Der Satz von Cramér (1938)

Ausarbeitung zu einem Vortrag im Seminar Große Abweichungen am 04.12.2010

Maren Urner

In diesem Vortrag soll der Satz von Cramér als ein Prinzip großer Abweichungen (LDP) vorgestellt werden. Die Idee des Satzes von Cramér ist es, die Wahrscheinlichkeiten seltener oder untypischer Ereignisse mittels einer exponentiellen Rate und einer Ratenfunktion I zu quantifizieren. So kann mit seiner Hilfe beispielsweise die Konvergenzgeschwindigkeit im schwachen Gesetz der großen Zahlen bestimmt werden.

Ich werde hier zunächst den Satz vorstellen und beweisen. Dabei werde ich mich vor allem auf die Darstellungen [Loe], [Hol] und [Kle] stützen. Danach werde ich genauer auf die Ratenfunktion I und ihre Eigenschaften eingehen und auch die allgemeine Definition einer Ratenfunktion geben. Zum Schluss werde ich noch kurz die Definition eines Prinzips großer Abweichungen nennen.

Satz von Cramér (1938)
Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten (i.i.d.) reellen Zufallsvariablen mit endlicher momenterzeugender Funktion

$$\varphi(t) := \mathbb{E}\, e^{tX_1} < \infty \;\forall t \in \mathbb{R} \tag{1.1.}$$

und sei $S_n := \sum_{i=1}^n X_i$. Dann gilt für alle $a > \mathbb{E}X_1$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -I(a),$$

wobei $I(a) := \sup_{t\in\mathbb{R}}[ta - \log\varphi(t)]$
die sogenannte Legendre-Transformierte von $\log\varphi(t)$ ist.

Beweis:
Zunächst machen wir einige Vorüberlegungen:
(i) Wir setzen $\log(0) = -\infty$ und beweisen

Lemma 1: Die folgenden Aussagen sind äquivalent:

1. $J(a) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -\infty$
2. $\mathbb{P}(X_1 \geq a) = 0$
3. $\mathbb{P}(S_n \geq na) = 0 \;\forall n \in \mathbb{N}$

Beweis durch Zirkelschluss:

1 ⟹ 2:

$$-\infty = J(a) \overset{\substack{Monotonie,\\ unabh.}}{\geq} \lim_{n\to\infty} \frac{1}{n} \log \prod_{i=1}^n \mathbb{P}(X_i \geq a) \overset{i.d.}{=} \lim_{n\to\infty} \frac{1}{n} \log([\mathbb{P}(X_1 \geq a)]^n)$$
$$= \lim_{n\to\infty} \log([\mathbb{P}(X_1 \geq a)]^{\frac{n}{n}}) = \lim_{n\to\infty} \log \mathbb{P}(X_1 \geq a) = \log \mathbb{P}(X_1 \geq a)$$
$$\Longrightarrow -\infty = \log \mathbb{P}(X_1 \geq a) \Longrightarrow \mathbb{P}(X_1 \geq a) = 0$$

2 ⟹ 3: Sei $n \in \mathbb{N}$ beliebig. Dann lässt sich $\mathbb{P}(S_n \geq na)$ wegen der Unabhängigkeit der X_i als Summe von Produkten schreiben, in denen jeweils für ein $i \in \{1, \dots, n\}$ der Faktor $\mathbb{P}(X_i \geq a)$ vorkommt. Wegen 2. und der identischen Verteilung der X_i ist aber $\mathbb{P}(X_i \geq a) = 0 \;\forall\, i \in \{1, \dots, n\}$. Damit folgt dann $\mathbb{P}(S_n \geq na) = 0$.

3 ⟹ 1: $\mathbb{P}(S_n \geq na) = 0 \;\forall n \in \mathbb{N} \;\Longrightarrow J(a) = \lim_{n\to\infty} \log(0^{\frac{1}{n}}) = \log 0 = -\infty$ □

(ii) Wir können vorwegnehmen, dass der Satz von Cramér für die Dirac-Verteilung (also für $\mathbb{P}(X_1 = \mathbb{E}X_1) = 1$) trivialerweise wahr ist, denn in diesem Fall gilt für $\mathbb{E}X_1 < a$, dass $\mathbb{P}(X_1 \geq a) = 0$ und damit nach Lemma 1

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -\infty = -\sup_{t\in\mathbb{R}}[t(a - \mathbb{E}X_1)] = -\sup_{t\in\mathbb{R}}[ta - \log(e^{t\mathbb{E}X_1})] = -I(a).$$

(iii) Wir betrachten nun einige Eigenschaften von $\varphi(t)$, die wir im Beweis benötigen werden:

Lemma 2:

Sei $\mu \coloneqq \mathbb{P}_{X_1}$ die Verteilung der Zufallsvariable X_1 (und damit natürlich auch die Verteilung der anderen Zufallsvariablen X_i) und sei $F(x) = \mathbb{P}(X_1 \leq x)$ die zugehörige Verteilungsfunktion.

Dann ist $\varphi(t) = \mathbb{E}\, e^{tX_1} = \int_{\mathbb{R}} e^{tx}\mu(dx) = \int_{\mathbb{R}} e^{tx} dF(x)$ und es gelten:

a) $\varphi(t)$ ist unendlich oft differenzierbar mit

$$\varphi^{(n)}(t) = \int_{\mathbb{R}} x^n e^{tx}\, \mu(dx) < \infty \;\forall t \in \mathbb{R}, \forall n \in \mathbb{N}$$

b) $\varphi(t)$ ist strikt konvex.

