

# GROSSE ABWEICHUNGEN FÜR EMPIRISCHE VERTEILUNGEN

Am Mathematischen Institut  
der Ludwig-Maximilians-Universität München eingereichte

## Bachelorarbeit

zur Erlangung des akademischen Grades Bachelor of Science (B.Sc.)  
im Fachgebiet Mathematik



vorgelegt von: Alexisz Tamás Gaál  
Matrikelnummer: 8075402  
Betreuer: Prof. Dr. Franz Merkl

9. August 2011



# Inhaltsverzeichnis

<b>1</b>	<b>Endlicher Phasenraum</b>	<b>3</b>
1.1	Eine Dimension . . . . .	3
1.2	Der m-dimensionale Phasenraum . . . . .	9
<b>2</b>	<b>Das Prinzip großer Abweichungen</b>	<b>21</b>
<b>3</b>	<b>Das Prinzip der Gibbs-Konditionierung</b>	<b>23</b>
<b>4</b>	<b>Der empirische Prozess</b>	<b>27</b>
<b>5</b>	<b>Der abzählbare Phasenraum</b>	<b>37</b>



## Zusammenfassung

Ziel dieser Bachelorarbeit ist die Herleitung des Satzes von Sanov mit kombinatorischen und approximativen Methoden. Wir zeigen, dass ein empirisches Maß, das von i. i. d. Zufallsvariablen erzeugt wird, in einem endlich-dimensionalen Phasenraum das Prinzip großer Abweichungen erfüllt. Wir zeigen einen Satz über die große Abweichung des empirischen Prozesses und erweitern das endlich-dimensionale Ergebnis durch Approximation auf den abzählbaren Fall. Es wird eine für die statistische Physik relevante Anwendung gegeben.

## 1 Endlicher Phasenraum

Für den gesamten ersten Abschnitt führen wir folgende Bezeichnungen ein. Wir legen unseren Überlegungen einen passenden Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  zugrunde und betrachten Zufallsvariablen  $(X_i)_{i \in \mathbb{N}}$  auf  $\Omega$  mit den Eigenschaften:

$$\begin{aligned} (X_i)_{i \in \mathbb{N}} \text{ sind i. i. d. Zufallsvariablen } \Omega \rightarrow \Gamma = \{1, 2, \dots, N\} \subset \mathbb{N}, \\ \text{die Verteilung von } X_1 \text{ besitzt die Zähldichte } (\rho_i)_{i \in \Gamma} \text{ mit } \rho_i > 0 \quad \forall i \in \Gamma. \end{aligned} \quad (1)$$

### 1.1 Eine Dimension

Das von den Zufallsvariablen  $X_1, X_2, \dots, X_n$  definierte *empirische Maß*  $L_n$  ist ein zufälliges Wahrscheinlichkeitsmaß auf dem Messraum  $(\Gamma, \mathcal{P}(\Gamma))$ . Es wird mit Hilfe der Diracmaße  $\delta_{X_i}$  in den Punkten  $X_i \quad i = 1, 2, \dots, n$  als

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

definiert. Wir werden die Wahrscheinlichkeitsmaße auf  $\Gamma$  mit ihren Zähldichten identifizieren bzw. das empirische Maß als ein zufälliges Element der Menge

$$\mathcal{M}_1(\Gamma) := \left\{ \nu = (\nu_1, \dots, \nu_N) \in [0, 1]^N : \sum_{i=1}^N \nu_i = 1 \right\}$$

auffassen. Dabei beschreibt  $\mathcal{M}_1(\Gamma)$  die Menge der Wahrscheinlichkeitsmaße auf  $\Gamma$ . Auf  $\mathcal{M}_1(\Gamma)$  definieren wir die Totalvariationsmetrik

$$d(\nu, \mu) := \frac{1}{2} \sum_{i=1}^N |\nu_i - \mu_i|.$$

**Lemma 1.1.** *Es gilt:  $d(L_n, \rho) \xrightarrow{n \rightarrow \infty} 0$   $\mathbb{P}$ -f.s..*

*Beweis.* Wir müssen zeigen, dass  $L_n(\{m\}) \xrightarrow{n \rightarrow \infty} \rho_m$   $\mathbb{P}$ -f.s. für alle  $m = 1, \dots, N$ . Es gilt:  $\mathbb{P}(X_i = m) = \rho_m$  für alle  $m = 1, \dots, N$  und  $i \in \mathbb{N}$  und die Ereignissysteme  $\{\{X_i = m\} : i \in \mathbb{N}\}$  sind unabhängig für alle  $m = 1, \dots, N$ . Die Aussage folgt aus dem starken Gesetz der großen Zahlen, denn

$$L_n(\{m\}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i=m\}} \xrightarrow{n \rightarrow \infty} \rho_m \quad \mathbb{P}\text{-f.s.} \quad \forall m = 1, \dots, N.$$

□

Nun können wir die große Abweichung des empirischen Maßes von  $\rho$  betrachten. Wir brauchen einige Vorbereitungen, um einen kombinatorischen Beweis für den Satz von Sanov zu geben, der diese große Abweichung beschreibt. Wir beginnen mit folgendem Lemma:

**Lemma 1.2.** (a) *Für alle  $\nu \in \mathcal{M}_1(\Gamma)$  existiert eine Folge  $\nu_n \in \frac{1}{n}K_n$  für  $n \in \mathbb{N}$ , so dass  $\lim_{n \rightarrow \infty} d(\nu_n, \nu) = 0$  mit*

$$K_n := \left\{ k = (k_1, \dots, k_N) \in \mathbb{N}_0^N : \sum_{i=1}^N k_i = n \right\}.$$

(b) *Die Abbildung  $I_\rho : \mathcal{M}_1(\Gamma) \rightarrow \mathbb{R}_{\geq 0}$  ist wohldefiniert, stetig und streng konvex.  $I_\rho(\nu) = 0$  genau dann, wenn  $\nu = \rho$ , wobei*

$$I_\rho(\nu) := \sum_{i=1}^N \nu_i \log \left( \frac{\nu_i}{\rho_i} \right).$$

*Bemerkung:* Hier machen wir von der Konvention  $0 \log(0) = 0$  Gebrauch. Den Wert  $I_\rho(\nu) = H(\nu|\rho)$  nennt man *relative Entropie von  $\nu$  bezüglich  $\rho$* .

*Beweis.* Wir zeigen, dass für alle  $\nu \in \mathcal{M}_1(\Gamma)$  gilt:  $d(\nu, \frac{1}{n}K_n) := \min_{\mu \in \frac{1}{n}K_n} d(\nu, \mu) \leq \frac{N}{2n}$ . Dies ergibt Teil (a).

Sei ein beliebiges  $\nu \in \mathcal{M}_1(\Gamma)$  gegeben. Wir definieren  $k' = (k'_1, \dots, k'_N) \in \mathbb{N}_0^N$  durch  $k'_i := \lfloor n\nu_i \rfloor$ , dann wählen wir  $m \in \mathbb{N}_0$  so, dass die Gleichung  $\sum_{i=1}^N k'_i + m = n$  gilt. Dann ist

$$m = n - \sum_{i=1}^N k'_i = \sum_{i=1}^N \underbrace{(n\nu_i - \lfloor n\nu_i \rfloor)}_{< 1} < N.$$

Wir definieren  $k \in K_n$  durch

$$k_i := \begin{cases} k'_i + 1, & i = 1, \dots, m \\ k'_i, & i = m + 1, \dots, N \end{cases}.$$

Dann ist  $|n\nu_i - k_i| \leq 1$  für alle  $i = 1, \dots, N$  und damit  $d(\nu, \frac{1}{n}k) \leq \frac{N}{2n}$ . Die Behauptung von (a) folgt.

Zu Teil (b): wir schreiben  $I_\rho(\nu) = \sum_{i=1}^N \nu_i \log\left(\frac{\nu_i}{\rho_i}\right) = \sum_{i=1}^N \rho_i h\left(\frac{\nu_i}{\rho_i}\right)$  mit  $h(x) := x \log(x)$  für  $x > 0$  und  $h(0) := 0$ . Die Funktion  $h : [0, \infty) \rightarrow \mathbb{R}$  ist stetig und streng konvex. Die Stetigkeit folgt aus der Regel von L'Hospital bzw. erhält man nach zweimaligem Ableiten  $h''(x) = \frac{1}{x} > 0$  für  $x > 0$ . Also ist  $h$  streng konvex auf dem offenem Intervall  $(0, \infty)$ . Da  $h(tx) = tx \log(tx) < tx \log(x) = th(x)$  für  $x > 0$  und  $t \in (0, 1)$ , folgt die strenge Konvexität auf ganz  $[0, \infty)$ . Wir folgern, dass  $I_\rho$  endlich, stetig und streng konvex ist, d.h.  $I_\rho(t\nu + (1-t)\mu) < tI_\rho(\nu) + (1-t)I_\rho(\mu)$  für alle  $\mu, \nu \in \mathcal{M}_1(\Gamma), \mu \neq \nu$  und  $t \in (0, 1)$ . Wir bemerken dabei, dass die konvexe Hülle zweier Punkte aus  $\mathcal{M}_1(\Gamma)$  ganz in der selben Menge liegt.

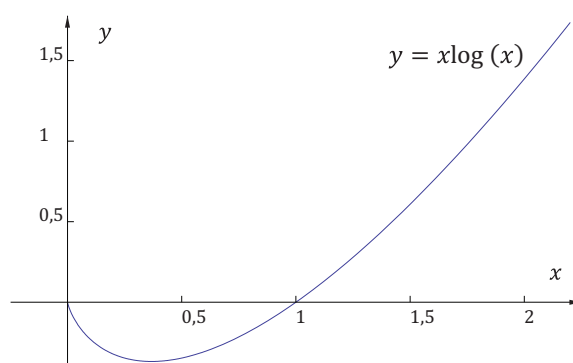


Abbildung 1: Graph der Funktion  $f(x) = x \log x$ .

Es bleibt zu zeigen, dass  $I_\rho(\nu) \geq 0$ , wobei Gleichheit genau dann besteht, wenn  $\nu = \rho$ . Das folgt aber aus der Jensenschen Ungleichung:

$$I_\rho(\nu) = \sum_{i=1}^N \rho_i h\left(\frac{\nu_i}{\rho_i}\right) \geq h\left(\sum_{i=1}^N \rho_i \frac{\nu_i}{\rho_i}\right) = h(1) = 0.$$

Gleichheit gilt genau dann, wenn  $\frac{\nu_i}{\rho_i} = C$  für alle  $i = 1, \dots, N$ . Dann ist jedoch  $C = 1$  und deshalb  $\nu = \rho$ .  $\square$

*Bemerkung:* Aus Teil (a) und aus der Beobachtung, dass  $K_n$  eine endliche Menge ist, folgt, dass  $\bigcup_{i=1}^{\infty} \frac{1}{n}K_n \subset \mathcal{M}_1(\Gamma)$  eine abzählbare dichte Teilmenge ist. Damit ist  $(\mathcal{M}_1(\Gamma), d)$  ein polnischer Raum, also ein vollständiger, separabler, metrischer Raum. Der Raum  $(\mathcal{M}_1(\Gamma), d)$  ist sogar kompakt, weil  $\mathcal{M}_1(\Gamma) \subset [0, 1]^N$  eine abgeschlossene Teilmenge ist.

Nun können wir den Satz von Sanov für einen endlichen Phasenraum  $\Gamma$  formulieren.

**Satz 1.3.** Seien  $X_1, X_2, \dots$  wie in (1). Weiter sei  $L_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  das empirische Maß und  $A \subset \mathcal{M}_1(\Gamma)$ . Dann gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in A) = - \inf_{\nu \in A} I_\rho(\nu),$$

wobei

$$I_\rho(\nu) = \sum_{i=1}^N \nu_i \log \left( \frac{\nu_i}{\rho_i} \right)$$

mit  $\nu = (\nu_1, \dots, \nu_N)$  und den Konventionen  $\inf \emptyset = \infty$  und  $\log 0 = -\infty$ .

*Beweis.* In dem Beweis stützen wir uns auf Ideen aus [3, Beweis von Theorem II. 2]. Falls  $A$  leer ist, dann gilt die Behauptung des Satzes. Wir setzen  $A \neq \emptyset$  voraus und definieren, wie in Lemma 1.2 die Menge

$$K_n := \left\{ k = (k_1, \dots, k_N) \in \mathbb{N}_0^N : \sum_{i=1}^N k_i = n \right\}.$$

Sei  $k = (k_1, k_2, \dots, k_N) \in K_n$ , dann gilt

$$\mathbb{P} \left( L_n = \frac{k}{n} \right) = n! \prod_{i=1}^N \frac{\rho_i^{k_i}}{k_i!},$$

da der Buchstabe  $m \in \Gamma$  in der unabhängigen Sequenz  $X_1, \dots, X_n$   $k_m$ -mal auftauchen muss. Wir teilen jedes  $\rho_s^{k_s}$  durch die Anzahl der Permutationen der  $k_s$  Elemente, also durch  $k_s!$ , damit wir diese Permutationen von demselben Buchstaben nicht mehrfach zählen. Wir multiplizieren noch das ganze mit  $n!$ , weil jede Reihenfolge der Buchstaben in  $X_1, \dots, X_n$  erlaubt ist und es nur auf ihre Anzahl ankommt.

