Startbeispiel

X_i, $i \in \mathbb{N}$ u.i.v. $\mathrm{Ber}(p)$, $p \in (0,1)$,

$S_n := X_1 + \ldots + X_n$ $(\sim \mathrm{Bin}(n,p))$.

Wir wissen:

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{} p \quad \text{f.s.} \quad (\text{starkes Ges. d. gr. Z.}),$$

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1) \quad (\text{zentraler Grenzwertsatz bzw. Satz von de Moivre – Laplace}),$$

d.h. „typischerweise" gilt $S_n = np \pm O(\sqrt{n})$.

Frage: Wie unwahrscheinlich sind untypische Ereignisse,

$$\mathbb{P}(S_n \geq an) \approx \,?$$

für $a > p$ (und wir denken an „großes" n).

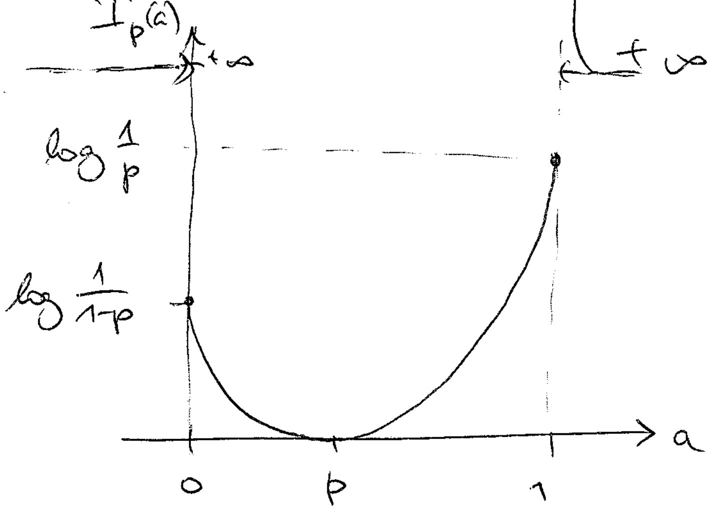
Beob. Für $p \le a \cancel{\le 1}$ ist

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

und für $\cancel{0 \le}\, a \le p$ ist

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

wobei $$I_p(a) = \begin{cases} a \log \frac{a}{p} + (1-a)\log\frac{1-a}{1-p}, & a \in [0,1] \\ +\infty, & \text{sonst.} \end{cases}$$



[grob gesprochen also $\mathbb{P}(S_n \ge n \cdot a) \approx e^{-nI_p(a)}$
diese W'keit ist also exponentiell klein in n]

Betrachte $a \in (p, 1)$:

Für $k > np$ ist

$$\frac{\mathrm{Bin}_{n,p}(\{k+1\})}{\mathrm{Bin}_{n,p}(\{k\})} = \frac{\binom{n}{k+1} p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{n-k}{k+1} \cdot \frac{p}{1-p} < 1$$

$(\Leftrightarrow (n-k)p < (k+1)(1-p)$
$\Leftrightarrow np + p - 1 < k)$

also

$$\mathbb{P}(S_n = \lceil na \rceil) \le \mathbb{P}(S_n \ge na) = \sum_{k=\lceil na \rceil}^{n} \mathbb{P}(S_n = k)$$

$$\le n(1-a)\, \mathbb{P}(S_n = \lceil na \rceil),$$

d.h. $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge a) = \lim \frac{1}{n} \log \mathbb{P}(S_n = \lceil na \rceil)$,

$$\mathbb{P}(S_n = \lceil na \rceil) = \binom{n}{\lceil na \rceil} p^{\lceil na \rceil} (1-p)^{n-\lceil na \rceil}$$

Stirling-Approx: $n! \sim \sqrt{2\pi n}\, n^n e^{-n}$ (im Sinne von $\frac{n!}{\sqrt{2\pi n}\, n^n e^{-n}} \xrightarrow{n\to\infty} 1$)

$$\sim \frac{\sqrt{2\pi n}\, n^n e^{-n} \cdot p^{\lceil na \rceil} (1-p)^{n-\lceil na \rceil}}{\sqrt{2\pi \lceil na \rceil}\, \lceil na \rceil^{\lceil na \rceil} e^{-\lceil na \rceil} \cdot \sqrt{2\pi (n-\lceil na \rceil)}\, (n-\lceil na \rceil)^{(n-\lceil na \rceil)} e^{-n+\lceil na \rceil}}$$

$$= \sqrt{\frac{n}{2\pi \lceil na \rceil (n-\lceil na \rceil)}}\; \frac{n^n \cdot p^{\lceil na \rceil} (1-p)^{n-\lceil na \rceil}}{\lceil na \rceil^{\lceil na \rceil} (n-\lceil na \rceil)^{n-\lceil na \rceil}}$$

$$= \frac{1}{\sqrt{2\pi n a(1-a)}} \underbrace{\left(\frac{p}{a}\right)^{na} \left(\frac{1-p}{1-a}\right)^{n(1-a)}}_{= \exp\left(-n\underbrace{\left\{a \log \frac{a}{p} + (1-a) \log \frac{1-a}{1-p}\right\}}_{= I_p(a)}\right)} \cdot K_n$$

mit Korrekturfaktor

$$K_n = \underbrace{\sqrt{\frac{n^2 a(1-a)}{\lceil na \rceil (n-\lceil na \rceil)}}}_{\xrightarrow{n\to\infty} 1} \; \underbrace{\frac{(na)^{na}}{\lceil na \rceil^{\lceil na \rceil}} \frac{(n-na)^{n-na}}{(n-\lceil na \rceil)^{n-\lceil na \rceil}}\, p^{\overbrace{\lceil na \rceil - na}^{\in [0,1)}} (1-p)^{na-\lceil na \rceil}}_{\in \left[\frac{1}{c}, c\right] \text{ für } n \text{ gen. groß mit einem } 1<c<\infty}$$

$$= \underbrace{\left(\frac{na}{\lceil na \rceil}\right)^{\lceil na \rceil}} \underbrace{\left(\frac{n-na}{n-\lceil na \rceil}\right)^{n-\lceil na \rceil}} \left(\frac{\cancel{n} a}{\cancel{n}(1-a)}\right)^{na-\lceil na \rceil}$$

$$= \underbrace{\left(1 + \frac{na - \lceil na \rceil}{\lceil na}\right)^{\lceil na \rceil}}_{\sim e^{na-\lceil na \rceil}} \cdot \underbrace{\left(1 + \frac{\lceil na \rceil - na}{n - \lceil na \rceil}\right)^{n-\lceil na \rceil}}_{\sim e^{\lceil na \rceil - na}}$$

Bem. Insbes. für $a > p$ und jedes $\varepsilon > 0$

$$\lim_{n\to\infty} \underbrace{\mathbb{P}(S_n < n(a+\varepsilon) \mid S_n \geq na)}_{} = 1$$

$$= \frac{\mathbb{P}(S_n \geq na) - \mathbb{P}(S_n \geq n(a+\varepsilon))}{\mathbb{P}(S_n \geq na)}$$

$$= 1 - \frac{\mathbb{P}(S_n \geq n(a+\varepsilon))}{\mathbb{P}(S_n \geq na)} = \cancel{1 - e^{-n(I_p(a+\varepsilon))}}$$

$$= 1 - \frac{e^{-n(I_p(a+\varepsilon) + o(1))}}{e^{-n(I_p(a) + o(1))}}$$

$$= 1 - \exp\Big(-n\big(\underbrace{I_p(a+\varepsilon) - I_p(a)}_{>0}\big) + o(n)\Big)$$

Also: Bedingt auf untypisch hohen Wert von S_n sehen wir mit hoher W'keit „den am wenigsten untypischen" unter den betrachteten Ausgängen.

Bem.: $I_p(a) = H(\mathrm{Ber}(a) \mid \mathrm{Ber}(p))$, die sog. relative Entropie.

1. Große Abweichungen für Summen von u.i.v. reellen ZVn

$X_1, X_2, \ldots$ u.i.v., reellwertige ZVn,

sei $\varphi(t) := \mathbb{E}[e^{tX_1}] \in (0,\infty) \quad \forall t \in \mathbb{R}$

(„Cramér-Bedingung")

(φ ist die Momentenerzeugende Fkt:
$$\frac{d^k}{dt^k}\varphi(t)\Big|_{t=0} = \mathbb{E}[X_1^k] \ ),$$

insbes. $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}[X_1]$ ex.,

sei $\sigma^2 > 0$.

$$S_n := X_1 + X_2 + \ldots + X_n$$

Satz 1 (Cramér, 1938; Chernoff 1952)

(Carl Harald Cramér, 1893–1985; Herman Chernoff, *1923)

Für $a > \mu$ ist
$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -I(a)$$
mit $I(z) = \sup_{t \in \mathbb{R}} \{tz - \log \varphi(t)\}$, $z \in \mathbb{R}$

(I ist die Legendre-Transformierte von φ)

(analog gilt für $a < \mu$: $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \leq na) = -I(a)$)

(einfache Beob.:
$$\mathbb{P}(S_n \geq na) = \mathbb{P}(e^{tS_n} \geq e^{tna}) \overset{\text{Markov-Ungl.}}{\leq} e^{-nta} \underbrace{\mathbb{E}[e^{tS_n}]}_{=(\mathbb{E}[e^{tX_1}])^n}$$
$$= \exp(-n(ta - \log\varphi(t))) \quad \text{für jedes } t \geq 0,$$

also $\frac{1}{n} \log \mathbb{P}(S_n \geq na)$
$$\leq -\sup_{t \geq 0} \{ta - \log \varphi(t)\}$$

<u>Lemma 2</u> X reelle ZV, $\varphi(t) := \mathbb{E}[e^{tX}] \in (0,\infty)\ \forall t \in \mathbb{R}$,

$\mu := \mathbb{E}[X]$, $\sigma^2 := \mathrm{Var}[X] > 0$,

$$I(x) := \sup_{t\in\mathbb{R}} \{tx - \log\varphi(t)\},\ x \in \mathbb{R} \qquad (*)$$

die Legendre-Transformierte von $\log\varphi$

Es gilt

1. $\varphi \in C^\infty(\mathbb{R})$, $\frac{d^k}{dt^k}\varphi(t) = \mathbb{E}[X^k e^{tX}]$,

   $\log\varphi$ ist strikt konvex,

   $\lim_{t\to\infty} \frac{1}{t}\log\varphi(t) = \operatorname{ess\,sup} X$, $\lim_{t\to-\infty}\frac{1}{t}\log\varphi(t) = \operatorname{ess\,inf} X$

2. a) $I \geq 0$, I ist konvex, nach unten halbstetig.

   Die Niveaumengen $\{x \in \mathbb{R} : I(x) \leq s\}$ sind kompakt für $s \in \mathbb{R}$

   b) Für $x \in (\operatorname{ess\,inf} X, \operatorname{ess\,sup} X)$ $(= \mathring{D}_I)$

   gibt es einen eindeutigen Maximierer t_x in $(*)$,

   bestimmt durch $\varphi(t_x)\cdot x = \mathbb{E}[Xe^{tX}]$ $(= \varphi'(t_x))$.

   I ist diff'bar und strikt konvex in $(\operatorname{ess\,inf} X, \operatorname{ess\,sup} X)$,

   $I \equiv \infty$ auf $\mathbb{R} \setminus [\operatorname{ess\,inf} X, \operatorname{ess\,sup} X]$

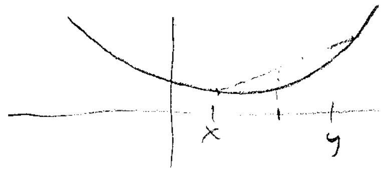
   c) $I(x) = 0 \iff x = \mu$, $I''(\mu) = \frac{1}{\sigma^2}$

   d) $\log\varphi(t) = \sup_{x\in\mathbb{R}} \{tx - I(x)\}$ (Umkehrformel)

(<u>Erinnerungen:</u> $f: \mathbb{R} \to \mathbb{R}$

- f konvex $:\Leftrightarrow$ $\forall x, y \in \mathbb{R}, \alpha \in [0,1]:$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y)$$



- f nach unten halbstetig $:\Leftrightarrow$

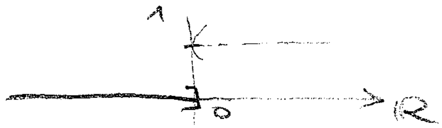
$$f^{-1}((t,\infty)) \text{ offen für jedes } t \in \mathbb{R}$$

$$\Leftrightarrow \quad f^{-1}((-\infty,t]) \text{ abgeschl. } \text{ —— " ——}$$

$$\left(\Leftrightarrow \quad \left(x_n \xrightarrow[n\to\infty]{} x \quad \Rightarrow \quad \liminf_n f(x_n) \geq f(x)\right)\right)$$

$$\Leftrightarrow \quad f = \sup\{g : g: \mathbb{R} \to \mathbb{R} \text{ stetig und } g \leq f\}$$

<u>Bsp.</u>



$f = 1_{(0,\infty)}$ ist n.u.h.,

$f = 1_{[0,\infty)}$ nicht (es ist nach oben halbstetig)

- X reelle ZV,

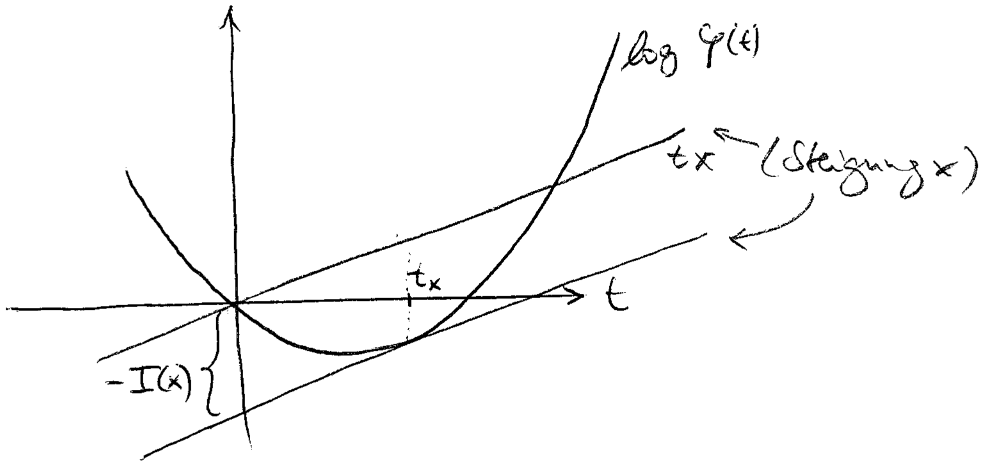
$$\operatorname{ess\,sup} X := \inf\{a \in \mathbb{R} : P(X \leq a) = 1\},$$

$$\operatorname{ess\,inf} X := \sup\{a \in \mathbb{R} : P(X \geq a) = 1\}$$

)

bis hierher am 30.10.

Geometrische Interpretation der Legendre-Transformierten



Bew. von Lemma 2:

1. Formel für $\frac{d^k}{dt^k}\varphi$ durch Differenzieren unter dem Erwartungswert.

(Gerechtfertigt (s.a. Klenke, W'theorie, Satz 6.28):

$$\frac{1}{h}\left(\mathbb{E}[X^k e^{(t+h)X}] - \mathbb{E}[X^k e^{tX}]\right)$$

$$= \mathbb{E}\Big[X^k e^{tX}\cdot \underbrace{\frac{1}{h}(e^{hX}-1)}_{\substack{= e^{\tilde h X} \text{ mit } \tilde h \in [-|h|,|h|], \\ |\dots| \le e^{hX}+e^{-hX}}}\Big] \xrightarrow[h\to 0]{} \mathbb{E}[X^{k+1}e^{tX}]$$

(dom. Konvergenz))

Zur Konvexität:

$u\in(0,1)$, $t_1<t_2$:

$$\log \mathbb{E}[e^{(ut_1+(1-u)t_2)X}] \le \log\left((\mathbb{E}[e^{t_1X}])^u\cdot(\mathbb{E}[e^{t_2X}])^{1-u}\right)$$

(Hölder-Ungl., ist strikt da X nicht auf einen Punkt konzentriert)

zum Verh. bei $t\to\infty$:

$\varphi(t)\le e^{t\cdot \operatorname{ess\,sup} X}$ $\forall t\ge 0$,

$\varphi(t)\ge \exp(t(\operatorname{ess\,sup} X-\varepsilon))\overbrace{\mathbb{P}(X\ge \operatorname{ess\,sup} X-\varepsilon)}^{>0}$ (bzw. leicht modifizierte Formulierung falls ess sup $X=\infty$)

analoges Argument für $t\to-\infty$.

2. a) $I(x) \geq 0 \cdot x - \log \varphi(0) = 0$.

Zur Konvexität: $u \in (0,1)$, $x_1 < x_2$

$$\sup_{t \in \mathbb{R}} \{ (ux_1 + (1-u)x_2)t - \log \varphi(t) \}$$
$$\leq \sup_t \{ ux_1 t - u \log \varphi(t) \} + \sup_t \{ (1-u)x_2 t - (1-u) \log \varphi(t) \}$$
$$= u I(x_1) + (1-u) I(x_2)$$

Für $s \geq 0$ ist

$$\{x : I(x) \leq s\} = \bigcap_{t > 0} \left(-\infty, \frac{s + \log \varphi(t)}{t}\right] \cap \bigcap_{t < 0} \left[\frac{s + \log \varphi(t)}{t}, \infty\right)$$
$$= \bigcap_{t > 0} \left[\frac{s + \log \varphi(-t)}{-t}, \frac{s + \log \varphi(t)}{t}\right]$$

ist kompakt (und damit insbes. abgeschlossen, d.h. I ist nach unten halbstetig)

b) $t \mapsto tx - \log \varphi(t)$ ist strikt konkav
$\Rightarrow$ höchstens ein Maximierer in $(*)$,

gilt $\frac{d}{dt}(tx - \log \varphi(t)) = x - \frac{\varphi'(t)}{\varphi(t)} = 0$ für ein $t = t_x$,
so ist dieses t_x der Maximierer.