Beweis:

a) Wir zeigen hierfür, dass $\varphi(t)$ für beliebiges $t_0 > 0$ auf $(-t_0, t_0)$ unendlich oft differenzierbar ist. Daraus folgt dann direkt die Differenzierbarkeit auf ganz $\mathbb{R}$. Wir verwenden im Folgenden das Differentiationslemma (eine Folgerung aus dem Satz über die dominierte Konvergenz; vgl. [Kle, Satz 6.28]) und zeigen mit vollständiger Induktion, dass $\forall n \in \mathbb{N}$ gilt:

$$\varphi^{(n-1)} \text{ ist differenzierbar mit } \varphi^{(n)}(t) = \int_{\mathbb{R}} x^n e^{tx}\, \mu(dx) < \infty \;\forall t \in (-t_0, t_0).$$

Induktionsanfang: Für $n = 1$ ist die Aussage wahr. Das zeigt man mit dem Differentiationslemma. Dieses gibt bestimmte Bedingungen an, unter denen φ differenzierbar ist und unter denen man die Ableitung in das Integral hineinziehen darf.
Prüfung dieser Bedingungen:

- $\forall t \in (-t_0, t_0)$ ist $(x \to e^{tx}) \in L^1(\mu)$ nach (1.1.)
- Für alle $x \in \mathbb{R}$ ist $f: (-t_0, t_0) \to \mathbb{R}$, $t \to e^{tx}$ differenzierbar. Dies folgt aus einfachen Ableitungsregeln und es ist $f'(t) = xe^{tx}$.
- $\sup_{t\in(-t_0,t_0)} |xe^{tx}| \in L^1(\mu)$, denn

$$\sup_{t\in(-t_0,t_0)} |xe^{tx}| \leq \mathbb{1}_{\{x|x<0\}} e^{-(1+t_0)x} + \mathbb{1}_{\{x|x\geq 0\}} e^{(1+t_0)x} \overset{(1.1.)}{\in} L^1(\mu)$$

Damit ist $\varphi(t)$ auf $(-t_0, t_0)$ differenzierbar mit

$$\varphi'(t) = \int_{\mathbb{R}} xe^{tx}\, \mu(dx) < \infty \;\forall\, t \in (-t_0, t_0)$$

Induktionsvoraussetzung: Für beliebiges, festes $n \in \mathbb{N}_0$ gilt

$$\varphi^{(n-1)} \text{ ist differenzierbar mit } \varphi^{(n)}(t) = \int_{\mathbb{R}} x^n e^{tx}\, \mu(dx) < \infty \;\forall t \in (-t_0, t_0)$$

<u>Induktionsschluss</u>: Zeige, dass mit der Induktionsvoraussetzung gilt:

$$\varphi^{(n)} \text{ ist differenzierbar mit } \varphi^{(n+1)}(t) = \int_{\mathbb{R}} x^{(n+1)} e^{tx}\, \mu(dx) \ < \infty\ \forall t \in (-t_0\,, t_0\,)$$

Wir benutzen wieder das Differentiationslemma.
Prüfung der Bedingungen:

- $\forall t \in (-t_0\,, t_0\,)$ ist $(x \to x^n e^{tx}) \in L^1(\mu)$ nach Induktionsvoraussetzung.
- Für alle $x \in \mathbb{R}$ ist $f\colon (-t_0\,, t_0\,) \to \mathbb{R}$, $t \to x^n e^{tx}$ differenzierbar.
  Dies folgt aus einfachen Ableitungsregeln und es ist $f^{'}(t) = x^{(n+1)} e^{tx}$.
- $\sup_{t\in(-t_0\,,t_0\,)} |x^{(n+1)} e^{tx}| \in L^1(\mu)$, denn

$$\sup_{t\in(-t_0\,,t_0\,)} |x^{(n+1)} e^{tx}| \leq \mathbb{1}_{\{x|x<0\}} e^{-(n+1+t_0\,)x} + \mathbb{1}_{\{x|x\geq 0\}} e^{(n+1+t_0)x} \overset{(1.1.)}{\in} L^1(\mu)$$

Damit ist $\varphi^{(n)}$ auf $(-t_0\,, t_0\,)$ differenzierbar mit

$$\varphi^{(n+1)}(t) = \int_{\mathbb{R}} x^{(n+1)} e^{tx}\, \mu(dx) \ < \infty\ \forall\ t \in (-t_0\,, t_0\,)$$

Hiermit ist gezeigt, dass $\varphi(t)$ unendlich oft differenzierbar ist mit

$$\varphi^{(n)}(t) = \int_{\mathbb{R}} x^n e^{tx}\, \mu(dx) < \infty\ \forall t \in \mathbb{R}, \forall n \in \mathbb{N}.$$

b) $\varphi(t)$ ist strikt konvex, da $\varphi''(t) = \int_{\mathbb{R}} x^2 e^{tx}\, \mu(dx) > 0\ \ \forall t \in \mathbb{R}$, falls $\mathbb{P}(X_1 = 0\,) < 1$
In unserem Fall ist aber $\mathbb{P}(X_1 = 0\,) = 1$ ausgeschlossen, da ja die Dirac-Verteilung hier nicht betrachtet wird.

□

(iv) Es sei bemerkt, dass die Existenz von $\mathbb{E}X_1$ aus (iii)a) folgt, da $\mathbb{E}X_1 = \varphi'(0) < \infty$ gilt.