Wir schätzen die Wahrscheinlichkeiten  $\log \left( \mathbb{P} \left( L_n = \frac{k}{n} \right) \right)$  mit Hilfe der Stirlingformel:  $\log(n!) = n \log n - n + \mathcal{O}(\log n)$  für  $n \rightarrow \infty$ .



$$\begin{aligned}
\log \left( \mathbb{P} \left( L_n = \frac{k}{n} \right) \right) &= \log \left( n! \prod_{i=1}^N \frac{\rho_i^{k_i}}{k_i!} \right) = \log(n!) + \sum_{i=1}^N \left( \log(\rho_i^{k_i}) - \log(k_i!) \right) \\
&= n \log n - n + \mathcal{O}(\log n) + \sum_{i=1}^N k_i \log(\rho_i) - \sum_{i=1}^N (k_i \log(k_i) - k_i + \mathcal{O}(\log k_i)) \\
&\stackrel{k_i \leq n, \sum_{i=1}^N k_i = n}{=} \sum_{i=1}^N k_i \log n - n + \sum_{i=1}^N k_i (\log(\rho_i) - \log(k_i)) + \underbrace{\sum_{i=1}^N k_i}_{=n} + \mathcal{O}(\log n) \\
&= n \sum_{i=1}^N \frac{k_i}{n} \log \left( \frac{\rho_i}{k_i/n} \right) + \mathcal{O}(\log n) = -n I_\rho(k/n) + \mathcal{O}(\log n), \quad (2)
\end{aligned}$$

wobei die asymptotische obere Schranke gleichmäßig auf  $K_n$  gilt. Weiterhin gilt, wegen  $L_n \in \frac{K_n}{n}$  und weil jedes Element in  $\frac{K_n}{n}$  ein empirisches Maß  $n$ -ter Stufe darstellt, die folgende Abschätzung:

$$Q_n(A) \leq \mathbb{P}(L_n \in A) \leq |K_n| Q_n(A). \quad (3)$$

Dabei ist  $Q_n(A)$  als

$$Q_n(A) = \max_{k \in K_n, \frac{k}{n} \in A} \left( n! \prod_{i=1}^N \frac{\rho_i^{k_i}}{k_i!} \right)$$

definiert und wir setzen  $Q_n(A) = 0$ , falls  $A \cap \frac{K_n}{n} = \emptyset$ . Für die Mächtigkeit der Menge  $K_n$  gilt:

$$|K_n| \stackrel{\substack{= \\ \uparrow \\ \text{Kombination mit Wiederholung}}}{=} \binom{n+N-1}{N-1} = \frac{(n+N-1)!}{(N-1)! n!} \leq \frac{1}{(N-1)!} (n+N-1)^{N-1}.$$

Daher ist  $|K_n| = \mathcal{O}(n^{N-1})$  für  $n \rightarrow \infty$ . Das heißt: es existieren ein  $n_0 \in \mathbb{N}$  und ein  $C \in \mathbb{R}$ , so dass  $|K_n| \leq C n^{N-1}$  für alle  $n \geq n_0$ . Wir folgern, dass  $\frac{1}{n} \log(|K_n|) = \mathcal{O}\left(\frac{\log n}{n}\right)$ . Wegen  $\frac{\log n}{n} \xrightarrow{n \rightarrow \infty} 0$  gilt, dass  $\frac{1}{n} \log(|K_n|) \rightarrow 0$  für  $n \rightarrow \infty$ . Aus (2) und (3) folgern wir nun

$$\frac{1}{n} \log \mathbb{P}(L_n \in A) \stackrel{(3)}{=} \frac{1}{n} \log(Q_n) + \mathcal{O}\left(\frac{\log n}{n}\right) \stackrel{(2)}{=} - \min_{k \in K_n, \frac{k}{n} \in A} I_\rho\left(\frac{k}{n}\right) + \mathcal{O}\left(\frac{\log n}{n}\right).$$

Es bleibt zu zeigen, dass  $\min_{k \in K_n} \min_{\frac{k}{n} \in A} I_\rho \left( \frac{k}{n} \right) \rightarrow \inf_{\nu \in A} I_\rho(\nu)$  für  $n \rightarrow \infty$ . Sei  $\nu \in A$  ein beliebiges Wahrscheinlichkeitsmaß. Dann existiert nach *Lemma 1.2 (a)* eine Folge  $\nu_n \in \frac{1}{n}K_n$ , so dass  $\lim_{n \rightarrow \infty} d(\nu_n, \nu) = 0$  und, weil  $A$  offen ist, existiert ein  $n_0 \in \mathbb{N}$ , so dass für alle  $n_0 \leq n : \nu_n \in A$ . Wegen der Stetigkeit von  $I_\rho$  aus *Lemma 1.2 (b)* gilt:  $I_\rho(\nu_n) \rightarrow I_\rho(\nu)$  für  $n \rightarrow \infty$ . Insgesamt haben wir also:

$$\limsup_{n \rightarrow \infty} \min_{k \in K_n} \min_{\frac{k}{n} \in A} I_\rho \left( \frac{k}{n} \right) \leq I_\rho(\nu) \quad \forall \nu \in A.$$

Durch optimieren über  $\nu$  erhalten wir

$$\limsup_{n \rightarrow \infty} \min_{k \in K_n} \min_{\frac{k}{n} \in A} I_\rho \left( \frac{k}{n} \right) \leq \inf_{\nu \in A} I_\rho(\nu).$$

Die umgekehrte Ungleichung

$$\inf_{\nu \in A} I_\rho(\nu) \leq \liminf_{n \rightarrow \infty} \min_{k \in K_n} \min_{\frac{k}{n} \in A} I_\rho \left( \frac{k}{n} \right)$$

gilt aber offensichtlich. Damit ist die Konvergenz der Folge gegen  $\inf_{\nu \in A} I_\rho(\nu)$  gezeigt.  $\square$

**Beispiel** (*relative Entropie bei Gleichverteilung*). Sei  $\rho$  die Gleichverteilung auf  $\Gamma$  und  $M \subset \Gamma$  eine nichtleere Teilmenge. Sei  $\nu$  die Gleichverteilung auf der Menge  $M$ . Wir berechnen die relative Entropie von  $\nu$  bezüglich  $\rho$ . Zur Vereinfachung der Notation sei  $M = \{1, 2, \dots, r\}$  mit  $r \in \{1, 2, \dots, N\}$ . Dann hat  $I_\rho = H(\nu|\rho)$  den Wert

$$H(\nu|\rho) = \sum_{i=1}^r \nu_i \log \left( \frac{\nu_i}{\rho_i} \right) = \sum_{i=1}^r \frac{1}{r} \log \left( \frac{N}{r} \right) = \log \left( \frac{N}{r} \right) = \log \left( \frac{N}{|M|} \right).$$

Es gilt sogar, dass  $\nu$  ein Minimierer der relativen Entropie  $H(\tilde{\nu}|\rho)$  unter allen Wahrscheinlichkeitsmaßen  $\tilde{\nu}$  auf  $M$  ist. Mit der Jensen'schen Ungleichung und mit der Setzung der konvexen Funktion  $h(x) = x \log(x)$  folgt nämlich:

$$\begin{aligned} H(\tilde{\nu}|\rho) &= \sum_{i=1}^r \tilde{\nu}_i \log \left( \frac{\tilde{\nu}_i}{\rho_i} \right) = \sum_{i=1}^r \rho_i h \left( \frac{\tilde{\nu}_i}{\rho_i} \right) = \frac{r}{N} \sum_{i=1}^r \frac{N}{r} \rho_i h \left( \frac{\tilde{\nu}_i}{\rho_i} \right) \geq \frac{r}{N} h \left( \sum_{i=1}^r \frac{N}{r} \tilde{\nu}_i \right) \\ &= \frac{r}{N} h \left( \frac{N}{r} \right) = \log \left( \frac{N}{r} \right) = H(\nu|\rho). \end{aligned}$$

## 1.2 Der m-dimensionale Phasenraum

In diesem Abschnitt betrachten wir  $\Gamma^m$  für ein fixes  $m \in \mathbb{N}$  als den zugrundeliegenden Phasenraum. Wir definieren  $\mathcal{M}_1(\Gamma^m)$  als die Menge der Wahrscheinlichkeitsmaße auf  $\Gamma^m$  und  $\widetilde{\mathcal{M}}_1(\Gamma^m)$  als die Menge der *zyklischen Wahrscheinlichkeitsmaße* auf  $\Gamma^m$ :

$$\widetilde{\mathcal{M}}_1(\Gamma^m) := \left\{ \nu = (\nu_{s_1, \dots, s_m}) \in \mathcal{M}_1(\Gamma^m) : \sum_{s_m} \nu_{s_1, \dots, s_{m-1}, s_m} = \sum_{s_m} \nu_{s_m, s_1, \dots, s_{m-1}} \right. \\ \left. \forall (s_1, \dots, s_{m-1}) \in \Gamma^{m-1} \right\}.$$

Nach der Definition gilt insbesondere, dass alle eindimensionalen Randverteilungen eines zyklischen Maßes übereinstimmen. Das m-dimensionale empirische Maß mit periodischen Randbedingungen ist als

$$L_n^m := \frac{1}{n} \sum_{i=1}^n \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})}$$

definiert mit  $\tilde{X}_s = X_{[s]}$ , wobei  $[s] = \min\{i \in \mathbb{N} : i \equiv s \pmod{n}\}$ . Damit ist das m-dimensionale empirische Maß  $L_n^m \in \widetilde{\mathcal{M}}_1(\Gamma^m)$  ein zyklisches Maß. Wir definieren auf der Menge  $\widetilde{\mathcal{M}}_1(\Gamma^m)$  die Totalvariationsmetrik

$$d_m(\nu, \mu) := \frac{1}{2} \sum_{s_1, \dots, s_m} |\nu_{s_1, \dots, s_m} - \mu_{s_1, \dots, s_m}|.$$

Um die Notation zu vereinfachen, lassen wir in diesem Abschnitt den unteren Index weg und schreiben  $d(\mu, \nu)$  statt  $d_m(\mu, \nu)$ . Dies verursacht keine Mehrdeutigkeiten, weil wir  $m \in \mathbb{N}$  fixiert haben.

Wir gehen analog zum eindimensionalen Fall vor. Zuerst betrachten wir die Konvergenzaussage für das empirische Maß.

**Lemma 1.4.** *Es gilt:  $d(L_n^m, \rho^{\otimes m}) \xrightarrow{n \rightarrow \infty} 0$   $\mathbb{P}$ -f.s. für alle  $m \in \mathbb{N}$ .*

*Beweis.* Wir definieren  $J_n^m := \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, \dots, X_{i+m-1})}$  als das empirische Maß *ohne Randbedingung*. Wir spalten den Teil von  $L_n^m$  ab, der die Periodizität sicherstellt und schreiben  $L_n^m$  als die konvexe Summe zweier Wahrscheinlichkeitsmaße. Für  $n \geq m$  gilt:

$$\begin{aligned}
L_n^m &= \frac{1}{n} \sum_{i=1}^n \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})} = \frac{1}{n} \left( \sum_{i=1}^{n-m} \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})} + \sum_{i=n-m+1}^n \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})} \right) \\
&= \frac{n-m}{n} \frac{1}{n-m} \sum_{i=1}^{n-m} \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})} + \frac{m}{n} \frac{1}{m} \sum_{i=n-m+1}^n \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})} \\
&= \left( \frac{n-m}{n} \right) J_{n-m}^m + \frac{m}{n} \frac{1}{m} \sum_{i=n-m+1}^n \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})}.
\end{aligned}$$

Weil  $\frac{m}{n}$  für  $n \rightarrow \infty$  gegen 0 konvergiert, genügt es, die Aussage des Satzes für  $J_n^m$  zu zeigen. Wir zeigen zunächst die Aussage für  $J_{nm}^m$ . Dazu spalten wir dieses Wahrscheinlichkeitsmaß in die konvexe Summe von  $m$  empirischen Maßen auf, die jeweils nur von unabhängigen Zufallsvariablen erzeugt werden:

$$\begin{aligned}
J_{nm}^m &= \frac{1}{nm} \sum_{i=1}^{nm} \delta_{(X_i, \dots, X_{i+m-1})} = \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=1}^m \delta_{(X_{mi+j}, X_{mi+j+1}, \dots, X_{mi+j+(m-1)})} \\
&= \frac{1}{m} \sum_{j=1}^m \left( \frac{1}{n} \sum_{i=0}^{n-1} \delta_{(X_{mi+j}, X_{mi+j+1}, \dots, X_{mi+j+(m-1)})} \right).
\end{aligned}$$

Für jedes  $s \in \{1, 2, \dots, m\}$  enthält die Menge  $\{(X_{mi+s}, X_{mi+s+1}, \dots, X_{mi+s+(m-1)}) : i \in \mathbb{N}_0\}$  nur i. i. d. Zufallsvektoren, deren Einträge auch i. i. d. sind. Aus dem starken Gesetz der Großen Zahlen folgt, wie in *Lemma 1.1*, dass  $\frac{1}{n} \sum_{i=0}^{n-1} \delta_{(X_{mi+j}, X_{mi+j+1}, \dots, X_{mi+j+(m-1)})}$  für  $n \rightarrow \infty$   $\mathbb{P}$ -f.s. gegen  $\rho^{\otimes m}$  konvergiert. Daher konvergiert auch  $J_{nm}^m$   $\mathbb{P}$ -f.s. gegen dieselbe Verteilung für  $n \rightarrow \infty$ . Nun folgt die Behauptung des Satzes aus folgender Beobachtung:

$$\forall m, n \in \mathbb{N} \forall s \in \{0, 1, \dots, m-1\} : d(J_{nm}^m, J_{nm+s}^m) \leq \frac{1}{2} \frac{2}{n} s \leq \frac{m}{n}.$$

Für  $k \in \{1, 2, \dots, nm\}$ ,  $a \in \{1, 2, \dots, s\}$  mit  $s \leq m-1$  gilt nämlich

$$\left| \frac{k}{nm} - \frac{k+a}{nm+s} \right| = \left| k \underbrace{\left( \frac{1}{nm} - \frac{1}{nm+s} \right)}_{\frac{s}{nm(nm+s)}} - \frac{a}{nm+s} \right| \leq k \frac{s}{nm(nm+s)} + \frac{a}{nm+s} \leq \frac{2}{n}.$$

Da  $\frac{m}{n}$  für  $n \rightarrow \infty$  gegen 0 konvergiert, konvergiert auch  $d(J_n^m, \rho^{\otimes m})$  für  $n \rightarrow \infty$  gegen 0.  $\square$

Als Nächstes überlegen wir, dass die Menge der m-dimensionalen empirischen Maße mit periodischen Randbedingungen dicht in der Menge der zyklischen Maße  $\widetilde{\mathcal{M}}_1(\Gamma^m)$  liegt. Ein empirisches Maß mit periodischen Randbedingungen  $L_n^m$  ist ein Element der Menge  $\frac{1}{n}\widetilde{K}_n$  mit

$$\widetilde{K}_n := \left\{ k = (k_{s_1, s_2, \dots, s_m}) \in \mathbb{N}_0^{(N^m)} : \sum_{s_m} k_{s_1, \dots, s_{m-1}, s_m} \right. \\ \left. = \sum_{s_m} k_{s_m, s_1, \dots, s_{m-1}} \quad \forall (s_1, \dots, s_{m-1}) \in \Gamma^{m-1}, \quad \sum_{s_1, s_2, \dots, s_m} k_{s_1, \dots, s_m} = n \right\}.$$

Aber nicht alle Elemente von  $\frac{1}{n}\widetilde{K}_n$  stellen ein empirisches Maß  $L_n^m$  dar. Dafür, dass sie auch zusammenhängend sind, müssen sie noch eine weitere Bedingung erfüllen. Diese können wir uns durch folgendes Beispiel erschließen. Wir nehmen ein Element  $k \in K_{10}$  für  $m = 4$ . Dieses wird durch Diracmaße in den Tupeln

1221	3434	4434
2212	4344	4343
2122	3443	3434
		4343

definiert und generiert ein zyklisches Wahrscheinlichkeitsmaß  $\frac{1}{10}k \in \widetilde{\mathcal{M}}_1(\Gamma^4)$ , das aber von keiner Sequenz  $X_1(\omega), X_2(\omega), \dots, X_{10}(\omega)$  erzeugt wird.