Es ist $\lim_{t \to \infty} (\log \varphi)'(t) = \lim_{t \to \infty} \frac{\mathbb{E}[e^{tX} \cdot X]}{\mathbb{E}[e^{tX}]} = \operatorname{ess\,sup} X$

(Argument im Fall $\beta := \operatorname{ess\,sup} X < \infty$ [$\beta = \infty$ mit leichter Modifikation]:
Für $\varepsilon > 0$ ist $\frac{\mathbb{E}[e^{tX} 1_{\{X \leq (1-\varepsilon)\beta\}}]}{\mathbb{E}[e^{tX} 1_{\{X > (1-\varepsilon)\beta\}}]} \leq \frac{e^{(1-\varepsilon)\beta t}}{e^{(1-\varepsilon/2)\beta t} \underbrace{P(X \geq (1-\frac{\varepsilon}{2})\beta)}_{>0}} \xrightarrow[t \to \infty]{} 0$,

also $\frac{\mathbb{E}[e^{tX} X]}{\mathbb{E}[e^{tX}]} \geq \frac{(1-\varepsilon)\beta\, \mathbb{E}[e^{tX} 1_{\{X > (1-\varepsilon)\beta\}}]}{\mathbb{E}[e^{tX} 1_{\{X > (1-\varepsilon)\beta\}}](1 + o(1))} \xrightarrow[t \to \infty]{} (1-\varepsilon)\beta$)

analog: $\lim_{t\to-\infty}(\log\varphi)'(t) = \operatorname{ess\,inf} X$

also: für $x \in \mathring{D}_I$ gibt es genau einen Maximierer (in (*))

Zeige: $x \mapsto t_x$ ist diff'bar in $\mathring{D}_I$,

wende dazu Satz v.d. impliziten Fkt an auf

$$F(x,t) = x - (\log\varphi)'(t)$$

(beachte: $\frac{\partial}{\partial t}F(x,t) = -(\log\varphi)''(t) \neq 0$,

d.h. wenn $F(x_0,t_0)=0$ gilt, so kann $F(x,t)=0$ in einer Umgebung von x_0 „nach t auflösen")

demnach $\frac{d}{dx}t_x = \frac{1}{(\log\varphi)''(t_x)}$ (>0, $\log\varphi$ ist strikt konvex)

(denn $\underset{\overset{\shortparallel}{x-(\log\varphi)'(t_x)}}{F(x,t_x)}=0$, also $1-(\log\varphi)''(t_x)\cdot\frac{d}{dx}t_x = 0$)

weiter ist

$$\frac{d}{dx}I(x) = \frac{d}{dx}\left(x t_x - \log\varphi(t_x)\right)$$

$$= t_x + \underbrace{x}_{=(\log\varphi)'(t_x)}\frac{d}{dx}t_x - (\log\varphi)'(t_x)\cdot\frac{d}{dx}t_x = t_x$$

also ist I strikt konvex und diff'bar auf $\mathring{D}_I$.

Für $x > \operatorname{ess\,sup} X$ ist

$$I(x) \geq \limsup_{t\to\infty} t\,\underbrace{\left(x-\frac{1}{t}\log\varphi(t)\right)}_{\geq c>0 \text{ falls } t \text{ gen. groß}} = +\infty$$

analog für $x < \operatorname{ess\,inf} X$.

c) $(\log \varphi)'(0) = \mu$, d.h. $t_\mu = 0$ und $I(\mu) = 0$

wegen strikter Konvexität und $I \geq 0$
ist μ die einzige Nullstelle

$$I''(\mu) = \frac{d}{dx} t_x \Big|_{x=\mu} = \frac{1}{(\log\varphi)''(\underbrace{t_\mu}_{=0})} = \frac{1}{(\log\varphi)''(0)} = \frac{1}{\sigma^2}$$

d) $I(x) \geq tx - \log\varphi(t) \quad \forall\, t, x$,

also $\log\varphi(t) \geq \sup_x \{tx - I(x)\} \quad \forall t$,

mit $x_t := (\log\varphi)'(t) \left(= \frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}\right)$ ist

$I(x_t) = tx_t - \log\varphi(t)$, also gilt die beh. Gleichheit

Bew. (d. Satzes von Cramér)

o.E. $a = 0 > \mu$ (sonst betrachte $\tilde{X}_i := X_i - a$)

d.h. z.z. $\lim \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \cancel{\log}\, \rho$

mit $\rho := \inf_{t\in\mathbb{R}} \log\varphi(t) \;(= -I(0) = -\sup_t \{0.t - \log\varphi(t)\})$

Wir nehmen an, dass $\mathbb{P}(X_1 < 0) \wedge \mathbb{P}(X_1 > 0) > 0$

(wenn $\mathbb{P}(X_1 < 0) = 1$, so ist $\mathbb{P}(S_n \geq 0) = 0$ und $\rho = 0$,

wenn $\mathbb{P}(X_1 \leq 0) = 1$, $\mathbb{P}(X_1 = 0) > 0$, so ist $\rho = \mathbb{P}(X_1 = 0)$
und $\mathbb{P}(S_n \geq 0) = \rho^n$,

wegen $\mu < 0$ muss $\mathbb{P}(X_1 < 0) > 0$ sein)

also $\lim_{t\to\pm\infty} \varphi(t) = \infty$, $\exists!\, t_0$ mit $\varphi(t_0) = \min_t \varphi(t)$,

es gilt $\varphi'(t_0) = 0$
und $t_0 > 0$, da $\varphi'(0) = \mu < 0$

Obere Schranke:

$$P(S_n \geq 0) = P(e^{t_0 S_n} \geq 1) \leq \frac{1}{1} \mathbb{E}[e^{t_0 S_n}]$$

(Markov-Ungl.)

$$= \exp(n \log \varphi(t_0)) = \rho^n$$

Untere Schranke:

$\hat{X}_i$ seien u.i.v. mit Dichte $\frac{e^{t_0 x}}{e^{\rho}}$ bzgl. $\mathcal{L}(X_1)$

(d.h. $P(\hat{X}_1 \in dx) = \frac{e^{t_0 x}}{e^{\rho}} P(X_1 \in dx)$)

$$\hat{\varphi}(t) := \mathbb{E}[e^{t \hat{X}_1}] = \frac{\varphi(t+t_0)}{\varphi(t_0)}$$

$$\mathbb{E}[\hat{X}_1] = \hat{\varphi}'(0) = \frac{\varphi'(t_0)}{\varphi(t_0)} = (\log \varphi)'(t_0) = 0,$$

somit $\hat{\varphi}''(0) = \mathbb{E}[\hat{X}_1^2] = \mathrm{Var}[\hat{X}_1] = \frac{\varphi''(t_0)}{\varphi(t_0)} =: \hat{\sigma}^2 \in (0, \infty)$

[die Vert. von $\hat{X}_1$ ist mit einem exponentiellen Faktor „gekippt" (engl. tilted) bzgl. der von X_1, sie heißt in diesem Kontext auch die Cramér-transformierte oder Esscher-transformierte Vert.]

$$\hat{S}_n := \sum_{i=1}^n X_i$$

$$P(S_n \geq 0) = \overset{e^{n\rho}}{\cancel{\rho^n}}\, \mathbb{E}\Big[e^{-t_0 S_n} \Big(\prod_{i=1}^n \frac{1}{e^{\rho}} e^{t_0 X_i}\Big) 1_{\{S_n \geq 0\}}\Big]$$

$$= \underset{e^{n\rho}}{\cancel{\rho^n}}\, \underbrace{\mathbb{E}[e^{-t_0 \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}]}$$

$$\geq \mathbb{E}[e^{-t_0 \hat{S}_n} 1_{\{0 \leq \hat{S}_n \leq \hat{\sigma}\sqrt{n}\}}]$$

$$\geq e^{-t_0 \hat{\sigma}\sqrt{n}} \underbrace{P\Big(\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0,1]\Big)}_{\xrightarrow[\text{ZGWS}]{n\to\infty} \Phi(1) - \Phi(0) > 0}$$

[Bem.: Man sieht hier bereits, dass eine feinere Argumentation mit dem lokalen ZGWS eine untere Schranke der Form $\ldots \geq \frac{C}{\sqrt{n}} \rho^n$ ergäbe, was auf den Satz von Bahadur-Rao hinweist.]

bis hierher am 6.11

Beispiele für Ratenfunktionen I (im Satz von Cramér)

1) $X_i \sim \mathrm{Ber}(p)$, $\log\varphi(t) = \log(pe^t + (1-p)\cdot 1)$,

$(\log\varphi)'(t) = \dfrac{pe^t}{pe^t+1-p} \overset{!}{=} x \in [0,1] \Leftrightarrow pe^t(1-x) = (1-p)x$,

$t_x = \log\dfrac{(1-p)x}{p(1-x)}$,

$I(x) = x\log\dfrac{x}{p} + (1-x)\log\dfrac{1-x}{1-p}$ bzw. $I = \infty$ auf $\mathbb{R}\setminus[0,1]$

(was wir bereits im Startbeispiel gesehen hatten)

1b) $X_i \sim \mathrm{Bin}(n,p)$,

$I(x) = n\left(x\log\dfrac{x}{p} + (1-x)\log\dfrac{1-x}{1-p}\right)$

(analoge Rechnung oder verwende Darst. $X \overset{d}{=} Y_1 + \dots + Y_n$ mit $Y_j \sim \mathrm{Ber}(p)$)

2) $X_i \sim \mathcal{N}(0,\sigma^2)$

$\varphi(t) = \mathbb{E}[e^{tX}] = e^{\frac{1}{2}t^2\sigma^2}$, $\log\varphi(t) = \frac{1}{2}t^2\sigma^2$, $(\log\varphi)'(t) = \sigma^2 t \overset{!}{=} x$

$I(x) = t_x x - \log\varphi(t_x) = \dfrac{1}{2}\dfrac{x^2}{\sigma^2}$, $x\in\mathbb{R}$ $\Leftrightarrow t = t_x = \dfrac{x}{\sigma^2}$

3) $X_i \sim \mathrm{Poi}(\alpha)$: $\log\mathbb{E}[e^{tX}] = \log\left(\sum_{k=0}^\infty e^{tk}e^{-\alpha}\dfrac{\alpha^k}{k!}\right) = \alpha(e^t-1)$,

$(\log\varphi)'(t) = \alpha e^t \overset{!}{=} x \Leftrightarrow t = t_x = \log\dfrac{x}{\alpha}$, (für $x>0$)

$I(x) = x\log\dfrac{x}{\alpha} - x + \alpha$ (für $x \geq 0$, $I \equiv +\infty$ auf $(-\infty,0)$)

4) $X_i \sim \mathrm{Exp}(\alpha)$, $\log\varphi(t) = \log\int_0^\infty \alpha e^{-\alpha x}e^{tx}\,dx = \begin{cases}\log\dfrac{\alpha}{\alpha-t}, & t<\alpha\\ +\infty, & t\geq\alpha\end{cases}$

[passt also nicht wörtlich in die bisher entwickelte Theorie]

$(\log\varphi)'(t) = \dfrac{1}{\alpha-t} \overset{!}{=} x\ (>0)$, $t_x = \alpha - \dfrac{1}{x}$,

$I(x) = \alpha x - 1 - \log(\alpha x)$, $x>0$ (und $\equiv +\infty$ auf $(-\infty,0]$)

Anwendungsbeispiele für den Satz von Cramér:

[• Asymptotik von $\mathbb{E}[(\frac{1}{n}S_n)^n]$]

• Asymptotik der Warteschlangenlänge

- (damit motiviert): Maximum einer Irrfahrt mit neg. Drift

  (und Beob.: dies gibt eine (gewisse) Antwort auf das Ruinproblem)

• Wartezeiten auf Länge untypische Stücke in Irrfahrtspfaden

• „Spielzeugbeispiel“ lokales Alignment

„Spielzeuganwendung" des Satzes von Cramér
(nach den Hollander, Ch. I.5)

$X_1, X_2, \ldots$ u.i.v. , $P(X_i = \frac{1}{2}) = P(X_i = \frac{3}{2}) = \frac{1}{2}$,

$S_n := X_1 + \ldots + X_n$

$\mathbb{E}[(\frac{1}{n} S_n)^n] \approx ?$ für $n \gg 1$.

„Naive" Idee: $\frac{1}{n} S_n \to 1$ (Ges. d. gr. Z.),

also $\mathbb{E}[(\frac{1}{n} S_n)^n] \approx 1^n = 1$.

$$\mathbb{E}[(\tfrac{1}{n} S_n)^n] = \int_0^{\infty (\wedge \frac{3}{2})} a^n \, P(\tfrac{1}{n} S_n \in da)$$

$$= \int_0^{\infty (\wedge \frac{3}{2})} n a^{n-1} P(\tfrac{1}{n} S_n \geq a) \, da$$

$$\simeq \int_0^{\infty (\wedge \frac{3}{2})} \exp(n(\log a - I(a))) \, da$$

„Gleichheit auf der exponentiellen Skala"

$$\simeq \exp\left(n \sup_{a>0} \{\log a - I(a)\}\right)$$

mit $I(a) = (a - \frac{1}{2}) \log \frac{a - \frac{1}{2}}{\frac{1}{2}} + (\frac{3}{2} - a) \log \frac{\frac{3}{2} - a}{\frac{1}{2}}$

(dies ist die um $\frac{1}{2}$ verschobene Ratenfkt. von Ber($\frac{1}{2}$) aus Satz 1) ,

es gilt $\frac{1}{n} \log \mathbb{E}[(\frac{1}{n} S_n)^n] = \sup_{a>0} \{\log a - I(a)\} =: b$

(es stellt sich heraus: $b = \log(3)$, sup wird für $a = \frac{3}{2}$ erreicht

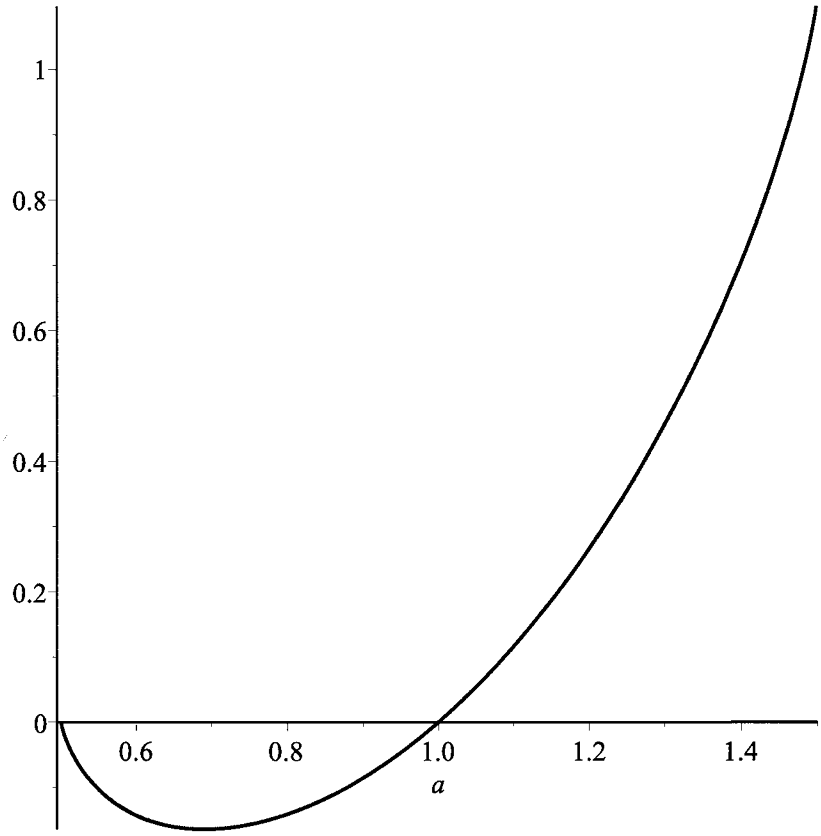
[Bem.: Die allgemeine Version dieses Gedankengangs fasst das Lemma von Varadhan, das wir später kennen lernen werden.]

15a

> $f := \log(a) + \left(a - \frac{1}{2}\right)\cdot\log\left(a - \frac{1}{2}\right) + \left(\frac{3}{2} - a\right)\cdot\log\left(\frac{3}{2} - a\right) + \log(2)$

$$f := \ln(a) + \left(a - \frac{1}{2}\right)\ln\left(a - \frac{1}{2}\right) + \left(\frac{3}{2} - a\right)\ln\left(\frac{3}{2} - a\right) + \ln(2) \qquad (1)$$

> $plot\left(f, a = \frac{1}{2} .. \frac{3}{2}\right)$



> $diff(f, a)$

$$\frac{1}{a} + \ln\left(a - \frac{1}{2}\right) - \ln\left(\frac{3}{2} - a\right) \qquad (2)$$

> $subs\left(a = \frac{3}{2}, f\right)$

$$\ln\left(\frac{3}{2}\right) + \ln(1) + \ln(2) \qquad (3)$$

> $simplify\left(subs\left(a = \frac{3}{2}, f\right)\right)$

$$\ln(3) \qquad (4)$$

>

Anwendung: Eine (zeitdiskrete) Warteschlange

Szenario: Ein Bediener,

A_t ... neu anfallende Arbeit im Intervall $(t-1,t]$, $t\in\mathbb{Z}$,

$(A_t)_{t\in\mathbb{Z}}$ u.i.v.,

Zu jedem Zeitpkt. t ($\in\mathbb{Z}$) kann max. C (>0, fest) viel abgearbeitet werden

Q_t ... Länge der Warteschlange zum (bzw. direkt nach) Zeitpunkt t.

$$Q_t = (Q_{t-1} + A_t - C) \vee 0, \quad t\in\mathbb{Z}$$

Wir nehmen an: $\mathbb{E}[A_0] < C$ („Stabilitätsbedingung")

(falls $\mathbb{E}[A_0] > C$ so divergiert Q_t für $t\to\infty$)

Frage: Starte mit $Q_{-T}=0$, $T \gg 1$,

$\mathbb{P}(Q_0 \geq q) \approx ?$ für $q \gg 1$

Beob.: $Q_0 = \max\{Q_{-1} + A_0 - C, 0\}$

$$= \max\{\max\{Q_{-2}+A_{-1}-C, 0\} + A_0 - C, 0\}$$

$$= \max\{Q_{-2}+A_{-1}+A_0-2C, A_0 - C, 0\}$$

$$= \dots = \max\{Q_{-T}+S_T - CT, S_{T-1}-C(T-1), S_{T-2}-C(T-2), \dots, S_1 - C, 0\}$$

mit $S_i := A_0 + A_{-1} + \dots + A_{-i+1}$, $i\in\mathbb{N}_0$,

also $\mathcal{L}(Q_0 \mid Q_{-T}=0) = \mathcal{L}(\max_{n=0,1,\dots,T}\{S_n - nC\})$

$$\xrightarrow[T\to\infty]{(d)} \mathcal{L}(\max_{n\in\mathbb{N}_0}\{S_n - nC\})$$ (die Gleichgewichtsverteilung von Q_0)

<u>Satz 3</u>$^{(*)}$ Sei $\Lambda(\theta) := \log \mathbb{E}[e^{\theta A_0}] \in (-\infty,\infty)\ \forall \theta \in \mathbb{R}$,