(v) Wir nehmen im folgenden Beweis oBdA $a = 0$ und $\mathbb{E}X_1 < 0$ an.
Alle anderen Fälle folgen direkt daraus, da mit Substitution $Y_i = X_i - a$ gilt:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum\nolimits_{i=1}^{n} X_i \geq na\right) &= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum\nolimits_{i=1}^{n} (X_i - a) \geq 0\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum\nolimits_{i=1}^{n} Y_i \geq 0\right) = -\sup_{t\in\mathbb{R}}[-\log(\mathbb{E}\, e^{tY_1})] \\
&= -\sup_{t\in\mathbb{R}}[-\log(\mathbb{E}\, e^{tX_1} e^{-ta})] \\
&= -\sup\nolimits_{t\in\mathbb{R}}[ta - \log(\mathbb{E}\, e^{tX_1})] = -I(a)
\end{aligned}$$

(vi) Für den nun folgenden, eigentlichen Beweis führen wir die Schreibweisen
$\psi(t) \coloneqq \log \varphi(t)$ und $g \coloneqq \inf_{t\in\mathbb{R}} \varphi(t)$ ein.
$\psi(t)$ heißt auch <u>kumulantenerzeugende Funktion</u> von X_1.
Natürlich gilt $g \geq 0$ und wegen der Monotonie des Logarithmus gilt
$I(0) = \sup_{t\in\mathbb{R}}[-\log \varphi(t)] = -\inf_{t\in\mathbb{R}}[\log \varphi(t)] = -\log[\inf_{t\in\mathbb{R}} \varphi(t)] = -\log g$
und $I(0) = \infty$ für $g = 0$.

Nach diesen Vorüberlegungen führen wir nun den eigentlichen Beweis:

Mit (v) und (vi) vereinfacht sich die zu zeigende Gleichheit. Wir müssen zeigen:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \log g$$

Dafür unterscheiden wir drei Fälle, je nachdem, wo $\mathbb{P}$ Masse hat.
Die ersten beiden Fälle können recht schnell abgehandelt werden, der dritte erfordert die Technik der exponentiellen Maßtransformation, die die zentrale Idee des Beweises darstellt.

a) $\mathbb{P}(X_1 < 0) = 1$

In diesem Fall ist $\varphi'(t) = \int_{\mathbb{R}} xe^{tx}\,\mu(dx) < 0\ \forall\, t \in \mathbb{R}$ und damit φ streng monoton fallend. Folglich gilt unter Benutzung des Satzes von der monotonen Konvergenz $g = \lim_{t\to\infty} \varphi(t) = 0.$ Dann ist mit Lemma 1

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \lim_{n\to\infty} \log(0^{\frac{1}{n}}) = \log 0 = \log g.$$

b) $\mathbb{P}(X_1 \leq 0) = 1$ und $1 > \mathbb{P}(X_1 = 0) > 0$

Auch hier ist, wie in a), φ streng monoton fallend und es gilt

$g = \lim\limits_{t\to\infty} \varphi(t) = \mathbb{P}(X_1 = 0).$

Wegen der Unabhängigkeit und identischen Verteilung der X_i folgt dann

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_1 = X_2 = \cdots = X_n = 0) = \prod_{i=1}^{n} \mathbb{P}(X_i = 0) = g^n$$

und damit $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \lim_{n\to\infty} \frac{1}{n} \log(g^n) = \log g$

c) $\mathbb{P}(X_1 < 0) > 0$ und $\mathbb{P}(X_1 > 0) > 0$

Dann gilt offensichtlich $\lim_{t\to\pm\infty} \varphi(t) = \infty$.

Aus der strikten Konvexität von φ folgt dann, dass es ein eindeutiges $\tau \in \mathbb{R}$ gibt, so dass φ in τ minimal wird, also dass $\varphi'(\tau) = 0$ und $\varphi(\tau) = g > 0$.

Da $\varphi'(0) = \mathbb{E}X_1 < 0$ ist, gilt $\tau > 0$.

Damit kann man die exponentielle Chebyshev-Ungleichung anwenden:

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(e^{\tau S_n} \geq 1) \leq \mathbb{E}e^{\tau S_n} = \mathbb{E}\prod_{i=1}^{n} e^{\tau X_i} \underbrace{=}_{i.i.d} [\varphi(\tau)]^n = g^n$$

So erhält man eine obere Schranke

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \leq \limsup_{n\to\infty} \frac{1}{n} \log(g^n) = \log g.$$

Nun bleibt zu zeigen

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq \log g.$$

Dafür benutzen wir die <u>exponentielle Maßtransformation</u>:

Wir führen dabei eine neue Folge $(Y_i)_{i\in\mathbb{N}}$ von unabhängigen und identisch verteilten Zufallsvariablen ein, die die Verteilung

$$\hat{\mu}(dx) := \frac{1}{g} e^{\tau x}\,\mu(dx)$$

besitzen.

$\hat{\mu}$ heißt auch die <u>Cramér-Transformierte</u> von μ.

Nun benötigen wir drei Lemmata:

Lemma 3:

Es gilt $\mathbb{E}Y_1 = 0$ und $\mathbb{V}Y_1 \in (0,\infty)$.

Beweis: Sei $\widehat{\varphi}(t) = \mathbb{E}e^{tY_1}$. Dann gilt wegen (1.1.):

$$\widehat{\varphi}(t) = \int_{\mathbb{R}} e^{tx}\hat{\mu}(dx) = \frac{1}{g}\int_{\mathbb{R}} e^{tx}e^{\tau x}\mu(dx) = \frac{1}{g}\int_{\mathbb{R}} e^{(t+\tau)x}\,\mu(dx) = \frac{1}{g}\,\varphi(t+\tau) < \infty$$

Folglich ist auch $\widehat{\varphi}$ eine endliche momenterzeugende Funktion, die unendlich oft differenzierbar ist, und es ergibt sich, wenn man sich erinnert, dass $\varphi''(t) > 0$ $\forall t \in \mathbb{R}$:

$$\mathbb{E}Y_1 = \widehat{\varphi}'(0) = \frac{1}{g}\,\varphi'(\tau) = 0$$

$$\mathbb{V}Y_1 = \mathbb{E}(Y_1 - \mathbb{E}Y_1)^2 = \mathbb{E}{Y_1}^2 = \widehat{\varphi}''(0) = \frac{1}{g}\varphi''(\tau) \in (0,\infty)$$