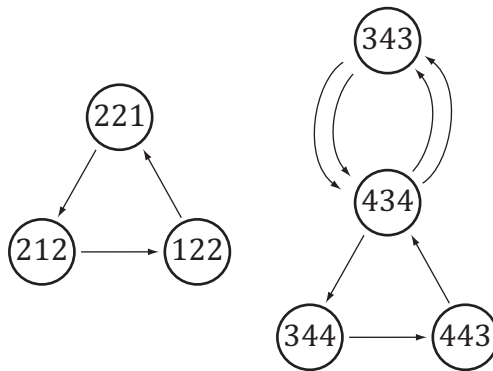


Abbildung 2: Der zum obigen Beispiel zugehörige Graph  $G(k)$ .

Um die gesuchte Bedingung zu formulieren, definieren wir für ein  $k \in \tilde{K}_n$  den gerichteten Graphen  $G(k)$  mit der Knotenmenge  $V \subset \Gamma^{m-1}$  so, dass wir für jedes  $(s_1, s_2, \dots, s_m) \in \Gamma^m$   $k_{s_1, s_2, \dots, s_m}$  Pfeile von  $(s_1, s_2, \dots, s_{m-1})$  nach  $(s_2, s_3, \dots, s_m)$  zeichnen. Die Knotenmenge bestimmen wir dann so, dass  $G(k)$  keine isolierten Knoten hat. Wir lassen also die Elemente aus  $\Gamma^{m-1}$ , von denen kein Pfeil ausgeht, weg. Man beachte, dass wegen der Bedingung  $\sum_{s_m} k_{s_1, \dots, s_{m-1}, s_m} = \sum_{s_m} k_{s_m, s_1, \dots, s_{m-1}} \forall (s_1, \dots, s_{m-1}) \in \Gamma^{m-1}$  jeder Knoten die gleiche Anzahl von aus- und einlaufenden Pfeilen hat. Die gesuchte Bedingung lautet dann, dass  $G(k)$  stark zusammenhängend sein soll. Das bedeutet, dass wir von jedem Knoten aus jeden anderen Knoten entlang der Pfeile erreichen können, wobei wir die Richtung der Pfeile respektieren. Wie man anhand von *Abbildung 2* sieht, ist der Graph  $G(k)$  zum obigen Beispiel nicht stark zusammenhängend. Wir formulieren diese Erkenntnis in folgendem Lemma:

**Lemma 1.5.** *Es gilt folgende Gleichheit von Mengen:  $\hat{K}_n = B_n$  mit*

$$\begin{aligned} \hat{K}_n &= \left\{ k \in \tilde{K}_n : G(k) \text{ ist stark zusammenhängend} \right\} \text{ und} \\ B_n &= \left\{ \hat{L}_n^m(x_1, x_2, \dots, x_n) := \sum_{i=1}^n \delta_{(\tilde{x}_i, \dots, \tilde{x}_{i+m-1})} : (x_1, \dots, x_n) \in \Gamma^n, \right. \\ &\quad \left. \tilde{x}_s = x_{[s]} \text{ mit } [s] = \min\{i \in \mathbb{N} : i \equiv s \pmod{n}\} \right\}. \end{aligned}$$

*Beweis.* Die Elemente der Mengen sind Maße mit Gesamtmasse  $n$ . Die Inklusion  $B_n \subset \hat{K}_n$  wird offensichtlich, wenn man sich die Elemente von  $B_n$  anschaut. Da wir jedes Element  $\mu \in B_n$  aus einer Sequenz  $x_1, x_2, \dots, x_n$  durch periodische Fortsetzung konstruieren, wird der zugehörige gerichtete Graph  $G(\mu)$  stark zusammenhängend. Wir haben sogar einen Eulerkreis in  $G(\mu)$ , also einen geschlossenen Weg im Graphen, in dem jeder Pfeil des Graphen genau einmal benutzt wird.

Umgekehrt sei  $\mu \in \hat{K}_n$ . Wenn wir einen Eulerkreis in  $G(\mu)$  finden, können wir eine Sequenz  $x_1, x_2, \dots, x_n$  so konstruieren, dass  $\mu = \hat{L}_n^m(x_1, x_2, \dots, x_n)$ . Um einen Eulerkreis zu finden, gehen wir wie folgt vor.

*Behauptung:* In Jedem stark zusammenhängenden, gerichteten Graphen, in dem in jedem Knoten die Anzahl auslaufender Kanten (Ausgangsgrad) gleich der Anzahl einlaufender Kanten (Eingangsgrad) ist, existiert ein Eulerkreis.

Sei  $G = (V, E)$  ein gerichteter, stark zusammenhängender Graph. Wir wählen einen Knoten  $v \in V$  aus. Ausgehend von  $v$  laufen wir entlang der Pfeile so, dass wir jeden Pfeil höchstens einmal benutzen. Wenn wir in einem Knoten ankommend den Weg nicht fortsetzen können, sind wir in dem Anfangsknoten  $v$ . Wären wir nämlich in einem anderen Knoten, so gäbe es einen Ausweg, weil der Ausgangsgrad gleich dem Eingangsgrad ist. Wir wählen einen solchen geschlossenen Weg  $W$  mit maximaler Kantenzahl aus. Dann enthält  $W$  alle Kanten, die in  $v$  starten oder dort enden. Wir nehmen an, dass  $W$  kein Eulerkreis ist. Wir betrachten

den nichtleeren Teilgraphen  $T = (\tilde{V}, \tilde{E})$  von  $G$ , der durch Weglassen der Kanten von  $W$  entsteht, und lassen dabei auch entstehende isolierte Knoten weg. Dann ist  $v \notin T$  und der Ausgangsgrad ist in jedem Knoten von  $T$  gleich dem Eingangsgrad. Sei  $\tilde{v} \in \tilde{V}$  beliebig, dann existiert ein gerichteter Weg  $S$  von  $v$  nach  $\tilde{v}$  in  $G$ , weil  $G$  stark zusammenhängt. Wir wählen den ersten Knoten  $s$  in  $S$  mit der Eigenschaft  $s \in \tilde{V}$  aus. Dann ist aber  $s$  in  $W$  enthalten, denn sonst wäre der Knoten vor  $s$  in  $S$  auch in  $\tilde{V}$ . Wie vorhin finden wir von  $s$  ausgehend einen geschlossenen Weg  $\tilde{W}$  in  $T$ . Wir definieren dann, einen geschlossenen Weg wie folgt: wir starten bei  $v$  und laufen entlang  $W$ , bis wir  $s$  zum ersten Mal erreichen, dann laufen wir entlang  $\tilde{W}$  und anschließend den Weg  $W$  weiter von der Stelle an, an der wir vorhin aufgehört haben. Dieser neue geschlossene Weg enthält  $v$  bzw. jede Kante von  $G$  höchstens einmal und hat mehr Kanten als  $W$ . Wegen der maximalen Kantenzahl von  $W$  ist das ein Widerspruch. Also muss  $W$  ein Eulerkreis sein.  $\square$

**Lemma 1.6.** *Für einen stark zusammenhängenden, gerichteten Graphen  $G = (V, E)$  mit  $V = \{1, 2, \dots, N\}$ , dessen Ausgangsgrad  $\alpha_s > 0$  in jedem Knoten  $s \in V$  gleich dem Eingangsgrad in  $s$  ist, gilt folgende Abschätzung für  $\mathcal{E}(G)$ , die Anzahl der Eulerkreise in  $G$ :*

$$\prod_{i=1}^N (\alpha_i - 1)! \leq \mathcal{E}(G) \leq \prod_{i=1}^N \alpha_i!$$

*Beweis.* Den Beweis entnehmen wir aus [3, Lemma II.10]. Nach der Behauptung im Beweis von Lemma 1.5 gibt es in  $G$  einen Eulerkreis.

*untere Schranke:* Wir nehmen einen Eulerkreis in  $G$  und markieren einen Knoten als Anfangsknoten. Weiter markieren wir in jedem Knoten die Kante, die in dem Eulerkreis bezüglich des Anfangsknotens als letzte benutzt wird. Als Nächstes nummerieren wir in jedem Knoten die nicht markierten Kanten durch. So erhalten wir für jede Nummerierung einen Eulerkreis, in dem wir von dem Anfangspunkt aus der Nummerierung nach die Kanten entlanglaufen. Wir werden den Weg immer fortsetzen können. Sollte nämlich in einem Knoten keine nummerierten Kanten mehr vorhanden sein, nehmen wir in diesem Knoten die markierte Kante als Ausweg. Wenn die markierte Kante in einem Knoten einmal benutzt wurde, werden wir nicht mehr in diesen Knoten zurückkehren. Wenn alle Kanten aufgebraucht sind, müssen wir demnach wieder im Anfangspunkt angekommen sein. Die Anzahl der möglichen Nummerierungen ergibt die untere Schranke.

*obere Schranke:* In jedem Knoten nummerieren wir die Kanten durch und markieren einen Knoten als Anfangsknoten. So erhalten wir einen Weg in  $G$ , der nicht unbedingt ein Eulerkreis ist. Alle Eulerkreise entstehen jedoch durch eine geeignete Nummerierung. Die Anzahl der möglichen Nummerierungen ergibt die obere Schranke.  $\square$

**Lemma 1.7.** (a) Für alle  $\nu \in \widetilde{\mathcal{M}}_1(\Gamma^m)$  existiert eine Folge  $\nu_n \in \frac{1}{n}\widetilde{K}_n$  für  $n \in \mathbb{N}$ , so dass  $\lim_{n \rightarrow \infty} d(\nu_n, \nu) = 0$ .

$$\widetilde{K}_n := \left\{ k = (k_{s_1, s_2, \dots, s_m}) \in \mathbb{N}_0^{(N^m)} : \sum_{s_m} k_{s_1, \dots, s_{m-1}, s_m} = \sum_{s_m} k_{s_m, s_1, \dots, s_{m-1}}, \right. \\ \left. \sum_{s_1, s_2, \dots, s_m} k_{s_1, \dots, s_m} = n \right\}.$$

(b) Die Abbildung  $I_\rho^m : \widetilde{\mathcal{M}}_1(\Gamma^m) \rightarrow \mathbb{R}_{\geq 0}$  ist wohldefiniert, stetig und streng konvex, mit Ausnahme von Strecken  $\{t\nu + (1-t)\mu : t \in [0, 1]\}$  mit  $\nu_{s_1, s_2, \dots, s_m} / \bar{\nu}_{s_1, \dots, s_{m-1}} = \mu_{s_1, s_2, \dots, s_m} / \bar{\mu}_{s_1, \dots, s_{m-1}}$  für alle  $(s_1, \dots, s_m) \in \Gamma^m$ , entlang denen  $I_\rho^m$  linear ist, das heißt  $I_\rho^m(t\nu + (1-t)\mu) = tI_\rho^m(\nu) + (1-t)I_\rho^m(\mu)$ . Weiter gilt:  $I_\rho^m(\nu) = 0$  genau dann, wenn  $\nu = \rho^{\otimes m}$ , wobei

$$I_\rho^m(\nu) := \sum_{s_1, s_2, \dots, s_m} \nu_{s_1, \dots, s_m} \log \left( \frac{\nu_{s_1, \dots, s_m}}{\bar{\nu}_{s_1, \dots, s_{m-1}} \rho_{s_m}} \right)$$

mit  $\nu = (\nu_{s_1, s_2, \dots, s_m})$  und  $\bar{\nu}_{s_1, \dots, s_{m-1}} := \sum_{s_m} \nu_{s_1, \dots, s_m}$ .

*Beweis.* Zu Teil (a): wir definieren den affinen Unterraum von  $\mathbb{R}^{(N^m)}$

$$\widetilde{E} = \left\{ x = (x_{s_1, s_2, \dots, s_m}) \in \mathbb{R}^{(N^m)} : \sum_{s_m} x_{s_1, \dots, s_{m-1}, s_m} = \sum_{s_m} x_{s_m, s_1, \dots, s_{m-1}}, \right. \\ \left. \sum_{s_1, s_2, \dots, s_m} x_{s_1, \dots, s_m} = 1 \right\}.$$

Weil die Variablen  $x_{s_1, s_2, \dots, s_m}$  in den linearen Bestimmungsgleichungen ganzzahlige Koeffizienten haben, existiert eine Basis  $(v_1, v_2, \dots, v_{N^m})$  von  $\mathbb{R}^{(N^m)}$  mit  $v_i \in \mathbb{Z}^{(N^m)}$  für alle  $i = 1, \dots, N^m$  und es gilt, dass  $\widetilde{E}$  ein affiner Raum über  $\text{Span}(v_1, \dots, v_q)$  mit  $q < N^m$  ist. Da  $\emptyset \neq \frac{1}{n}\widetilde{K}_n \subset \widetilde{E}$  für alle  $n \in \mathbb{N}$ , existiert ein  $x_n \in \frac{1}{n}\widetilde{K}_n$  und für alle  $z_1, z_2, \dots, z_q \in \mathbb{Z}$  gilt  $x_n + \sum_{i=1}^q \frac{z_i}{n} v_i \in \frac{1}{n}\widetilde{Z}_n$  mit  $\widetilde{Z}_n = n\widetilde{E} \cap \mathbb{Z}^{(N^m)}$ . Für ein beliebiges  $x \in \widetilde{E}_n$  rechnen wir:



$$\begin{aligned}
\left\| x_n + \sum_{i=1}^q \frac{z_i}{n} v_i - x \right\|_1 &= \frac{1}{n} \left\| \underbrace{n(x_n - x)}_{\in \text{Span}(v_1, \dots, v_q)} + \sum_{i=1}^q z_i v_i \right\|_1 = \frac{1}{n} \left\| \sum_{i=1}^q \alpha_i v_i + z_i v_i \right\|_1 \\
&\leq \frac{1}{n} \sum_{i=1}^q \|\alpha_i v_i + z_i v_i\|_1 = \frac{1}{n} \sum_{i=1}^q \underbrace{|\alpha_i + z_i|}_{\leq 1} \|v_i\|_1 \\
&\leq \frac{q}{n} \max\{\|v_i\|_1 : i = 1, 2, \dots, q\} =: \frac{1}{n} C,
\end{aligned}$$

wobei wir  $n(x_n - x)$  in die Basis  $(v_1, v_2, \dots, v_q)$  entwickelt haben mit Entwicklungskoeffizienten  $\alpha_i \in \mathbb{R}$  für  $i = 1, 2, \dots, q$ , und wir bestimmen die ganzen Zahlen  $z_1, z_2, \dots, z_q$  so, dass  $|\alpha_i + z_i| \leq 1$  für alle  $i = 1, 2, \dots, q$ .

Wir folgern, dass  $d_1(x, \frac{1}{n} \tilde{Z}_n) \leq \frac{C}{n}$  für eine Konstante  $C > 0$  und mit der Metrik  $d_1$ , die von der 1-Norm erzeugt wird. Das impliziert die Aussage aus (a) für Maße  $\nu \in \tilde{\mathcal{M}}_1(\Gamma^m)$ , deren Zähldichte positiv ist, d.h.  $\nu_{s_1, s_2, \dots, s_m} > 0$  für alle  $(s_1, s_2, \dots, s_m) \in \Gamma^m$ .

