiet in der statistischen Mechanik vorgestellt werden und zwar die Jarzynski-Gleichung. Benannt ist diese Gleichung nach dem amerikanischen Physiker und Universitätsprofessor Christopher Jarzynski, der diese im Jahr 1997 entdeckte. Diese Gleichung aus der statistischen Mechanik bringt die Differenz der freien Energie F mit der Arbeit W, die in sogenannten Nicht-Gleichgewichtsprozessen an einem System geleistet wird, in Zusammenhang. Beim Beweis dieser Gleichung spielten Abschätzungen der Jensen-Differenz eine wichtige Rolle.

Freie Energie (auch Helmholtz-Energie genannt) beschreibt dabei ein thermodynamisches Potential, welches die Arbeit misst, die von einem thermodynamischen System bei konstanter Temperatur und konstantem Volumen (isotherm und isochor) geleistet werden kann.

In der Thermodynamik steht die Differenz der freien Energie $\Delta F = F_B - F_A$ zwischen zwei Zuständen A und B mit der an einem System geleisteten Arbeit W über die Ungleichung

$$\Delta F = F_B - F_A \leq W$$

in Zusammenhang.
Gleichheit gilt hier nur im Falle eines quasi-statischen Prozesses. Dabei heißt ein Prozess quasi-statisch, wenn die Parameter des Prozesses so langsam verändert werden, dass das System zu jedem Zeitpunkt im Gleichgewicht verharrt bzw. sich beliebig nahe davon befindet. Ein solcher Prozess durchläuft also eine Folge von Gleichgewichtszuständen, d.h. infinitesimal kleine Zustandsänderungen.

Im Gegensatz zur obigen Aussage, gilt die Jarzynski-Gleichung für beliebige Prozesse. Die Gültigkeit der Jarzynski-Gleichung ist also unabhängig davon, wie schnell der Prozess abläuft. Die Jarzynski-Gleichung lautet

$$e^{-\Delta F/(k\cdot T)} =< e^{-W/(k\cdot T)} >$$

Dabei ist

- $k = 1,380649 \cdot 10^{-23}$ $Joule/Kelvin$ die Boltzmann-Konstante
- T die Temperatur (in Kelvin) des Systems im Gleichgewichtszustand A

$< \ldots >$ bedeutet, dass der Durchschnitt über alle möglichen Realisierungen eines äußeren Prozesses, der das System vom Zustand A in den Zustand B überführt, genommen wird. Mit der Definition

$$\beta := \frac{1}{kT}$$

wird die Jarzynski-Gleichung zu

$$e^{-\Delta F\beta} =< e^{-W\beta} >$$

Der Quotient $\beta = \frac{1}{kT}$ wird *Energienormierung* genannt.

Für den Grenzfall eines unendlich langsamen Prozesses ist die Arbeit, die in jedem Schritt auf das System ausgewirkt wird, numerisch gesehen die gleiche. Daher wird in diesem Fall der Durchschnitt irrelevant und die Jarzynski-Gleichung vereinfacht sich zu $\Delta F = W$.

Weiters folgt aus der Jarzynski-Gleichung, dass es Prozesse mit strikt positiver Eintrittswahrscheinlichkeit gibt, für die gilt, dass

$$\Delta F > W$$

7.3 Physik

In der Physik werden die Jensen-Differenz und auch Abschätzungen davon bei der Ermittlung der Fluktuation von thermodynamischen Größen und deren Erwartungswert verwendet.

7.4 Machine learning

Im Bereich des machine learning wird häufig das stochastic gradient descent (SGD) Verfahren (Gradientenabstiegsverfahren) angewandt. Dieses iterative Verfahren dient unter anderem zur Optimierung von sogenannten loss und cost functions. Eine loss bzw. cost function ist eine Funktion, die einer fehlerhaften Entscheidung einen Schaden in Form einer Schätzung zuordnet, der durch eine vom korrekten Parameter abweichende Entscheidung entsteht.

Gegeben hat man ein statistisches Model, also ein Tripel $(X, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$ bestehend aus einer Grundmenge X, einer σ-Algebra $\mathcal{A}$ und einer Menge $(P_\theta)_{\theta\in\Theta}$ von Wahrscheinlichkeitsmaßen auf dem Messraum $(X, \mathcal{A})$. Weiters hat man noch einen Entscheidungsraum (Ω, Σ) also einen Messraum, bei dem die σ-Algebra alle einelementigen Mengen enthält, z.B. $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann heißt eine Funktion $L : \Theta \times \Omega \to [0, +\infty]$ Verlustfunktion oder loss function, wenn für jedes fixe $\theta \in \Theta$ die Funktion $L(\theta, \cdot)$ messbar ist bzgl. $\Sigma - \mathcal{B}([0, +\infty])$.