□

Lemma 4: Es sei $T_n = \sum_{i=1}^n Y_i$. Dann gilt

$$\mathbb{P}(S_n \geq 0) = g^n\mathbb{E}\big(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\big).$$

Beweis: Es ist

$$\mu(dx) = ge^{-\tau x}\hat{\mu}(dx).$$

Daher gilt:

$$\begin{aligned}\mathbb{P}(S_n \geq 0) = \mathbb{E}\big(\mathbb{1}_{\{S_n\geq 0\}}\big) &= \int_{\mathbb{R}} \dots \Big(\int_{\mathbb{R}} \mathbb{1}_{\{\sum_{i=1}^n x_i\geq 0\}}\,\mu(dx_1)\Big)\dots\mu(dx_n)\\ &= \int_{\mathbb{R}} \dots \Big(\int_{\mathbb{R}} \mathbb{1}_{\{\sum_{i=1}^n x_i\geq 0\}}\,ge^{-\tau x_1}\hat{\mu}(dx_1)\Big)\dots ge^{-\tau x_n}\hat{\mu}(dx_n)\\ &= g^n\int_{\mathbb{R}} \dots \Big(\int_{\mathbb{R}} \mathbb{1}_{\{\sum_{i=1}^n x_i\geq 0\}}\,e^{-\tau\sum_{i=1}^n x_i}\,\hat{\mu}(dx_1)\Big)\dots\hat{\mu}(dx_n)\\ &= g^n\,\mathbb{E}\big(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\big)\end{aligned}$$

□

Lemma 5: Es gilt

$$\liminf_{n\to\infty}\frac{1}{n}\log\,\mathbb{E}\big(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\big) \geq 0.$$

Beweis:

Sei $C > 0$. Dann erhält man wegen der Monotonie des Erwartungswertes eine untere Schranke

$$\begin{aligned}\mathbb{E}\big(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\big) &\geq \mathbb{E}\left(e^{-\tau T_n}\mathbb{1}_{\{0\leq T_n\leq C\sqrt{n\mathbb{V}Y_1}\}}\right) \geq \mathbb{E}\left(e^{-\tau C\sqrt{n\mathbb{V}Y_1}}\mathbb{1}_{\{0\leq T_n\leq C\sqrt{n\mathbb{V}Y_1}\}}\right)\\ &= e^{-\tau C\sqrt{n\mathbb{V}Y_1}}\,\mathbb{E}\left(\mathbb{1}_{\{0\leq T_n\leq C\sqrt{n\mathbb{V}Y_1}\}}\right) = e^{-\tau C\sqrt{n\mathbb{V}Y_1}}\,\mathbb{P}\left(\frac{T_n}{\sqrt{n\mathbb{V}Y_1}} \in [0,C]\right).\end{aligned}$$

Wegen der Monotonie des Logarithmus gilt dann auch

$$\frac{1}{n}\log\mathbb{E}\big(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\big) \geq \frac{-\tau C\sqrt{n\mathbb{V}Y_1}}{n} + \frac{1}{n}\log\left[\mathbb{P}\left(\frac{T_n}{\sqrt{n\mathbb{V}Y_1}} \in [0,C]\right)\right].$$

Auf Grund von Lemma 3 kann man nun den Zentralen Grenzwertsatz auf T_n anwenden und erhält

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\right)\geq\liminf_{n\to\infty}\left(\frac{-\tau C\sqrt{n\mathbb{V}Y_1}}{n}+\frac{1}{n}\log\left[\mathbb{P}\left(\frac{T_n}{\sqrt{n\mathbb{V}Y_1}}\in[0,C]\right)\right]\right)$$

$$=\liminf_{n\to\infty}\frac{-\tau C\sqrt{n\mathbb{V}Y_1}}{n}+\liminf_{n\to\infty}\frac{1}{n}\log\underbrace{\left[\frac{1}{\sqrt{2\pi}}\int_0^C e^{-\frac{x^2}{2}}dx\right]}_{\in(0,1)}=0.$$

□

Aus Lemma 4 und 5 ergibt sich nun

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(S_n\geq 0)=\liminf_{n\to\infty}\frac{1}{n}\log\left[g^n\mathbb{E}\left(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\right)\right]$$

$$=\log g+\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left(e^{-\tau T_n}\mathbb{1}_{\{T_n\geq 0\}}\right)\geq\log g.$$

Damit ist der Satz gezeigt.

□

Bemerkung 6:
Mit dem Satz von Cramér kann man auch die Wahrscheinlichkeiten $\mathbb{P}(S_n\leq na)$ für $a<\mathbb{E}X_1$ durch $\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(S_n\leq na)=-I(a)$ abschätzen. Dies sieht man leicht durch Übergang von $(X_i)_{i\in\mathbb{N}}$ auf $(-X_i)_{i\in\mathbb{N}}$.

Beispiele 7:
Wir berechnen die durch die Legendre-Transformierte gegebene Ratenfunktion I für eine Folge unabhängiger Zufallsvariablen $(X_i)_{i\in\mathbb{N}}$, die standardnormalverteilt sind:

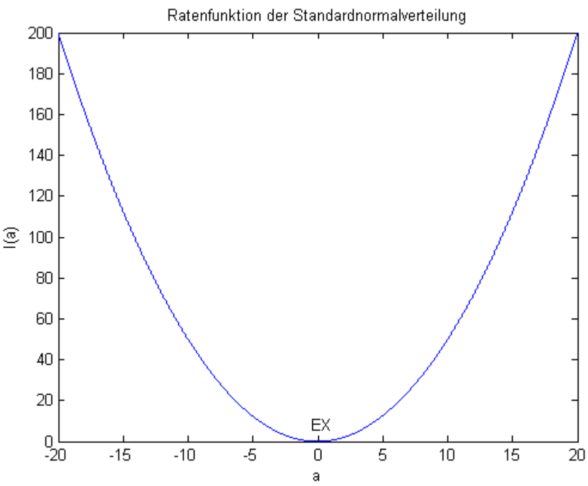
$$\varphi(t)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}e^{tx}\,e^{-\frac{1}{2}x^2}dx$$

$$=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}e^{tx}\,e^{-\frac{1}{2}x^2}e^{-\frac{1}{2}t^2}e^{\frac{1}{2}t^2}dx\overset{bin.Formel}{\cong}e^{\frac{1}{2}t^2}\overbrace{\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}e^{-\frac{1}{2}(x-t)^2}dx}^{=1}=e^{\frac{1}{2}t^2}$$