Sei nun  $\nu \in \tilde{\mathcal{M}}_1(\Gamma^m)$  beliebig, dann existiert ein  $M \in \mathbb{N}$  und eine Folge  $\nu^n \in \tilde{\mathcal{M}}_1(\Gamma^m)$  für  $n \geq M$  so, dass für die Zähldichte  $\nu_{s_1, s_2, \dots, s_m}^n \geq \frac{C}{n}$  für alle  $(s_1, s_2, \dots, s_m) \in \Gamma^m$  gilt und  $\lim_{n \rightarrow \infty} d(\nu, \nu^n) = 0$ . Diese Folge können wir zum Beispiel auf der Strecke zwischen der Gleichverteilung und  $\nu$  wählen. Zu  $\nu^n$  existiert nach obiger Überlegung ein  $k_n \in \tilde{Z}_n$  so, dass  $d(\nu^n, \frac{1}{n} k_n) \leq \frac{C}{2n}$ . Damit liegt  $k_n$  wegen der Bedingung an die Zähldichte von  $\nu^n$  sogar in  $\tilde{K}_n$ . Dann konvergiert  $\frac{1}{n} k_n$  gegen  $\nu$ , denn für jedes  $\epsilon > 0$

$$d(\nu, \frac{1}{n} k_n) \leq d(\nu, \nu^n) + d(\nu^n, \frac{1}{n} k_n) \leq \frac{\epsilon}{2} + \frac{C}{n} \leq \epsilon,$$

falls  $n$  groß genug gewählt wird.

Um Teil (b) zu zeigen, nehmen wir ein  $(\lambda_{s_1, s_2, \dots, s_m})$  aus dem Translationsraum von  $\tilde{\mathcal{M}}_1(\Gamma^m)$ , also  $\sum_{s_1, s_2, \dots, s_m} \lambda_{s_1, s_2, \dots, s_m} = 0$  und  $\sum_{s_m} \lambda_{s_1, s_2, \dots, s_m} = \sum_{s_m} \lambda_{s_m s_1, \dots, s_{m-1}}$ . Zu zeigen ist, dass  $\phi(t) = I_p^m(\nu + t\lambda)$  für ein genügend kleines positives  $t$  konvex ist für ein beliebiges  $\nu \in \tilde{\mathcal{M}}(\Gamma^m)$ . Zu einem  $\nu \in \tilde{\mathcal{M}}(\Gamma^m) \subset [0, 1]^{N^m}$  am Rand von  $[0, 1]^{N^m}$  dürfen wir natürlich  $\lambda$  nur so wählen, dass  $\nu + t\lambda \in [0, 1]^{N^m}$  für genügend kleines positive  $t$  gilt. Es genügt offensichtlich zu zeigen, dass  $\psi_{s_1, \dots, s_m}(t) = (\nu_{s_1, \dots, s_m} + t\lambda_{s_1, s_2, \dots, s_m}) \log \left( \frac{\nu_{s_1, \dots, s_m} + t\lambda_{s_1, s_2, \dots, s_m}}{\bar{\nu}_{s_1, \dots, s_{m-1}} + t\lambda_{s_1, s_2, \dots, s_m}} \right)$  für alle  $(s_1, \dots, s_m) \in \Gamma^m$  konvex ist, wobei wir den linearen Teil der Ratenfunktion abgespalten haben. Um die Notation zu vereinfachen setzen wir  $\psi(t) = (\nu + t\lambda) \log((\nu + t\lambda)/(\bar{\nu} + t\lambda))$  und leiten diese reelle Funktion zweimal ab:

$$\begin{aligned}
\psi'(t) &= \lambda \log \left( \frac{\nu + t\lambda}{\bar{\nu} + t\bar{\lambda}} \right) + \lambda - \bar{\lambda} \frac{\nu + t\lambda}{\bar{\nu} + t\bar{\lambda}} \\
\psi''(t) &= \lambda \left( \frac{\lambda}{\nu + t\lambda} - \frac{\bar{\lambda}}{\bar{\nu} + t\bar{\lambda}} \right) - \bar{\lambda} \left( \frac{\lambda(\bar{\nu} + t\bar{\lambda}) - \bar{\lambda}(\nu + t\lambda)}{(\bar{\nu} + t\bar{\lambda})^2} \right) \\
&= \frac{(\lambda\bar{\nu} - \bar{\lambda}\nu)^2}{(\bar{\nu} + t\bar{\lambda})^2(\nu + t\lambda)} \geq 0,
\end{aligned}$$

wobei die Ungleichung für  $t$  klein genug gilt. Ist nämlich  $\nu = 0$ , dann betrachten wir nach obiger Bemerkung nur nichtnegatives  $\lambda$ . Für  $\lambda = 0$  spielt dieser Summand wegen  $0 \log 0 = 0$  keine Rolle. Gleichheit besteht genau dann, wenn  $\nu/\bar{\nu} = \lambda/\bar{\lambda}$ . Also ist  $I_\rho$  genau entlang der Strecken linear, wo  $\nu_{s_1, s_2, \dots, s_m} / \bar{\nu}_{s_1, \dots, s_{m-1}}$  konstant ist für alle  $(s_1, \dots, s_m) \in \Gamma^m$ , denn  $\frac{d}{dt}((\nu + t\lambda)/(\bar{\nu} + t\bar{\lambda})) = (\lambda\bar{\nu} - \bar{\lambda}\nu)/(\bar{\nu} + t\bar{\lambda})^2$ .

Die Aussage über den Wertebereich erhalten wir mit Anwendung der Jenschen'schen Ungleichung für die konvexe Funktion  $h(x) = x \log(x)$ :

$$I_\rho^m(\nu) = \sum_{s_1, s_2, \dots, s_m} \bar{\nu}_{s_1, \dots, s_{m-1}} \rho_{s_m} h \left( \frac{\nu_{s_1, \dots, s_m}}{\bar{\nu}_{s_1, \dots, s_{m-1}} \rho_{s_m}} \right) \geq h \left( \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \right) = h(1) = 0.$$

Gleichheit gilt genau dann, wenn  $\frac{\nu_{s_1, \dots, s_m}}{\bar{\nu}_{s_1, \dots, s_{m-1}} \rho_{s_m}} = C \in \mathbb{R}$  für alle  $(s_1, \dots, s_m)$  mit  $\nu_{s_1, \dots, s_m} \neq 0$ . Dann gilt nämlich  $C = 1$  und daher  $\nu_{s_1, \dots, s_m} = \bar{\nu}_{s_1, \dots, s_{m-1}} \rho_{s_m}$  für alle  $(s_1, \dots, s_m) \in \Gamma^m$ .

Sei  $\mu \in \widetilde{\mathcal{M}}(\Gamma^{M+1})$ , dann ist das Wahrscheinlichkeitsmaß  $\bar{\mu} \in \mathcal{M}_1(\Gamma^M)$  zyklisch, denn

$$\begin{aligned}
\sum_{s_M} \bar{\mu}_{s_1, s_2, \dots, s_M} &= \sum_{s_M, s_{M+1}} \mu_{s_1, s_2, \dots, s_M, s_{M+1}} = \sum_{s_{M+1}, s_M} \mu_{s_{M+1}, s_1, \dots, s_M} = \sum_{s_{M+1}} \bar{\mu}_{s_{M+1}, s_1, \dots, s_{M-1}} \\
&= \sum_{s_M} \bar{\mu}_{s_M, s_1, \dots, s_{M-1}},
\end{aligned}$$

also  $\bar{\mu} \in \widetilde{\mathcal{M}}_1(\Gamma^M)$ .

Wir zeigen durch Iteration, dass  $\nu_{s_1, \dots, s_m} = \rho_{s_1} \rho_{s_2} \dots \rho_{s_m}$ . Dazu setzen wir  $\nu_{s_1, s_2, \dots, s_n}^n = \sum_{s_{n+1}, s_{n+2}, \dots, s_m} \nu_{s_1, s_2, \dots, s_m}$ . Dann gilt  $\nu^n \in \widetilde{\mathcal{M}}_1(\Gamma^n)$ . Der Anfangsschritt der Iteration ist die bereits gezeigte Gleichheit  $\nu_{s_1, \dots, s_m}^m = \rho_{s_m} \nu_{s_1, \dots, s_{m-1}}^{m-1}$ , dann gilt:

$$\begin{aligned}
\nu_{s_1, \dots, s_m}^m &\stackrel{(*)}{=} \rho_{s_m} \nu_{s_1, \dots, s_{m-1}}^{m-1} = \rho_{s_m} \sum_t \nu_{s_1, \dots, s_{m-1}, t}^m = \rho_{s_m} \sum_t \nu_{t, s_1, \dots, s_{m-1}}^m \\
&\stackrel{(*)}{=} \rho_{s_m} \sum_t \nu_{t, s_1, \dots, s_{m-2}}^{m-1} \rho_{s_{m-1}} = \rho_{s_{m-1}} \rho_{s_m} \nu_{s_1, \dots, s_{m-2}}^{m-2}.
\end{aligned}$$

Somit folgt, dass  $\nu_{s_1, \dots, s_{m-1}}^{m-1} = \rho_{s_{m-1}} \nu_{s_1, \dots, s_{m-2}}^{m-2}$ . In dem  $n$ -ten Schritt rechnet man

$$\begin{aligned} \nu_{s_1, \dots, s_{m-n+1}}^{m-n+1} &\stackrel{(**)}{=} \rho_{s_{m-n+1}} \nu_{s_1, \dots, s_{m-n}}^{m-n} = \rho_{s_{m-n+1}} \sum_t \nu_{s_1, \dots, s_{m-n}, t}^{m-n+1} \\ &= \rho_{s_{m-n+1}} \sum_t \nu_{t, s_1, \dots, s_{m-n}}^{m-n+1} \stackrel{(**)}{=} \rho_{s_{m-n+1}} \sum_t \rho_{s_{m-n}} \nu_{t, s_1, \dots, s_{m-n-1}}^{m-n} \\ &= \rho_{s_{m-n}} \rho_{s_{m-n+1}} \sum_t \nu_{s_1, \dots, s_{m-n-1}, t}^{m-n} = \rho_{s_{m-n}} \rho_{s_{m-n+1}} \nu_{s_1, \dots, s_{m-n-1}}^{m-n-1}. \end{aligned}$$

Daher impliziert  $I_\rho^m(\nu) = 0$ , dass  $\nu = \rho^{\otimes m}$ . Die umgekehrte Richtung folgt aus der Definition von  $I_\rho^m$ . Die Stetigkeit von  $I_\rho^m$  folgt aus der Stetigkeit von  $h$ .  $\square$

Nun zeigen wir eine Verallgemeinerung von Satz 1.3 für den Produktphasenraum  $\Gamma^m$ .

**Satz 1.8.** *Seien  $X_1, X_2, \dots$  wie in (1). Weiter sei  $L_n^m = \frac{1}{n} \sum_{i=1}^n \delta_{(\tilde{X}_i, \dots, \tilde{X}_{i+m-1})}$  das  $m$ -dimensionale empirische Maß mit periodischen Randbedingungen ( $\tilde{X}_s = X_{[s]}$  mit  $[s] = \min\{i \in \mathbb{N} : i \equiv s \pmod{n}\}$ ) und  $A \subset \widetilde{\mathcal{M}}_1(\Gamma^m)$ . Dann gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in A) = - \inf_{\nu \in A} I_\rho^m(\nu),$$

wobei

$$I_\rho^m(\nu) := \sum_{s_1, s_2, \dots, s_m} \nu_{s_1, \dots, s_m} \log \left( \frac{\nu_{s_1, \dots, s_m}}{\bar{\nu}_{s_1, \dots, s_{m-1}} \rho_{s_m}} \right)$$

mit  $\nu = (\nu_{s_1, s_2, \dots, s_m})$  und  $\bar{\nu}_{s_1, \dots, s_{m-1}} := \sum_{s_m} \nu_{s_1, \dots, s_m}$ .

*Beweis.* In dem Beweis stützen wir uns auf Ideen aus [3, Beweis von *Theorem II. 8*] und definieren, wie in Lemma 1.5 die Menge

$$\widehat{K}_n = \left\{ k \in \widetilde{K}_n : G(k) \text{ ist stark zusammenhängend} \right\}.$$

Dann gilt für ein  $k = (k_{s_1, \dots, s_m}) \in \widehat{K}_n$ , dass die Wahrscheinlichkeit

$$\mathbb{P} \left( L_n^m(\{(s_1, \dots, s_m)\}) = \frac{k_{s_1, \dots, s_m}}{n} \quad \forall (s_1, \dots, s_m) \in \Gamma^m \right) = \kappa \prod_{i=1}^N \rho_i^{\hat{k}_i}$$

mit  $\hat{k}_i = \sum_{s_1, \dots, s_{m-1}} k_{s_1, \dots, s_{m-1}, i}$  und einem kombinatorischen Faktor  $\kappa$ , der die Anzahl der Sequenzen  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  zählt, die alle dasselbe Wahrscheinlichkeitsmaß  $\frac{k}{n}$  erzeugen. Wir wollen diesen Faktor  $\kappa$  abschätzen. Für  $\kappa$  gilt:

$$\kappa = \gamma \frac{\mathcal{E}(G(k))}{\prod_{s_1, \dots, s_m} k_{s_1, \dots, s_m}!},$$

wobei  $\mathcal{E}(G(k))$  die Anzahl der Eulerkreise in  $G(k)$  bezeichnet, die durch die Anzahl der Permutationen der parallelen Kanten geteilt wird, um diese nicht zu unterscheiden. Der Faktor  $\gamma$  zählt die Anzahl der verschiedenen möglichen Anfangspunkte eines Eulerkreises, so dass die zugehörigen Sequenzen  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  unterschiedlich sind. Daher ist  $\gamma \in [1, n]$ . Weil  $\bar{k}_{s_1, \dots, s_{m-1}}$  der Ausgangsgrad von  $(s_1, \dots, s_{m-1})$  ist, gilt nach *Lemma 1.6*:

$$\prod_{s_1, \dots, s_{m-1}} (\bar{k}_{s_1, \dots, s_{m-1}} - 1)! \leq \mathcal{E}(G(k)) \leq \prod_{s_1, \dots, s_{m-1}} \bar{k}_{s_1, \dots, s_{m-1}}!.$$