$\mathbb{E}[A_0] < C$ (und $\mathrm{Var}[A_0] > 0$),

betrachte Q_0 im Gleichgewicht.

$$\lim_{q\to\infty} \frac{1}{q} \log P(Q_0 \geq q) = -\inf_{a>0} a\Lambda^*(C + \tfrac{1}{a})$$

$$\left(= -\sup\{\theta > 0 : \Lambda(\theta) < C\cdot\theta\}\right)$$

mit $\Lambda^*(a) := \sup_{\theta\in\mathbb{R}} \{\theta a - \Lambda(\theta)\}$

(die Legendre-Transformierte von Λ)

<u>Bew.</u>: Bem.: $\tilde{\Lambda}(\theta) := \log \mathbb{E}[e^{\theta(A_0 - C)}] = \Lambda(\theta) - C\cdot\theta$,
hat Legendre-Transformierte $\tilde{\Lambda}^*(a) = \Lambda^*(a+C)$ (per Inspektion)

<u>Untere Schr.</u>: $P(Q_0 \geq q) = P(\exists n \in \mathbb{N} : S_n - C\cdot n \geq q)$

$$\geq P(S_{\lceil aq\rceil} - C_{\lceil aq\rceil} \geq \lceil qa\rceil \cdot \tfrac{1}{a}) \quad \forall a > 0$$

mit Satz 1 (Cramér)

$$\liminf_{q\to\infty} \frac{1}{q}\log P(Q_0 \geq q) \geq \liminf a\frac{1}{qa}\log P(S_{\lceil aq\rceil} - C_{\lceil aq\rceil} \geq \lceil qa\rceil \tfrac{1}{a})$$

$$= -a\Lambda^*(C + \tfrac{1}{a}) \quad \forall a > 0$$

also $\liminf \frac{1}{q}\log P(Q_0 \geq q) \geq -\inf_{a>0} a\Lambda^*(C + \frac{1}{a})$

<u>obere Schr.</u>:

$$P(Q_0 \geq q) \leq \sum_{t=0}^{\infty} P(S_t - C_t \geq q) \underset{\text{Chebyshev-Ungl.}}{\leq} e^{-\theta q} \sum_{t=0}^{\infty} e^{t(\Lambda(\theta) - C\theta)} \quad \text{(für jedes } \theta > 0)$$

$$= e^{-\theta q} \cdot \frac{1}{1 - e^{\Lambda(\theta) - C\theta}} \quad \text{(sofern } \Lambda(\theta) < C\cdot\theta)$$

(*) Satz 3 ist eine Folgerung aus dem nachfolgenden Satz 4.

also

$$\limsup \frac{1}{q}\log P(Q_0 \geq q) \leq -\sup\{\theta > 0 : \Lambda(\theta) < C\cdot\theta\}$$

z.z. bleibt

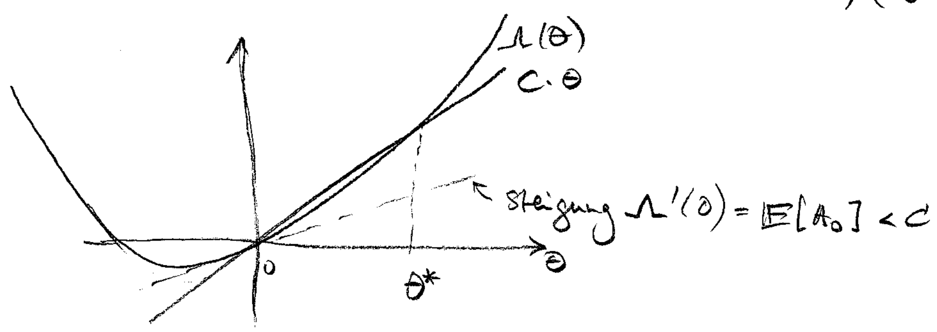
$$\theta^* := \sup\{\theta > 0 : \Lambda(\theta) < C\cdot\theta\} \geq \inf_{a>0}\{a\Lambda^*(C+\tfrac{1}{a}) : a>0\}$$

o.E. $\theta^* < \infty$ (sonst ist die Ungl. trivial)

und somit $\Lambda(\theta^*) = C\cdot\theta^*$ (Stetigk. von Λ)

Konvexität von Λ:

$$(*) \qquad \Lambda(\theta) \geq C\theta^* + \Lambda'(\theta^*)(\theta - \theta^*) \quad \forall\, \theta$$



$$\inf_{a>0}\{a\underbrace{\Lambda^*(C+\tfrac{1}{a})}_{=\sup_{\theta>0}\{\theta(C+\frac{1}{a})-\Lambda(\theta)\}}\} = \inf_{a>0}\sup_{\theta>0}\{\theta(1+C) - a\Lambda(\theta)\}$$

sup darf auf $\theta > 0$ eingeschr. werden, da Argument $\geq C > \mathbb{E}[A_0] = \Lambda'(0)$

setze $(*)$ ein

$$\leq \inf_{a>0}\underbrace{\sup_{\theta>0}\{\theta(1+Ca-a\Lambda'(\theta^*)) + a\theta^*(\Lambda'(\theta^*)-C)\}}_{=\begin{cases} a\theta^*(\Lambda'(\theta^*)-C) & \text{falls } 1+Ca-a\Lambda'(\theta^*) \leq 0 \\ +\infty & \text{sonst}\end{cases}}$$

$$= \inf_{a>0,\ 1+a(\underbrace{\Lambda'(\theta^*)-C}_{>0 \text{ (siehe Skizze)}})\leq 0}\{a\theta^*(\Lambda'(\theta^*)-C)\} = \theta^*.$$

Beispiel: Poissonverteilte anfallende Arbeit

$A_i \sim \mathrm{Poi}(\alpha)\ ,\ \alpha < C$

$$\Lambda(\theta) = \log \mathbb{E}[e^{\theta A_1}] = \log \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} e^{\theta k}$$

$$= \alpha(e^{\theta} - 1)$$

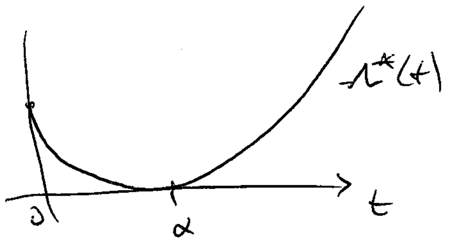
$t \geq 0$:

$$\Lambda^*(t) = \sup_{\theta} \{\theta t - \Lambda(\theta)\}$$

$$= \theta_t \cdot t - \Lambda(\theta_t)\ ,$$ θ_t Lösung von $t = \Lambda'(\theta)\ (= \alpha e^{\theta})$,

$$= \alpha - t + t \log \frac{t}{\alpha}$$ d.h. $\theta_t = \log \frac{t}{\alpha}$

$(\Lambda^*(t) = +\infty$ für $t < 0)$

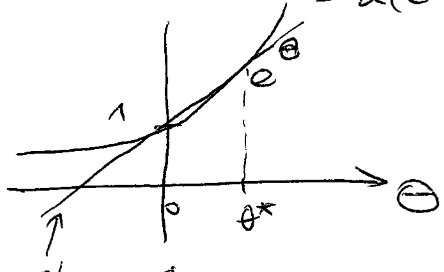


Somit

$$\lim_{q \to \infty} \frac{1}{q} \log \mathbb{P}(Q_0 > q) = -\theta^*$$

mit $\theta^* = \sup\{\theta > 0 : \Lambda(\theta) < C \cdot \theta\}$,

θ^* Lösung von $\underbrace{\Lambda(\theta)}_{= \alpha(e^{\theta} - 1)} = C \cdot \theta$



Steigung $\frac{C}{\alpha} > 1$

Zur Asymptotik der Schwänze des Maximums einer Irrfahrt mit negativer Drift

$Y_1, Y_2, \dots$ u.i.v. reelle ZVn, $\varphi(t) := \mathbb{E}[e^{tY_1}] \in (0,\infty)$ $\forall t \in \mathbb{R}$,

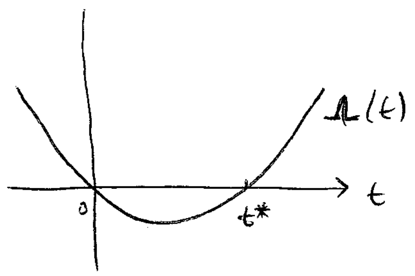
$\mu := \mathbb{E}[Y_1] < 0$, $\mathbb{P}(Y_1 > 0) > 0$,

$S_n := Y_1 + \dots + Y_n$, $n \in \mathbb{N}$.

$\Lambda(t) := \log \varphi(t)$

erfüllt $\Lambda(0) = 0$, $\Lambda'(0) = \mathbb{E}[Y_1] < 0$,

ist strikt konvex, $\Lambda(t) \to \infty$ für $|t| \to \infty$ (Lemma 2)



demnach $\exists_1\, t^* > 0$ mit $\Lambda(t^*) = 0$

(und $t^* = \sup\{t \in \mathbb{R} : \Lambda(t) = 0\}$
$= \sup\{t > 0 : \Lambda(t) = 0\}$)

Sei $M := \max_{n \in \mathbb{N}_0} S_n$ (≥ 0) ($S_n \to -\infty$ f.s., also $M < \infty$ f.s.)

Satz 4

$$\frac{1}{x} \log \mathbb{P}(M \geq x) \xrightarrow[x \to \infty]{} -t^* = -\inf_{a > 0} a \Lambda^*\left(\frac{1}{a}\right)$$

(mit $\Lambda^*(x) = \sup_{t \in \mathbb{R}} \{tx - \Lambda(t)\}$, die Legendre-Transformierte von Λ)

Obere Schranke:

$$\mathbb{P}(M \geq x) \leq \sum_{n=1}^{\infty} \mathbb{P}(S_n \geq x)$$

$$\leq \sum_{n=1}^{\infty} e^{-tx} \underbrace{\mathbb{E}[e^{tS_n}]}_{=\varphi(t)^n = e^{n\Lambda(t)}} = e^{-tx} \cdot \frac{e^{\Lambda(t)}}{1-e^{\Lambda(t)}}$$

für jedes $t>0$ (mit $\Lambda(t)<0$),

also $\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}(M \geq x) \leq -\sup\{t : t>0, \Lambda(t)<0\}$
$= -t^*$.

Untere Schranke:

$$\mathbb{P}(M \geq x) \geq \mathbb{P}\left(S_{\lceil ax \rceil} \geq \frac{1}{a}\lceil ax \rceil\right) \quad \text{(für jedes } a>0)$$

$$\geq \exp\left(-\lceil ax \rceil \left(\Lambda^*\left(\tfrac{1}{a}\right) + o(1)\right)\right)$$

(Satz v. Cramér)

eine Funktion, die für $x \to \infty$ gegen 0 konvergiert

also

$$\liminf_{x\to\infty} \frac{1}{x} \log \mathbb{P}(M \geq x) \geq -\inf_{a>0} \left\{ a\Lambda^*\left(\tfrac{1}{a}\right) \right\}$$

die beiden Schranken zeigen insbesondere, dass $t^* \leq \inf_{a>0} a\Lambda^*(\frac{1}{a})$,

z.z. bleibt $\inf_{a>0} aI(\frac{1}{a}) \leq t^*$ $(*)$.

$$\inf_a aI\left(\tfrac{1}{a}\right) = \inf_{a>0} a \cdot \sup_{t\in\mathbb{R}} \left\{ \tfrac{1}{a}t - \Lambda(t) \right\}$$

$$= \inf_{a>0} \sup_{t>0} \{t - a\Lambda(t)\}$$

(sup darf auf $t>0$ eingeschränkt werden, denn $\Lambda(t)>0$ für $t<0$, so dass solche t für das sup nicht in Frage kommen)

$\Lambda(t) \geq \overset{0}{\overbrace{\Lambda(t^*)}} + (t-t^*)\Lambda'(t^*)$ (Konvexität von Λ)

$$\leq \inf_{a>0} \underbrace{\sup_{t>0} \{t - a(t-t^*)\Lambda'(t^*)\}}_{= \begin{cases} +\infty & \text{wenn } 1-a\Lambda'(t^*)>0 \\ at^*\Lambda'(t^*) & \text{wenn } 1-a\Lambda'(t^*) \leq 0 \end{cases}}$$

$$= \frac{1}{\Lambda'(t^*)} t^* \Lambda'(t^*) = t^* \qquad \square$$

Bem.: Es gilt $\mathbb{P}(M \geq x) \leq e^{-t^* x}$ für $x > 0$:

$Z_n := e^{t^* S_n}$, $(Z_n)_{n \in \mathbb{N}_0}$ ist Martingal,

~~Bew.:~~

$\tau_x := \inf\{n \in \mathbb{N}_0 : S_n \geq x\}$ ist Stoppzeit,

$$1 = \mathbb{E}[Z_0] = \mathbb{E}[Z_{\tau_x \wedge n}] \quad (\text{für jedes } n \in \mathbb{N})$$

$$\geq e^{t^* x}\, \mathbb{P}(\tau_x \leq n) \xrightarrow[n\to\infty]{} e^{t^* x} \underbrace{\mathbb{P}(\tau_x < \infty)}_{= \mathbb{P}(M \geq x)}$$

[und wenn $\mathbb{P}(Y_1 \in (-\mathbb{N}_0) \cup \{1\}) = 1$,
so gilt Gleichheit]

(siehe z.B. Klenke, Übung 10.2.2)

Bem. Die untere Schranke zeigt auch, dass die asymptotisch optimale Strategie für $\{M \geq x\}$ darin besteht, für $a \cdot x$ Schritte mit Geschwindigkeit $\approx \frac{1}{a}$ nach oben zu laufen (wo $a = \frac{1}{\Lambda'(t^*)}$)

Williams Pfad-Zerlegungen...

Bem. · (Interpretation für das Ruinproblem einer Versicherung)

$Y_n :=$ Schadenszahlungen − Prämieneinnahmen im n-ten Jahr, (seien u.i.v., $\mathbb{E}[Y_1] < 0$, ..)

$K_0 :=$ Startkapital (> 0)

so ist $\{M > K_0\} = \{\exists\, n \in \mathbb{N} : K_0 - (Y_1 + \dots + Y_n) < 0\}$

das Ereignis, dass (in diesem Modell) die Versicherungsgesellschaft bankrott geht.

Lange untypische Stücke in Irrfahrtspfaden

(vgl. DZ98, Ch. 3.2)

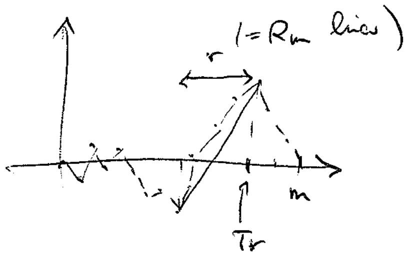
$$S_n = \sum_{i=1}^n X_i \quad , X_i \text{ u.i.v.} \quad (\text{mit Cramér-Bed. o.ä.} \ldots)$$

$A \subset \mathbb{R}$ (denke an Halbintervalle)

$$R_m := \max\left\{ \ell - k : 0 \le k < \ell \le m, \ \frac{S_\ell - S_k}{\ell - k} \in A \right\},$$

(mit Setzung $\max \emptyset = -\infty$)

$$T_r := \min\left\{ \ell : \frac{S_\ell - S_k}{\ell - k} \in A \ \text{ für ein } k \in \{0, 1, \ldots, \ell - r\} \right\}$$



(R_m ist die Länge des längsten Teilstücks bis zum m-ten Schritt, dessen Steigung in A liegt,

T_r gibt an, wie lange man warten muss, bis man zum ersten Mal ein Teilstück der Länge $\ge r$ findet, dessen Steigung in A liegt.

[T_r ist eine Stoppzeit für die Irrfahrt.]

Bem.: $\{R_m \ge r\} = \{T_r \le m\}$

Sei $I_A := -\lim_{n\to\infty} \frac{1}{n} \log \underbrace{P(S_n \in A)}_{= \mu_n(A)}$ (existiert nach Satz v. Cramér) [in der in Satz 1 bewiesenen Version zumindest für den Fall, dass A ein Halbintervall ist

Satz Es gilt

$$\lim_{m\to\infty} \frac{R_m}{\log m} = \lim_{r\to\infty} \frac{r}{\log T_r} = \frac{1}{I_A} \quad \text{f.s.}$$

[also grob gesprochen $R_m \approx \frac{\log m}{I_A}$, $T_r \approx e^{I_A \cdot r}$]

[wir nehmen $I_A \in (0,\infty)$ an ...]

1) Zeige: $\liminf_{r\to\infty} \frac{1}{r}\log T_r \geq I_A$ f.s. $(*)$

$$\{T_r \leq m\} \subset \bigcup_{k=0}^{m-r} \bigcup_{\ell=k+r}^{m} \underbrace{\left\{ \frac{S_\ell - S_k}{\ell - k} \in A \right\}}_{=: C_{k,\ell}} \subset \bigcup_{k=0}^{m-1} \bigcup_{\ell=k+r}^{\infty} C_{k,\ell}$$

also $\mathbb{P}(T_r \leq m) \leq \sum_{k=0}^{m-1} \sum_{\ell=k+r}^{\infty} \underbrace{\mathbb{P}(C_{k,\ell})}_{=\mu_{\ell-k}(A)} \leq m \cdot \sum_{n=r}^{\infty} \mu_n(A)$,

für $\varepsilon \in (0, I_A)$:

$$\sum_{r=1}^{\infty} \mathbb{P}\Big(\underbrace{T_r \leq e^{r(I_A-\varepsilon)}}_{(T_r \leq \lfloor e^{r(I_A-\varepsilon)} \rfloor \text{ da } T_r \text{ ganzzahlig})} \Big) \leq \sum_{r=1}^{\infty} e^{r(I_A-\varepsilon)} \sum_{n=r}^{\infty} c \cdot e^{-n(I_A-\frac{\varepsilon}{2})}$$

(mit $c = c(\varepsilon) < \infty$, $c' = c'(\varepsilon) < \infty$)

$$\leq c' \sum_{r=1}^{\infty} e^{-r\varepsilon/2} < \infty,$$

mit Borel-Cantelli also

$T_r > e^{r(I_A-\varepsilon)}$ für $r \geq r_0 \; (= r_0(\omega))$ f.s.,

d.h. $\liminf \frac{1}{r}\log T_r \geq I_A - \varepsilon$, mit $\varepsilon \downarrow 0$ folgt $(*)$.

Wegen $\{R_m \geq r\} = \{T_r \leq m\}$ impliziert $(*)$ auch $\limsup_{m\to\infty} \frac{R_m}{\log m} \leq \frac{1}{I_A}$ f.s.

(Detailargument z.B.

