Große Abweichungen bei moderat abhängigen Folgen
Das Gärtner-Ellis Theorem

Thomas Rupp*
Version 1.0

22. Dezember 2000

Man sagt: eine Folge (P_n) von Wahrscheinlichkeitsmaßen genügt dem Großen Abweichungsprinzip [1] mit der Rate n und Ratenfunktion I, wenn folgende Punkte erfüllt sind:

- Eigenschaften einer Ratenfunktion
  - $I \not\equiv \infty$ (D1)
  - I ist unterhalbstetig (D2)[2]
  - I besitzt kompakte Niveaumengen (D3)
- $\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -I(C) \quad \forall C \subset \mathcal{X}$ abgeschlossen
- $\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \geq -I(O) \quad \forall O \subset \mathcal{X}$ offen

Dabei ist $\mathcal{X}$ ein polnischer Raum und $I : \mathcal{X} \to [0, \infty]$.
Im Folgenden wollen wir klären, wie wir diese Ratenfunktion näher spezifizieren bzw. in speziellen Fällen berechnen können.

0.1 Voraussetzungen

Sei mit (Z_n) eine Folge von $\mathbb{R}^d$-wertigen, borelmessbaren Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{B}(\mathbb{R}^d), \mathbb{P})$; also in der mit einem Wahrscheinlichkeitsmaß $\mathbb{P}$ ausgestatteten Borel-Algebra über dem $\mathbb{R}^d$. Weiterhin sei (die Kumulantenerzeugende Funktion)

$$\varphi_n(t) := \mathbb{E}e^{\langle t, Z_n \rangle}, \quad t \in \mathbb{R}^d, n \in \mathbb{N} \tag{1}$$

und es gelte

$$\lim_{n\to\infty} \frac{1}{n} \log \varphi_n(nt) =: \Lambda(t) \in [-\infty, \infty] \text{ existiert} \tag{2}$$

$$0 \in \operatorname{int}(\mathcal{D}_\Lambda) \text{ wobei } \mathcal{D}_\Lambda := \{t \in \mathbb{R}^d : \Lambda(t) < \infty\} \tag{3}$$

Wir logarithmieren und teilen durch n, weil wir uns nur für die Geschwindigkeit der Abnahme der Wahrscheinlichkeiten interessieren (also den Exponenten ohne die Multiplikation mit n).
(3) bedeutet, dass $\Lambda(\cdot)$ um den Nullpunkt herum nur endliche Werte annimmt.

Im Folgenden sei mit $P_n(\cdot) := \mathbb{P}(Z_n \in \cdot)$ eine Famillie von Wahrscheinlichkeitsmaßen definiert und f^* bezeichne stets die Legendretransformierte von f; also $f^*(x) = \sup_t (\langle x, t \rangle - f(t))$ mit, wenn nicht explizit erwähnt, $t \in \mathbb{R}^d$.

*thomas@7t7.de

[1] LDP, *large deviation principle*, es gibt an, mit welcher Geschwindigkeit die Wahrscheinlichkeit für untypische Ereignisse abfällt

[2] wird durch (D3) impliziert

Lemma 0.1 *Λ hat die Eigenschaften*

1. *Λ ist konvex und $\Lambda > -\infty$*
2. *Λ^* ist eine Ratenfunktion und ist konvex*

Beweis

1. *Wenn $\log\varphi_n$ konvex ist, dann gilt dies auch für Λ. Wegen $\Lambda(0) = 0$ und da Λ konvex ist gilt $\Lambda > -\infty$.*

   Mit der Hölder'schen Ungleichung und $Y := e^{\lambda t_1 X}, Z := e^{(1-\lambda)t_2 X}$ folgt:

$$\begin{aligned}
\mathbb{E}(Y \cdot Z) &\leq (\mathbb{E}(Y^{\frac{1}{\lambda}}))^{\lambda}(\mathbb{E}(Z^{\frac{1}{1-\lambda}}))^{1-\lambda} \\
\Leftrightarrow \left[\mathbb{E}\left(e^{(\lambda_1 t_1 + (1-\lambda)t_2)X}\right)\right] &\leq \left[\mathbb{E}\left(e^{t_1 X}\right)\right]^{\lambda}\left[\mathbb{E}\left(e^{t_2 X}\right)\right]^{(1-\lambda)} \\
\Leftrightarrow \log(\varphi(\lambda t_1 + (1-\lambda)t_2) &\leq \lambda \log\varphi(t_1) + (1-\lambda)\log\varphi(t_2) \\
\Leftrightarrow & \quad \log\varphi \text{ konvex und dann auch } \varphi_n.
\end{aligned}$$

2. **D1** $\Lambda^* \not\equiv \infty$

   Da Λ konvex ist gibt es eine affin-lineare Minorante: $\langle x_0, t\rangle$. Und demnach $\Lambda(t) > \langle x_0, t\rangle$. Und da $\Lambda(0) = 0$ folgt $\Lambda^*(x_0) = \sup_t(\langle x_0, t\rangle - \Lambda(t)) = 0$.

   D2 Λ^* *ist unterhalbstetig*

   Die Legendretransformierte einer konvexen Funktion ist konvex und unterhalbstetig.

   D3 Λ^* *besitzt kompakte Niveaumengen: $\{x : \Lambda^*(x) \leq y\}$ kompakt $\forall y \in \mathbb{R}^d$*

   Vollständigkeit: Da Λ^* unterhalb stetig und $\mathbb{R}^d$ vollständig ist, sind sie vollständig, denn einzelne Punkte dürfen ja nur nach unten abweichen und liegen somit in der Niveaumenge.

   Beschränktheit: Nach (3) gibt es eine δ-Kugel um 0, so dass Λ dort ein endliches Maximum c annimmt. Dann folgt damit

$$\begin{aligned}
\Lambda^*(x) &\geq \sup_{t \in B_\delta(0)} (\langle x, t\rangle - \Lambda(t)) \\
&\geq \sup_{t \in B_\delta(0)} (\langle x, t\rangle - c) = \sup_{t \in B_\delta(0)} \|x\| \cdot \|t\| - c = \delta|x| - c.
\end{aligned}$$

   Also übertrifft $\Lambda^*(x)$ jedes geforderte y, wenn $|x|$ groß genug wird. Wegen $\leq y$ und Vollständigkeit sind sie abgeschlossen.

□

Bemerkung 0.2 *Für alle positiven Folgen a_n, b_n gilt:*

$$\lim_{n\to\infty} \frac{\log(a_n + b_n)}{n} = \lim_{n\to\infty} \frac{\log\max\{a_n, b_n\}}{n}$$

Definition 0.3 $\Lambda^*(S) = \inf_{x \in S} \Lambda^*(x), \quad S \subset \mathbb{R}^d$

Definition 0.4 (exponierter Punkt) *Ein Punkt $x \in \mathbb{R}^d$ heisst exponierter Punkt bezüglich der Funktion Λ^* gdw. ein Punkt $t \in \mathbb{R}^d$ existiert, so dass $\Lambda^*(y) - \Lambda^*(x) > \langle y - x, t\rangle \quad \forall y \neq x$.*

*Also wenn die Steigungstangente an x mit der Funktion nur genau einen Schnittpunkt (nämlich x) besitzt. Solch ein t heisst die **exponierende Hyperebene** zu x.*

Im weiteren bezeichnet E die Menge der exponierten Punkte bzgl. Λ^* mit den exponierenden Hyperebenen aus $\mathrm{int}(\mathcal{D}_\Lambda)$.