Daher ist  $\mathcal{E}(G(k)) = e^{\mathcal{O}(\log(n))} \prod_{s_1, \dots, s_{m-1}} \bar{k}_{s_1, \dots, s_{m-1}}!$  und wegen  $\gamma \in [1, n]$  gilt insgesamt:

$$\kappa = e^{\mathcal{O}(\log(n))} \frac{\prod_{s_1, \dots, s_{m-1}} \bar{k}_{s_1, \dots, s_{m-1}}!}{\prod_{s_1, \dots, s_m} k_{s_1, \dots, s_m}!} \quad \text{für } n \rightarrow \infty,$$

wobei die asymptotische obere Schranke gleichmäßig auf  $\widehat{K}_n$  ist. Wir schätzen die Wahrscheinlichkeiten  $\log(\mathbb{P}(L_n^m = k/n))$  mit Hilfe der Stirlingformel:

$$\begin{aligned} \log \left( \mathbb{P} \left( L_n^m = \frac{k}{n} \right) \right) &= \log \left( e^{\mathcal{O}(\log(n))} \frac{\prod_{s_1, \dots, s_{m-1}} \bar{k}_{s_1, \dots, s_{m-1}}!}{\prod_{s_1, \dots, s_m} k_{s_1, \dots, s_m}!} \prod_{i=1}^N \rho_i^{\hat{k}_i} \right) \\ &= \mathcal{O}(\log n) + \sum_{s_1, \dots, s_{m-1}} \log(\bar{k}_{s_1, \dots, s_{m-1}}!) + \sum_{i=1}^N \log(\rho_i^{\hat{k}_i}) - \sum_{s_1, \dots, s_m} \log(k_{s_1, \dots, s_m}!) \\ &= \sum_{s_1, \dots, s_{m-1}} (\bar{k}_{s_1, \dots, s_{m-1}} \log \bar{k}_{s_1, \dots, s_{m-1}} - \bar{k}_{s_1, \dots, s_{m-1}} + \mathcal{O}(\log \bar{k}_{s_1, \dots, s_{m-1}})) \\ &\quad - \sum_{s_1, \dots, s_m} (k_{s_1, \dots, s_m} \log k_{s_1, \dots, s_m} - k_{s_1, \dots, s_m} + \mathcal{O}(\log k_{s_1, \dots, s_m})) + \sum_{i=1}^N \hat{k}_i \log \rho_i + \mathcal{O}(\log n) \\ &= \sum_{s_1, \dots, s_m} k_{s_1, \dots, s_m} (\log \bar{k}_{s_1, \dots, s_{m-1}} + \log \rho_{s_m} - \log k_{s_1, \dots, s_m}) + \mathcal{O}(\log n) \\ &= n \sum_{s_1, \dots, s_m} k_{s_1, \dots, s_m} / n \log \left( \frac{\bar{k}_{s_1, \dots, s_{m-1}} / n}{k_{s_1, \dots, s_m} / n} \rho_{s_m} \right) + \mathcal{O}(\log n) = -n I_\rho^m(k/n) + \mathcal{O}(\log n) \end{aligned}$$

Aus *Lemma 1.5* folgt folgende Abschätzung:

$$Q_n(A) \leq \mathbb{P}(L_n^m \in A) \leq |\widehat{K}_n| Q_n(A),$$

wobei  $Q_n(A)$  als

$$Q_n(A) = \max_{k \in \widehat{K}_n, \frac{k}{n} \in A} \mathbb{P} \left( L_n^m = \frac{k}{n} \right)$$

definiert ist. Zudem setzen wir  $Q_n(A) = 0$ , falls  $A \cap \frac{\widehat{K}_n}{n} = \emptyset$ . Für die Mächtigkeit der Menge  $\widehat{K}_n$  gilt:

$$|\widehat{K}_n| \leq \left| \left\{ k = (k_{s_1, s_2, \dots, s_m}) \in \mathbb{N}_0^{(N^m)} : \sum_{s_1, s_2, \dots, s_m} k_{s_1, \dots, s_m} = n \right\} \right| = \binom{n + N^m - 1}{N^m - 1}.$$

Wir folgern wie im Beweis von *Satz 1.3*, dass  $\frac{1}{n} \log(|\widehat{K}_n|) = \mathcal{O} \left( \frac{\log n}{n} \right)$  für  $n \rightarrow \infty$ . Aus *Lemma 1.7 (a)* folgern wir, dass für jedes  $\nu \in \widetilde{\mathcal{M}}_1(\Gamma^m)$  eine Folge  $\hat{\nu}_n \in \frac{1}{n} \widehat{K}_n$  für  $n \in \mathbb{N}$  so existiert, dass  $\lim_{n \rightarrow \infty} d(\hat{\nu}_n, \nu) = 0$ . Wir können nämlich zu der Knotenmenge  $\Gamma^{m-1}$  gerichtete Kanten  $E$  nehmen, so dass der Graph  $(\Gamma^{m-1}, E)$  stark zusammenhängt und das zu  $(\Gamma^{m-1}, E)$  gehörende Maß  $k$  in  $\frac{1}{r} \widehat{K}_r$  liegt für ein  $r \in \mathbb{N}$ . Wir nehmen eine Folge  $\nu_n \in \frac{1}{n} \widehat{K}_n$  wie in *Lemma 1.7 (a)* und definieren  $\hat{\nu}_{n+r}$  durch  $\hat{\nu}_{n+r} := (n\nu_n + k)/(n+r)$ . Dann liegt  $\hat{\nu}_{n+r}$  in  $\frac{1}{n+r} \widehat{K}_{n+r}$  und  $\lim_{n \rightarrow \infty} d(\hat{\nu}_{n+r}, \nu) = 0$ . Den Beweis beenden wir wie den von *Satz 1.3*, wozu wir noch die Stetigkeit von  $I_\rho^m$  aus *Lemma 1.7 (b)* brauchen.  $\square$



## 2 Das Prinzip großer Abweichungen

In dem ersten Abschnitt haben wir zwei typische Sätze aus der Theorie von großen Abweichungen gesehen. Jetzt formulieren wir allgemeiner das Prinzip großer Abweichungen für polnische Räume und zeigen anschließend, dass das  $m$ -dimensionale empirische Maß  $L_n^m$  mit periodischen Randbedingungen dieses Prinzip erfüllt. Ab jetzt sei  $(E, d)$  ein polnischer Raum.

**Definition.** Eine Funktion  $f : E \rightarrow [-\infty, \infty]$  nennen wir *unterhalbstetig*, falls  $\liminf_{n \rightarrow \infty} f(x_n) \geq f(x)$  für jede Folge  $(x_n)$  in  $E$  mit  $\lim_{n \rightarrow \infty} x_n = x$ .

In der reellen Analysis zeigt man, dass eine Funktion genau dann unterhalbstetig ist, wenn ihre Niveaumengen abgeschlossen sind, also  $f^{-1}([-\infty, c])$  abgeschlossen ist für jedes  $c \in \mathbb{R}$ .

**Definition.** Eine Funktion  $I : E \rightarrow [0, \infty]$  nennen wir *Ratenfunktion*, wenn

- (i)  $I \not\equiv \infty$ .
- (ii)  $I$  besitzt kompakte Niveaumengen.

Insbesondere sind Ratenfunktionen unterhalbstetig. Nun können wir das Prinzip großer Abweichungen formulieren.

**Definition.** Eine Folge von Borel-Wahrscheinlichkeitsmaßen  $(P_n)_{n \in \mathbb{N}}$  auf dem polnischen Raum  $(E, d)$  erfüllt *das Prinzip großer Abweichungen* mit Rate  $n$  und Ratenfunktion  $I$ , wenn

- (i)  $I$  eine Ratenfunktion ist.
- (ii)  $\limsup_{n \rightarrow \infty} \frac{1}{n} \log(P_n(A)) \leq -I(A) \quad \forall A \subset E$  abgeschlossen.
- (iii)  $\liminf_{n \rightarrow \infty} \frac{1}{n} \log(P_n(O)) \geq -I(O) \quad \forall O \subset E$  offen,

mit  $I(S) = \inf_{x \in S} I(x)$  für eine Teilmenge  $S \subset E$ .

**Satz 2.1.** *Mit den Voraussetzungen aus (1) erfüllt  $(\mathbb{P}(L_n^m \in \cdot))_{n \in \mathbb{N}}$ , das Bildmaß des  $m$ -dimensionalen empirischen Maßes  $(L_n^m)_{n \in \mathbb{N}}$  mit periodischen Randbedingungen, das Prinzip großer Abweichungen mit Rate  $n$  und Ratenfunktion  $I_\rho^m$ .*

*Bemerkung:* Diesen Satz werden wir auch als der Satz von Sanov bezeichnen. Mit Satz 1.8 wissen wir außerdem, dass für offene Mengen  $O \subset \widetilde{\mathcal{M}}_1(\Gamma^m)$  die obere und die untere Schranke übereinstimmen. In diesem Fall haben wir sogar Konvergenz gegen  $-I_\rho^m(O)$ . Die Folge  $\frac{1}{n} \log(P_n(S))$  konvergiert auch, falls  $I_\rho^m(\text{int}(S)) = I_\rho^m(\overline{S})$ . Solche Mengen  $S$  nennen wir  $I_\rho^m$  stetig.

*Beweis.* Aus  $\widetilde{\mathcal{M}}_1(\Gamma^m)$  kompakt und aus *Lemma 1.7* folgern wir, dass  $I_\rho^m$  eine Ratenfunktion auf dem polnischen Raum  $(\widetilde{\mathcal{M}}_1(\Gamma^m), d)$  ist. Die untere Schranke (iii) folgt aus *Satz 1.8*. Sei nun  $A \subset \widetilde{\mathcal{M}}_1(\Gamma^m)$  abgeschlossen, dann ist  $A$  kompakt. Sei  $\delta > 0$ . Wegen der Stetigkeit von  $I_\rho^m$  aus *Lemma 1.7 (b)* finden wir für jedes  $\nu \in A$  ein  $\delta_\nu$ , so dass

$$I_\rho^m(B_{\delta_\nu}(\nu)) = \inf_{\mu \in B_{\delta_\nu}(\nu)} I_\rho^m(\mu) \geq I_\rho^m(\nu) - \delta \quad (4)$$

mit  $B_{\delta_\nu}(\nu) = \{\mu \in \widetilde{\mathcal{M}}_1(\Gamma^m) : d(\mu, \nu) < \delta_\nu\}$  der offene Ball von Radius  $\delta_\nu$  um  $\nu$ . Wegen der Kompaktheit von  $A$  besitzt die offene Überdeckung  $(B_{\delta_\nu}(\nu))_{\nu \in A}$  von  $A$  eine endliche Teilüberdeckung  $(B_{\delta_{\nu_i}}(\nu_i))_{1 \leq i \leq k}$ .

Zum nächsten Schritt brauchen wir folgende Überlegung. Sind  $(a_n)$  und  $(b_n)$  Folgen positiver reellen Zahlen, dann gilt

$$\log(a_n + b_n) \leq \log(2 \max\{a_n, b_n\}) = \log 2 + \log(\max\{a_n, b_n\})$$

und daher

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log(a_n + b_n) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log(\max\{a_n, b_n\}) \\ &= \max\left\{ \limsup_{n \rightarrow \infty} \frac{1}{n} \log(a_n), \limsup_{n \rightarrow \infty} \frac{1}{n} \log(b_n) \right\}. \end{aligned}$$

Somit folgt mit  $B_i = B_{\delta_{\nu_i}}(\nu_i)$  und mit *Satz 1.8*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in A) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in \cup_{i=1}^k B_i) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^k \mathbb{P}(L_n^m \in B_i) \\ &\leq \max_{i=1,2,\dots,k} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in B_i) = \max_{i=1,2,\dots,k} -I_\rho^m(B_i). \end{aligned}$$

Mit (4) folgern wir  $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in A) \leq -\inf_{\nu \in A} I_\rho^m(\nu) + \delta$ . Anschließend folgt mit  $\delta \rightarrow 0$  die Behauptung des Satzes.  $\square$



### 3 Das Prinzip der Gibbs-Konditionierung

In diesem Abschnitt geben wir eine Anwendung für den Satz von Sanov *Satz 1.3* beziehungsweise *Satz 2.1*, die in der statistischen Mechanik von großer Bedeutung ist. Sei nun wieder (1) vorausgesetzt. Wir stellen uns die Elemente von  $\Gamma = \{1, 2, \dots, N\}$  als die möglichen Zustände eines Teilchens vor.  $X_i$  beschreibt den Zustand des  $i$ -ten Teilchens. Jedes Teilchen befindet sich mit der Wahrscheinlichkeit  $\rho_s > 0$  in dem Zustand  $s \in \Gamma$ .

Sei  $f : \Gamma \rightarrow \mathbb{R}$  eine Funktion.  $f(s)$  interpretieren wir als die Energie des Zustands  $s \in \Gamma$ . Wir setzen  $Y_i = f(X_i)$  und  $\hat{S}_n = \frac{1}{n} \sum_{i=1}^n Y_i$  als die mittlere Energie. Eine für die statistische Mechanik wichtige Fragestellung ist folgendes: sei  $A \subset \mathbb{R}$  eine Menge. Was ist die bedingte Verteilung von  $X_1$  gegeben die mittlere Energie  $\hat{S}_n \in A$  für großes  $n$ ? Wir interessieren uns also für die Häufungspunkte der Folge  $(\mu_n)_{n \in \mathbb{N}}$  mit  $\mu_n \in \mathcal{M}_1(\Gamma)$  ein deterministisches Wahrscheinlichkeitsmaß definiert durch bedingte Wahrscheinlichkeiten

$$\mu_n(\{s\}) = \mathbb{P}(X_1 = s | \hat{S}_n \in A) \quad s \in \Gamma.$$

Wir setzen dabei voraus, dass das Ereignis  $\{\hat{S}_n \in A\}$  positive Wahrscheinlichkeit hat. Offensichtlich haben wir folgenden Zusammenhang:

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^n f(X_i) = \int_{\Gamma} f \, dL_n$$

mit dem empirischen Maß  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ . Wir erinnern uns, dass  $L_n$  ein zufälliges Wahrscheinlichkeitsmaß ist, also eine Abbildung  $\Omega \rightarrow \mathcal{M}_1(\Gamma)$ . Es handelt sich demnach um ein zufälliges Integral auf der rechten Seite. Da  $(X_i)_{i \leq n}$  i. i. d. sind, haben wir für jede Testfunktion  $\phi : \Gamma \rightarrow \mathbb{R}$

$$\begin{aligned} \int_{\Gamma} \phi \, d\mu_n &= \mathbb{E}[\phi(X_1) | \hat{S}_n \in A] = \mathbb{E}[\phi(X_2) | \hat{S}_n \in A] \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \phi(X_i) | \hat{S}_n \in A\right] = \mathbb{E}\left[\int_{\Gamma} \phi \, dL_n \mid \int_{\Gamma} f \, dL_n \in A\right]. \end{aligned}$$

Setzt man jetzt  $\phi = \delta_s$  für  $s \in \Gamma$  und fasst  $L_n$  als ein zufälliges und  $\mu_n$  als ein deterministisches Element in  $[0, 1]^N$  auf, so erhält man

$$\mu_n = \mathbb{E}[L_n | L_n \in \Sigma]$$

mit  $\Sigma = \{\nu \in \mathcal{M}_1(\Gamma) : \int_{\Gamma} f \, d\nu \in A\}$ .