Damit ist $\varphi(t)<\infty\ \forall\, t\in\mathbb{R}$ und

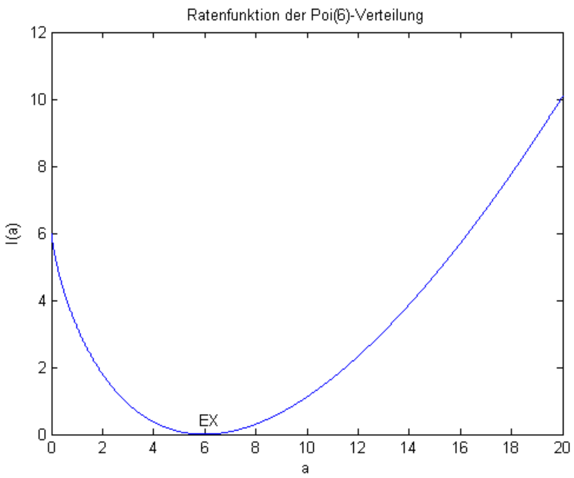
$$I(a)=\sup_{t\in\mathbb{R}}\left[ta-\frac{t^2}{2}\right]=\frac{a^2}{2},$$

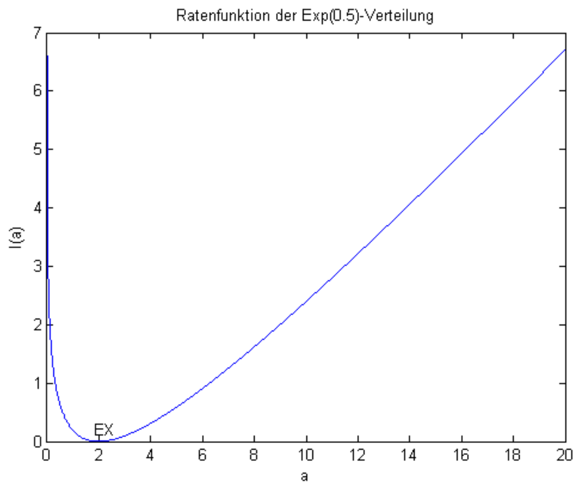
wie man sofort erkennt, wenn man $\left[ta-\frac{t^2}{2}\right]$ zweimal nach t ableitet.



Weitere Beispiele für durch die Legendre-Transformierte gegebene Ratenfunktionen I sollen hier nur angegeben werden (nach [Koe, Beispiel 1.4.2.]).

Für eine Folge $(X_i)_{i\in\mathbb{N}}$ unabhängiger, Poi(λ)-verteilter Zufallsvariablen ist
$I(a) = \lambda - a + a\log\frac{a}{\lambda}$ für $a > 0$, $I(0) = \lambda$, sonst $I(a) = \infty$.
Für eine Folge $(X_i)_{i\in\mathbb{N}}$ unabhängiger, Exp(λ)-verteilter Zufallsvariablen ist
$I(a) = \lambda a - 1 - \log(\lambda a)$ für $a > 0$, sonst $I(a) = \infty$.
In letzterem Fall ist zu beachten, dass $\varphi(t)$ **nicht** für alle $t \in \mathbb{R}$ endlich ist. Hier liegt ein Fall wie im noch folgenden Satz 14 vor.





Im Folgenden sollen einige wichtige Eigenschaften der Legendre-Transformierten I bewiesen werden.

Lemma 8:

Unter den Bedingungen des Satzes von Cramér gilt:

a) I ist von unten halbstetig und konvex auf $\mathbb{R}$.
b) Für $a \in {}^\circ D_I$ existiert ein eindeutiges $t_a \in \mathbb{R}$ mit $t_a a - \psi(t_a) = \max_{t\in\mathbb{R}}[ta - \psi(t)]$, wobei ${}^\circ D_I$ das Innere von $D_I := \{z \in \mathbb{R} | I(z) < \infty\}$ ist.
c) I ist stetig und strikt konvex auf ${}^\circ D_I$.
d) $I(x) \geq 0\ \forall x \in \mathbb{R}$ mit $I(x) = 0$ genau dann, wenn $x = \mathbb{E}X_1$.
e) I hat für alle $L \geq 0$ kompakte Niveaumengen $N_L := \{z \in \mathbb{R} | I(z) \leq L\}$.

Beweis:

a) Konvexität:

$\forall 0 \leq t \leq 1$ und $x, y \in \mathbb{R}$ gilt:

$$\begin{aligned} tI(x) + (1-t)I(y) &= \sup_{s\in\mathbb{R}}[tsx - t\psi(s)] + \sup_{s\in\mathbb{R}}[(1-t)sy - (1-t)\psi(s)] \\ &\geq \sup_{s\in\mathbb{R}}[\ tsx - t\psi(s) + (1-t)sy - (1-t)\psi(s)] \\ &= \sup_{s\in\mathbb{R}}[(tx + (1-t)y)s - \psi(s)] = I(tx + (1-t)y) \end{aligned}$$

Damit ist I konvex.