$[R_m \leq R_{m'}$ für $m \leq m']$

$R_m = R_{\exp(\frac{1}{I_A}\log m) I_A} < (1+\varepsilon)(\frac{1}{I_A}\log m)$ für $m \geq m_0$ f.s.,

sonst wäre $T_r \leq e^{r \cdot \frac{I_A}{1+\varepsilon}}$ für unendlich viele r (für jedes $\varepsilon > 0$)
)

2) zeige $\limsup\limits_{r\to\infty} \frac{1}{r} \log T_r \le I_A$ f.s. (∗∗)

$r \in \mathbb{N}$.

Sei $B_{\ell,r} := \left\{ \frac{S_{\ell \cdot r} - S_{(\ell-1)r}}{r} \in A \right\}$,

$B_{1,r}, B_{2,r}, \dots$ sind u.a. mit $\mathbb{P}(B_{i,r}) = \mu_r(A)$,

$\bigcup\limits_{\ell=1}^{\lfloor m/r \rfloor} B_{\ell,r} \subset \{T_r \le m\}$, also

$$\mathbb{P}(T_r > m) \le 1 - \mathbb{P}\Big(\bigcup_{\ell=1}^{\lfloor m/r \rfloor} B_{\ell,r}\Big) = \big(1 - \mathbb{P}(B_{1,r})\big)^{\lfloor m/r \rfloor} \underset{(1-x \le e^{-x})}{\le} e^{-\lfloor m/r \rfloor \mathbb{P}(B_{1,r})} = e^{-\lfloor \frac{m}{r} \rfloor \mu_r(A)}$$

Für $\varepsilon > 0$ (wähle $m = \lfloor e^{r(I_A+\varepsilon)} \rfloor$)

$$\sum_{r=1}^{\infty} \mathbb{P}\big(T_r > e^{r(I_A+\varepsilon)}\big) \le \sum_{r=1}^{\infty} \exp\Big(-\frac{e^{r(I_A+\varepsilon)}}{r} \cdot c_1 e^{-r(I_A+\frac{\varepsilon}{2})}\Big),$$

(mit $c_1 = c_1(\varepsilon, A) < \infty$)

$$\le \sum_{r=1}^{\infty} e^{-\frac{c_1}{r} e^{+r\frac{\varepsilon}{2}}} < \infty$$

mit Borel-Cantelli also

$$\limsup_{r\to\infty} \frac{1}{r} \log T_r \le I_A + \varepsilon,$$ mit $\varepsilon \downarrow 0$ folgt (∗∗)

(wiederum wegen $\{R_m \ge r\} = \{T_r \le m\}$ impliziert (∗∗) auch $\liminf\limits_{m} \frac{R_m}{\log m} \ge \frac{1}{I_A}$,

denn $R_m = R_{\exp(\frac{\log m}{I_A} \cdot I_A)} \ge (1-\varepsilon) \frac{1}{I_A} \log m$ für $m \ge m_0$ f.s.,

sonst wäre $T_r \ge e^{r \frac{I_A}{1-\varepsilon}}$ für unendlich viele r)

„Spielzeug"-Anwendung auf eine Fragestellung im Zusammenhang mit (lokalen) Sequenz-Alignments

(vgl. auch M. Waterman, Introduction to computational biology, Chapman-&-Hall 1995, Ch. 11)

$X_1, X_2, \ldots$ u.i.v. mit Werten in einem endlichen „Alphabet" E,

$\tilde{X}_1, \tilde{X}_2, \ldots$ ——"——

$s: E \times E \to \mathbb{R}$ „Scorefunktion" (z.B. $s(x,y) = \begin{cases} +1, & x=y \\ -1, & x \neq y \end{cases}$)

$Y := s(X_1, \tilde{X}_1)$ erfülle $\mathbb{E}[Y] < 0$, $P(Y>0) > 0$

$\Lambda(t) := \log \mathbb{E}[e^{tY}]$ (Erwartung existiert, da Y nur endl. viele Werte annimmt)

$$H_n := \max\left\{ \sum_{k=1}^{\ell} s(X_{i+k}, \tilde{X}_{j+k}) : 0 \le i,j \le n-\ell \right\}$$

(bester lokaler Score)

Beob.:

$$\mathbb{P}\left(H_n \le \frac{1+\varepsilon}{t^*} \log(n^2) \right) \xrightarrow[n\to\infty]{} 1 \quad \text{für jedes } \varepsilon > 0$$

Bew.:

$$S_m(i,j) := \sum_{k=1}^{m} s(X_{i+k}, \tilde{X}_{j+k})$$

(für n gen. groß)

$$\underbrace{\mathbb{P}\left(H_n > \frac{1+\varepsilon}{t^*} \log(n^2) \right)}_{= \bigcup_{0 \le i,j \le n-1} \bigcup_{m=1}^{n-(i \vee j)} \{S_m(i,j) > \ldots\}} \le \sum_{i,j=0}^{n-1} \overbrace{\mathbb{P}\left(\max_{m \in \mathbb{N}} S_m(i,j) > \frac{1+\varepsilon}{t^*} \log(n^2) \right)}^{\le C e^{-(t^* - \frac{\varepsilon}{2}) \frac{1+\varepsilon}{t^*} \log(n^2)}}$$

$$\le n^2 C e^{-(1+\varepsilon')\log(n^2)} \xrightarrow[n\to\infty]{} 0.$$

2. Allgemeine Formulierung von Prinzipien großer Abweichungen

[Vorbem.] Y_n ZVn mit $Y_n \xrightarrow[n\to\infty]{P} y_0$, wir möchten für $A \not\ni y_0$ quantifizieren $\mathbb{P}(Y_n \in A) \approx e^{-n r_A}$

[vgl. Satz 1.1 (Cramér) für $Y_n = \frac{1}{n}(X_1 + \dots + X_n)$ und A ein Halbintervall]

Beob.: $a_1, a_2, \dots, a_k \geq 0$, so ist $\lim_{n\to\infty} \frac{1}{n} \log\Big(\overbrace{\sum_{i=1}^{k} e^{-n a_i}}^{\leq k e^{-n \cdot \min_i a_i}}\Big) = -\min_i a_i$

„naive" Forderung:

$$\mathbb{P}(Y_n \in A) \overset{!}{=} e^{-n[\inf_{y \in A} I(y) + o(1)]} \quad \text{für jedes } A$$

[aber dies ist i. A. zu viel verlangt: kann z.B. nicht erfüllt sein, wenn Y_n eine Dichte besitzt und A Lebesgue-Maß 0.]

E (hausdorffscher, regulärer)* topologischer Raum (wir denken meist an einen metrischen Raum (E, d) mit der von der Metrik erzeugten Topologie)

$\mu_n \in \mathcal{M}_1(E)$, $n \in \mathbb{N}$, $(\gamma_n) \subset \mathbb{R}_+$ mit $\gamma_n \to \infty$,

$I: E \to [0, \infty]$ (und $I \not\equiv \infty$)

*) d.h. für $F \subset E$ abgeschl. und $x \in E \setminus F$ gibt es offene Mengen $G_1, G_2 \subset E$ mit $F \subset G_1$, $x \in G_2$ und $G_1 \cap G_2 = \emptyset$

Def. 1 (μ_n) erfüllt ein <u>Prinzip großer Abweichungen</u> (PGA, englisch LDP, large deviation principle) mit Rate γ_n (auch: Skala γ_n) und Ratenfunktion I, falls gilt

1) $\forall s \geq 0$: Die Niveaumenge $\Phi_I(s) := \{x \in E : I(x) \leq s\}$ ist kompakt

2) $\forall G \subset E$ offen: $\liminf\limits_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(G) \geq -\inf\limits_{x\in G} I(x)$

3) $\forall F \subset E$ abgeschlossen: $\limsup\limits_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(F) \leq -\inf\limits_{x\in F} I(x)$

Entsprechend erfüllt eine Familie X_n von E-wertigen ZVn ein PGA, falls dies für $(\mathbb{P}(X_n \in \cdot))_n$ gilt.

Wenn man statt 1) und 3) nur

1') $\Phi_I(s)$ abgeschlossen und 3') $\forall F \subset E$ kompakt...

fordert, spricht man von einem „schwachen PGA“ (weak LDP)

[In der Lit. ist auch der Begriff „gute Ratenfunktion“ („good rate function“), falls I kompakte Niveaumengen hat, verbreitet.]

<u>Bem. 2</u> a) 1) impliziert: I ist nach unten halbstetig

b) Es ist $\inf_{x \in E} I(x) = 0$ $\left(= \min_{x\in E} I(x)\right.$ wegen Kompaktheit der Niveaumengen$\left.\right)$,
[wähle $F = E$ in 3)]

falls der Minimierer x_0 eindeutig ist, so erfüllt (μ_n) ein (schwaches) Gesetz d. gr. Zahlen:

$$\mu_n(B_\varepsilon(x_0)) \underset{n\to\infty}{\longrightarrow} 1$$

[und falls zudem $\gamma_n / \log n \underset{n\to\infty}{\longrightarrow} \infty$, so gilt für eine Folge von ZV'n (X_n) mit $X_n \sim \mu_n$ auch $X_n \to x_0$ f.s.]
(verwende Borel-Cantelli: $\sum_{n=1}^{\infty} \mu_n(B_\varepsilon(x_0)^c) < \infty \quad \forall \varepsilon > 0$)

c) Es gilt $\mu_n(A) = \exp\left(-\gamma_n\left(\inf_{x\in A} I(x) + o(1)\right)\right)$

Sofern $\inf_{x\in \mathring{A}} I(x) = \inf_{x \in \bar{A}} I(x)$ (insbes. erfüllt, wenn I stetig und $A \subset \bar{\mathring{A}}$)
(A ist eine „I-Stetigkeitsmenge")

[$\mathring{A}$ = offenes Inneres von A $(= E \setminus \overline{A^c})$
$\bar{A}$ = Abschluss von A]

d) Falls existent ~~so~~, ist die Ratenfunktion eindeutig best.

Seien I, J Ratenfunktionen für (μ_n) zur Rate γ_n, $x \in E$:

$$-I(x) \le -\inf_{y \in B_\varepsilon(x)} I(x) \le \liminf_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(B_\varepsilon(x))$$

$$\le \limsup_{n\to\infty} \ldots \le \limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n\left(\overline{B_\varepsilon(x)}\right)$$

$$\le -\inf_{y \in \overline{B_\varepsilon(x)}} J(y) \underset{\varepsilon\downarrow 0}{\searrow} -J(x)$$

(denn J ist nach unten halbstetig)

Bem. 3 (Äquivalente Formulierung eines PGA) [im metrischen Fall]

(μ_n) erfüllt PGA mit Rate γ_n und Ratenfunktion I g.d.w.

I erf. 1)

2') $\forall x \in E, \varepsilon > 0$: $\liminf_n \frac{1}{\gamma_n} \log \mu_n(B_\varepsilon(x)) \geq -I(x)$

3') $\forall s \geq 0, \varepsilon > 0$:

$$\limsup \frac{1}{\gamma_n} \log \mu_n(\{y : d(\Phi_I(s), y) \geq \varepsilon\}) \leq -s$$

[2') zeigt, in welchem Sinne die untere Schranke „lokal" ist]

Bew.: „⇒": 2) ⇒ 2') $\left(B_\varepsilon(x) \text{ ist offen}, \inf_{B_\varepsilon(x)} I \leq I(x) \right)$

3) ⇒ 3') $\left((\Phi_I(s)^\varepsilon)^c \text{ ist abgeschl.}, \inf_{(\Phi_I(s)^\varepsilon)^c} I \geq s \right)$

„⇐": 2') ⇒ 2) $\liminf \frac{1}{\gamma_n} \log \mu_n(B_\varepsilon(x)) \geq -I(x)$ für jedes $x \in G$ mit $B_\varepsilon(x) \subset G$,

also auch $\ldots \geq -\inf_{x \in G} I(x)$

1) + 3') ⇒ 3)

Sei F abg., $s_0 := \inf_F I > 0$ (o.b.d.A.), $s \in (0, s_0)$:

$F \cap \Phi_I(s) = \emptyset$, also $\exists \varepsilon > 0$ mit $\Phi_I(s)^\varepsilon \cap F \neq 0$ (nutze Kompaktheit von $\Phi_I(s)$)

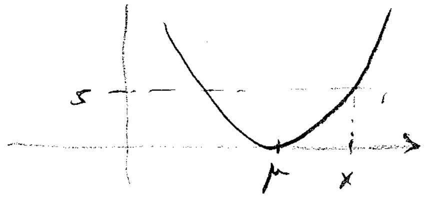
$$\Rightarrow \limsup \frac{1}{\gamma_n} \log \mu_n(F) \leq \limsup \frac{1}{\gamma_n} \mu_n((\Phi_I(s)^\varepsilon)^c) \overset{3')}{\leq} -s,$$

dann $s \nearrow s_0$.

Bem.

$X_1, X_2, \dots$ u.i.v. reelle ZVn, unter den Voraussetzungen von Satz 1.1 (Satz v. Cramér) $\left[\text{d.h. } \varphi(t) := \mathbb{E}[e^{tX_1}] < \infty \ \forall t \in \mathbb{R}\right]$ erfüllen $S_n := \frac{1}{n}(X_1 + \dots + X_n)$, $n \in \mathbb{N}$ ein PGA mit Rate n und Ratenfkt. $I(x) = \sup_{t \in \mathbb{R}} \{tx - \log \varphi(t)\}$

Wir hatten gesehen: I ist ≥ 0, strikt konvex, eind. Minimum $I(\mu) = 0$ bei $\mu = \mathbb{E}[X_1]$, I hat kompakte Niveaumengen (Lemma 1.2)



1) ✓

Zeige 2') aus Bem. 3: $x \in \mathbb{R}$, o.E. $x \neq \mu$, $\varepsilon > 0$: (betrachte den Fall $x > \mu$)

$$\liminf_n \frac{1}{n} \log \underbrace{P\left(\frac{S_n}{n} \in B_\varepsilon(x)\right)} \geq -I\left(x - \tfrac{\varepsilon}{2}\right) \geq -I(x)$$

$$\geq P\left(x - \tfrac{\varepsilon}{2} \leq \tfrac{S_n}{n} < x + \tfrac{\varepsilon}{2}\right)$$

$$= \underbrace{P\left(\frac{S_n}{n} \geq x - \frac{\varepsilon}{2}\right)}_{= e^{-n(I(x - \frac{\varepsilon}{2}) + o(1))}} - \underbrace{P\left(\frac{S_n}{n} \geq x + \frac{\varepsilon}{2}\right)}_{= e^{-n(I(x + \frac{\varepsilon}{2}) + o(1))}}$$

zu 3') aus Bem. 3:

Sei $s > 0$, $\varepsilon > 0$. $\Phi_I(s) = [a(s), b(s)]$ mit $a(s) < \mu < b(s)$ (wo $I(a(s)) = I(b(s)) = s$)

$$\limsup \frac{1}{n} \log \underbrace{P\left(\frac{S_n}{n} \leq a(s) - \varepsilon \text{ oder } \frac{S_n}{n} \geq b(s) + \varepsilon\right)} \leq -s$$

$$\leq \underbrace{P\left(\frac{S_n}{n} \leq a(s) - \varepsilon\right)} + \underbrace{P\left(\frac{S_n}{n} \geq b(s) + \varepsilon\right)}$$

$$= \exp(-n(\underbrace{I(a(s) - \varepsilon)}_{> s} + o(1))) \qquad (= \exp(-n(\overbrace{I(b(s) + \varepsilon)}^{> s} + o(1)))$$

<u>Def. 4</u> $(\mu_n) \subset \mathcal{M}_1(E)$ heißt <u>exponentiell straff</u> (zur Rate γ_n) $:\Longleftrightarrow$

$\forall s > 0 \; \exists K \subset E$ kompakt mit

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(K^c) \le -s.$$

<u>Beob. 5</u> (μ_n) erfüllt schwaches PGA (zu (γ_n), I) und (μ_n) exponentiell straff
$\Longleftrightarrow$ (μ_n) erfüllt PGA.

„$\Leftarrow$" ✓

„$\Rightarrow$": Zeige 3): Sei $F \subset E$ abgeschl., $s > 0$,
wähle $K_s \subset E$ kompakt mit
$$\limsup_n \frac{1}{\gamma_n} \log \mu_n(K_s^c) \le -s.$$

$$\limsup \frac{1}{\gamma_n} \log \underbrace{\mu_n(F)}_{\le \mu_n(F\cap K_s) + \mu_n(K_s^c)} \le -\left(\inf_{F\cap K_s} I \wedge s\right)$$

$$\le -\left(\inf_F I \wedge s\right), \text{ dann } s \nearrow \infty.$$

Zeige 1): $s > 0$, K_{2s} (wie oben) kompakt

K_{2s}^c ist offen,

$$\inf_{K_{2s}^c} I \underset{2)}{\ge} -\liminf \frac{1}{\gamma_n} \log \mu_n(K_{2s}^c) \ge 2s > s,$$

also ist $\Phi_I(s) \subset K_{2s}$ kompakt.