Der folgende Satz aus [2, *Theorem 3.3.3*] charakterisiert die Häufungspunkte der Folge  $(\mu_n)$  unter den Voraussetzungen  $\Sigma \neq \emptyset$  und  $\Sigma$  ist  $I_\rho$ -stetig, d.h.

$$I_\Sigma := \inf_{\nu \in \text{int}(\Sigma)} I_\rho(\nu) = \inf_{\nu \in \overline{\Sigma}} I_\rho(\nu).$$

**Satz 3.1** (Das Gibbs-Prinzip). *Sei  $\mathcal{M} = \{\nu \in \overline{\Sigma} : I_\rho(\nu) = I_\Sigma\}$ , wobei  $\emptyset \neq \Sigma \subset \mathcal{M}_1(\Gamma)$   $I_\rho$ -stetig ist. Dann gilt:*

- (a) *Jeder Häufungspunkt von  $(\mu_n)$  liegt in  $\overline{\text{conv}}(\mathcal{M})$ , in dem Abschluss der konvexen Hülle von  $\mathcal{M}$ .*
- (b) *Ist  $\Sigma$  konvex mit nichtleerem Inneren, dann besteht  $\mathcal{M}$  aus einem einzigen Punkt, gegen den  $(\mu_n)$  für  $n \rightarrow \infty$  konvergiert.*

*Beweis.* In dem Beweis folgen wir den Darlegungen in [2, Beweis von *Theorem 3.3.3*]. Nach der Bemerkung, die auf den Beweis von *Lemma 1.2* folgt, ist  $\mathcal{M}_1(\Gamma)$  und daher auch  $\overline{\Sigma}$  kompakt. Nun folgt aus der Stetigkeit von  $I_\rho$  aus *Lemma 1.2 (b)*, dass  $\mathcal{M} \neq \emptyset$ . Sei  $U \subset \mathcal{M}_1(\Gamma)$ , dann folgt mit  $\{L_n \in \Sigma\} = \{L_n \in \Sigma \cap U\} \dot{\cup} \{L_n \in \Sigma \cap U^c\}$

$$\begin{aligned} & \mathbb{E}[L_n | L_n \in \Sigma] - \mathbb{E}[L_n | L_n \in \Sigma \cap U] \\ &= \mathbb{P}(L_n \in U^c | L_n \in \Sigma) (\mathbb{E}[L_n | L_n \in \Sigma \cap U^c] - \mathbb{E}[L_n | L_n \in \Sigma \cap U]). \end{aligned}$$

Wir zeigen jetzt  $\mathbb{E}[L_n | L_n \in \Sigma \cap U] \in \overline{\text{conv}}(U)$ . Sei dazu  $\nu \in \mathcal{M}_1(\Gamma) \cap (\overline{\text{conv}}(U))^c$ . Wir zeigen  $\mathbb{E}[L_n | L_n \in \Sigma \cap U] \neq \nu$ . Da die Mengen  $\{\nu\}$  und  $\overline{\text{conv}}(U)$  kompakt, konvex und disjunkt sind, existiert eine separierende Hyperebene  $H$  in  $\mathbb{R}^N$ , die die beiden Mengen streng separiert, das heißt es gilt:  $\nu \notin H$ ,  $\overline{\text{conv}}(U) \cap H = \emptyset$  und  $\nu$  und  $\overline{\text{conv}}(U)$  liegen auf verschiedenen Seiten von  $H$ . Dreht man nun das Integral mit einer Matrix  $D \in O(N)$ , wobei  $D(H)$  senkrecht auf dem ersten kanonischen Basisvektor des  $\mathbb{R}^N$  steht, so folgern wir die Behauptung leicht aus der Beobachtung, dass das Integral und die lineare Abbildung  $D$  kommutieren, und mit Anwendung der Monotonieeigenschaft des Integrals für die erste Komponente.

Da  $\mathbb{E}[L_n | L_n \in \Sigma \cap U] \in \overline{\text{conv}}(U)$  und  $\mu_n = \mathbb{E}[L_n | L_n \in \Sigma]$  folgt, dass

$$\begin{aligned} d(\mu_n, \overline{\text{conv}}(U)) &\leq \mathbb{P}(L_n \in U^c | L_n \in \Sigma) d(\mathbb{E}[L_n | L_n \in \Sigma \cap U^c], \mathbb{E}[L_n | L_n \in \Sigma \cap U]) \\ &\leq \mathbb{P}(L_n \in U^c | L_n \in \Sigma), \end{aligned} \tag{5}$$

weil  $d$  auf  $\mathcal{M}_1(\Gamma) \times \mathcal{M}_1(\Gamma)$  durch 1 von oben beschränkt ist. Als Nächstes zeigen wir mit Hilfe des Satzes von Sanov *Satz 2.1*, dass für  $\delta > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(L_n \in \mathcal{M}^\delta | L_n \in \Sigma) = 1 \quad (6)$$

mit  $\mathcal{M}^\delta = \{\nu \in \mathcal{M}_1(\Gamma) : d(\nu, \mathcal{M}) < \delta\}$ . Mit dem Satz von Sanov folgt, dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Sigma) = -I_\Sigma$$

und, weil  $\mathcal{M}^\delta$  offen ist, folgern wir für jedes  $\delta > 0$  mit dem Satz von Sanov, dass

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Sigma \cap (\mathcal{M}^\delta)^c) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in \bar{\Sigma} \cap (\mathcal{M}^\delta)^c) \\ &\leq - \inf_{\nu \in \bar{\Sigma} \cap (\mathcal{M}^\delta)^c} I_\rho(\nu). \end{aligned}$$

Da  $\bar{\Sigma} \cap (\mathcal{M}^\delta)^c$  kompakt ist, existiert ein  $\tilde{\nu} \in \bar{\Sigma} \cap (\mathcal{M}^\delta)^c$ , so dass

$$\inf_{\nu \in \bar{\Sigma} \cap (\mathcal{M}^\delta)^c} I_\rho(\nu) = I_\rho(\tilde{\nu}) > I_\Sigma,$$

wobei die strikte Ungleichung aus der Tatsache folgt, dass das Infimum auf der linken Seite außerhalb von  $\mathcal{M}$  aber in  $\bar{\Sigma}$  angenommen wird.

Nun folgt (6), denn es gilt

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in (\mathcal{M}^\delta)^c | L_n \in \Sigma) \\ &= \limsup_{n \rightarrow \infty} \left( \frac{1}{n} \log \mathbb{P}(L_n \in \Sigma \cap (\mathcal{M}^\delta)^c) - \frac{1}{n} \log \mathbb{P}(L_n \in \Sigma) \right) \leq -I_\rho(\tilde{\nu}) + I_\Sigma < 0. \end{aligned}$$

Wenn man jetzt in (5)  $U = \mathcal{M}^\delta$  setzt, dann folgt  $d(\mu_n, \overline{\text{conv}}(\mathcal{M}^\delta)) \xrightarrow{n \rightarrow \infty} 0$ . Ist  $(\mu_{n_i})_{i \in \mathbb{N}}$  eine konvergente Teilfolge mit  $\lim_{i \rightarrow \infty} \mu_{n_i} = \mu \in \mathcal{M}_1(\Gamma)$ , dann folgt für jedes  $\delta > 0$  mit der Dreiecksungleichung  $d(\mu, \overline{\text{conv}}(\mathcal{M}^\delta)) = 0$ , weil

$$\begin{aligned} d(\mu, \overline{\text{conv}}(\mathcal{M}^\delta)) &= \inf_{\nu \in \overline{\text{conv}}(\mathcal{M}^\delta)} d(\mu, \nu) \leq \inf_{\nu \in \overline{\text{conv}}(\mathcal{M}^\delta)} (d(\mu, \mu_{n_i}) + d(\mu_{n_i}, \nu)) \\ &= d(\mu, \mu_{n_i}) + d(\mu_{n_i}, \overline{\text{conv}}(\mathcal{M}^\delta)) \xrightarrow{i \rightarrow \infty} 0. \end{aligned}$$

Dann gibt es für jedes  $\delta > 0$  ein  $\mu_\delta \in \text{conv}(\mathcal{M})$ , so dass  $d(\mu, \mu_\delta) < 2\delta$ . Damit liegt  $\mu$  in  $\overline{\text{conv}}(\mathcal{M})$ . Das war die Aussage von Teil (a).

Zu Teil (b): es genügt zu zeigen, dass für  $\Sigma$  konvex mit nichtleerem Inneren  $\mathcal{M}$  aus einem einzigen Punkt besteht. Die Konvergenz folgt dann aus (5) und (6), wie in Teil (a). Aus dem Beweis von Teil (a) folgt, dass  $\mathcal{M}$  nicht leer ist. Da der Abschluss einer konvexen Menge wieder konvex ist, liegt für  $\nu, \mu \in \bar{\Sigma}$  auch  $\frac{\nu+\mu}{2}$  wieder in derselben Menge. Angenommen  $\nu, \mu \in \mathcal{M}$  und  $\nu \neq \mu$ , dann folgt aus der strengen Konvexität der Ratenfunktion  $I_\rho$ , dass  $I_\rho\left(\frac{\nu+\mu}{2}\right) < \frac{1}{2}(I_\rho(\nu) + I_\rho(\mu)) = I_\Sigma$ . Das ist aber ein Widerspruch zur Minimalität von  $I_\Sigma$ . Somit ist  $\nu = \mu$ .  $\square$

## 4 Der empirische Prozess

In diesem Abschnitt werden wir die große Abweichung einer empirischen Verteilung auf  $\Gamma^{\mathbb{N}}$ , des so genannten empirischen Prozesses  $R_n$ , von der Verteilung  $\rho^{\otimes \mathbb{N}}$  untersuchen. Wie im ersten Abschnitt wird auch in diesem folgendes vorausgesetzt:

$$\begin{aligned} (X_i)_{i \in \mathbb{N}} \text{ sind i.i.d. Zufallsvariablen } \Omega \rightarrow \Gamma = \{1, 2, \dots, N\} \subset \mathbb{N}, \\ \text{die Verteilung von } X_1 \text{ besitzt die Zahldichte } (\rho_i)_{i \in \Gamma} \text{ mit } \rho_i > 0 \quad \forall i \in \Gamma. \end{aligned} \quad (1)$$

Wir wollen ein empirisches Ma  $R_n$  auf  $\Gamma^{\mathbb{N}}$  definieren. Empirisch soll dabei, wie im ersten Abschnitt, bedeuten, dass das Ma  $R_n$  von einem Index  $n \in \mathbb{N}$  abhangt und nur Information aus  $n$  unabhangigen Zufallsexperimenten enthalt, deren Ergebnisse  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  sind. Wir wollen ebenfalls den periodischen Charakter aus *Abschnitt 1.2* beibehalten. Wir definieren die Zufallsvariable  $\mathbb{X}^n : (\Omega, \mathcal{A}) \rightarrow (\Gamma^{\mathbb{N}}, \mathcal{P}(\Gamma)^{\otimes \mathbb{N}})$

$$\mathbb{X}^n := (\underbrace{X_1, X_2, \dots, X_n}_{\text{1. Experiment}}, \underbrace{X_1, X_2, \dots, X_n}_{\text{2. Experiment}}, \underbrace{X_1, X_2, \dots, X_n}_{\text{3. Experiment}}, \dots)$$

und den Linksshiftoperator  $\sigma : \Gamma^{\mathbb{N}} \rightarrow \Gamma^{\mathbb{N}}$  mit  $(\sigma x)_i = x_{i+1}$  fur alle  $i \in \mathbb{N}$  und  $x \in \Gamma^{\mathbb{N}}$ . Der *empirische Prozess* ist dann durch Diracmae in den Punkten definiert, die als Linksshift von  $\mathbb{X}^n$  entsteht:

$$R_n := \frac{1}{n} \sum_{i=1}^n \delta_{\sigma^i(\mathbb{X}^n)}.$$

Dann ist  $R_n$  ein zufalliges,  $\sigma$ -invariantes Wahrscheinlichkeitsma auf  $(\Gamma^{\mathbb{N}}, \mathcal{P}(\Gamma)^{\otimes \mathbb{N}})$ , das heit  $R_n \sigma^{-1} = R_n$ , denn fur eine beliebige Teilmenge  $A \subset \Gamma^{\mathbb{N}}$  gilt:

$$\begin{aligned} R_n \sigma^{-1}(A) = R_n(\Gamma \times A) &= \frac{1}{n} \sum_{i=1}^n \delta_{\sigma^i(\mathbb{X}^n)}(\Gamma \times A) \\ &= \frac{1}{n} \sum_{i=2}^{n+1} \delta_{\sigma^i(\mathbb{X}^n)}(A) = \frac{1}{n} \sum_{i=1}^n \delta_{\sigma^i(\mathbb{X}^n)}(A) = R_n(A). \end{aligned}$$

Die Menge der  $\sigma$ -invarianten Wahrscheinlichkeitsmae auf  $(\Gamma^{\mathbb{N}}, \mathcal{P}(\Gamma)^{\otimes \mathbb{N}})$  werden wir mit

$$\widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) := \{\nu \in \mathcal{M}_1(\Gamma^{\mathbb{N}}) : \nu \sigma^{-1} = \nu\}$$

bezeichnen, wobei  $\mathcal{M}_1(\Gamma^{\mathbb{N}})$  die Menge der Wahrscheinlichkeitsmae auf  $(\Gamma^{\mathbb{N}}, \mathcal{P}(\Gamma)^{\otimes \mathbb{N}})$  angibt. Wir definieren noch fur  $x = (x_1, x_2, x_3, \dots) \in \Gamma^{\mathbb{N}}$  die Projektion  $\pi^k$  auf die  $k$ -te

Komponente durch  $\pi^k(x) = x_k$  und durch  $\pi_m(x) = (x_1, x_2, \dots, x_m) \in \Gamma^m$  die Projektion auf die ersten  $m$ -Komponenten. Wir definieren eine Metrik  $d$  auf der Menge  $\widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}})$  wie folgt

$$d(\mu, \nu) := \sum_{m \in \mathbb{N}} 2^{-m} d_m(\pi_m \mu, \pi_m \nu).$$

Nun untersuchen wir die Konvergenzeigenschaft von  $R_n$  für  $n \rightarrow \infty$ .