Halbstetigkeit von unten:

Sei $x \in \mathbb{R}$ und $(x_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen mit $x_n \to x$.
Da man für beliebiges $s \in \mathbb{R}$ die Abschätzung

$$\liminf_{x_n\to x} I(x_n) \geq \liminf_{x_n\to x}[s\,x_n - \psi(s)] = sx - \psi(s)$$

machen kann, ergibt sich

$$\liminf_{x_n \to x} I(x_n) \geq \sup_{s \in \mathbb{R}} [sx - \psi(s)] = I(x)$$

und damit ist gezeigt, dass I von unten halbstetig ist.

b) Eindeutigkeit: Sei $t_a \in \mathbb{R}$ ein Maximierer, d.h. $t_a a - \psi(t_a) = \max_{t \in \mathbb{R}} [ta - \psi(t)]$. Nach der Kettenregel ist $\psi'(t) = \frac{\varphi'(t)}{\varphi(t)}$. Dann gilt $0 = a - \frac{\varphi'(t_a)}{\varphi(t_a)}$.
Also ist $a = \frac{\varphi'(t_a)}{\varphi(t_a)}$. Um $\psi''(t)$ zu bestimmen, definieren wir die Verteilungsfunktion

$$F_t(x) = \frac{1}{\varphi(t)} \int_{-\infty}^{x} e^{ty}\, dF(y).$$

Der Erwartungswert einer Zufallsvariable, die diese Verteilung besitzt, ist dann

$$\int_{\mathbb{R}} x d\, F_t(x) = \frac{1}{\varphi(t)} \int_{\mathbb{R}} x\, e^{tx} dF(x) = \frac{\varphi'(t)}{\varphi(t)}.$$

Dann erhält man nach der Kettenregel und der Jensen'schen Ungleichung

$$\psi''(t) = \frac{d}{dt} \frac{\varphi'(t)}{\varphi(t)} = \frac{\varphi''(t)}{\varphi(t)} - \left(\frac{\varphi'(t)}{\varphi(t)}\right)^2 = \int_{\mathbb{R}} x^2\, dF_t(x) - \left(\int_{\mathbb{R}} x\, dF_t(x)\right)^2 \geq 0\,. \quad (1.2)$$

Da man die Dirac-Verteilung in $\mathbb{E}X_1$ nicht betrachten muss, da für diese $°D_I = \{\}$ ist, ist diese Ungleichung sogar strikt. Damit ist $\frac{\varphi'(t)}{\varphi(t)}$ streng monoton wachsend und folglich ist t_a eindeutig.

Existenz: Hierfür zeigt man, dass $a, b \in [-\infty, \infty]$ mit $a \leq z \leq b\ \forall z \in D_I$ die Asymptoten der Abbildung $t \to \psi'(t)$ bilden, denn dann kann man aus der Stetigkeit von ψ' folgern, dass zu jedem $a \in °D_I$ ein $t_a \in \mathbb{R}$ existiert für das $\psi'(t_a) = a$ gilt. Damit ist $0 = \frac{d}{dt}(t_a a - \psi(t_a))$ und folglich ist t_a mit (1.2.) ein Maximierer. Für einen genauen Beweis siehe [Geb, Lemma 1.10 und Lemma 1.14]

c) Wir zeigen zunächst die Stetigkeit auf $°D_I$:
Es folgt direkt aus der Konvexität von I, dass $°D_I$ ein offenes Intervall ist.
Desweiteren folgt aus der Konvexität von I, dass $\forall\, a, b, c \in °D_I$ mit $a < b < c$ gilt, dass

$$\frac{I(b) - I(a)}{b - a} \leq \frac{I(c) - I(b)}{c - b} \quad (1.3.)$$

(vgl. [Ama, S. 339])
Sei nun $x \in °D_I$ beliebig gewählt und sei auch $y \in °D_I$ mit $x \neq y$ beliebig. Dann existieren $a, b, c, d \in °D_I$ mit $a < b < x, y < c < d$ und man erhält durch wiederholte Anwendung von (1.3.)

$$\frac{I(b) - I(a)}{b - a} \leq \frac{I(y) - I(x)}{y - x} \leq \frac{I(d) - I(c)}{d - c} \quad \text{für } y > x$$

und

$$\frac{I(b) - I(a)}{b - a} \leq \frac{I(x) - I(y)}{x - y} \leq \frac{I(d) - I(c)}{d - c} \quad \text{für } y < x.$$

Ist $L := \max\left\{\left|\frac{I(b) - I(a)}{b - a}\right|, \left|\frac{I(d) - I(c)}{d - c}\right|\right\}$ so ist $\left|\frac{I(x) - I(y)}{x - y}\right| \leq L$.

Nun sieht man die Stetigkeit direkt, denn ist $\varepsilon > 0$ gegeben, so gilt für $\delta = \frac{\varepsilon}{2(L+1)} > 0$

$$|x - y| < \delta \Longrightarrow |I(x) - I(y)| \leq L\delta = \frac{\varepsilon L}{2(L + 1)} < \varepsilon.$$

Wir zeigen nun noch die strikte Konvexität von I auf $°D_I$:
Wie bereits in a) gezeigt gilt $\forall\ 0 < t < 1$ und $\forall\ x, y \in \mathbb{R}$ mit $x \neq y$

$$tI(x) + (1-t)I(y) = \sup_{s\in\mathbb{R}}[tsx - t\psi(s)] + \sup_{s\in\mathbb{R}}[(1-t)sy - (1-t)\psi(s)]$$
$$\geq \sup_{s\in\mathbb{R}}[(tx + (1-t)y)s - \psi(s)] = I(tx + (1-t)y).$$