Definition 0.5 (relatives Inneres) *Wenn $A \subset \mathbb{R}^d$ nichtleer und konvex ist, dann ist das relative Innere von A definiert als*

$$rint(A) := \{x \in A : \forall y \in A \exists \varepsilon > 0 : x - \varepsilon(y - x) \in A\}\,.$$

Das relative Innere ist sozusagen das Innere aus der Menge selbst heraus betrachtet; unabhängig vom Raum in dem es eingebettet ist. Insbesondere ist $\mathrm{rint}(A) \supset \mathrm{int}(A)$ und $\mathrm{rint}(\{x\}) = \{x\}$.

1 Das Gärtner-Ellis Theorem

Satz 1.1 *Wenn (2) und (3) erfüllt sind, dann gilt*

1. $\limsup\limits_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\Lambda^*(C) \quad \forall C \in \mathbb{R}^d$ *abgeschlossen*
2. $\liminf\limits_{n\to\infty} \frac{1}{n} \log P_n(O) \geq -\Lambda^*(O \cap E) \quad \forall O \in \mathbb{R}^d$ *offen,*
3. *Wenn Λ zusätzlich noch die folgenden Eigenschaften besitzt,*
   - *(a) Λ ist auf $\mathbb{R}^d$ unterhalbstetig.*
   - *(b) Λ ist innerhalb von $int(\mathcal{D}_\Lambda)$ differenzierbar*
   - *(c) Entweder $int(\mathcal{D}_\Lambda) = \mathbb{R}^d$ oder Λ ist steil; d.h. die Steigung am Rand von $int(\mathcal{D}_\Lambda)$ ist unendlich: $\lim_{t\to\partial\mathcal{D}_\Lambda : t\in\mathcal{D}_\Lambda} |\nabla\Lambda(t)| = \infty$.*

   dann können wir $O \cap E$ durch O ersetzen und (P_n) genügt dem LDP mit Rate n und Ratenfunktion Λ^*.

Beweis *Der Beweis des Gärtner-Ellis-Theorems gliedert sich in 3 Teile:*

1 Obere Schranke

Zuerst werden wir die Schranke für kompakte Mengen beweisen und dann auf abgeschlossene ausgedehnen. Das wir mit der oberen Schranke anfangen ist kein Zufall, wie wir später feststellen werden.
Sei dazu $\delta > 0$ und $\Lambda^*_\delta := \min\{\Lambda^*(x) - \delta, \frac{1}{\delta}\}$ für $x \in \mathbb{R}^d$; also eine durch $\frac{1}{\delta}$ nach oben beschränkte, sich an Λ^* annähernde Minorante von Λ^*.
Da $\Lambda^*_\delta(x) \leq \Lambda^*(x) = \sup\left(\langle x, t\rangle - \Lambda(t)\right)$ gibt es ein $t_x \in \mathbb{R}^d$, so dass

$$\langle x, t_x\rangle - \Lambda(t_x) \geq \Lambda^*_\delta(x). \tag{4}$$

Sei A_x eine beliebige Umgebung um x, so dass

$$\inf_{y\in A_x} \langle y, t_x\rangle \geq \langle x, t_x\rangle - \delta \Leftrightarrow \inf_{y\in A_x} \langle y - x, t_x\rangle \geq -\delta. \tag{5}$$

erfüllt ist. Aus der exponentiellen Chebyshev-Ungleichung

$$\mathbb{P}(X \geq \varepsilon) = \mathbb{P}(e^{nX} \geq e^{n\varepsilon}) \leq \mathbb{E}(e^{n(X-\varepsilon)})$$

folgt mit (5)

$$\begin{aligned} P_n(A_x) &= \mathbb{P}(Z_n \in A_x) \leq \mathbb{P}\left(\langle Z_n - x, t_x\rangle \geq -\delta\right) \\ &\leq \mathbb{E}\left(e^{n\langle Z_n - x, t_x\rangle + n\delta}\right) = e^{\delta n}\mathbb{E}\left(e^{n\langle Z_n - x, t_x\rangle}\right) = e^{\delta n} e^{-n\langle x, t_x\rangle}\mathbb{E}\left(e^{n\langle t_x, Z_n\rangle}\right) \end{aligned}$$

Schreiben wir das mit Hilfe von (1) um, so erhalten wirn schliesslich die Ungleichung

$$P_n(A_x) \leq e^{\delta n} e^{-n\langle x, t_x \rangle} \varphi_n(n t_x) \tag{6}$$

Wir kommen nun zur eigentlichen Abschätzung für eine beliebige kompakte Menge $K \subset \mathbb{R}^d$*. Da* K *kompakt ist, besitzt* K *eine endliche Teilüberdeckung* $\bigcup_{i=1,\ldots,N} A_{x_i}$*. Damit kommen wir mit Hilfe von (6) für jedes* n *zu folgender Abschätzung:*

$$\begin{aligned}
\frac{1}{n}\log P_n(K) &\leq \frac{1}{n}\log\left[N \max_{i=1,\ldots,N} P_n(A_{x_i})\right] \\
&= \frac{1}{n}\log N + \frac{1}{n}\log e^{\delta n} + \max_{i=1,\ldots,N}\left(\frac{1}{n}\log e^{-n\langle x, t_{x_i}\rangle} + \frac{1}{n}\log \varphi_n(n t_{x_i})\right) \\
&= \frac{1}{n}\log N + \delta - \min_{i=1,\ldots,N}\left(\langle x, t_{x_i}\rangle - \frac{1}{n}\log \varphi_n(n t_{x_i})\right)
\end{aligned}$$

Da N *konstant ist folgt mit* $n \to \infty$*, (2), (4) und Definition 0.3*

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n}\log P_n(K) &\overset{2}{\leq} \delta - \min_{i=1,\ldots,N}\left(\langle x, t_{x_i}\rangle - \Lambda(t_{x_i})\right) \\
&\overset{4}{\leq} \delta - \min_{i=1,\ldots,N} \Lambda^*_\delta(x_i) \\
&\overset{Def.0.3}{\leq} \delta - \Lambda^*_\delta(K)
\end{aligned}$$