Satz 6 (Varadhans Lemma)

(X_n) erfüllt PGA auf E mit Rate γ_n, Ratenfunktion I,

$f: E \to \mathbb{R}$ sei stetig und erfülle

$$(*) \quad \underbrace{\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n f(X_n)} \mathbb{1}_{\{f(X_n)\geq M\}}\right]}_{} = -\infty.$$

Dann ist $\lim_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n f(X_n)}\right] = \sup_{x\in E} \{f(x) - I(x)\}$

Bem.: $(*)$ ist erfüllt falls f nach oben beschr. ✓ oder falls

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{a\gamma_n f(X_n)}\right] < \infty \quad \text{für ein } a > 1$$

[denn $\frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n f(X_n)} \mathbb{1}_{\{f(X_n)\geq M\}}\right] = M + \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n (f(X_n)-M)} \mathbb{1}_{\{f(X_n)\geq M\}}\right]$

$$\leq M + \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{a\gamma_n (f(X_n)-M)}\right] \leq -(a-1)M + \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{a\gamma_n f(X_n)}\right]$$

Bew.: Untere Schranke

Sei $x \in E$, $\varepsilon > 0$. Wähle $G \ni x$, G offen mit $\inf_G f \geq f(x) - \varepsilon$:

$$\liminf_n \frac{1}{\gamma_n} \log \underbrace{\mathbb{E}\left[e^{\gamma_n f(X_n)}\right]}_{\geq \mathbb{E}\left[e^{\gamma_n \inf_G f} \cdot \mathbb{1}_{\{X_n \in G\}}\right]} \geq f(x) - \varepsilon - \inf_G I \geq f(x) - \varepsilon - I(x)$$

zunächst $\varepsilon \downarrow 0$, dann optimiere über $x \in E$, erhalte „$\geq$"

Obere Schranke

nehme zunächst an, $\sup_E f \leq M < \infty$. $\varepsilon > 0$

zu $x \in E$ wähle $G_x \ni x$ offen mit $\sup_{y\in G_x} f(y) \leq f(x) + \varepsilon$

und $\inf_{y\in G_x} I(y) \geq I(x) - \varepsilon$

[Stetigkeit von f
bzw. nach unten Halbstetigkeit von I]

Sei $s > 0$, $\overbrace{\Phi_I(s)}^{=\{x : I(x) \le s\}} \subset \bigcup_{i=1}^{m} G_{x_i}$ für geeign. $x_1, \dots, x_m$
(Kompaktheit von $\Phi_I(s)$)

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}[\underbrace{e^{\gamma_n f(X_n)}}_{\le \sum_{i=1}^{m} e^{\gamma_n (f(x_i)+\varepsilon)} 1_{\{X_n \in \bar{G}_{x_i}\}} + e^{\gamma_n M} 1_{\{X_n \notin \Phi_I(s)\}}}]$$

$$\le \max_{i=1,\dots,m} \{ f(x_i) + \varepsilon - I(x_i) + \varepsilon \} \vee (\underbrace{-s}_{+M})$$

$$\le \left(\sup_{x \in E} \{ f(x) - I(x) + 2\varepsilon \} \right) \vee (\underbrace{-s}_{+M})$$

nun $s \nearrow \infty$, dann $\varepsilon \searrow 0$, erhalte „$\le$".

im allg. Fall: $f_M(x) := f(x) \wedge M$,

$$\limsup_{n} \frac{1}{\gamma_n} \log \mathbb{E}[\underbrace{e^{\gamma_n f(X_n)}}_{\le e^{\gamma_n f_M(X_n)} + e^{\gamma_n f(X_n)} 1_{\{f(X_n) \ge M\}}}]$$

$$\le \left(\sup_{x \in E} \{ f_M(x) - I(x) \} \right) \vee \Big(\underbrace{\limsup_{n} \frac{1}{\gamma_n} \log \mathbb{E}[e^{\gamma_n f(X_n)} 1_{\{f(X_n) \ge M\}}]}_{\xrightarrow[M\to\infty]{} -\infty \text{ wg. } (*)} \Big)$$

Bem.: Die untere Schranke benötigt eigentlich nur, dass f nach unten halbstetig ist (wir haben verwendet, dass Mengen $\{x : f(x) > c\}$ offen sind),

die obere Schranke benötigt nur, dass f nach oben halbstetig ist (Mengen $\{x : f(x) < c\}$ sind offen).

Korollar 7 (PGA für exponentiell transformierte Maße, gekipptes / getiltetes PGA)

(X_n), γ_n, I, f wie in Satz 6, setze

$$\mu_n(B) := \frac{\mathbb{E}[1_B(X_n) e^{\gamma_n f(X_n)}]}{\mathbb{E}[e^{\gamma_n f(X_n)}]}, \quad B \in \mathcal{B}(E)$$

Dann erfüllt (μ_n) ein PGA mit Rate γ_n und Ratenfunktion $I_f(x) = I(x) - f(x) - \inf_{y \in E} \{I(y) - f(y)\}$

Bew.:

1.) Für $G \subset E$ offen gilt

$$(**) \quad \liminf_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}[e^{\gamma_n f(X_n)} 1_G(X_n)] \geq \sup_{x \in G} \{f(x) - I(x)\}$$

(Argument analog zur unteren Schr. in Satz 6)

2) Für $F \subset E$ abgeschl. gilt

$$(***) \quad \limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}[e^{\gamma_n f(X_n)} 1_F(X_n)] \leq \sup_{x \in F} \{f(x) - I(x)\}$$

(betr. zunächst den Fall $\sup_x f(x) \leq M < \infty$, wähle zu $\varepsilon > 0$ [wie in der oberen Schr. in Satz 6]

$x_1, \dots, x_m \in F$, offene Umg. $G_{x_i} \ni x_i$ mit $\sup_{G_{x_i}} f \leq f(x_i) + \varepsilon$, $\inf_{G_{x_i}} I \geq I(x_i) - \varepsilon$

und $F \cap \Phi_I(s) \subset G_{x_1} \cup \dots \cup G_{x_m}$

$\Rightarrow \limsup \frac{1}{\gamma_n} \log[\dots]$

$$\leq \dots \leq \left(\sup_{x \in F} \{f(x) - I(x)\} + 2\varepsilon\right) \vee (-s),$$

dann $s \nearrow \infty$, $\varepsilon \searrow 0$.

im allg. Fall betr. $f_M(x) = f(x) \wedge M$, etc.

1.) + Satz 6 zeigen untere Schranke im PGA:
G offen

$$\liminf \frac{1}{\gamma_n} \log \mu_n(G) = \liminf \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n f(X_n)} 1_G(X_n)\right]$$

$$- \lim \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n f(X_n)}\right]$$

$$= \sup_G \{f - I\} - \sup_E \{f - I\} = -\inf_G \{I - f\}$$

2.) + Satz 6 zeigen obere Schranke im PGA
(Argument analog)

Zum Curie-Weiss-Modell

[ein mikroskopisches Modell für Ferromagnetismus,
insbes. für die Existenz der Curie-Temperatur (P. Curie, 1895),
(Pierre Curie, 1859–1906, Pierre-Ernest Weiss, 1865–1940)]

$$\sigma = (\sigma_i)_{i=1,\dots,N} \in \{\pm 1\}^N =: \Omega_N$$

[Interpretation: σ_i ... Ausrichtung des i-ten Elementarmagneten]

$$H_N(\sigma) = \underbrace{-\frac{1}{2N}\sum_{i,j=1}^{N} \sigma_i \sigma_j}_{= \frac{1}{2}\sum_{i=1}^{N}\sigma_i \frac{1}{N}\sum_{j=1}^{N}\sigma_j} - h\sum_{i=1}^{N}\sigma_i \qquad (h \in \mathbb{R})$$

[Stärke des externen Magnetfelds]

[Interpretation: $H_N(\sigma)$ ... innere Energie (des Systems im Mikrozustand σ),
(ist extensiv: proportional zur Systemgröße N)

$$P_N(\{\sigma\}) = \frac{1}{2^N} \quad (P_N \text{ ist die uniforme Vert. auf } \Omega_N)$$

$\beta \geq 0$

$$\mu_{N,\beta,h}(\{\sigma\}) = \frac{1}{Z_{N,\beta,h}} e^{-\beta H_N(\sigma)} P_N(\{\sigma\}), \quad Z_{N,\beta,h} = \mathbb{E}_{P_N}\left[e^{-\beta H_N(\sigma)}\right]$$

[„Zustandssumme", Partitionsfkt.]

[β ... „inverse Temperatur", $\beta = \frac{k}{T}$,

$\mu_{N,\beta,h}$ ... Gibbs-Maß
oder Boltzmann-Verteilung
(zu Energie H_N, inv. Temp. β)
beschreibt (die Verteilung der mikroskop. Zustände) ein(es) Systems im thermischen Gleichgewicht.
(ein Postulat der statistischen Physik)]

[Ludwig Boltzmann, 1844–1906, Josiah Willard Gibbs, 1839–1903]

$$M_N = M_N(\sigma) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i$$ [mittlere Magnetisierung, eine [gewissermaßen: „die"] „makroskopische" Größe]

Frage(n) Wie verhält sich $\mu_{N,\beta,h}$ (bzw. M_N unter $\mu_{N,\beta,h}$) für $N \gg 1$?

Gibt es „besondere" Punkte (β, h) ?

Bem. 1) Unter P_N ist M_N vert. wie $\frac{1}{N}(2X-N)$ mit $X \sim \mathrm{Bin}_{N,1/2}$

2) $H_N(\sigma) = -N\left\{\frac{1}{2}(M_N(\sigma))^2 + h\, M_N(\sigma)\right\}$

(„mean field"-Eigenschaft [des Modells] [dt. Molekularfeldtheorie oder Molekularfeldnäherung)

Beob. M_N erfüllt unter P_N ein PGA (mit Rate N und) Ratenfunktion

$$I_{0,0}(m) = \frac{1+m}{2}\log\frac{1+m}{2} + \frac{1-m}{2}\log\frac{1-m}{2}$$

(für $m \in [-1,1]$, sonst $= +\infty$)

(denn $M_N = 2\frac{X_N}{N} - 1$ mit $X_N \sim \mathrm{Bin}_{N,1/2}$

$\Leftrightarrow \frac{X_N}{N} = \frac{M_N+1}{2}$, $P(M_N \in A) = P\left(\frac{X_N}{N} \in \{x = \frac{m+1}{2} : m \in A\}\right)$

und $\frac{X_N}{N}$ erfüllt PGA nach Satz 1.1 (Cramér))

Varadhans Lemma liefert

$$g(\beta,h) := \lim \frac{1}{N} \log Z_{N,\beta,h}$$

$$= \sup_m \left\{ \tfrac{1}{2}\beta m^2 + \beta h m - I_{0,0}(m) \right\}$$

$$= -\beta \inf_m \underbrace{\left\{ -\frac{m^2}{2} - hm + \frac{1}{\beta} I_{0,0}(m) \right\}}_{F_{\beta,h}(m)}$$

[in der thermodynamischen Interpretation ist $g(\beta,h)$ die Gibbs'sche freie Energie, $F_{\beta,h}(m)$ das Helmholtz-Potential oder die Helmholtz'sche freie Energie

$$(F = \underset{\uparrow \text{ innere Energie}}{U} - T \cdot \underset{\leftarrow \text{ Entropie}}{S})$$

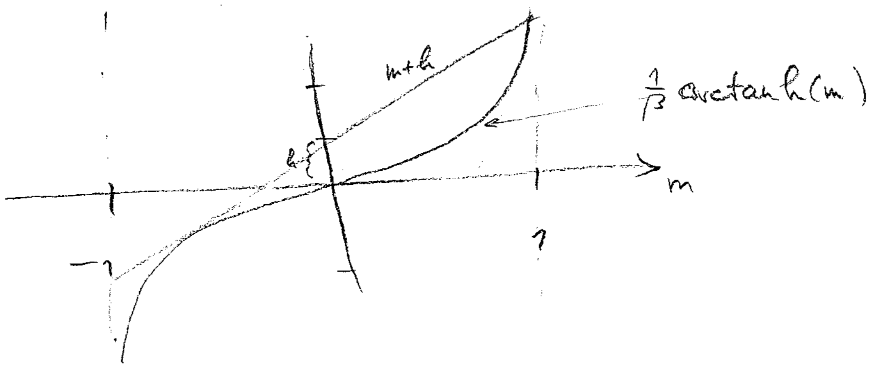
(Hermann Ludwig Ferdinand von Helmholtz, 1821–1894)]

$$0 \overset{!}{=} \frac{\partial}{\partial m} F_{\beta,h}(m) \quad \Big(= -m - h + \frac{1}{\beta}\Big(\frac{1}{2}\log\frac{1+m}{2} + \overbrace{\frac{1+m}{2}\cdot\frac{1}{\frac{1+m}{2}}\cdot\frac{1}{2}}^{=1}$$

$$- \frac{1}{2}\log\frac{1-m}{2} + \underbrace{\frac{1-m}{2}\,\frac{1}{\frac{1-m}{2}}\left(-\frac{1}{2}\right)}_{=-1}\Big)\Big)$$

$$= -m - h + \frac{1}{\beta}\underbrace{\frac{1}{2}\log\frac{1+m}{1-m}}_{=\operatorname{arctanh}(m)}$$

d.h. $\tanh(\beta(m+h)) = m$



$h=0$: ($m=0$ ist stets Lösung)

für $\beta \leq 1$: $m=0$ ist einzige Lösung, ist Minimum von $F_{\beta,0}(\cdot)$

für $\beta > 1$: $\exists$ zwei Minima

$$m_{\beta,0,+} = - m_{\beta,0,-} \quad (\in (0,1))$$

(Lösungen von $\beta \cdot m = \operatorname{arctanh}(m)$)

$h \neq 0$: für jedes $\beta \neq 0$ $\exists_1$ globales Minimum $m_{\beta,h}$, dies hat dasselbe Vorz. wie h

[u.U. existiert ein weiteres lokales Minimum, das das entgegengesetzte Vorzeichen hat]

PGA unter exponentieller Transformation (Kor. 7) zeigt:

Unter $\mu_{N,\beta,h}$ erfüllt M_N PGA mit Ratenfunktion

$$I_{\beta,h}(m) = \beta \underbrace{\Big(F_{\beta,h}(m) - \inf_{m' \in [-1,1]} F_{\beta,h}(m') \Big)}_{\tilde F_{\beta,h}(m)} \quad \left(\begin{array}{l} m \in [-1,1], \\ \text{sonst } +\infty \end{array} \right)$$

[Bem.: Für β groß ist dies eine nicht-konvexe Ratenfunktion]

$\tilde\mu_{N,\beta,h} = \mu_{N,\beta,h} \circ M_N^{-1}$, Verteilung von M_N unter $\mu_{N,\beta,h}$

Korollar:

$\beta \leq 1,\ h = 0$: $\quad \bar{\mu}_{N,\beta,0} \xrightarrow{w} \delta_0 \quad \left(\mu_{N,\beta,0}(|M_N| > \varepsilon) \xrightarrow[N\to\infty]{} 0 \right)$

[sogar exponentiell schnell]

$\beta > 1,\ h = 0$: $\quad \bar{\mu}_{N,\beta,0} \xrightarrow{w} \frac{1}{2}\delta_{m_{\beta,0,-}} + \frac{1}{2}\delta_{m_{\beta,0,+}}$

$\big(\mu_{N,\beta,0}(|M_N - m_{\beta,0,+}| > \varepsilon$ oder $|M_N - m_{\beta,0,-}| > \varepsilon) \longrightarrow 0,$

dann benutze Symmetrie $\big)$ [sog. „spontane Magnetisierung"]

[$\beta_c = 1$, „Phasenübergang"]

$\beta > 0,\ h \neq 0$: $\quad \bar{\mu}_{N,\beta,h} \xrightarrow{w} \delta_{m_{\beta,h}}$

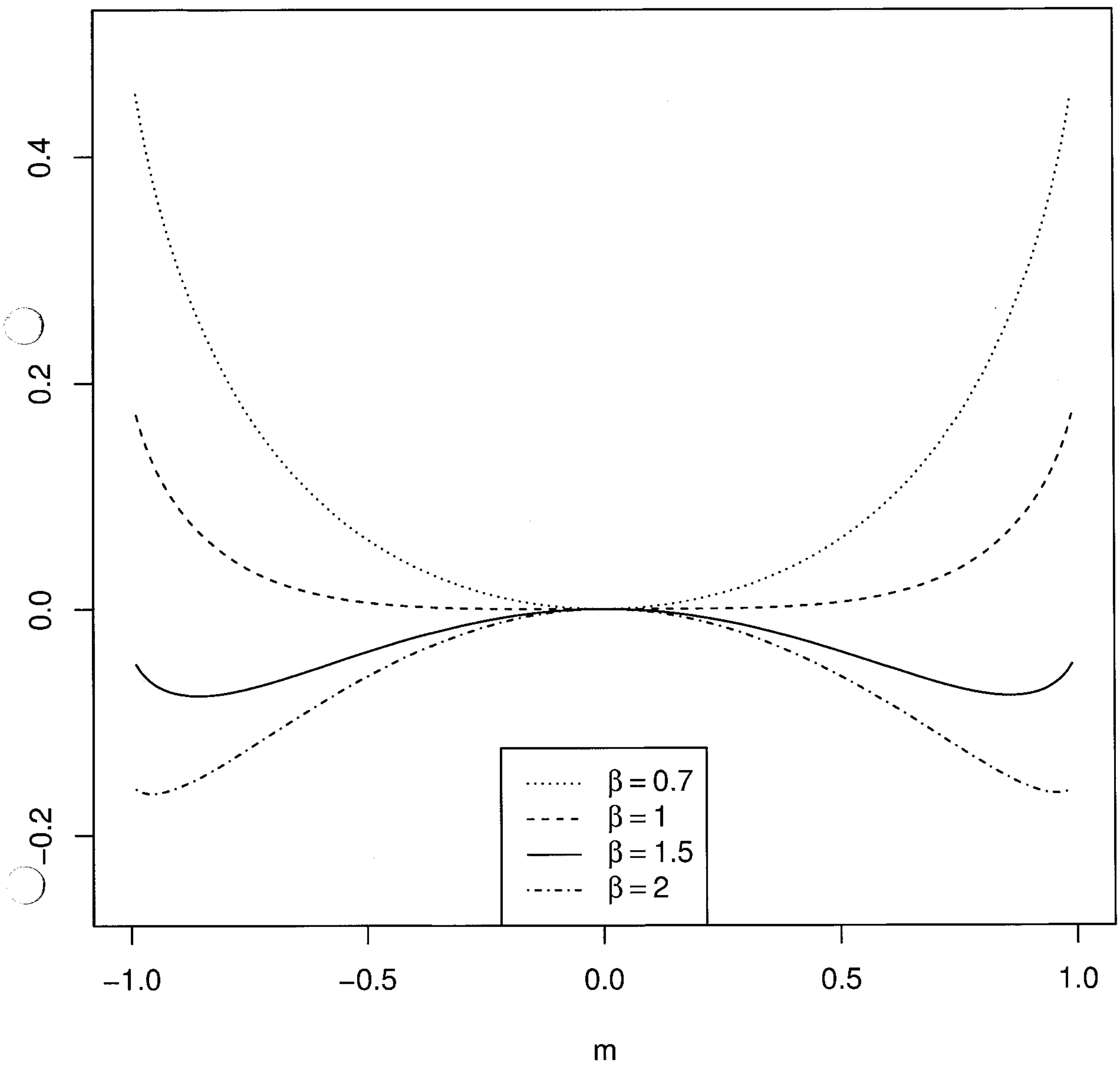
[Plots $\beta \mapsto m_{\beta,h}$, $h \mapsto m_{\beta,h}$ zeigen.]