Das Gleichheitszeichen steht in der 2. Zeile genau dann, wenn beide Suprema in der 1. Zeile an der selben Stelle $s_0 \in \mathbb{R}$ angenommen werden. Man weiß nun aus b), dass beide Suprema Maxima für je einen eindeutigen Maximierer s_x bzw. s_y sind. Damit in der 2. Zeile das Gleichheitszeichen steht, muss also gelten $s_x = s_y$. Es ist aber $\psi'(s_x) = x \neq y = \psi'(s_y)$ und damit $s_x \neq s_y$, da ja ψ' streng monoton wachsend ist, falls F nicht die Dirac-Verteilung ist, was wir jedoch wieder annehmen können, da für die Dirac-Verteilung $°D_I = \{\}$ ist.
Es gilt also $tI(x) + (1-t)I(y) > I(tx + (1-t)y)$ $\forall 0 < t < 1$ und $\forall\ x, y \in \mathbb{R}$ mit $x \neq y$ und damit ist I strikt konvex.

d) Da $\psi(0) = \log \mathbb{E}(1) = 0$, folgt direkt $I(x) \geq 0x - \psi(0) = 0\ \forall x \in \mathbb{R}$.
Um $I(\mathbb{E}X_1) = 0$ zu zeigen, beachte, dass $-\log$ eine konvexe Funktion ist und benutze die Jensen'sche Ungleichung:

$$\psi(t) = \log \mathbb{E}\ e^{tX_1} \geq \mathbb{E}(\log\ e^{tX_1}) = t\ \mathbb{E}X_1$$

Dann ergibt sich $0 \leq I(\mathbb{E}X_1) \leq t\ \mathbb{E}X_1 - t\ \mathbb{E}X_1 = 0$.
Da $I(x) \geq 0\ \forall x \in \mathbb{R}$, ist $\mathbb{E}X_1$ also Minimalstelle. Aus der Konvexität von I auf $\mathbb{R}$ bzw. der strikten Konvexität von I auf $°D_I$ folgt dann, dass $\mathbb{E}X_1$ die einzige Minimalstelle von I ist.

e) Sei $L \geq 0$ beliebig. Um die Kompaktheit der Niveaumengen N_L zu zeigen, zeigen wir, dass diese beschränkt und abgeschlossen sind.
Auf den Niveaumengen ist I nach c) stetig. Da die Stetigkeit äquivalent dazu ist, dass Urbilder abgeschlossener Mengen abgeschlossen sind, folgt die Abgeschlossenheit von N_L direkt aus der Abgeschlossenheit der Menge

$$\{I(z) | z \in \mathbb{R}\ \text{ und } I(z) \leq L\} = [0, L].$$

Die Beschränktheit von N_L folgt aus der Tatsache, dass einerseits $I(z) \geq z - \psi(1)\ \forall z \in \mathbb{R}$ und folglich $z \leq L + \psi(1) < \infty\quad \forall z \in N_L$, und andererseits $I(z) \geq -z - \psi(-1)\ \forall z \in \mathbb{R}$ und folglich $z \geq -L - \psi(-1) > -\infty\quad \forall z \in N_L$.

□

Als weitere wichtige Eigenschaft der Legendre-Transformierten, deren Beweis hier nicht gegeben werden soll, findet man, dass die Legendre-Transformierte I unendlich oft differenzierbar ist und dass gilt $I''(\mathbb{E}X_1) = \frac{1}{\mathbb{V}X_1}$ (vgl. [Win, Lemma 1.12]).

Bemerkung 9:

Aus Lemma 8 a) und d) kann man folgern, dass, wenn $a > \mathbb{E}X_1$ gilt, $I(z) \geq I(a)\ \forall z \geq a$ ist. Damit gilt unter den Bedingungen des Satzes von Cramér auch die umgeschriebene Gleichung

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\tfrac{1}{n} S_n \in A) = -\inf_{z\in A} I(z) \qquad \text{mit } A = [a, \infty)$$

für alle $a > \mathbb{E}X_1$

Wir haben bislang I einfach Ratenfunktion genannt, ohne den Begriff „Ratenfunktion" zu definieren. Das soll hier nachgeholt werden:

Definition 10:

Sei E ein polnischer Raum mit vollständiger Metrik d. Eine von unten halbstetige Funktion $I: E \to [0, \infty]$ heißt Ratenfunktion. Sind die Niveaumengen

$$N_L \coloneqq \{z \in \mathbb{R} | I(z) \leq L\}$$

für alle $L \geq 0$ kompakt , so heißt I gute Ratenfunktion.

Es ist offensichtlich, dass die Legendre-Transformierte aus dem Satz von Cramér eine gute Ratenfunktion nach Definition 10 ist.

Korollar 11:

Unter den Bedingungen aus dem Satz von Cramér genügt die Folge $(X_i)_{i \in \mathbb{N}}$ dem starken Gesetz der großen Zahlen.

Beweis:

Sei oBdA $\mathbb{E}X_1 = 0$.
Bemerke, dass $\forall \delta > 0$

$$\mathbb{P}\left(\frac{|S_n|}{n} \geq \delta\right) = \mathbb{P}\left(\frac{S_n}{n} \geq \delta\right) + \mathbb{P}\left(\frac{S_n}{n} \leq -\delta\right)$$

gilt.
Aus dem Satz von Cramér und Bemerkung 6 folgt dann, dass für genügend große $n \in \mathbb{N}$

$$\mathbb{P}\left(\frac{S_n}{n} \geq \delta\right) \leq e^{-\frac{1}{2}nI(\delta)}$$

und

$$\mathbb{P}\left(\frac{S_n}{n} \leq -\delta\right) \leq e^{-\frac{1}{2}nI(\delta)}$$

gelten, da ja $e^{-nI(\delta)} < e^{-\frac{1}{2}nI(\delta)}$ ist.