Lassen wir nun $\delta \downarrow 0$ *gehen, erhalten wir wegen* $\Lambda^*_\delta(x) \to \min\{\Lambda^*(x) - 0, \infty\}$ *die gewünschte Abschätzung für kompakte Teilmengen:*

$$\limsup_{n\to\infty} \frac{1}{n}\log P_n(K) \leq -\Lambda^*(K)$$

Sei C *eine abgeschlossene Teilmenge des* $\mathbb{R}^d$*. Dann ist* $C \cap [-N, N]^d$ *kompakt: Vollständigkeit wird durch den* $\mathbb{R}^d$ *gewährleistet, Beschränkheit durch den Schnitt. Wenn wir dann die Ungleichung für den Grenzwert* $N \to \infty$ *zeigen sind wir fertig, denn wir wissen, dass* $\lim_{N\to\infty}\left(C \cap [-N, N]^d\right) = C$*.*
Sei $\hat{N} := [-N, N]^d$*. Mit der obigen Ungleichung, der Tatsache, dass* $(C \cap \hat{N})$ *kompakt ist und somit Satz 1.1.1 schon anwendbar,* $(C \cap \hat{N}), (\mathbb{R}^d \backslash \hat{N})$ *disjunkt sind und zusammen* C *enthalten, folgt mit Bemerkung 0.2:*

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n}\log P_n(C) &\leq \limsup_{n\to\infty} \frac{1}{n}\log P_n\left[(C \cap \hat{N}) \cup (\mathbb{R}^d \backslash \hat{N})\right] \\
&= \limsup_{n\to\infty} \frac{1}{n}\log\left(P_n[(C \cap \hat{N})] + P_n[(\mathbb{R}^d \backslash \hat{N})]\right) \\
&\overset{Bem.0.2}{=} \limsup_{n\to\infty} \frac{1}{n}\log\max\{P_n[(C \cap \hat{N})], P_n[(\mathbb{R}^d \backslash \hat{N})]\} \\
&= \max\{-\Lambda^*(C \cap \hat{N}), -M_N\} \qquad \forall N > 0
\end{aligned}$$

Dabei ist $-M_N := \limsup_{n\to\infty} \frac{1}{n}\log P_n(\mathbb{R}^d \backslash [-N, N]^d)$*.*
Nun zeigen wir, dass $\lim_{N\to\infty} -M_N = -\infty$ *ist und somit Satz 1.1.1 bewiesen ist:*
Seien $e^{(1)}, \ldots, e^{(d)}$ *die Einheitsvektoren des* $\mathbb{R}^d$*. Wegen (3) gibt es einen Vektor nahe der Null, für den gilt:* $\exists \delta > 0 : \forall i : \Lambda(\pm\delta e^{(i)}) < \infty$*. Mit der exponentiellen Chebyshev-Ungleichung und (1) folgt dann für alle* i

$$\begin{aligned}
\mathbb{P}(Z_n^{(i)} \leq -N) = \mathbb{P}\left(e^{n\delta(-Z_n^{(i)})} \geq e^{n\delta N}\right) &\leq \mathbb{E}\left(e^{-n\delta Z_n^{(i)} - n\delta N}\right) = e^{-n\delta N}\varphi_n(-n\delta e^{(i)}) \\
\mathbb{P}(Z_n^{(i)} \geq N) = \mathbb{P}\left(e^{n\delta Z_n^{(i)}} \geq e^{n\delta N}\right) &\leq \mathbb{E}\left(e^{n\delta Z_n^{(i)} - n\delta N}\right) = e^{-n\delta N}\varphi_n(n\delta e^{(i)})
\end{aligned}$$

Dadurch lässt sich nun abschätzen, wie wahrscheinlich es ist, dass eine Komponente der Zufallsvariable nicht in das Intervall $[-N,N]$ fällt. $-M_N$ beschreibt das Verhalten, dass die ganze Zufallsvariable ausserhalb liegt; diese ist sicherlich kleiner, als wenn ≥ 1 Komponenten ausserhalb liegen (also ist auch der Logarithmus kleiner):

$$\begin{aligned}
-M_N &\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\exists i \in \{1,\ldots,d\} : Z_n^{(i)} \notin [-N,N]\right) \\
&\leq \max_{i=1,\ldots,d} \limsup_{n\to\infty} \frac{1}{n} \log P_n\left(\max\{e^{-n\delta N}\varphi_n(-n\delta e^{(i)}), e^{-n\delta N}\varphi_n(n\delta e^{(i)})\}\right) \\
&= -\delta N + \max_{i=1,\ldots,d} \max\{\Lambda(-\delta e^{(i)}), \Lambda(\delta e^{(i)})\}
\end{aligned}$$

Bei der mittleren Zeile wurde Bemerkung 0.2 benutzt. Da $\Lambda(\pm\delta e^{(i)}) < \infty)$ vorausgesetzt wurde folgt: $\lim_{N\to\infty} -M_N < -\infty$, womit die obere Schranke gezeigt wäre.

2 Untere Schranke

Für jede offene Teilmenge O, jedes $x \in O \cap E$ und alle $\varepsilon \leq \varepsilon_0(x)$ gilt

$$P_n(O) \geq P_n(B_\varepsilon(x))$$

Dabei ist ε_0 so klein, dass $O \supset B_{\varepsilon_0}(x)$.
Um Satz 1.12 zu beweisen, reicht es aus zu zeigen, dass

$$\lim_{\varepsilon\downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log P_n(B_\varepsilon(x)) \geq -\Lambda^*(x) \quad \forall x \in E. \tag{7}$$

Denn dann gilt die Ungleichung auch für alle $x \in O \cap E$.
Wir fixieren dazu $x \in E$. Dann gibt es dazu eine exponierende Hyperebene $\tau \in int(\mathcal{D}_\Lambda)$ zu x. Und aus $\Lambda(\tau) = \lim \frac{1}{n} \log \varphi_n(n\tau) < \infty$ folgt $\lim \varphi_n(n\tau) < \infty$ und somit ist $\varphi_n(n\tau)$ für genügend großes n endlich und wir können folgendermaßen ein gekipptes Wahrscheinlichkeitsmaß[3] *definieren: ($\hat{P}_n$ hat die Dichte $\frac{e^{n\langle y,\tau\rangle}}{\varphi_n(n\tau)}$ bezüglich P_n)*

$$\hat{P}_n(A) := \int \frac{1_A}{\varphi_n(n\tau)} e^{n\langle y,\tau\rangle} P_n(dy) \quad \forall A \in \mathcal{B}(\mathbb{R}^d), y \in \mathbb{R}^d$$

Umgeschrieben bedeutet dies: $P_n(dy) = \varphi_n(n\tau)e^{-n\langle y,\tau\rangle}\hat{P}_n(dy)$. Dadurch können wir nun geschickt eine Abschätzung nach unten herleiten:

$$\begin{aligned}
\frac{1}{n} \log P_n(B_\varepsilon(x)) &= \frac{1}{n} \log \int\limits_{B_\varepsilon(x)} P_n(dy) = \frac{1}{n} \log \int\limits_{B_\varepsilon(x)} \varphi_n(n\tau)e^{-n\langle y,\tau\rangle}\hat{P}_n(dy) \\
&= \frac{1}{n} \log \varphi_n(n\tau) + \frac{1}{n} \log \int\limits_{B_\varepsilon(x)} e^{-n\langle y,\tau\rangle}\hat{P}_n(dy) \\
&\overset{4}{\geq} \frac{1}{n} \log \varphi_n(n\tau) + \frac{1}{n} \log[e^{-n(\langle x,\tau\rangle + \varepsilon|\tau|)} \int\limits_{B_\varepsilon(x)} \hat{P}_n(dy)] \\
&= \frac{1}{n} \log \varphi_n(n\tau) - \langle x,\tau\rangle - \varepsilon|\tau| + \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x))
\end{aligned}$$