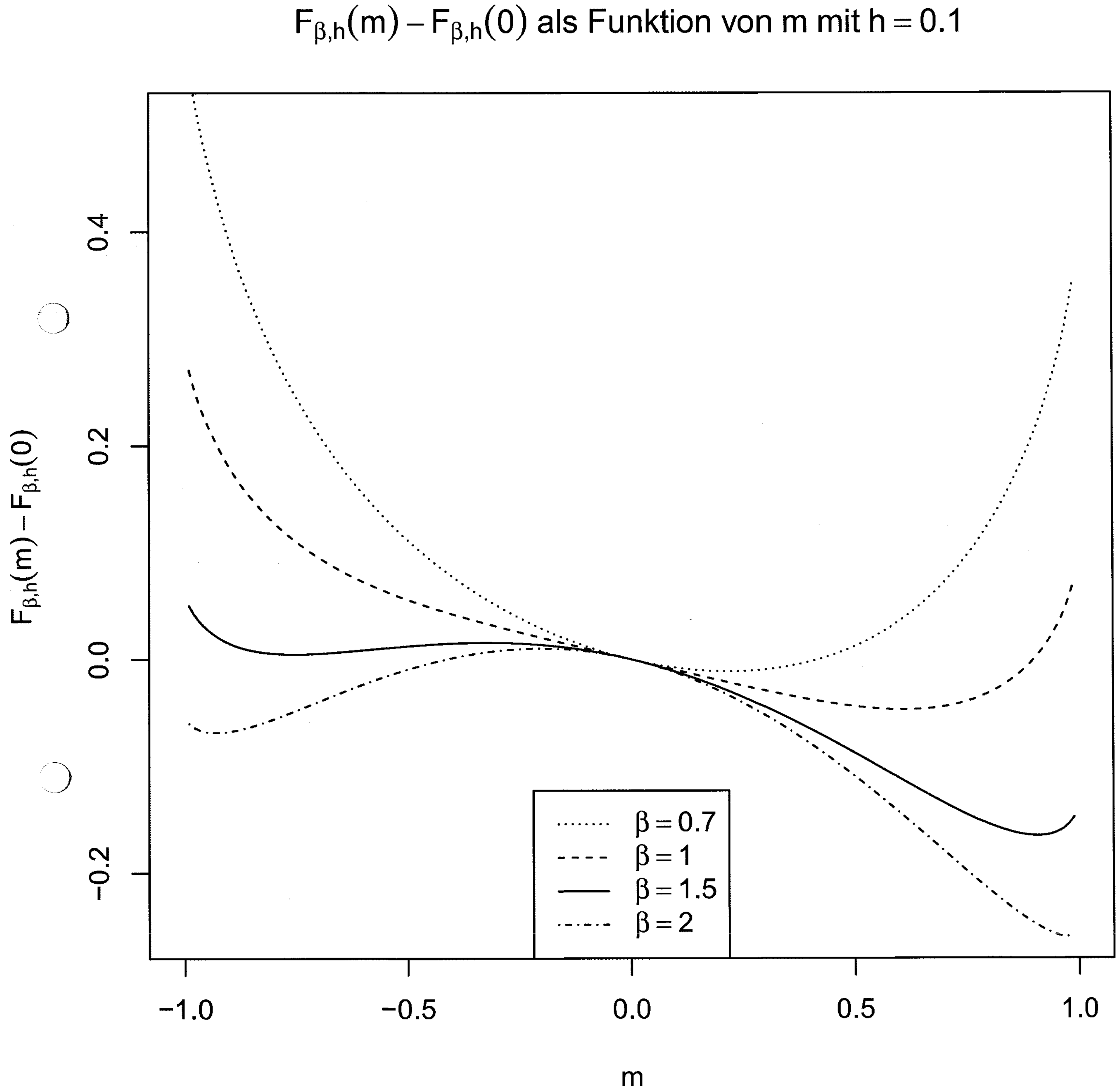
[Bem.: Curie-Weiss'sches Gesetz:

$$m_{\beta,h} \sim \frac{h}{\frac{1}{\beta} - 1} = \frac{h}{T - T_c} \text{ für } \beta \downarrow 0 \quad (\text{d.h. } T \uparrow \infty) \Big]$$

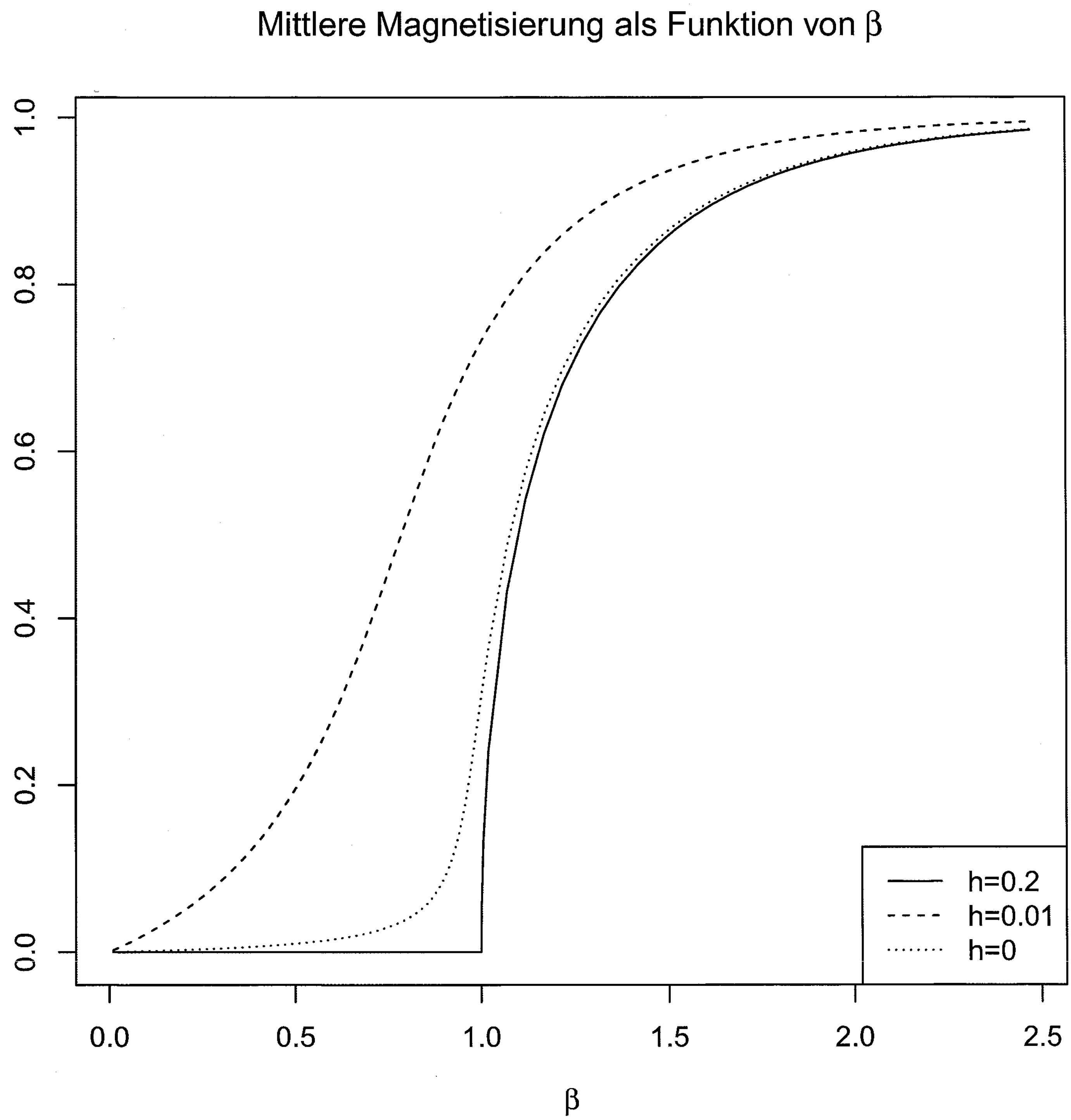
für $h > 0$

$F_{\beta,h}(m) - F_{\beta,h}(0)$ als Funktion von m mit $h = 0$

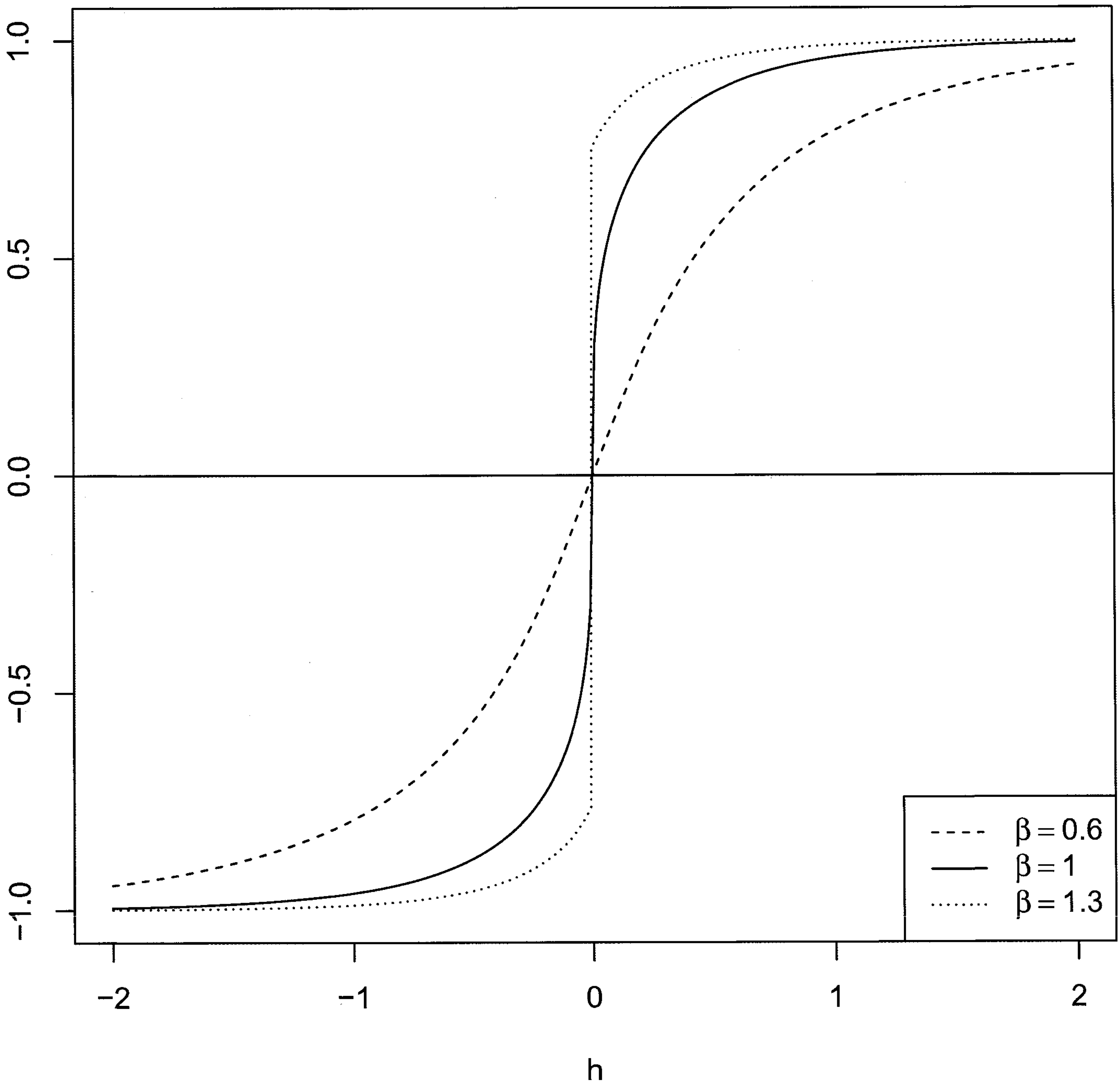




$F_{\beta,h}(m) - F_{\beta,h}(0)$ als Funktion von m mit $h = 0.1$
$F_{\beta,h}(m) - F_{\beta,h}(0)$
0.4
0.2
0.0
−0.2
−1.0
−0.5
0.0
0.5
1.0
m
$\beta = 0.7$
$\beta = 1$
$\beta = 1.5$
$\beta = 2$


Mittlere Magnetisierung als Funktion von β
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
β
h=0.2
h=0.01
h=0

Mittlere Magnetisierung als Funktion von h



3. Große Abweichungen der empirischen Verteilung einer u.i.v. Folge von ZVn

E polnischer Raum (metrischer, separabler u. vollständiger Raum)

$X_1, X_2, \dots$ u.i.v., $X_i \sim \mu \in \mathcal{M}_1(E)$

$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ (n-tes) empirisches Maß

(ZV mit Werten in $\mathcal{M}_1(E)$)

Wir statten $\mathcal{M}_1(E)$ mit der schwachen Topologie aus

(d.h. $\nu_n \xrightarrow[n\to\infty]{w} \nu \iff \forall f \in C_b(E): \int f \, d\nu_n \to \int f \, d\nu$)

<u>Bem.</u> $L_n \xrightarrow[n\to\infty]{w} \mu$ f.s.

(denn $\int f \, dL_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i) \xrightarrow[n\to\infty]{} \mathbb{E}[f(X_1)] = \int f \, d\mu$ f.s.
nach starkem Ges. d. gr. Z)

<u>Satz 1</u> (Sanov, 1957)
(für den Fall $|E| < \infty$)

[Ivan Nicolaevich Sanov, On the probability of large deviations of random variables (Russian), Mat. Sb. 42, 11–44 (1957)]

Die Folge $(L_n)_n$ erfüllt ein PGA mit Rate n und Ratenfunktion

$$I_\mu(\nu) = H(\nu|\mu) := \begin{cases} \int_E \left(\log \frac{d\nu}{d\mu}\right) d\nu, & \text{falls } \nu \ll \mu \\ +\infty, & \text{sonst.} \end{cases}$$

$H(\nu|\mu)$ heißt relative Entropie (von ν bzgl. μ), auch „Kullback-Leibler-Abstand"

[Erinnerung: $\nu \ll \mu$ („ν ist absolut stetig bzgl. μ")
$:\iff \forall B \in \mathcal{B}(E): \mu(B) = 0 \Rightarrow \nu(B) = 0$,
nach Satz von Radon-Nikodym ist dies äquivalent zur Existenz einer Dichte $\frac{d\nu}{d\mu} \in \mathcal{L}^1(\mu)$: $\nu(B) = \int_B \frac{d\nu}{d\mu}(x) \, \mu(dx)$

Bemerkung:
Satz von Sanov „von Hand“ im Fall $|E| < \infty$.

$$E = \{1, \dots, d\}, \quad \nu \in \mathcal{M}_1(E) \quad \left(\text{mit } \nu(i) := \nu(\{i\}) > 0 \;\; \forall\, i = 1, \dots, d\right)$$

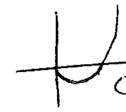
$X_1, X_2, \dots$ u.i.v. $\sim \nu$,

$$L_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i} \quad \text{(empirische Verteilung)}$$

Für $\mu \in \mathcal{M}_1(E)$ sei $H(\mu|\nu) = \sum_{x \in E} \mu(x) \log \frac{\mu(x)}{\nu(x)}$ (relative Entropie von μ bzgl. ν)

(die Form $H(\mu|\nu) = \sum_{x \in E} \nu(x) \cdot \underbrace{\frac{\mu(x)}{\nu(x)} \log \frac{\mu(x)}{\nu(x)}}_{= f(\mu(x)/\nu(x))}$ mit $f(x) = x \log x$

zeigt: $H(\mu|\nu) \geq 0$
und $H(\mu|\nu) = 0$ g.d.w. $\mu = \nu$)



(strikt konv.)

$$\boxed{\begin{array}{l} L_N,\ N \in \mathbb{N} \text{ erfüllt PGA} \\ \text{mit Rate } N \text{ und Ratenfunktion } H(\cdot|\nu) \end{array}}$$

Skizze:
Sei $\mu_N = \sum_{i=1}^{d} \frac{k_i}{N} \delta_i \quad \left(\text{mit } \left|\frac{k_i}{N} - \mu(i)\right| < \frac{1}{N}\right)$

$$\mathbb{P}(L_N = \mu_N) = \binom{N}{k_1\, k_2 \dots k_d} \prod_{i=1}^{d} \nu(i)^{k_i}$$

$$\underset{\text{Stirling}}{\sim} (2\pi)^{\frac{1-d}{2}} \left(\frac{N^{d-1}}{\frac{k_1}{N} \cdot \frac{k_2}{N} \cdots \frac{k_d}{N}} \right)^{1/2} \frac{N^N}{k_1^{k_1} \cdots k_d^{k_d}} \prod_{i=1}^{d} \nu(i)^{k_i}$$

$$= c_N(\mu) \exp\Big[\underbrace{\sum_{i=1}^{d} k_i \left(\log \nu(i) - \log \frac{k_i}{N} \right)}_{\sim -N \sum_{i=1}^{d} \mu(i) \log \frac{\mu(i)}{\nu(i)} = -N H(\mu|\nu)} \Big]$$

subexponentiell $\longrightarrow$ $c_N(\mu)$

[und beachte:
$$\#\{(k_1, \dots, k_d) \in \mathbb{N}_0^d : k_1 + \dots + k_d = N\} = \binom{N+d-1}{d-1} \leq c_d \cdot N^{d-1},$$ was subexp. wächst]

Bem. $H(\nu|\mu) \in [0,\infty]$, $H(\nu|\mu) = 0 \iff \nu = \mu$

(denn falls es eine Dichte $\frac{d\nu}{d\mu} \in \mathcal{L}^1(\mu)$ gibt, so ist

$$H(\nu|\mu) \underset{\uparrow}{=} \int \log\frac{d\nu}{d\mu}(x)\, \frac{d\nu}{d\mu}(x)\, \nu(dx) \geq \int \frac{d\nu}{d\mu} d\nu \cdot \log\Big(\overbrace{\int \frac{d\nu}{d\mu} d\nu}^{=1}\Big) = 0,$$

Integral stets wohldef., da $[0,\infty) \ni y \mapsto y\log y$ nach unten beschränkt

Jensensche Ungl.

die Ungl. ist scharf, es sei denn $\frac{d\nu}{d\mu} = 1$ ν-f.ü.

und in diesem Fall ist $\nu = \mu$)

Lemma 2

$$H(\nu|\mu) = \sup_{\substack{f: E \to \mathbb{R} \\ \text{beschr., m-b.}}} \Big\{\int f d\nu - \log \int e^f d\mu\Big\} = \sup_{f \in C_b(E)} \Big\{\int f d\nu - \log \int e^f d\mu\Big\}$$

die Abb. $\nu \mapsto H(\nu|\mu)$ ist konvex und nach unten halbstetig, hat kompakte Niveaumengen.

Bem.