**Lemma 4.1.** *Der empirische Prozess  $R_n$  konvergiert für  $n \rightarrow \infty$  schwach gegen  $\rho^{\otimes \mathbb{N}}$ , das heißt für jedes  $m \in \mathbb{N}$  und jede Menge  $Z \subset \Gamma^{\mathbb{N}}$  der Form  $Z = \{\pi^1 \in A_1, \pi^2 \in A_2, \dots, \pi^m \in A_m\}$  mit Teilmengen  $A_1, A_2, \dots, A_m \subset \Gamma$  gilt, dass  $\lim_{n \rightarrow \infty} R_n(Z) = \rho^{\otimes \mathbb{N}}(Z)$   $\mathbb{P}$ -f.s.*

*Beweis.* Für ein  $n \in \mathbb{N}$  und für das Bildmaß  $\pi_m R_n = R_n(\pi_m)^{-1}$  gilt  $\pi_m R_n = L_n^m$ , mit dem  $m$ -dimensionalen empirischen Maß  $L_n^m$  mit periodischen Randbedingungen. Für  $Z = \{\pi^1 \in A_1, \pi^2 \in A_2, \dots, \pi^m \in A_m\}$  gilt dann mit  $A = A_1 \times A_2 \times \dots \times A_m$

$$\begin{aligned} R_n(Z) &= R_n(\pi^1 \in A_1, \pi^2 \in A_2, \dots, \pi^m \in A_m) = R_n(\pi_m \in A_1 \times A_2 \times \dots \times A_m) \\ &= R_n(\pi_m \in A) = \pi_m R_n(A) = L_n^m(A) \xrightarrow{n \rightarrow \infty} \rho^{\otimes m}(A) = \pi_m \rho^{\otimes \mathbb{N}}(A) = \rho^{\otimes \mathbb{N}}(Z), \end{aligned}$$

wobei die Konvergenz  $\mathbb{P}$ -f.s. gilt nach *Lemma 1.4*. □

Wir formulieren den Satz über die große Abweichung des empirischen Prozesses  $R_n$  aus [3, *Theorem II.23*].

**Satz 4.2.** *Seien  $(X_i)_{i \in \mathbb{N}}$  i.i.d. Zufallsvariablen wie in (1). Für  $a > 0$  setzen wir  $B_a^c(\rho^{\otimes \mathbb{N}}) = \{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) : d(\nu, \rho^{\otimes \mathbb{N}}) > a\}$  und definieren*

$$J(a) = \inf_{\nu \in B_a^c(\rho^{\otimes \mathbb{N}})} I_\rho^\infty(\nu),$$

mit  $I_\rho^\infty(\nu) = \sup_{m \geq 2} I_\rho^m(\pi_m \nu)$ . Dann gilt:

- (a)  $\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(R_n \in B_a^c(\rho^{\otimes \mathbb{N}})) \geq -J(a)$
- (b)  $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(R_n \in B_a^c(\rho^{\otimes \mathbb{N}})) \leq -J(a-),$

wobei  $J(a-) = \lim_{\delta \downarrow 0} J(a - \delta)$ .

Als Vorbereitung zu dem Beweis von *Satz 4.2* betrachten wir die in dem Satz definierten Funktionen und untersuchen deren Eigenschaften.

**Lemma 4.3.** (a) Die Abbildung  $m \mapsto I_\rho^m(\pi_m \nu)$  ist monoton steigend für alle  $\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}})$ .

(b) Die Abbildung  $a \mapsto J(a)$  ist monoton steigend und rechtsseitig stetig.

(c) Für alle  $m \in \mathbb{N}$  und  $\mu \in \widetilde{\mathcal{M}}_1(\Gamma^m)$  gilt:

$$\inf_{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}): \pi_m \nu = \mu} I_\rho^\infty(\nu) = I_\rho^m(\mu).$$

*Beweis des Lemmas 4.3.* Zu Teil (a): wir setzen  $\pi_m \nu(\{(s_1, s_2, \dots, s_m)\}) = \nu_{s_1, s_2, \dots, s_m}$  für  $(s_1, s_2, \dots, s_m) \in \Gamma^m$  und  $m \in \mathbb{N}$  und rechnen mit Anwendung der Jensen'schen Ungleichung für die konkave Logarithmusfunktion

$$\begin{aligned} I_\rho^m(\pi_m \nu) &= - \sum_{s_1, s_2, \dots, s_m} \nu_{s_1, s_2, \dots, s_m} \log \left( \frac{\nu_{s_1, s_2, \dots, s_{m-1}} \rho_{s_m}}{\nu_{s_1, s_2, \dots, s_m}} \right) \\ &= - \sum_{s_2, s_3, \dots, s_m} \nu_{s_2, s_3, \dots, s_m} \underbrace{\sum_{s_1} \frac{\nu_{s_1, s_2, \dots, s_m}}{\nu_{s_2, s_3, \dots, s_m}}}_{=1} \log \left( \frac{\nu_{s_1, s_2, \dots, s_{m-1}} \rho_{s_m}}{\nu_{s_1, s_2, \dots, s_m}} \right) \\ &\geq - \sum_{s_2, s_3, \dots, s_m} \nu_{s_2, s_3, \dots, s_m} \log \left( \sum_{s_1} \frac{\nu_{s_1, s_2, \dots, s_{m-1}} \rho_{s_m}}{\nu_{s_2, s_3, \dots, s_m}} \right) = I_\rho^{m-1}(\pi_{m-1} \nu). \end{aligned}$$

Zu Teil (b): wir nehmen positive reelle Zahlen  $a' \leq a$ , dann gilt  $B_a^c(\rho^{\otimes \mathbb{N}}) \subset B_{a'}^c(\rho^{\otimes \mathbb{N}})$ . Das Infimum über eine größere Menge ist kleiner, daher ist  $J$  nach Definition monoton steigend. Um die rechtsseitige Stetigkeit von  $J$  in  $a > 0$  zu zeigen, wählen wir eine beliebige Folge  $a_n > a$  für  $n \in \mathbb{N}$  mit  $\lim_{n \rightarrow \infty} a_n = a$ . Dann existiert für alle  $\nu \in B_a^c(\rho^{\otimes \mathbb{N}})$  ein  $n_0$ , so dass  $\nu \in B_{a_n}^c(\rho^{\otimes \mathbb{N}})$  für alle  $n \geq n_0$ . Deshalb existiert eine Folge  $\nu_n \in B_{a_n}^c(\rho^{\otimes \mathbb{N}})$ , so dass  $\lim_{n \rightarrow \infty} I_\rho^\infty(\nu_n) = J(a)$ . Aus  $J(a) \leq J(a_n) \leq I_\rho^\infty(\nu_n)$  folgt nun  $\lim_{n \rightarrow \infty} J(a_n) = J(a)$ . Insgesamt folgt die rechtsseitige Stetigkeit von  $J$ .

Zu Teil (c): wir bemerken zuerst, dass nach Teil (a) für alle  $\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}})$  und  $m \in \mathbb{N}$  gilt, dass  $I_\rho^\infty(\nu) \geq I_\rho^m(\pi_m \nu)$  und somit die Ungleichung  $\inf_{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}): \pi_m \nu = \mu} I_\rho^\infty(\nu) \geq I_\rho^m(\mu)$  für alle  $\mu \in \widetilde{\mathcal{M}}_1(\Gamma^m)$  besteht.

Um zu der umgekehrten Ungleichung zu gelangen, konstruieren wir ein  $\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}})$  so, dass  $\pi_m \nu = \mu$  und  $I_\rho^\infty(\nu) = I_\rho^m(\mu)$  für ein gegebenes  $\mu \in \widetilde{\mathcal{M}}_1(\Gamma^m)$ . Dazu betrachten wir die rekursiv definierte Folge  $\nu^n \in \widetilde{\mathcal{M}}_1(\Gamma^n)$  für  $n \geq m$  mit folgenden Eigenschaften:

- (i)  $\nu^m = \mu$
- (ii)  $\pi_n \nu^{n+1} = \nu^n$  für alle  $n \geq m$
- (iii)  $I_\rho^{n+1}(\nu^{n+1}) = I_\rho^n(\nu^n)$  für alle  $n \geq m$ .

Hierzu wird nun der Projektionsoperator  $\pi_n$  auf der Menge  $\Gamma^{n+1}$  definiert wird.

Man erhält diese Eigenschaften, wenn man  $\nu^m = \mu$  setzt und  $\nu^{n+1}$  über die Zähldichte  $\nu_{s_1, \dots, s_n}^{n+1} = \nu^{n+1}(\{(s_1, \dots, s_{n+1})\})$  wie folgt definiert (vgl. [3, S. 24]):

$$\nu_{s_1, \dots, s_{n+1}}^{n+1} := \frac{\nu_{s_1, \dots, s_n}^n \nu_{s_2, \dots, s_{n+1}}^n}{\nu_{s_2, \dots, s_n}^{n-1}} \quad \forall n \geq m,$$

wobei  $\nu_{s_2, \dots, s_m}^{m-1} := \sum_{s_1=1}^N \nu_{s_1, \dots, s_m}^m$  und  $\frac{0}{0} = 0$  gesetzt wird. Dann besitzt die Folge  $(\nu^n)_{n \geq m}$  in der Tat die Eigenschaften (i), (ii), (iii), denn durch Induktion über  $n$  folgt

$$\begin{aligned} \sum_{s_{n+1}} \nu_{s_1, \dots, s_{n+1}}^{n+1} &= \sum_{s_{n+1}} \frac{\nu_{s_1, \dots, s_n}^n \nu_{s_2, \dots, s_{n+1}}^n}{\nu_{s_2, \dots, s_n}^{n-1}} = \nu_{s_1, \dots, s_n}^n \\ &= \sum_{s_{n+1}} \frac{\nu_{s_{n+1}, s_1, \dots, s_{n-1}}^n \nu_{s_1, \dots, s_n}^n}{\nu_{s_1, \dots, s_{n-1}}^{n-1}} = \sum_{s_{n+1}} \nu_{s_{n+1}, s_1, \dots, s_n}^{n+1}. \end{aligned}$$

Wir folgern (ii) und, dass  $\nu^n \in \widetilde{\mathcal{M}}_1(\Gamma^n)$ . Um (iii) zu zeigen rechnen wir

$$\begin{aligned} I_\rho^{n+1}(\nu^{n+1}) &= \sum_{s_1, s_2, \dots, s_{n+1}} \nu_{s_1, s_2, \dots, s_{n+1}}^{n+1} \log \left( \frac{\nu_{s_1, s_2, \dots, s_{n+1}}^{n+1}}{\nu_{s_1, s_2, \dots, s_n}^n \rho_{s_{n+1}}} \right) \\ &= \sum_{s_1, s_2, \dots, s_{n+1}} \frac{\nu_{s_1, \dots, s_n}^n \nu_{s_2, \dots, s_{n+1}}^n}{\nu_{s_2, \dots, s_n}^{n-1}} \log \left( \frac{\nu_{s_2, \dots, s_{n+1}}^n}{\nu_{s_2, \dots, s_n}^{n-1} \rho_{s_{n+1}}} \right) \\ &= \sum_{s_2, s_3, \dots, s_{n+1}} \nu_{s_2, \dots, s_{n+1}}^n \log \left( \frac{\nu_{s_2, \dots, s_{n+1}}^n}{\nu_{s_2, \dots, s_n}^{n-1} \rho_{s_{n+1}}} \right) \\ &= I_\rho^n(\nu^n). \end{aligned}$$

Wir folgern die Behauptung des Lemmas aus dem nachfolgenden Kolmogorow'schen Existenzsatz.

Nach Eigenschaft (ii) ist die Folge  $(\nu_n)_{n \geq m}$  eine Kolmogorow-konsistente Folge bezüglich der Filtration  $\mathcal{F}_n = \sigma(\pi_n)$  für  $n \geq m$ , wobei wir  $\nu_n$  als ein Wahrscheinlichkeitsmaß auf dem Messraum  $(\Gamma^{\mathbb{N}}, \mathcal{F}_n)$  auffassen. Da die kompakte Klasse  $\mathcal{C} = \{\prod_{n \in \mathbb{N}} A_n : A_n \subset \Gamma\}$  alle  $\sigma$ -Algebren  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  enthält, folgern wir aus *Satz 4.4 (Kolmogorow'scher Existenzsatz)*, dass ein  $\nu \in \mathcal{M}_1(\Gamma^{\mathbb{N}})$  existiert mit  $\pi_n \nu = \nu^n$  für alle  $n \geq m$ . Dann ist  $\nu$   $\sigma$ -invariant, da  $\nu \circ \sigma^{-1}(Z) = \nu(Z)$  für Zylindermengen  $Z$  und daher wegen der  $\sigma$ -Stetigkeit von  $\nu$  und  $\nu \circ \sigma^{-1}$  gilt, dass  $\nu \circ \sigma^{-1}(C) = \nu(C)$  für jedes  $C \in \mathcal{C}$ . Da  $\{A \subset \mathcal{P}(\Gamma)^{\otimes \mathbb{N}} : \nu(A) = \nu \circ \sigma^{-1}(A)\}$



ein Dynkin-System und  $\mathcal{C}$  ein  $\cap$ -stabiler Erzeuger von  $\mathcal{P}(\Gamma)^{\otimes \mathbb{N}}$  ist, folgt aus dem Dynkin-Lemma, dass  $\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}})$   $\sigma$ -invariant ist. Wir folgern aus Teil (a) und aus (iii) und (i), dass

$$I_\rho^\infty(\nu) = \lim_{n \rightarrow \infty} I_\rho^n(\pi_n \nu) = \lim_{n \rightarrow \infty} I_\rho^n(\nu^n) = I_\rho^m(\mu).$$

□

**Definition.** Ein Mengensystem  $\mathcal{C}$  über einen Ergebnisraum  $\Omega$  heißt *kompakte Klasse*, wenn für jede Folge  $C_m \in \mathcal{C}$   $m \in \mathbb{N}$  mit leerem Durchschnitt  $\bigcap_{m \in \mathbb{N}} C_m = \emptyset$  ein  $M \in \mathbb{N}$  existiert, so dass  $\bigcap_{m \leq M} C_m = \emptyset$ .

**Definition.** Ein Netz  $(P_t)_{t \in I}$  von Wahrscheinlichkeitsmaßen nennt man *Kolmogorow-konsistentes Netz* bezüglich der Filtration  $(\mathcal{F}_t)_{t \in I}$ , wenn  $P_t$  ein Wahrscheinlichkeitsmaß auf  $\mathcal{F}_t$  ist und die Konsistenzbedingung  $P_t|_{\mathcal{F}_s} = P_s$  für alle  $s, t \in I$  mit  $s \leq t$  erfüllt ist.