Es gilt also

$$\mathbb{P}\left(\frac{|S_n|}{n} \geq \delta\right) \leq 2e^{-\frac{1}{2}nI(\delta)}.$$

Daraus wiederum folgt, dass für ein beliebiges, genügend großes $m \in \mathbb{N}$

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{|S_n|}{n} \geq \delta\right) = \sum_{n=1}^{m-1} \underbrace{\mathbb{P}\left(\frac{|S_n|}{n} \geq \delta\right)}_{\leq 1} + \sum_{n=m}^{\infty} \mathbb{P}\left(\frac{|S_n|}{n} \geq \delta\right) \leq (m-1) + 2\sum_{n=m}^{\infty} e^{-\frac{1}{2}nI(\delta)}$$

$$= (m-1) + 2\sum_{n=m}^{\infty} \underbrace{\left(e^{-\frac{1}{2}I(\delta)}\right)}_{\in(0,1)}^{n} < \infty$$

gilt. Nun kann man das Borel-Cantelli-Lemma anwenden und erhält

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{|S_n|}{n}\geq\delta\right)=0\ \forall\,\delta>0,$$

woraus direkt das starke Gesetz der großen Zahlen folgt:

$$\mathbb{P}(\limsup_{n\to\infty}\left|\frac{\sum_{i=1}^{n}(X_i-\mathbb{E}X_i)}{n}\right|=0)=\mathbb{P}\left(\limsup_{n\to\infty}\frac{|S_n|}{n}=0\right)=1$$

Es soll nun zum Schluss noch gesagt werden, dass der Satz von Cramér ein Prinzip großer Abweichungen nach dem zweiten Teil folgender Definition ist:

Definition 12:

Sei E ein polnischer Raum mit vollständiger Metrik d, sei I eine Ratenfunktion und $(\mu_\varepsilon)_{\varepsilon>0}$ eine Familie von Wahrscheinlichkeitsmaßen auf E. Wir sagen, dass $(\mu_\varepsilon)_{\varepsilon>0}$ ein Prinzip großer Abweichungen (kurz LDP für Large Deviations Principle) mit Ratenfunktion I erfüllt, falls

(LDP1) $\liminf_{\varepsilon\to 0}\varepsilon\log(\mu_\varepsilon(U))\geq-\inf I(U)$ für jedes offene $U\subset E$

(LDP2) $\limsup_{\varepsilon\to 0}\varepsilon\, log(\mu_\varepsilon(C))\leq-\inf I(C)$ für jedes abgeschlossene $C\subset E$

Wir sagen, dass eine Familie $(\mathbb{P}_n)_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf E ein LDP mit Rate $r_n\uparrow\infty$ und Ratenfunktion I erfüllt, falls (LDP1) und (LDP2) für die Folge $\varepsilon_n=\frac{1}{r_n}$ und für $\mu_{\frac{1}{r_n}}=\mathbb{P}_n$ gelten.

Satz 13:
Gelten die Bedingungen aus dem Satz von Cramér, so erfüllt $\mathbb{P}_n\coloneqq\mathbb{P}_{S_n/n}$ ein LDP mit Rate n und guter Ratenfunktion $I(a)\coloneqq\sup_{t\in\mathbb{R}}[ta-\log\varphi(t)]$.

Beweis: siehe [Kle, Beispiel 23.10]

Eine Verstärkung dieses Satzes schließt den Fall mit ein, dass die momenterzeugende Funktion φ nicht endlich ist:

Satz 14:
Sei $(X_i)_{\,i\in\mathbb{N}}$ eine Folge von i.i.d. reellen Zufallsvariablen. Dann erfüllt $(\mathbb{P}_{S_n/n})_{n\in\mathbb{N}}$ ein LDP mit Ratenfunktion $I(a)=\sup_{t\in\mathbb{R}}[ta-\log\varphi(t)]$.

Beweis: siehe [Dem, Theorem 2.2.3]

Bemerkung 15:
Die Ratenfunktion aus Satz 14 muss nicht unbedingt eine gute Ratenfunktion sein.

Literaturverzeichnis:

[Ama] H. AMANN, J. ESCHER; *Analysis I*. 2. Auflage, Birkhäuser Verlag, Basel – Boston – Berlin (2002).

[Dem] A. DEMBO, O. ZEITOUNI, *Large deviations techniques and applications.* 2. Auflage, Applications of Mathematics Vol. 38; Springer-Verlag, New York (1998).

[Geb] S. GEBENNUS, *Das asymptotische Verhalten der gewichteten Höhe des zweifach gewichteten Verzweigungsprozesses*. Diplomarbeit, WWU Münster (2006); Download als pdf-Datei unter http://wwwmath.unimuenster.de/statistik/alsmeyer/Diplomarbeiten/Gebennus.pdf.

[Hol] F. DEN HOLLANDER, *Large Deviations*. Fields Institute Monographs Vol. 14, AMS, Providence, RI (2000).

[Kle] A. KLENKE, *Wahrscheinlichkeitstheorie*. 2. Auflage, Springer-Verlag, Berlin – Heidelberg (2008).

[Koe] W. KÖNIG, *Große Abweichungen, Techniken und Anwendungen*. Vorlesungsskript, Universität Leipzig (Wintersemester 2004/05 und Sommersemester 2006); Download als pdf-Datei unter http://www.wias-berlin.de/people/koenig/www/GA.pdf.

[Loe] M. LÖWE, *Wahrscheinlichkeitstheorie II*. Vorlesungsskript, WWU Münster; Download als pdf-Datei unter http://wwwmath.uni-muenster.de/statistik/loewe/skript.pdf.

[Win] A. WINTER, *Die Theorie der großen Abweichungen und Anwendungen. Basierend auf dem gleichnamigen Skript von Wolfgang König*. Vorlesungsskript, TU München (Wintersemester 2009/10); Download als pdf-Datei unter http://www-m5.ma.tum.de/pers/winter/ld01.pdf.