1) Sei $\nu \ll \mu$. Für $f \in \underset{\text{beschr., m.b. Fkt.nen}}{B_b(E)}$ definiere

$\mu_f \in \mathcal{M}_1(E)$ via $\frac{d\mu_f}{d\mu} = \frac{e^f}{\int e^f d\mu}$,

es ist

$$H(\nu|\mu) = \int \log\frac{d\nu}{d\mu}(x)\,\nu(dx) = \int \log\Big(\frac{d\nu}{d\mu_f}(x)\,\frac{d\mu_f}{d\mu}(x)\Big)\nu(dx)$$

$$= \underbrace{H(\mu_f|\nu)}_{\geq 0} + \int f d\nu - \log\int e^f d\mu$$

d.h. $H(\nu|\mu) \geq \sup_{f \in B_b(E)} \{\cdots\} \underset{\uparrow}{\geq} \sup_{f\in C_b(E)} \{\cdots\}$ und falls

trivialerweise

$f = \log\frac{d\nu}{d\mu} \in C_b(E)$ so gilt Gleichheit

[Im Allgemeinen ist $x \mapsto \log \frac{d\nu}{d\mu}(x)$ weder stetig noch beschränkt, und wir benötigen ein Approximations-argument]

2) Wir zeigen: Falls $J := \sup_{f \in C_b(E)} \{ \int f \, d\nu - \log \int e^f \, d\mu \} < \infty$,

so ist $\nu \ll \mu$ und $J \geq H(\nu|\mu)$ (+)

[beachte: falls $J = \infty$, so ist nichts mehr zu zeigen]

n. Def. gilt $\int f \, d\nu - \log \int e^f \, d\mu \overset{(*)}{\leq} J \quad \forall f \in C_b(E)$,

falls f_n Ungl. (*) erfüllen und $f_n \to f$ punktweise mit $\sup_n \|f_n\|_\infty < \infty$,
so erfüllt auch f (*) [verwende dominierte Konvergenz]

demnach: (*) gilt für $f \in B_b(E)$
[approximiere z.B. zunächst Indikatorfunktionen abgeschlossener Mengen ...]

Sei $A \in \mathcal{B}(E)$ mit $\mu(A) = 0$, mit $f = r \cdot 1_A$, $r > 0$
folgt $r\nu(A) - \log 1 = r \cdot \nu(A) \leq J < \infty$ in (*)
mit $r \to \infty$ also $\nu(A) = 0$, demnach $\nu \ll \mu$.

Sei $\varphi := \frac{d\nu}{d\mu}$ (ex. nach Satz von Radon-Nikodym)

Falls ~~log f~~ $\log \varphi \in B_b(E)$, so
folgt $H(\nu|\mu) = \int \log \varphi \, d\nu \leq J$ (nach (*))
und somit (+).

Falls $\log \varphi$ nach unten beschr. (aber nicht notw. nach oben)

$$H(\nu|\mu) = \int \log \varphi \, d\nu \underset{\text{(Fatou)}}{\leq} \liminf_{n\to\infty} \int \log(\varphi \wedge n) \, d\nu$$

$$\leq J + \underbrace{\liminf_{n\to\infty} \log\left(\int f \wedge n \, d\mu\right)}_{\overset{\uparrow}{=} \log(\int f d\mu) = \log(1) = 0}$$

dominierte Konvergenz

d.h. (+) gilt in diesem Fall.

Falls $\log\varphi$ nicht nach unten beschränkt:

$\varepsilon \in [0,1]$, $\varphi_\varepsilon := \varepsilon + (1-\varepsilon)\varphi$, ν_ε def. durch $\frac{d\nu_\varepsilon}{d\mu} = \varphi_\varepsilon$,

n. obigem ist

$$(**) \quad H(\nu_\varepsilon|\mu) = \int \log \varphi_\varepsilon \, d\nu_\varepsilon \leq J_\varepsilon := \sup_{f \in C_b(E)} \left\{ \int f \, d\nu_\varepsilon - \log \int e^f d\mu \right\}$$

$[0,1] \ni \varepsilon \mapsto J_\varepsilon$ ist konvex, beschränkt, nach unten halbstetig,

$\Longrightarrow$ J_ε ist stetig

[zeige $H(\nu_\varepsilon|\mu) \underset{\varepsilon\downarrow 0}{\longrightarrow} H(\nu|\mu)$:]

$$H(\nu_\varepsilon|\mu) = \int \underbrace{\varphi_\varepsilon \log \varphi_\varepsilon}_{\underset{\text{(Jensen)}}{\leq} \varepsilon \cdot 1 \log 1 + (1-\varepsilon)\varphi \log \varphi} d\mu \leq (1-\varepsilon) \int \varphi \log \varphi \, d\mu = (1-\varepsilon) H(\nu|\mu)$$

Sowie

$$H(\nu_\varepsilon|\mu) = \varepsilon \int \underbrace{\log \varphi_\varepsilon}_{\geq \log \varepsilon} d\mu + (1-\varepsilon) \int \varphi \underbrace{\log \varphi_\varepsilon}_{\geq \varepsilon \log 1 + (1-\varepsilon)\log\varphi} d\mu$$

(Jensen, log ist konkav)

$$\geq \varepsilon \log \varepsilon + (1-\varepsilon)^2 \underbrace{\int \varphi \log \varphi \, d\mu}_{= H(\nu|\mu)}$$

mit $\varepsilon \downarrow 0$ folgt aus $(**)$: $\quad H(\nu|\mu) \le J_0 = J$,

d.h. $(+)$ gilt.

Konvexität und u. unten Halbstetigkeit von $\nu \mapsto H(\nu|\mu)$ folgen aus der Darstellung $H(\nu|\mu) = J$ als Supremum linearer, stetiger Fktnen.

Sei $s > 0$, zeige $\{\nu : H(\nu|\mu) \le s\}$ $(\subset \mathcal{M}_1(E))$ ist kompakt.

$\varepsilon > 0$. Wähle $K \subset E$ kompakt mit $\mu(E \setminus K) < \varepsilon$,

setze $f = M \cdot 1_{K^c}$ mit $M > 0$ in $(*)$,

für ν mit $H(\nu|\mu) \le s$ ist

$$s \geq H(\nu|\mu) \geq \int f\,d\nu - \log \int e^f\,d\mu$$

$$= M\nu(K^c) - \log\left(\mu(K) + e^M \mu(K^c)\right),$$

also $$\nu(K^c) \le \frac{s + \log(1 + \varepsilon e^M)}{M} \underset{\substack{\uparrow \\ \text{wähle} \\ M = \log\frac{1}{\varepsilon}}}{=} \frac{s + \log(2)}{\log(\frac{1}{\varepsilon})}$$

d.h. die s-Niveaumenge ist straff und somit relativkompakt (nach Satz von Prohorov),

sie ist auch abgeschl. (denn $\nu \mapsto H(\nu|\mu)$ ist nach unten halbstetig),

somit ist sie kompakt. $\qquad\lrcorner$

Bis hierher am 4.12.2014

Bew. (des Satzes von Sanov)

<u>Untere Schranke:</u>

$G \subset \mathcal{M}_1(E)$ offen, $\nu \in G$ mit $H(\nu|\mu) < \infty$.

Nehme zunächst an, dass $\operatorname{ess\,inf} \frac{d\nu}{d\mu} \geq c > 0$ für ein $c \in \mathbb{R}$ (wegen $H(\nu|\mu) < \infty$ gilt $\nu \ll \mu$, d.h. es gibt die Dichte $\frac{d\nu}{d\mu}$)

dann gilt auch $\mu \ll \nu$ (also $\mu \sim \nu$)

und $\varphi := \frac{d\mu}{d\nu}$ ist (nach oben) beschränkt.

$\hat{X}_1, \hat{X}_2, \ldots$ u.i.V. $\sim \nu$.

$$\mathbb{P}(L_n(X) \in G] = \mathbb{E}\Big[\mathbb{1}_{\{L_n(\hat{X}) \in G\}} \prod_{i=1}^n \varphi(\hat{X}_i)\Big]$$

$$\geq \mathbb{E}\Big[\mathbb{1}_{\{L_n(\hat{X}) \in G,\ \prod_{i=1}^n \varphi(\hat{X}_i) \geq e^{-n(H(\nu|\mu)+\varepsilon)}\}} \prod_{i=1}^n \varphi(\hat{X}_i)\Big]$$

$$\geq e^{-n(H(\nu|\mu)+\varepsilon)} \underbrace{\mathbb{P}\Big(L_n(\hat{X}) \in G,\ \frac{1}{n}\sum_{i=1}^n \log \varphi(\hat{X}_i) \geq -H(\nu|\mu) - \varepsilon\Big)}_{\xrightarrow[n\to\infty]{} 1}$$

(denn $L_n(\hat{X}) \xrightarrow[n\to\infty]{w} \nu \in G$ f.s.

und $\frac{1}{n}\sum_{i=1}^n \varphi(\hat{X}_i) \xrightarrow[n\to\infty]{} \mathbb{E}[\log \varphi(\hat{X}_1)]$

[nutze Ges. d. gr. Z.] $\qquad = -H(\nu|\mu)$)

Somit $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n(X) \in G) \geq -H(\nu|\mu) - \varepsilon$

(dann $\varepsilon \downarrow 0$, optimiere über $\nu \in G$)

Für den allg. Fall setze $\nu_\varepsilon := \varepsilon\mu + (1-\varepsilon)\nu$,

es gilt $\frac{d\nu_\varepsilon}{d\mu} \geq \varepsilon > 0$, $\nu_\varepsilon \xrightarrow[\varepsilon\downarrow 0]{w} \nu$ und $H(\nu_\varepsilon|\mu) \xrightarrow[\varepsilon\downarrow 0]{} H(\nu|\mu)$

(vgl. Bew. von Lemma 2)

Obere Schranke:

Sei $F \subset \mathcal{M}_1(E)$ kompakt, o.E. $\inf_{\nu \in F} H(\nu|\mu) > 0$.

Wähle $\alpha \in (0, \inf_{\nu \in F} H(\nu|\mu))$, ~~für~~

zu $\nu \in F$ ein $f_\nu \in C_b(E)$ mit $\int f_\nu \, d\nu - \log \int e^{f_\nu} d\mu > \alpha$

(nach Lemma 2 möglich)

$\mathcal{U}_\nu := \{\eta \in \mathcal{M}_1(E) : \int f_\nu \, d\eta - \log \int e^{f_\nu} d\mu > \alpha\} \ni \nu$

ist offene Umg.

$$\mathbb{P}(L_n(X) \in \mathcal{U}_\nu) = \mathbb{P}\left(e^{n \int f_\nu \, dL_n(X)} > e^{n\alpha + n \log \int e^{f_\nu} d\mu}\right)$$

$$\underset{\text{(Markov-Ungl.)}}{\leq} e^{-n\alpha} \underbrace{\left(\int e^{f_\nu} d\mu\right)^{-n} \mathbb{E}\left[e^{n \int f_\nu \, dL_n(X)}\right]}_{=1} = e^{-n\alpha}$$

also $\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(L_n(X) \in \mathcal{U}_\nu) \leq -\alpha$.

F kompakt $\Rightarrow F \subset \mathcal{U}_{\nu_1} \cup \dots \cup \mathcal{U}_{\nu_k}$ für geeign. $\nu_1, \dots, \nu_k \in F$,

$$\Rightarrow \limsup \frac{1}{n} \log \underbrace{\mathbb{P}(L_n(X) \in F)}_{\leq \mathbb{P}(L_n \in \mathcal{U}_{\nu_1}) + \dots + \mathbb{P}(L_n \in \mathcal{U}_{\nu_k}) \leq k e^{-n\alpha}} \leq -\alpha$$

[dies zeigt soweit ein <u>schwaches</u> PGA.]

Zeige: (L_n) ist exponentiell straff (zur Rate n)

[dann zeigt Beob. 2.5 das volle PGA]

Für $\varepsilon > 0$, $K \subset E$ kompakt sei $A_{K,\varepsilon} = \{\nu \in \mathcal{M}_1(E) : \nu(K^c) \leq \varepsilon\}$

($A_{K,\varepsilon} \subset \mathcal{M}_1(E)$ ist abgeschl. (vgl. Portmanteau-Theorem)

Seien $\varepsilon_m \downarrow 0$, $B_m \subset E$ kompakt mit $\mu(B_m^c) =: b_m < \varepsilon_m$

[wir spezifizieren ε_m, b_m später]

$\mathcal{K} := \bigcap_m A_{B_m, \varepsilon_m}$ ist kompakt (n. Satz von Prohorov, offenb. ist $\mathcal{K}$ straff)

$$\mathbb{P}(L_n \notin \mathcal{K}) = \mathbb{P}\Big(\exists\, m : \overbrace{\sum_{i=1}^n 1_{B_m^c}(X_i)}^{\sim \mathrm{Bin}(n, b_m)} > n\varepsilon_m\Big)$$

$$\leq \sum_{m=1}^\infty \underbrace{\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n 1_{B_m^c}(X_i) > \varepsilon_m\Big)}_{\leq \exp\left(-n \cdot \sup_{t \geq 0}\{\varepsilon_m t - \log(e^t b_m + 1 - b_m)\}\right)}$$

$$\underbrace{\sup_{t \geq 0}\{\varepsilon_m t - \log(e^t b_m + 1 - b_m)\}}_{= I_{\mathrm{Ber}(b_m)}(\varepsilon_m) = \varepsilon_m \log \frac{\varepsilon_m}{b_m} + (1-\varepsilon_m)\log\frac{1-\varepsilon_m}{1-b_m}}$$

(Satz 1.1 (Cramér))

Zu $s > 0$ wähle (B_m), (ε_m) so dass $I_{\mathrm{Ber}(b_m)}(\varepsilon_m) > 2ms$

$$\mathbb{P}(L_n \notin \mathcal{K}) \leq \sum_{m=1}^\infty e^{-2ms \cdot n} = e^{-2sn}\frac{1}{1-e^{-2sn}} \leq e^{-sn}$$

$\lrcorner$

Beispiel: Prinzip der Gibbs-Konditionierung

$X_1, X_2, \ldots$ u.i.v. $\sim \mu \in \mathcal{M}_1(E)$, $|E| < \infty$ (der Einfachheit der Darstellung halber)

(o.E. sei $E = \{1, \ldots, d\}$, $\mu(\{i\}) > 0$ für $i = 1, \ldots, d$)

wir fassen $\mathcal{M}_1(E)$ als (kompakte) Teilmenge von $\mathbb{R}^d$ auf, mit entspr. Relativtopologie [ist in diesem Fall mit der schwachen Topologie auf $\mathcal{M}_1(E)$ identisch: jede Funktion $f: E \to \mathbb{R}$ ist stetig und beschränkt]

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

Frage: Sei $f: E \to \mathbb{R}$, $A \in \mathcal{B}(\mathbb{R})$

$$\mathcal{L}\left(X_1 \,\middle|\, \int f \, dL_n \in A\right) \approx ? \quad \text{für } n \text{ groß}$$

(speziell: falls $\int f \, d\mu \notin A$)

Beachte: Bedingt auf $\{\int f \, dL_n \in A\}$ sind $X_1, \ldots, X_n$ zwar austauschbar, aber <u>nicht</u> unabhängig.

[Frage ist z.B. im Kontext der statistischen Physik bedeutsam: Wie verhält sich ein „typisches Teilchen" eines großen Systems, das eine „globale" (und „untypische") Randbedingung erfüllen muss?]

Sei $\Gamma \subset \mathcal{M}_1(E)$, $\Gamma \neq \emptyset$ mit

$$\inf_{\nu \in \Gamma^\circ} H(\nu|\mu) = \inf_{\nu \in \overline{\Gamma}} H(\nu|\mu) =: I_\Gamma ,$$

$$\mathcal{M}_1(E) \ni \tilde{\mu}_n := \mathcal{L}(X_1 \mid L_n \in \Gamma) \ \Big(\underset{\uparrow}{=} \mathbb{E}[L_n \mid L_n \in \Gamma] \Big),$$

(fasse L_n als Vektor $(L_n(\{1\}), L_n(\{2\}), \dots, L_n(\{d\}))$ auf)

$$\tilde{\mathcal{M}} := \{\nu \in \overline{\Gamma} : H(\nu|\mu) = I_\Gamma\}$$

<u>Satz 3</u> (Gibbs-Konditionierung)

1) Alle Häufungspunkte von $(\tilde{\mu}_n)_n$ liegen im Abschluss der konvexen Hülle von $\tilde{\mathcal{M}}$.

2) Falls Γ konvex ist [mit $\mathring{\Gamma} \neq \emptyset$], so gilt $\tilde{\mathcal{M}} = \{\mu^*\}$ und $\tilde{\mu}_n \xrightarrow[n\to\infty]{w} \mu^*$ für ein $\mu^* \in \mathcal{M}_1(E)$.