---

[3] wir "kippen" das Wahrscheinlichkeitsmaß so, dass die derzeit sehr unwahrscheinlichen Ereignisse zu typischen Ereignissen werden, die wir besser handhaben können

[4] wegen $\langle y,\tau\rangle = \langle x,\tau\rangle + \langle y-x,\tau\rangle \leq \langle x,\tau\rangle + \varepsilon|\tau|$ weil $y \in B_\varepsilon(x)$

Zusammengefasst ergibt sich also die folgende Ungleichung:

$$\lim_{\varepsilon \downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log P_n(B_\varepsilon(x)) \geq [\Lambda(\tau) - \langle x, \tau\rangle] + \lim_{\varepsilon \downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x))$$

Wegen $\Lambda^*(x) \geq [\langle x,\tau\rangle - \Lambda(\tau)]$ ist $[\Lambda(\tau) - \langle x,\tau\rangle] \geq -\Lambda^*(x)$ müssen wir noch, um (7) zu beweisen, zeigen, dass

$$\lim_{\varepsilon \downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x)) = 0 \text{ ist.} \tag{8}$$

Bilden wir nun die Kumulantenerzeugende Funktion $\hat{\varphi}_n$ von $\hat{P}_n$:

$$\hat{\varphi}_n(nt) = \int \left(e^{\langle nt, y\rangle} \cdot \frac{e^{n\langle y,\tau\rangle}}{\varphi_n(n\tau)} P_n(dy) \right) = \frac{\int \left(e^{n\langle t+\tau, y\rangle} P_n(dy) \right)}{\varphi_n(n\tau)} = \frac{\varphi_n(n(t+\tau))}{\varphi_n(n\tau)}$$

Da $\lim\limits_{n\to\infty} \frac{1}{n} \log \hat{\varphi}_n(nt) = \Lambda(t+\tau) - \Lambda(\tau) =: \hat{\Lambda}(t)$ und da nach Konstruktion $\Lambda(\tau) < \infty$ und nach Voraussetzung $\Lambda > -\infty$ erfüllt $\hat{\varphi}_n$ die Bedingung (2) und $\hat{\Lambda}(t)$ die Bedingung (3), da $\hat{\Lambda}(0) = 0$ und $\Lambda(x)$ konvex ist. Dadurch erhalten wir die Beziehung

$$\begin{aligned}
\hat{\Lambda}^*(x) \quad &:= \quad \sup_{t\in\mathbb{R}^d} [\langle x,t\rangle - \hat{\Lambda}(t)] = \sup_{t\in\mathbb{R}^d} [\langle x,t\rangle - \Lambda(t+\tau) + \Lambda(\tau)] \\
&\overset{t':=t+\tau}{=} \quad \sup_{t'\in\mathbb{R}^d} [\langle x, t'-\tau\rangle - \Lambda(t')] + \Lambda(\tau) = \sup_{t'\in\mathbb{R}^d} [\langle x,t'\rangle - \Lambda(t')] - \langle x,\tau\rangle + \Lambda(\tau) \\
&= \quad \Lambda^*(x) - \langle x,\tau\rangle + \Lambda(\tau)
\end{aligned} \tag{9}$$

Lemma 0.1 sagt uns, dass $\hat{\Lambda}^*$ eine Ratenfunktion ist. Also können wir bereits 1.11 auf $(\hat{P}_n)$ anwenden und erhalten folgende Abschätzung nach oben:

$$\limsup_{n\to\infty} \frac{1}{n} \log \hat{P}_n(\mathbb{R}^d \backslash B_\varepsilon(x)) \leq -\hat{\Lambda}^*(\mathbb{R}^d \backslash B_\varepsilon(x)) \qquad \forall \varepsilon > 0$$

Da $\hat{\Lambda}^*$ kompakte Niveaumengen besitzt und auf diesen ihre Minima annimmt (wegen der Unterhalbstetigkeit), gilt mit Definition 0.3:

$$\hat{\Lambda}^*(\mathbb{R}^d \backslash B_\varepsilon(x)) = \hat{\Lambda}^*(x_0) \text{ für ein } x_0 \neq x.$$

Wir wissen aber auch, dass x ein exponierter Punkt und $\tau \in int(\mathcal{D}_\Lambda)$ eine exponierende Hyperebene für x ist. Also können wir Definition 0.4 mit dem obigen x_0 auf (9) anwenden:

$$\begin{aligned}
\hat{\Lambda}^*(x_0) \quad &\overset{(9)}{=} \quad \Lambda^*(x_0) - \langle x_0,\tau\rangle + \Lambda(\tau) \\
&\overset{5}{\geq} \quad [\Lambda^*(x_0) - \langle x_0,\tau\rangle] + \langle x,\tau\rangle - \Lambda^*(x) \\
&\overset{Def.\ 0.4^6}{>} \quad 0
\end{aligned}$$

Damit wissen wir, dass

$$\limsup_{n\to\infty} \frac{1}{n} \log \hat{P}_n(\mathbb{R}^d \backslash B_\varepsilon(x)) \leq -\hat{\Lambda}^*(\mathbb{R}^d \backslash B_\varepsilon(x)) = -\hat{\Lambda}^*(x_0) < 0 \qquad \forall \varepsilon > 0$$

und da, unabhängig von n, $\hat{P}_n(\mathbb{R}^d) = 1$ ist, impliziert dies (8). Denn angenommen für ein festes ε würde gelten: $\liminf_{n\to\infty} \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x)) = -c < 0$ mit $c > 0$. Dann gäbe

[5] da $\Lambda^*(x) \geq \langle x,\tau\rangle - \Lambda(\tau) \Leftrightarrow \Lambda(\tau) \geq \langle x,\tau\rangle - \Lambda^*(x)$

[6] denn $\Lambda^*(x_0) - \Lambda^*(x) > \langle x_0 - x, \tau\rangle$ für alle $x_0 \in \mathbb{R}^d, x_0 \neq x$

es ein N, so dass für alle $n > N$ gilt: $\hat{P}_n(B_\varepsilon(x)) \leq e^{-\frac{nc}{2}} \downarrow 0$.Wir wir aber gerade gesehen haben, gilt genauso $\hat{P}_n(\mathbb{R}^d \backslash B_\varepsilon(x)) \leq e^{-\frac{n\hat{\Lambda}^*(x_0)}{2}} \downarrow 0$. Und da $1 = \hat{P}_n(\mathbb{R}^d) = \hat{P}_n(B_\varepsilon(x)) + \hat{P}_n(\mathbb{R}^d \backslash B_\varepsilon(x)) \downarrow 0$ haben wir den gewünschten Widerspruch.

3 Schritt von $O \cap E$ nach O

Nun werden wir sehen, dass wenn Λ die schärferen Bedingungen erfüllt, die Abschätzung nach unten für alle offenen Mengen gilt.