> "Der Gradientenabstieg lässt sich vom Prinzip her leicht erklären: Angenommen, man stünde im Gebirge im dichten Nebel. Das Tal, und somit der Weg nach Hause, ist vom Nebel verdeckt. Wohin laufen wir? Wir können das Ziel zwar nicht sehen, tasten uns jedoch so heran, dass unser Gehirn den Gradienten (den Unterschied der Höhen beider Füße) berechnet, somit die Steigung des Bodens kennt und sich entgegen dieser Steigung unser Weg fortsetzt."(Aunkofer. *Gradientverfahren.* aus [4])

Mathematisch gesehen, beginnt man mit einer *random initialization* α, berechnet dann die Ausgabe (=*forwardpropagation*) und vergleicht diese mithilfe einer Verlustfunktion (z.B. mittels mean squared error) mit dem tatsächlichen Wert. Wegen der random initialization macht man mit großer Wahrscheinlichkeit einen Verlust. Für die (differenzierbare) Verlustfunktion berechnen wir den Gradienten für gegebene Eingabewerte und begeben uns entgegen dem Gradienten in Richtung des Minimums unserer Verlustfunktion. Ist dieses Minimum gefunden, sagt man, dass der Lernalgorithmus konvergiert.

8 Anhang

Landau-Notation: Im Beweis der oberen und der unteren Schranke sowie in weiteren Teilen dieser Seminararbeit haben wir die Landau-Notation verwendet, welche hier näher erläutert werden soll. f und g seien Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ und der Grenzwert a sei in den erweiterten reellen Zahlen: $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

Notation	Definition	Mathematische Definition
$f \in o(g)$	asymptotisch gegenüber g vernachlässigbar	$\lim_{x\to a} \left\lvert\frac{f(x)}{g(x)}\right\rvert = 0$
$f \in \mathcal{O}(g)$	asymptotische obere Schranke	$\limsup_{x\to a} \left\lvert\frac{f(x)}{g(x)}\right\rvert < \infty$
$f \in \Omega(g)$	asymptotische untere Schranke, $g \in \mathcal{O}(f)$	$\liminf_{x\to a} \left\lvert\frac{f(x)}{g(x)}\right\rvert > 0$
$f \in \omega(g)$	asymptotisch dominant, $g \in o(f)$	$\lim_{x\to a} \left\lvert\frac{f(x)}{g(x)}\right\rvert = \infty$
$f \in \Theta(g)$	asymptotisch scharfe Schranke, $f \in \mathcal{O}(g)$ und $f \in \Omega(g)$	$0 < \liminf_{x\to a} \left\lvert\frac{f(x)}{g(x)}\right\rvert \leq \limsup_{x\to a} \left\lvert\frac{f(x)}{g(x)}\right\rvert < \infty$

9 Literatur- und Abbildungsverzeichnis

Literatur:

1. N. KUSOLITSCH: *Maß- und Wahrscheinlichkeitstheorie. Eine Einführung.* Springer Verlag, Wien 2013.

2. L. RÜSCHENDORF: *Mathematische Statistik.* Springer Verlag, Berlin 2014.

3. X. GAO, M. SITHARAM, A. ROITBERG: *Bounds on the Jensen gap, and implications for mean-concentrated distributions* https://arxiv.org/pdf/1712.05267.pdf (letzter Zugriff: 15.02.2020)

4. B. AUNKOFER: *Deep Learning.* Training eines Neurons mit dem Gradientverfahren. https://data-science-blog.com/blog/2019/01/13/training-eines-neurons-mit-dem-gradientenverfahren/ (letzter Zugriff: 21.12.2019)

5. Jarzynski equality. https://en.wikipedia.org/wiki/Jarzynski_equality (letzter Zugriff: 22.12.2019)

6. Landau Symbole. https://de.wikipedia.org/wiki/Landau-Symbole (letzter Zugriff: 11.01.2020)

Abbildungen:
Alle Abbildungen wurden selbstständig mit Python 3.7 und der Bibliothek *matplotlib* erstellt.