Bew.: Metrisiere $\mathcal{M}_1(E)$ mit Totalvariationsabstand

$$d_{TV}(\nu, \nu') := \frac{1}{2} \sum_{i=1}^{d} |\nu(\{i\}) - \nu'(\{i\})| \quad (\in [0,1])$$

1) Sei $\tilde{\mathcal{M}}^\delta := \{\nu' \in \mathcal{M}_1(E) : d_{TV}(\nu', \tilde{\mathcal{M}}) < \delta\}$, $\delta > 0$,

es gilt

$$\mathbb{P}(L_n \in \tilde{\mathcal{M}}^\delta \mid L_n \in \Gamma) \underset{n\to\infty}{\longrightarrow} 1 \qquad (*)$$

[Bem.: Für eine wesentlich allgemeinere Version von Satz 3, für den E ein allg. polnischer Raum sein kann, siehe z.B. Dembo & Zeitouni, Thm. 7.3.3] denn $\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(L_n \in \Gamma) = I_\Gamma$,

(nach Satz von Sanov)

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(L_n \in \Gamma \cap (\tilde{\mathcal{M}}^\delta)^c)$$

$$\leq \limsup \frac{1}{n}\log \mathbb{P}(L_n \in \underbrace{\overline{\Gamma} \cap (\tilde{\mathcal{M}}^\delta)^c}_{\text{ist abgeschl.}})$$

$$\underset{(\text{Sanov})}{\leq} - \inf_{\nu \in \overline{\Gamma}\cap(\tilde{\mathcal{M}}^\delta)^c} H(\nu|\mu) < -I_\Gamma .$$

($\overline{\Gamma}\cap(\tilde{\mathcal{M}}^\delta)^c$ ist kompakt und $H(\cdot|\mu)$ u.u.h., also wird Infimum angenommen in einem ν', $\nu' \notin \tilde{\mathcal{M}}$)

(Somit $\mathbb{P}(d_{TV}(L_n, \tilde{\mathcal{M}}) \geq \delta \,|\, \Gamma) \xrightarrow[n\to\infty]{} 0$ [sogar exponentiell schnell])

Zeige:

$$d_{TV}(\tilde{\mu}_n, co(\tilde{\mathcal{M}}^\delta)) \xrightarrow[n\to\infty]{} 0 \qquad (**)$$

(mit $co(\tilde{\mathcal{M}}^\delta) = \{\alpha\nu + (1-\alpha)\nu' : \nu, \nu' \in \tilde{\mathcal{M}}^\delta, \alpha \in [0,1]\}$ konvexe Hülle von $\tilde{\mathcal{M}}^\delta$)

$$\tilde{\mu}_n - \overbrace{\mathbb{E}[L_n \,|\, L_n \in \tilde{\mathcal{M}}^\delta \cap \Gamma]}^{\in co(\tilde{\mathcal{M}}^\delta)}$$

$$= \mathbb{E}[L_n \,|\, L_n \in \Gamma] - \mathbb{E}[L_n \,|\, L_n \in \tilde{\mathcal{M}}^\delta \cap \Gamma]$$

$$\left(= \frac{\mathbb{E}[L_n 1_{\{L_n \in \Gamma\cap\tilde{\mathcal{M}}^\delta\}}] + \mathbb{E}[L_n 1_{\{L_n\in\Gamma\cap(\tilde{\mathcal{M}}^\delta)^c\}}]}{\mathbb{P}(L_n\in\Gamma)} - \frac{\mathbb{E}[L_n 1_{\{L_n\in\Gamma\cap\tilde{\mathcal{M}}^\delta\}}]}{\mathbb{P}(L_n\in\Gamma\cap\tilde{\mathcal{M}}^\delta)} \right.$$

$$= \frac{\mathbb{P}(L_n\in\Gamma\cap(\tilde{\mathcal{M}}^\delta)^c)}{\mathbb{P}(L_n\in\Gamma)} \cdot \mathbb{E}[L_n \,|\, L_n \in \Gamma\cap(\tilde{\mathcal{M}}^\delta)^c] \; -$$

$$- \left(1 - \frac{P(L_n \in \Gamma \cap \tilde{\mathcal{M}}^\delta)}{P(L_n \in \Gamma)} \right) \mathbb{E}[L_n \mid L_n \in \Gamma \cap \tilde{\mathcal{M}}^\delta] \Big)$$

$$= \underbrace{\mathbb{P}(L_n \in \Gamma \cap (\tilde{\mathcal{M}}^\delta)^c \mid L_n \in \Gamma)}_{\xrightarrow[n\to\infty]{} 0}$$

$$\cdot \Big(\underbrace{\mathbb{E}[L_n \mid L_n \in \Gamma \cap (\tilde{\mathcal{M}}^\delta)^c] - \mathbb{E}[L_n \mid L_n \in \Gamma \cap \tilde{\mathcal{M}}^\delta]}_{\text{haben Abstand } \le 1 \text{ (stets)}} \Big)$$

d.h. $(**)$ gilt.

[bis hierher am 11.12.2014]

2) $\tilde{\mathcal{M}} \neq \emptyset$ (denn $H(\cdot|\mu)$ hat kompakte Niveaumengen),

ang. $\nu_1, \nu_2 \in \tilde{\mathcal{M}}$ mit $\nu_1 \neq \nu_2$,

so wären auch $\tilde{\nu}_\theta := \theta \nu_1 + (1-\theta)\nu_2$, $\theta \in [0,1]$,

$\tilde{\nu}_\theta \in \Gamma$ (Konvexität) und $H(\tilde{\nu}_\theta|\mu) < \theta H(\nu_1|\mu) + (1-\theta) H(\nu_2|\mu) = I_\Gamma$

(strikte Konvexität von $H(\cdot|\mu)$), $\lightning$

Bem. I.A. wird in Satz 3, 1) die konvexe Hülle $\mathrm{Co}(\tilde{\mathcal{M}})$ tatsächlich gebraucht, d.h. 1) gilt nicht für $\tilde{\mathcal{M}}$ selbst.

Behr. z.B. $E = \{0,1\}$, $\mu = \mathrm{Ber}(\frac{1}{2})$,

$\Gamma = \{ \nu \in \mathcal{M}_1(\{0,1\}) : \nu(\{1\}) \in [0,\frac{1}{4}] \cup [\frac{3}{4}, 1] \}$,

$\mathcal{L}(X_1 \mid L_n \in \Gamma) \to \mathrm{Ber}(\frac{1}{2}) \in \mathrm{Co}(\tilde{\mathcal{M}})$ aber $\notin \tilde{\mathcal{M}} = \{\mathrm{Ber}(\frac{1}{4}), \mathrm{Ber}(\frac{3}{4})\}$

Zusammenhang zu Gibbs-Maßen

$\varphi : E \to \mathbb{R}$ „Energiefunktion" (betr. $|E| < \infty$)

$\Gamma = \{ \nu \in \mathcal{M}_1(E) : \int \varphi \, d\nu = z \}$

für ein $z \in [\min(\varphi), \max(\varphi)]$,

So ist $\widetilde{\mathcal{M}} = \{ \nu \in \Gamma : H(\nu|\mu) \text{ ist minimal} \}$

$= \{ \mu_\beta \}$

mit $\mu_\beta(\{i\}) = \dfrac{e^{\beta\varphi(i)}}{Z_\beta} \mu(\{i\})$ mit $Z_\beta = \sum_{i=1}^{d} e^{\beta\varphi(i)} \mu(\{i\})$

und $\beta \, (= \beta(z))$ bestimmt durch $\int \varphi \, d\mu_\beta = z$.

Argument: (schreibe $\mu(\{i\}) =: \mu_i$, $\nu(\{i\}) =: \nu_i$)

Minimiere $F(\nu) = H(\nu|\mu) = \sum_{i=1}^{d} \nu_i (\log \nu_i - \log \mu_i)$

unter Nebenbed. $G_1(\nu) = \sum_{i=1}^{d} \nu_i = 1$

und $G_2(\nu) = \sum_{i=1}^{d} \nu_i \varphi(i) = z$.

Für ein minimierendes ν^* (existiert: $\mathcal{M}_1(E)$ ist kompakt, F u.u. halbstetig)

gibt es Lagrange-Multiplikatoren $\lambda_1, \lambda_2 \in \mathbb{R}$

mit $\Big(\underbrace{\frac{\partial}{\partial \nu_i} H(\nu|\mu)(\nu^*)}_{\log \nu_i^* - \log \mu_i + 1} \Big)_{i=1,\dots,d} = \lambda_1 \Big(\underbrace{\big(\frac{\partial}{\partial \nu_i} G_1\big)(\nu^*)}_{=(1,\dots,1)} \Big)_{i=1,\dots,d}$

$+ \lambda_2 \Big(\underbrace{\big(\frac{\partial}{\partial \nu_i} G_2\big)(\nu^*)}_{=(\varphi(1),\dots,\varphi(d))} \Big)_{i=1,\dots,d}$

also $\log \nu_i^* - \log \mu_i + 1 = \lambda_1 + \lambda_2 \varphi(i)$, $i = 1, \dots, d$,

$$\nu_i^* = \mu_i e^{\lambda_2 \varphi(i) + \lambda_1 - 1}.$$

Alternativ schreibe für $\nu \in \Gamma$

$$H(\nu|\mu) = \int \log \frac{d\nu}{d\mu}\, d\nu = \int \log \frac{d\nu}{d\mu_\beta} + \overbrace{\log \frac{d\mu_\beta}{d\mu}}^{= \beta\varphi(\cdot) - \log Z_\beta}\, d\nu$$

$$= \underbrace{\int \log \frac{d\nu}{d\mu_\beta}\, d\nu}_{= H(\nu|\mu_\beta),\ \geq 0 \text{ stets und } = 0 \text{ im Fall } \nu = \mu_\beta} + \beta \underbrace{\int \varphi\, d\nu}_{= z} - \log Z_\beta$$

Beob. 4

Kontraktionsprinzip (abstrakt)

$E, \tilde{E}$ metr. Räume, $f : E \to \tilde{E}$ stetig,

$(\mu_n)_n \subset \mathcal{M}_1(E)$ erfülle PGA mit Raten γ_n
und Ratenfunktion I.

Die Familie $\tilde{\mu}_n := \mu_n \circ f^{-1}$ ($\in \mathcal{M}_1(\tilde{E})$, die Bildmaße)
erfüllt PGA mit Raten γ_n und Ratenfunktion

$$\tilde{I}(y) = \inf\{ I(x) : x \in E, f(x) = y \}, \quad y \in \tilde{E}$$

(bzw. $\tilde{I}(y) = +\infty$ falls $f^{-1}(\{y\}) = \emptyset$)

Bew.
1) $\{\tilde{I} \leq s\} = \{ f(x) : x \in E, I(x) \leq s\}$

$= f(\{I \leq s\})$ ist kompakt

(Bild einer kompakten Menge unter stetiger Abb.)

2) $\tilde{G} \subset \tilde{E}$ offen, $\tilde{\mu}_n(\tilde{G}) = \mu_n(\underbrace{f^{-1}(\tilde{G})}_{\text{ist offen}})$

3) $\tilde{F} \subset \tilde{E}$ abgeschl., $\tilde{\mu}_n(\tilde{F}) = \mu_n(\underbrace{f^{-1}(\tilde{F})}_{\text{ist abg.}})$

Beh. folgt mit Bem. 2.3.

Bem. 5

Mittels des Kontraktionsprinzips kann man den Satz von Cramér aus dem Satz von Sanov „zurückgewinnen":

[Wir betrachten der Einfachheit halber hier nur den Fall beschränkter Summanden, i.A. benötigt man danach ein weiteres Approximationsargument.]

$X_1, X_2, \ldots$ u.i.v. beschränkte reelle ZVn, $\sim \mu \in \mathcal{M}_1([a,b])$
$(-\infty < a < b < \infty)$

$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ empir. Vert.,

$f : \mathcal{M}_1([a,b]) \to [a,b]$
$\nu \mapsto \int x\, \nu(dx)$ ist stetig (dies gilt i.A. nicht, wenn der Wertebereich unbeschränkt ist)

$\tilde{S}_n := \frac{1}{n} \sum_{i=1}^{n} X_i = f(L_n)$

erfüllt somit PGA mit Rate n und Ratenfunktion

$$\tilde{I}(y) = \inf \{ H(\nu | \mu) : \nu \in \mathcal{M}_1([a,b]),\ \int x\, \nu(dx) = y \},$$
$$y \in [a,b]$$

Das optimale $\nu^* (= \nu^*_y)$ ist geg. durch $\frac{d\nu^*}{d\mu}(x) = \frac{e^{t^* x}}{\int e^{t^* x} \mu(dx)}$ mit $t^* (= t_y)$ festgel. durch $\int x\, \nu^*(dx) = y$,

also
$$\tilde{I}(y) = H(\nu^* | \mu) = \int \log \frac{d\nu^*}{d\mu}\, d\nu^*$$
$$= \int t^* x\, \nu^*(dx) - \log \int e^{t^* x} \mu(dx) = t^* y - \log \int e^{t^* x} \mu(dx)$$
$$= \sup_t \{ t y - \log \int e^{tx} \mu(dx) \}$$
(vgl. Satz 1.1)

Anwendungsbeispiel: Fehlerwahrscheinlichkeiten bei (Neyman-Pearson) Tests

$\mu_0, \mu_1 \in \mathcal{M}_1(E)$, $\mu_0 \neq \mu_1$, $\mu_0 \ll \mu_1$ und $\mu_1 \ll \mu_0$

$X_1, X_2, \ldots, X_n$ u.i.v. („n Beobachtungen", Werte in E)

$T_n : E^n \to \{0,1\}$ ein „Test", Interpretation

$T_n(X_1, \ldots, X_n) = 0$, so nehme Hypothese H_0: „die Beob. entstammen Vert. μ_0" an,

$T_n(X_1, \ldots, X_n) = 1$, so lehne H_0 ab zugunsten H_1: „die Beob. entstammen Vert. μ_1"

$$\alpha_n = \mathbb{P}_{\mu_0}(\underset{\downarrow}{\overset{= T_n(X_1, \ldots, X_n)}{T_n}} = 1) \quad \text{(W'keit für) „Fehler 1. Art"}$$

$$\beta_n = \mathbb{P}_{\mu_1}(T_n = 0) \quad \text{(— " —) „Fehler 2. Art"}$$

$$Y_i := \log \frac{d\mu_1}{d\mu_0}(X_i) \quad \text{(„log-Likelihood-Quotient" [für } i\text{-te Beob.]))}$$

Neyman-Pearson-Test: (1933) [Jerzy Neyman, 1894–1981, Egon Pearson, 1895–1980]

$$T_n(X_1, \ldots, X_n) = \mathbb{1}_{\{\frac{1}{n}\sum_{i=1}^{n} Y_i > \gamma\}}$$

(für einen Schwellwert γ)

$e^{Y_1} \ge 1$ für $Y_1 \ge 0$,
$e^{Y_1} \in (0,1)$ für $Y_1 < 0$, Striktheit da Y_1 nicht μ_0-f.s. konstant

<u>Bem.</u>:

$$\gamma_0 := \underbrace{\mathbb{E}_{\mu_0}[Y_1]}_{= -H(\mu_0|\mu_1)\ (\le 0)} < \mathbb{E}_{\mu_0}\left[e^{Y_1} Y_1\right] = \underbrace{\mathbb{E}_{\mu_1}[Y_1]}_{= H(\mu_1|\mu_0)\ (\ge 0)} =: \gamma_1$$

d.h. sinnvolle Wahlen sind $\gamma \in (\gamma_0, \gamma_1)$.

<u>Bem./Bericht</u> (zur Optimalität von N-P-Tests)

N-P-Test, Schwellenwert γ

T_n' Test mit $\mathbb{P}_{\mu_0}(T_n'(X_1,\dots,X_n) = 1) =: \alpha_n' \le \mathbb{P}_{\mu_0}(T_n = 1) =: \alpha_n$,

so gilt $\beta_n' := \mathbb{P}_{\mu_1}(T_n' = 0) \ge \mathbb{P}_{\mu_1}(T_n = 0) =: \beta_n$

(denn $\beta_n' - \beta_n = \mathbb{E}_{\mu_1}[T_n - T_n']$

$$= \mathbb{E}_{\mu_1}\Big[\underbrace{(T_n - T_n')\frac{e^{Y_1+\dots+Y_n}}{e^{Y_1+\dots+Y_n}}}_{\ge (T_n - T_n')\frac{e^{\gamma n}}{e^{Y_1+\dots+Y_n}}}\Big] \ge e^{\gamma n}\underbrace{\mathbb{E}_{\mu_0}[T_n - T_n']}_{= \alpha_n - \alpha_n' \ge 0}\Big)$$

(falls $T_n = 1$ so ist $Y_1 + \dots + Y_n > \gamma n$,
falls $T_n = 0$ so ist $\frac{e^{\gamma n}}{e^{Y_1+\dots+Y_n}} > 1$
und $-T_n'$ ist u.U. negativ)

[bis hierher am 18.12.14]

Satz 6

Sei $\gamma \in (\gamma_0, \gamma_1)$, $\varphi_0(t) := \mathbb{E}_{\mu_0}[e^{tY_1}]$, $T_n = 1_{\{\frac{1}{n}\sum_{i=1}^n Y_i > \gamma\}}$

$\alpha_n = P_{\mu_0}(T_n = 1)$, $\beta_n = P_{\mu_1}(T_n = 0)$,

so ist

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n = -I_0(\gamma) < 0,$$
$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n = -(I_0(\gamma) - \gamma) < 0$$

mit $I_0(z) = \sup_{t\in\mathbb{R}} \{zt - \log \varphi_0(t)\}$

(Legendre-Transformierte von $\log \varphi_0$)

Bew.:

(Vorbem.: Für $t \in [0,1]$ ist

$$0 < \mathbb{E}_{\mu_0}[e^{tY_1}] \le \int_{\{\frac{d\mu_1}{d\mu_0} < 1\}} 1 \, d\mu_0 + \int_{\{\frac{d\mu_1}{d\mu_0} \ge 1\}} \underbrace{\left(\frac{d\mu_1}{d\mu_0}\right)^t}_{\le \frac{d\mu_1}{d\mu_0}} d\mu_0$$

$$\le \int 1 \, d\mu_0 + \int 1 \, d\mu_0 = 2,$$

d.h. φ_0 ist endl. in $[0,1]$)

$$\frac{1}{n} \log \alpha_n = \frac{1}{n} \log P_{\mu_0}\left(\frac{1}{n}\sum_{i=1}^n Y_i > \gamma\right) \longrightarrow I_0(\gamma)$$

(Satz v. Cramér, Satz 1.1)

[Durchsicht des Arguments zeigt, dass es hier genügt, dass φ_0 in einer „Rechts-Umgebung" der 0 endlich ist.]

(Beachte: $\gamma > \gamma_0 = \mathbb{E}_{\mu_0}[Y_1]$)

$$\varphi_1(t) := \mathbb{E}_{\mu_1}[e^{tY_1}] = \mathbb{E}_{\mu_0}[e^{Y_1} \cdot e^{tY_1}] = \varphi_0(t+1),$$

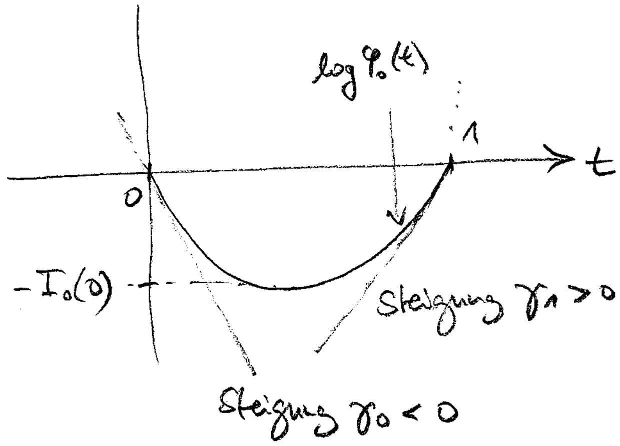
$$I_1(z) = \sup_{t \in \mathbb{R}} \{tz - \log \varphi_1(t)\} = -z + I_0(z)$$

analog $\frac{1}{n} \log \beta_n = \frac{1}{n} \log \mathbb{P}_{\mu_1}\left(\frac{1}{n} \sum_{i=1}^n Y_i \leq \gamma\right) \longrightarrow -I_1(\gamma)$

$= -(I_0(\gamma) - \gamma)$

(Satz v. Cramér,
beachte $\gamma < \gamma_1 = \mathbb{E}_{\mu_1}[Y_1]$)

$\lrcorner$



Bem.: Für die Wahl $\gamma = 0$ sind die Raten gleich, dies ist die beste Wahl, wenn $\max\{\alpha_n, \beta_n\}$ minimiert werden soll.

($I_0(0)$ heißt auch Chernoff-Information (von μ_0 und μ_1).)