Lemma 1.2 *Wenn Λ den Bedingungen 1.13a - 1.13c genügt, dann ist $E \supset rint(\mathcal{D}_{\Lambda^*})$.*

Zum Beweis siehe hierzu **Convex Analysis** *von* R.T. Rockafellar, *Princeton University Press, 1970 das Corollary 26.4.1, Seite 257.*
Hilfreich zu wissen ist, dass 'closedness' äquivalent zur Unterhalbstetigkeit (1.13a) ist und die Funktion wegen 1.13b, 1.13c und Lemma 0.1 'proper essentially smooth' ist.

Bemerkung 1.3 *Aus dem Lemma 1.2 folgen die Eigenschaften*

1. *Nach Lemma 0.1 ist Λ^* konvex und somit ist $\mathcal{D}_{\Lambda^*}$ eine konvexe Menge.*
2. *Nach Lemma 0.1 ist Λ^* eine Ratenfunktion und somit $\not\equiv \infty$; also ist $\mathcal{D}_{\Lambda^*} \neq \emptyset$ und somit $rint(\mathcal{D}_{\Lambda^*}) \neq \emptyset$.*

Wir zeigen nun, dass unter den gegebenen Bedingungen $\Lambda^*(O \cap E) = \Lambda^*(O)$ ist. Wegen Definition 0.3 ist $\Lambda^*(O \cap rint(\mathcal{D}_{\Lambda^*})) \geq \Lambda^*(O \cap E) \geq \Lambda^*(O)$ und wegen Lemma 1.2 reicht es aus zu zeigen, dass

$$\Lambda^*(O \cap rint(\mathcal{D}_{\Lambda^*})) \leq \Lambda^*(O) \text{ gilt.}$$

Da $O \cap \mathcal{D}_{\Lambda^*} = \emptyset \Rightarrow \Lambda^*(O) = \infty$ (weil $\mathcal{D}_{\Lambda^*}$ alle Punkte mit endlichen Bildern enthält), können wir davon ausgehen, dass $O \cap \mathcal{D}_{\Lambda^*} \neq \emptyset$. Sei $y \in O \cap \mathcal{D}_{\Lambda^*}$ und $z \in rint(\mathcal{D}_{\Lambda^*})$ (nach Bemerkung 1.3 existiert so ein z).
Entweder besteht $\mathcal{D}_{\Lambda^*}$ aus genau einem Punkt und somit $y = z$ oder aus unendlich vielen Punkten. In jedem Fall liegt die Strecke $\overline{yz}$ in $\mathcal{D}_{\Lambda^*}$ (da es sich nach Bemerkung 1.3 um eine konvexe Menge handelt) und somit gibt es ein $\delta > 0$, so dass $\delta z + (1-\delta)y \in O \cap rint(\mathcal{D}_{\Lambda^*})$ (entweder ist $y = z$ oder es gibt unendlich viele Punkte, die nahe an y sowohl in O als auch in $rint(\mathcal{D}_{\Lambda^*})$ liegen).
Also folgt mit Definition 0.3, dass

$$\begin{aligned}
\Lambda^*(O \cap rint(\mathcal{D}_{\Lambda^*})) &\overset{\textit{Def. } 0.3}{\leq} \Lambda^*(\delta z + (1-\delta)y) \\
&\leq \lim_{\delta \downarrow 0} \Lambda^*(\delta z + (1-\delta)y) \\
&\overset{\Lambda^* \textit{konvex}}{\leq} \lim_{\delta \downarrow 0} [\delta \Lambda^*(z) + (1-\delta)\Lambda^*(y)] = \Lambda^*(y)
\end{aligned}$$

In der letzten Zeile wird die Gleichheit durch die Stetigkeit von Λ^* auf $rint(\mathcal{D}_{\Lambda^*})$ gerechtfertigt.
Da die Ungleichung für alle $y \in O \cap \mathcal{D}_{\Lambda^*}$ erfüllt ist, gilt sie insbesondere auch für das Infimum und somit haben wir

$$\Lambda^*(O \cap rint(\mathcal{D}_{\Lambda^*})) \leq \Lambda^*(O \cap \mathcal{D}_{\Lambda^*}) \leq \Lambda^*(O)^7$$

Somit wäre das Theorem bewiesen.

□

[7] denn für alle $y \in O \backslash (O \cap \mathcal{D}_{\Lambda^*})$ gilt: $\Lambda^*(y) = \infty$

Um das Theorem noch zu komplettieren:

Satz 1.4 *Wenn eine Folge von Wahrscheinlichkeitsmaßen* (P_n) *dem LDP genügt, dann ist die dazugehörige Ratenfunktion eindeutig.*

Beweis *Seien* I *und* J *zwei zu* (P_n) *gehörige Ratenfunktionen. Sei* $x \in \mathcal{X}$ *beliebig und* $B_N := B_{\frac{1}{N}}(x), N \in \mathbb{N}$ *eine Folge von offenen Kugeln um* x*. Dann gilt*

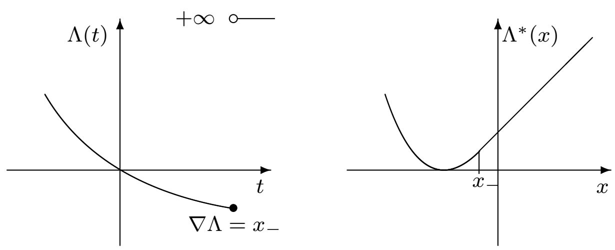
$$\begin{aligned} -I(x) &\leq -I(B_{N+1}) \leq \liminf_{n\to\infty} \frac{1}{n} \log P_n(B_{N+1}) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P_n(cl(B_{N+1})) \leq -J(cl(B_{N+1})) \leq -J(B_N) \end{aligned}$$

Die erste und letzte Ungleichung stimmt wegen Definition 0.3, die zweite und vierte aus der Definition der Ratenfunktion und die dritte ist offensichtlich wahr. Lassen wir nun $N \to \infty$ *gehen, dann folgt wegen der Unterhalbstetigkeit* $(\lim_{\varepsilon\downarrow 0} \inf_{y\in B_\varepsilon(x)} f(y) = f(x))$ *der Ratenfunktion, dass* $-J(B_N) \to -J(x)$*. Da* I *und* J *austauschbar sind, müssen sie gleich sein.*

$\square$

1.1 Bedingungen

Das im Schritt 3 des Beweises die Forderungen 1.13a-1.13c notwendig sind, sieht man daran, dass Λ an einer Knickstelle (Verletzung von 1.13b) der Gradient nicht eindeutig ist und die Legendretransformierte für diese ('dazwischen' liegenden) Gradienten linear (also nicht strikt konvex) ist. Diese Punkte (Gradienten) sind dann bezüglich Λ^* nicht exponiert.
Bei der Verletzung der Steilheit (1.13c) verläuft das Supremum $\sup(\langle x,t\rangle - \Lambda(t))$ ab x_- linear $(\langle x, x_-\rangle + |\Lambda(x_-)|)$, da immer $t \in \partial\mathcal{D}_\Lambda$ gilt, wie wir anhand von Bild 1.1 besser sehen können.



[Bild 1.1]

Natürlich stellt sich auch noch die Frage, was unter einer moderat abhängigen Folge von Zufallsvariablen zu verstehen ist. Dies hängt zum einen davon ab, ob die Eigenschaften (2) und (3) erfüllt sind, zum anderen davon, wie hilfreich die Grenzen sind, die das Gärtner-Ellis-Theorem liefert; sind diese zu grob oder sogar trivial, dann hilft uns diese Wahrheit auch nicht wirklich weiter.

2 Rückschau

In ein paar spezifischen Fällen können wir die Ratenfunktion (die Legendretransformierte von der herunterskalierten Kumulantenerzeugenden) mehr oder weniger explizit hinschreiben. Wir beschäftigen hier uns kurz mit unabhängig, identisch verteilten Zufallsvariablen und Markovketten.

2.1 Unabhängig, identisch verteilte, reelle Zufallsvariablen

Satz 2.1 (Cramérs Theorem) *Sei $(X_i)_i$ eine Folge von unbahängig, identisch verteilten Zufallsvariablen mit den unten stehenden Bedingungen 2.1.1 oder 2.1.2. Mit $S_n := \sum_{i=1}^n X_i$ gilt dann für alle $a > \mathbb{E}X_1$*[8]*:*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a) = -\sup_{t\in\mathbb{R}}(\langle a,t\rangle - \log\varphi(t)) = -\Lambda^*(a).$$

Sei nun $Z_n := \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Dann ist für alle $t \in \mathbb{R}$

$$\varphi_n(nt) = \mathbb{E}(e^{ntZ_n}) = \mathbb{E}(\prod_{i=1}^n e^{tX_i}) \overset{9}{=} (\mathbb{E}(e^{tX_1}))^n = \varphi(t)^n$$

Dabei ist $\varphi = \varphi_1$ die Kumulantenerzeugende Funktion von X_1. Also ist

$$\Lambda(t) = \lim_{n\to\infty} \frac{1}{n}\log\varphi_n(nt) = \log\varphi(t)^{10}$$

Je nachdem, welche Anforderungen wir stellen, können wir das Gärtner-Ellis Theorem in vollem Umfang anwenden:

2.1.1 $\varphi(t) < \infty \qquad \forall t \in \mathbb{R}$

Wenn wir obiges voraussetzen, dann nimmt Λ nur endliche Werte an und ist konvex. Offensichtlich sind dann die Bedingungen (2) (der Limes existiert wegen der Konvexität) und (3) erfüllt. Dann ist auch Λ auf ganz $\mathbb{R}$ (unterhalb) stetig (1.13a) und überall endlich (1.13c). Bedingung (1.13b) ist erfüllt, weil wir den Differentialoperator in das Integral ziehen können[11] und die Ableitung endlich ist.

2.1.2 $0 \in \mathbf{int}\mathcal{D}_\varphi$ und $\lim_{t\to\partial\mathcal{D}_\varphi : t\in\mathcal{D}_\varphi}(\log\varphi'(t)) = \infty$

Mit diesen Bedingungen sind natürlich (3) (da $0 \in \mathrm{int}\mathcal{D}_\varphi \Rightarrow 0 \in \mathrm{int}\mathcal{D}_{\log\varphi}$) und (1.13c) erfüllt und aus letzterem und der Konvexität folgt (1.13a). (2) ist erfüllt, weil $\Lambda^*(t) = \log\varphi(t)$ existiert. Wie bei 2.1.1 ist auch hier (1.13b) erfüllt.

Wenn wir 2.1.1 oder 2.1.2 voraussetzen, können wir nun wie gesagt das Theorem anwenden. Mit $a > \mathbb{E}(X_i)$ und $C := [a,\infty], O := (a,\infty)$ besagt das Theorem, dass

$$\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}(Z_n \in C) \leq -\Lambda^*(C) \text{ und } \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(Z_n \in O) \geq -\Lambda^*(O)$$

$$\begin{aligned} \text{da} \quad & \lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(Z_n \in C) = \lim_{n\to\infty}\frac{1}{n}\log\left(\mathbb{P}(Z_n \in O) + \mathbb{P}(Z_n \in \{a,\infty\})\right) \\ \Rightarrow \quad & -\Lambda^*(O) = -\Lambda^*(C) \end{aligned}$$

Da Λ^* konvex und um a herum stetig ist, sowie bei $\mathbb{E}(X_1)$ ihr Minimum hat[12], nimmt $\Lambda^*([a,\infty])$ das Minimum bei a an.
Damit wäre Cramérs Theorem als Spezialfall des Gärtner-Ellis Theorems gezeigt.

[8] wegen der Unabhängigkeit geht es aus Symmetriegründen analog mit $a < \mathbb{E}(X_1)$

[9] da die X_i identisch verteilt und unabhängig sind

[10] dabei ist aber immernoch $\Lambda(0) = 0$

[11] $\frac{1}{h}(\varphi(t+h) - \varphi(t)) = \int \frac{1}{h}\left(e^{x(t+h)} - e^{xt}\right)\mathbb{P}(dx) \to \int xe^{tx}\mathbb{P}(dx)$ da der Integrand endlich und durch die Steigungstangente von φ beschränkt ist, sowie x im Inneren des Endlichkeitsbereiches liegt

[12] folgt wieder aus dem Symmetrieargument und der Unabhängigkeit

2.2 Überleitung zu Sanovs Theorem

Seien (X_i) unabhängig, identisch verteilte Zufallsvariablen mit dem endlichen Wertebereich $\Gamma := \{1, \ldots, r\} \subset \mathbb{N}$. Die Verteilungsgewichte ρ_s seien alle strikt positiv. Sei $Z_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ mit $\delta_{X_i} := (0, \ldots, 1_{X_i}, 0, \ldots, 0) \in \{0,1\}^r$. Dann gilt für alle $t \in \mathbb{R}^r$

$$\varphi_n(nt) = \mathbb{E}\left(e^{\langle nt, Z_n\rangle}\right) = \mathbb{E}\left(e^{\left\langle nt, \frac{1}{n}\sum_{i=1}^n \delta_{X_i}\right\rangle}\right) \overset{13}{=} \left(\mathbb{E}(e^{\left\langle t, \delta_{X_1}\right\rangle})\right)^n = \left(\sum_{i=1}^n \rho_i e^{t_i}\right)^n$$

Demnach is $\Lambda(t) = \lim \frac{1}{n}\log(\sum \rho_s e^{t_s})^n = \log(\sum \rho_s e^{t_s})$. Also sind wegen $\Lambda(t) < \infty$ und der Differenzierbarkeit der Funktion die Voraussetzungen (2),(3) und (1.13a) bis (1.13c) erfüllt.
Dann ist die Legendretransformierte für alle $\nu \in \{m \in [0,1]^r \mid \sum_s m_s = 1\} =: \mathcal{M}(\Gamma)$ (die Menge aller Wahrscheinlichkeitsmaße auf Γ)

$$\Lambda^*(\nu) = \sup_{t\in\mathbb{R}^r}\left(\sum_s \nu_s t_s - \log\left(\sum_s \rho_s e^{t_s}\right)\right)$$

Da $\rho_s > 0 \quad \forall s$, ist $\Lambda(t)$ strikt konvex und wir können das Supremum einfach ausrechnen: der Gradient in jeder Komponente soll Null sein, also muss für jedes s gelten:

$$\frac{\partial\left(\sum_i \nu_i t_i - \log\left(\sum_s \rho_s e^{t_s}\right)\right)}{\partial t_s} = 0 \Leftrightarrow \nu_s = \frac{\rho_s e^{t_s}}{\sum_u \rho_u e^{t_u}}$$

Also ist $t_s = \log\left(\frac{\nu_s}{\rho_s}\right) + C$ mit $C := \log(\sum_u \rho_u e^{t_u})$. Somit ergibt sich

$$\begin{aligned}
\Lambda^*(\nu) &= \left(\sum_s \nu_s(\log\left(\frac{\nu_s}{\rho_s}\right) + C) - \log\left(\sum_s \rho_s e^{\log\left(\frac{\nu_s}{\rho_s}\right)+C}\right)\right) \\
&= \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) + \sum_s \nu_s C - \log\left(\sum_s \rho_s \frac{\nu_s}{\rho_s} e^C\right) \\
&\overset{14}{=} \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) + C - \log(e^C) = \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right)
\end{aligned}$$

Sei $a > 0$ und $d(\nu, \rho) := \frac{1}{2}\sum_s |\nu_s - \rho_s|$, sowie zur Abkürzung $C := \{\mu \in \mathcal{M}(\Gamma) | d(\mu, \rho) \geq a\}$ und $O := \{\nu \in \mathcal{M}(\Gamma) | d(\nu, \rho) > a\}$.

Sanovs Theorem besagt, mit den obigen Bezeichnungen, dass

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(Z_n \in O) = -\inf_{\nu\in O}\sum_{s=1}^r \nu_s \log\frac{\nu_s}{\rho_s}.$$

Wie eben gesehen, ist wegen Definition 0.3: $-\inf_{\nu\in O}\sum_{s=1}^r \nu_s \log\frac{\nu_s}{\rho_s} = -\Lambda^*(\nu)$. Mit $C \supset O$ wissen wir wegen Satz 1.1, dass

$$-\Lambda^*(O) \leq \liminf \frac{1}{n}\log \mathbb{P}(Z_n \in O) \leq \limsup \frac{1}{n}\log \mathbb{P}(Z_n \in C) \leq -\Lambda^*(C)$$

bereits gilt. Bleibt also noch zu prüfen, ob auch $-\Lambda^*(O) \geq -\Lambda^*(C)$ gilt.Liegt das Infimum im Inneren von C, dann ist die Angelegenheit klar. Liegt es am Rand, so benutzen wir die Stetigkeit von Λ^* um uns zu überzeugen, dass $\Lambda^*(O)$ gegen $\Lambda^*(C)$ konvergiert.

Damit wäre Sanovs Theorem als Spezialfall des Gärtner-Ellis Theorems gezeigt. Wie man sieht ist Sanovs Theorem eine mehrdimensionale Variante von Cramérs Theorem.

[13] da die X_i identisch verteilt und unabhängig sind
[14] weil $\sum_s \nu_s = \sum_s \rho_s = 1$

2.3 Markovketten

Sei (X_i) eine stationäre Γ-wertige[15] Markovkette mit Übergangsmatrix $P_{r\times r}$, die nur aus echt positiven Einträgen besteht. Sei $Z_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i,X_{i+1})}$ (das empirische Paarmaß L_n^2[16]) mit periodischen Grenzbedingungen (also $X_{n+1} = X_1$ bzw. $P_{x_n,x_{n+1}} = 1$). Z_n ist also eine $r\times r$-Matrix, welche die gemittelten Ausprägungen der Markovkette angibt. Die Startverteilung der stationären Markovette sei $\pi = (\pi_s)$ mit $\pi_s > 0 \forall s\in\Gamma$.

Wir wollen nun mit Hilfe des Gärtner-Ellis-Theorems zeigen, dass $P_n(\cdot) := \mathbb{P}(Z_n \in \cdot)$ dem LDP mit Rate n genügt. Die Ratenfunktion werden wir explizit ausrechnen.

Die Kumulantenerzeugende berechnet sich folgendermaßen:

$$\begin{aligned}
\varphi(nt) &= \mathbb{E}\left(e^{\langle nt, Z_n\rangle}\right) = \mathbb{E}\left(e^{\left\langle t, \sum\limits_{i=1}^{n}\delta_{(X_i,X_{i+1})}\right\rangle}\right)\\
&= \sum_{x_1,\ldots,x_{n+1}\in\Gamma^{n+1}} \mathbb{P}\left((X_1,\ldots,X_{n+1}) = (x_1,\ldots,x_n,x_1)\right)\cdot\prod_{k=1}^{n} e^{t_{x_k x_{k+1}}}\\
&= \sum_{x_1,\ldots,x_{n+1}\in\Gamma^{n+1}} \pi_{x_1}\cdot\prod_{k=1}^{n} P_{x_k x_{k+1}}\cdot\prod_{k=1}^{n} e^{t_{x_k x_{k+1}}}\\
&= \sum_{x_1,\ldots,x_{n+1}\in\Gamma^{n+1}} \pi_{x_1} e^{t_{x_n x_1}}\cdot\prod_{k=1}^{n-1} P_{x_k x_{k+1}} e^{t_{x_k x_{k+1}}}\\
&= \sum_{x_1,x_n} \pi_{x_1} e^{t_{x_n x_1}} \sum_{x_2,\ldots,x_{n-1}} \prod_{k=1}^{n-1} Q_{x_k x_{k+1}}(t)\\
&= \sum_{x_1,x_n} \pi_{x_1} e^{t_{x_n x_1}} (Q^{n-1})_{x_1 x_n}(t)\\
&= \sum_{i,j=1}^{r} \pi_i e^{t_{ji}} (Q^{n-1})_{ij}(t)
\end{aligned}$$

Dabei setzen wir $Q_{ij}(t) = P_{ij}e^{t_{ij}}$ zur Abkürzung[17].

Aus der Perron-Frobenius-Theorie erfahren wir, da alle Einträge nicht negativ sind, dass $\lim Q_{i,j}^n \sim r_i l_j \lambda^n$ für $r(t)$ und $l(t)$ als rechter und linker Eigenvektor zum eindeutig bestimmten größten Eigenwert λ von Q. $r(t), l(t)$ sind so normiert, dass ihr produkt den Wert Eins ergibt. Dann ist

$$\begin{aligned}
\Lambda(t) &= \frac{1}{n}\lim_{n\to\infty}\log\sum_{i,j=1}^{r}\pi_i e^{t_{ji}}(Q^{n-1})_{ij}(t)\\
&= \frac{1}{n}\lim_{n\to\infty}\log\max\{\pi_i e^{t_{ji}}(Q^{n-1})_{ij}(t) | i,j = 1\ldots r\}\\
&= \frac{1}{n}\lim_{n\to\infty}\log const\lambda^n(t) = \log\lambda(t)
\end{aligned}$$

Weiterhin erhalten wir die Feststellung, dass $\Lambda(t)$ analytisch (lokal als Potenzreihe entwickelbar) und $\Lambda(t)$ strikt konvex ist, und somit alle Bedingungen des Gärtner-Ellis-Theorems erfüllt werden. Das emprirische Paarmaß genügt also dem LDP mit der Legendretransformierten der Kumulantenerzeugenden Funktion als Ratenfunktion:

[15] $\Gamma = \{1,\ldots,r\}\subset\mathbb{N}$

[16] $L_n^2 \in \tilde{\mathcal{M}}(\Gamma\times\Gamma) := \{\nu\in\Gamma\times\Gamma : \sum_{k,l}\nu_{kl} = 1$ und $\sum_k \nu_{kl} = \sum_k \nu_{lk} \forall l\}$

[17] Auf $(Q^{n-1})_{ij}$ kommen wir, da $A_{ij}^2 = \sum_k a_{ik}a_{kj}$ und somit $A_{ij}^3 = \sum_k a_{ik}\sum_l a_{kl}a_{lj}$ usw.

Die Legendretransformierte ist dann

$$\Lambda^*(\nu) = \sup_{t\in\mathbb{R}^r\times\mathbb{R}^r}\left(\sum_{i,j}\nu_{ij}t_{ij} - \log\lambda(t)\right), \qquad \nu \in \mathcal{M}(\Gamma\times\Gamma)^{18}.$$

Da $\log\lambda(t)$ strikt konvex ist, wird das Supremum genau dann angenommen, wenn der Gradient Null ist:

$$v_{ij} - \frac{\partial}{\partial t_{ij}}\log\lambda(t) = 0 \Leftrightarrow v_{ij} = \frac{1}{\lambda(t)}\frac{\partial}{\partial t_{ij}}\lambda(t) \qquad \forall i,j.$$

Seien nun $r(t)$ und $l(t)$ der rechte und linke Eigenvektor von $Q(t)$ zu $\lambda(t)$ (sie sind auch analytisch in t), welche so normiert sind, dass $\sum_i l_i(t)r_i(t) = 1$ ist[19].
Mit diesen Hilfsmitteln können wir $\frac{\partial}{\partial t_{i,j}}\lambda(t)$ berechnen:

$$\begin{aligned}\frac{\partial}{\partial t_{ij}}\lambda(t) &= \frac{\partial}{\partial t_{ij}}l(t)Q(t)r(t) = \frac{\partial}{\partial t_{ij}}\sum_{k,l}l_k(t)Q_{kl}(t)r_l(t)\\ &= \sum_{k,l}l_k(t)\left(\frac{\partial}{\partial t_{ij}}Q_{kl}(t)\right)r_l(t) = l_i(t)P_{ij}e^{t_{ij}}r_j(t)\end{aligned}$$

In der letzten Zeile können wir die partiellen Ableitungen auf den Matrixeintrag Q_{kl} beschränken, weil $\sum_{k,l}\frac{\partial l_k(t)}{\partial t_{ij}}Q_{kl}(t)r_l(t) + l_k(t)Q_{kl}(t)\frac{\partial r_l(t)}{\partial t_{ij}} = \sum_k \frac{\partial l_k(t)}{\partial t_{ij}}\lambda r_k + \frac{\partial r_k(t)}{\partial t_{ij}}\lambda l_k = \lambda\frac{\partial}{\partial t_{ij}}\sum_k l_k r_k = \lambda\frac{\partial 1}{\partial t_{ij}} = 0$. Damit können wir nun $\Lambda^*(\nu)$ genau bestimmen:

$$\begin{aligned}\nu_{ij} &= \frac{1}{\lambda(t)}l_i(t)P_{ij}e^{t_{ij}}r_j(t) \qquad \forall i,j\\ \Leftrightarrow t_{ij} &= \log\left(\frac{\lambda(t)\nu_{ij}}{l_i(t)P_{ij}r_j(t)}\right)\end{aligned}$$

Setzen wir dieses nun in die Legendretransformierte ein, können wir diese explizit angeben:

$$\begin{aligned}\Lambda^*(\nu) &= \sum_{i,j}\nu_{ij}\log\left(\frac{\lambda(t)\nu_{ij}}{l_i(t)P_{ij}r_j(t)}\right) - \log\lambda(t)\\ &= \sum_{i,j}\nu_{ij}\left(\log\lambda(t) + \log\left(\frac{\nu_{ij}}{l_i(t)P_{ij}r_j(t)}\frac{r_i(t)}{r_i(t)}\right)\right) - \log\lambda(t)\\ &= \sum_{i,j}\nu_{ij}\left(\log r_i(t) - \log r_j(t) + \log\left(\frac{\nu_{ij}}{\hat{\nu}_i P_{ij}}\right)\right) \qquad \hat{\nu}_i := \sum_j \nu_{ij} = l_i(t)r_i(t)^{20}\\ &= \sum_{i,j}\nu_{ij}\log\left(\frac{\nu_{ij}}{\hat{\nu}_i P_{ij}}\right)\end{aligned}$$

In der vorletzten Zeile verschwinden 2 Terme, weil $\sum_{i,j}\nu_{ij}\log r_i = \sum_i \log r_i \sum_j \nu_{ij} = \sum_j \log r_j = \sum_j \log r_j \sum_i \nu_{ij} = \sum_{i,j}\nu_{ij}\log r_j$ (da $\sum_i \nu_{ij} = \sum_j \nu_{ij} = 1$).
Die gesuchte Ratenfunktion für das empirische Paarmaß bei stationären Markovketten mit periodischen Randbedingungen ist also

$$I(\nu) = \Lambda^*(\nu) = \sum_{i,j}\nu_{ij}\log\left(\frac{\nu_{ij}}{\hat{\nu}_i P_{ij}}\right)$$

[18] $\{m \in [0,1]^{r\times r} | \sum_i m_{ij} = \sum_j m_{ij}, \sum_{ij} m_{ij} = 1\}$
[19] also wenn $Q(t)r'(t) = \lambda(t)r'(t)$ etc. mit $\sum_i l'_i(t)r'_i(t) = x$, dann $r(t) := \frac{r'(t)}{\sqrt{x}}$ und $l(t) := \frac{l'(t)}{\sqrt{x}}$
[20] wieso diese Gleichheit gilt, soll hier nicht weiter erläutert werden

3 Bemerkungen

3.1 Begriffe

Folgende Begriffe wurden auf die angegebene Weise ins Deutsche übersetzt:

english		deutsch
exposed (point)	$\longrightarrow$	exponierter (Punkt)
moment (generating function)	$\longrightarrow$	Kumulanten(erzeugende Funktion)
rate-function	$\longrightarrow$	Ratenfunktion
steep(ness)	$\longrightarrow$	steil(heit)
tilted (probability measure)	$\longrightarrow$	gekipptes (Wahrscheinlichkeitsmaß)