**Satz 4.4** (Kolmogorow'scher Existenzsatz). *Sei  $(\Omega, \Sigma)$  ein Ereignisraum und  $\mathcal{F} = (\mathcal{F}_t)_{t \in I}$  eine Filtration in  $\Sigma$  über  $(I, \leq)$ . Weiter sei  $(P_t)_{t \in I}$  ein Kolmogorow-konsistentes Netz bezüglich  $\mathcal{F}$ . Es existiere weiterhin eine kompakte Klasse  $\mathcal{C} \subset \mathcal{P}(\Omega)$  mit der Eigenschaft*

$$\forall t \in I \quad \forall F \in \mathcal{F}_t : \quad P_t(F) = \sup\{P_t(C) : C \subset F, C \in \mathcal{C} \cap \mathcal{F}_t\}.$$

*Dann existiert genau ein Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $(\Omega, \sigma(\bigcup_{t \in I} \mathcal{F}_t))$  mit der Eigenschaft  $\mathbb{P}|_{\mathcal{F}_t} = P_t$  für alle  $t \in I$ .*

Um diesen Satz zu zeigen, zeigen wir zuerst folgendes Lemma aus [5, Satz 1.6.2]:

**Lemma 4.5.** *Sei  $P$  ein Wahrscheinlichkeitsinhalt auf einer Mengenalgebra  $\mathcal{A}$  und es existiere eine kompakte Subklasse  $\mathcal{C} \subset \mathcal{A}$ , so dass*

$$\forall A \in \mathcal{A} : \quad P(A) = \sup\{P(C) : C \subset A, C \in \mathcal{C}\},$$

*dann ist  $P$  ein Prämaß auf  $\mathcal{A}$ .*

*Beweis des Lemmas aus [5].* Es genügt die Nullstetigkeit von  $P$  zu zeigen. Wir betrachten dazu eine gegen die leere Menge absteigende Folge von Mengen  $A_n \searrow \emptyset$  in  $\mathcal{A}$  für  $n \rightarrow \infty$ . Zu zeigen ist, dass  $\lim_{n \rightarrow \infty} P(A_n) = 0$ . Sei  $\epsilon > 0$  beliebig. Wir wählen Mengen  $C_n \in \mathcal{C}$  mit  $C_n \subset A_n$ , so dass  $P(A_n) \leq P(C_n) + \epsilon 2^{-n}$  für alle  $n \in \mathbb{N}$ . Wegen  $C_n \subset A_n$  gilt auch  $\bigcap_{n \in \mathbb{N}} C_n \subset \bigcap_{n \in \mathbb{N}} A_n = \emptyset$ . Daher existiert wegen der Kompaktheit von  $\mathcal{C}$  ein  $M \in \mathbb{N}$ , so dass  $\bigcap_{n=1}^M C_n = \emptyset$ . Die Inklusion  $A_M = \bigcap_{n=1}^M A_n = \bigcap_{n=1}^M A_n - \bigcap_{n=1}^M C_n \subset \bigcup_{n=1}^M (A_n - C_n)$  und die Additivität und Subadditivität von  $P$  zeigen, dass

$$\lim_{n \rightarrow \infty} P(A_n) \leq P(A_M) \leq \sum_{n=1}^M P(A_n - C_n) \leq \sum_{n \in \mathbb{N}} [P(A_n) - P(C_n)] \leq \epsilon.$$

Lassen wir  $\epsilon \downarrow 0$  gehen, so erhalten wir  $\lim_{n \rightarrow \infty} P(A_n) = 0$ .  $\square$

*Beweis des Kolmogorow'schen Existenzsatzes.* In dem Beweis folgen wir den Darlegungen aus [1, Beweis von *Theorem 2*]. Wir definieren für  $P(F) = P_t(F)$  für alle  $F \in \mathcal{F}_t$  und setzen  $\mathcal{A} = \cup_{t \in I} \mathcal{F}_t$ . Dann ist  $P : \mathcal{A} \rightarrow [0, 1]$  wohldefiniert wegen der Kolmogorow-Konsistenz. Weiter gilt  $P(\emptyset) = 0$  und  $P(\Omega) = 1$  und, weil  $(\mathcal{F}_t)_{t \in I}$  eine Filtration ist, ist  $P$  eine additive Mengenfunktion. Insgesamt ergibt sich, dass  $P$  ein Wahrscheinlichkeitsinhalt auf der Mengenalgebra  $\mathcal{A} = \cup_{t \in I} \mathcal{F}_t$ . Nach *Lemma 4.5* ist  $P$  ein Prämaß auf  $\mathcal{A}$ . Wendet man nun den Fortsetzungssatz von Carathéodory an, so erhält man das eindeutige Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $\sigma(\mathcal{A})$  mit der Eigenschaft  $\mathbb{P}|_{\mathcal{F}_t} = P_t$  für alle  $t \in I$ .  $\square$

*Lemma 4.3* stellt zusammen mit *Satz 1.8* alle nötigen Werkzeuge für den Beweis von *Satz 4.2* bereit (vgl. [3, *Theorem II.23*]).

*Beweis des Satzes 4.2.* (vgl. [3]). Um die untere Schranke zu zeigen, benutzen wir, wie schon im Beweis von *Lemma 4.1*, dass  $\pi_m R_n = L_n^m$ . Aus der Definition der Metrik  $d$  folgt nun für alle  $n, M \in \mathbb{N}$  die Inklusion  $\{d(R_n, \rho^{\otimes \mathbb{N}}) > a\} \supset \{\sum_{m=1}^M 2^{-m} d_m(L_n^m, \rho^{\otimes m}) > a\}$  und deshalb  $\mathbb{P}(d(R_n, \rho^{\otimes \mathbb{N}}) > a) \geq \mathbb{P}(\sum_{m=1}^M 2^{-m} d_m(L_n^m, \rho^{\otimes m}) > a)$ . Die Menge  $\{\mu \in \widetilde{\mathcal{M}}_1(\Gamma^m) : \sum_{m=1}^M 2^{-m} d_m(\pi_m \mu, \rho^{\otimes m}) > a\}$  ist offen in  $\widetilde{\mathcal{M}}_1(\Gamma^m)$  und es gilt folgende Gleichheit von Mengen:

$$\left\{ \sum_{m=1}^M 2^{-m} d_m(L_n^m, \rho^{\otimes m}) > a \right\} = \left\{ L_n^M \in \left\{ \mu \in \widetilde{\mathcal{M}}_1(\Gamma^m) : \sum_{m=1}^M 2^{-m} d_m(\pi_m \mu, \rho^{\otimes m}) > a \right\} \right\},$$

da für alle  $M \geq m$  gilt  $\pi_m L_n^M = L_n^m$ , wobei  $\pi_m$ , die Projektion auf die ersten  $m$  Komponenten, auf  $\Gamma^M$  definiert wird. Daher folgt für alle  $M \geq 1$  mit Anwendung von *Satz 1.8* und der vorhergehenden Beobachtungen beziehungsweise durch *Lemma 4.3 (c)*:

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d(R_n, \rho^{\otimes \mathbb{N}}) > a) &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \sum_{m=1}^M 2^{-m} d_m(L_n^m, \rho^{\otimes m}) > a \right) \\ &= - \inf_{\mu \in \widetilde{\mathcal{M}}_1(\Gamma^m) : \sum_{m=1}^M 2^{-m} d_m(\pi_m \mu, \rho^{\otimes m}) > a} I_\rho^M(\mu) \\ &= - \inf_{\nu \in \widetilde{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) : \sum_{m=1}^M 2^{-m} d_m(\pi_m \nu, \rho^{\otimes m}) > a} I_\rho^\infty(\nu). \end{aligned}$$

Es gilt  $\{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) : d(\nu, \rho^{\otimes \mathbb{N}}) > a + 2^{-M}\} \subset \{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) : \sum_{m=1}^M 2^{-m} d_m(\pi_m \nu, \rho^{\otimes m}) > a\}$  wegen

$$\begin{aligned}
d(\nu, \rho^{\otimes \mathbb{N}}) - 2^{-M} &= \sum_{m=1}^{\infty} 2^{-m} d_m(\pi_m \nu, \rho^{\otimes m}) - \sum_{m=M+1}^{\infty} 2^{-m} \\
&\leq \sum_{m=1}^M 2^{-m} d_m(\pi_m \nu, \rho^{\otimes m}) + \sum_{m=M+1}^{\infty} 2^{-m} - \sum_{m=M+1}^{\infty} 2^{-m} \\
&= \sum_{m=1}^M 2^{-m} d_m(\pi_m \nu, \rho^{\otimes m}). \tag{7}
\end{aligned}$$

Hierzu haben wir die Ungleichung  $d_m(\mu, \nu) \leq 1$  für alle  $\mu, \nu \in \widetilde{\mathcal{M}}_1(\Gamma^m)$ ,  $m \in \mathbb{N}$  benutzt. Wir folgern aus dieser Mengeninklusion, dass

$$\begin{aligned}
- \inf_{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) : \sum_{m=1}^M 2^{-m} d_m(\pi_m \nu, \rho^{\otimes m}) > a} I_{\rho}^{\infty}(\nu) &\geq - \inf_{\nu \in \widehat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}) : d(\nu, \rho^{\otimes \mathbb{N}}) > a + 2^{-M}} I_{\rho}^{\infty}(\nu) \\
&= -J(a + 2^{-M})
\end{aligned}$$

und erhalten damit unter Anwendung der rechtsseitigen Stetigkeit von  $J$  aus *Lemma 4.3 (b)* die untere Schranke für  $M \rightarrow \infty$ .

Um zu der oberen Schranke zu gelangen, schließen wir aus Ungleichung (7) mit  $\nu = R_n$  die Ungleichung

$$\mathbb{P}(d(R_n, \rho^{\otimes \mathbb{N}}) > a) \leq \mathbb{P}\left(\sum_{m=1}^M 2^{-m} d_m(L_n^m, \rho^{\otimes m}) > a - 2^{-M}\right).$$

Analog zur unteren Schranke kommen wir auf die Ungleichung

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d(R_n, \rho^{\otimes \mathbb{N}}) > a) \leq -J(a - 2^{-M}) \quad \forall M \in \mathbb{N}.$$

Mit dem Grenzübergang  $M \rightarrow \infty$  erhalten wir die Behauptung des Satzes. □

Wir untersuchen nun die Ratenfunktion  $I_\rho^\infty$  aus *Satz 4.2*.

**Lemma 4.6.** *Mit den Voraussetzungen aus (1) gilt*

(a)  $I_\rho^\infty$  ist endlich und unterhalbstetig auf  $\widehat{\mathcal{M}}(\Gamma^\mathbb{N})$ .

(b)  $I_\rho^\infty(\nu) \geq 0$  und Gleichheit besteht genau dann, wenn  $\nu = \rho^{\otimes \mathbb{N}}$ .

*Beweis.* Aus *Lemma 4.3 (a)* folgt, dass  $I_\rho^\infty(\nu) = \lim_{m \rightarrow \infty} I_\rho^m(\pi_m \nu)$  und  $I_\rho^m(\pi_m \nu) \geq I_\rho^n(\pi_n \nu)$  für  $m \geq n$ . Mit den stetigen Funktionen  $f_m(\nu) = I_\rho^m(\pi_m \nu)$  und einer Folge  $\nu_n \rightarrow \nu$  in  $\widehat{\mathcal{M}}(\Gamma^\mathbb{N})$  haben wir dann  $\liminf_{n \rightarrow \infty} I_\rho^\infty(\nu_n) \geq \liminf_{n \rightarrow \infty} f_m(\nu_n) = f_m(\nu)$  für alle  $m \in \mathbb{N}$ . Durch optimieren über  $m$  erhält man  $\liminf_{n \rightarrow \infty} I_\rho^\infty(\nu_n) \geq I_\rho^\infty(\nu)$ , also die Unterhalbstetigkeit von  $I_\rho^\infty$ .

Als Nächstes zeigen wir, dass  $I_\rho^\infty(\nu) \leq \max_s \log(\frac{1}{\rho_s})$ . Dazu setzen wir  $(\pi_m \nu)_{s_1, \dots, s_m} = \nu_{s_1, \dots, s_m}$  und definieren die *relative Entropie*  $H(\nu|\mu) = \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \log\left(\frac{\nu_{s_1, \dots, s_m}}{\mu_{s_1, \dots, s_m}}\right)$  für  $\nu, \mu \in \widehat{\mathcal{M}}(\Gamma^m)$ . Dann gilt nach *Lemma 4.3 (a)*  $H(\pi_m \nu|\pi_{m-1} \nu \otimes \rho) = I_\rho^m(\pi_m \nu) \xrightarrow{m \rightarrow \infty} I_\rho^\infty(\nu)$ . Wir betrachten den Ausdruck  $H(\pi_m \nu|\rho^{\otimes m})$  und formen ihn durch Iteration um (vgl. [3, S. 25])

$$\begin{aligned} H(\pi_m \nu|\rho^{\otimes m}) &= H(\pi_m \nu|\pi_{m-1} \nu \otimes \rho) + H(\pi_{m-1} \nu|\rho^{\otimes m-1}) \\ &= \dots = \sum_{n=1}^m H(\pi_n \nu|\pi_{n-1} \nu \otimes \rho) = \sum_{n=1}^m I_\rho^n(\pi_n \nu). \end{aligned}$$

Wir folgern  $I_\rho^\infty(\nu) = \lim_{m \rightarrow \infty} I_\rho^m(\pi_m \nu) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=1}^m I_\rho^n(\pi_n \nu) = \lim_{m \rightarrow \infty} \frac{1}{m} H(\pi_m \nu|\rho^{\otimes m})$ , weil die Folge der Mittelwerte einer Folge den gleichen Grenzwert hat wie die Folge selbst, wobei auch bestimmte Divergenz erlaubt ist. Wir haben weiterhin die Zerlegung

$$\frac{1}{m} H(\pi_m \nu|\rho^{\otimes m}) = \frac{1}{m} \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \log(\nu_{s_1, \dots, s_m}) - \frac{1}{m} \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \log(\rho_{s_1} \dots \rho_{s_m}),$$

wobei der erste Summand negativ ist und wir den zweiten Summanden unter Anwendung der Beobachtung, dass für ein  $\nu \in \widehat{\mathcal{M}}(\Gamma^\mathbb{N})$  gilt, dass  $\pi_m \nu \in \widehat{\mathcal{M}}(\Gamma^m)$ , folgendermaßen von oben abschätzen:

$$\begin{aligned}
& -\frac{1}{m} \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \log(\rho_{s_1} \dots \rho_{s_m}) = -\frac{1}{m} \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \sum_i \log(\rho_{s_i}) \\
& = -\sum_i \frac{1}{m} \sum_{s_1, \dots, s_m} \nu_{s_1, \dots, s_m} \log(\rho_{s_i}) = -\sum_i \frac{1}{m} \sum_{s_i} \nu_{s_i} \log(\rho_{s_i}) = -\sum_s \nu_s \log(\rho_s) \\
& \leq \max_s \log\left(\frac{1}{\rho_s}\right).
\end{aligned}$$

Teil (b) folgt aus *Lemma 1.7 (b)*. Nach diesem Lemma gilt, dass  $I_\rho^m(\pi_m \nu) \geq 0$  für alle  $m \in \mathbb{N}$  und Gleichheit genau dann besteht, wenn  $\pi_m \nu = \rho^{\otimes m}$ . Stimmt das für alle  $m \in \mathbb{N}$ , dann stimmen die beiden Wahrscheinlichkeitsmaße  $\rho^{\otimes \mathbb{N}}$  und  $\nu$  auf der Menge der Zylindermengen, die einen schnittstabilen Erzeuger von  $\mathcal{P}(\Gamma)^{\otimes \mathbb{N}}$  bildet, überein. Deshalb ist nach dem Eindeutigkeitssatz für Wahrscheinlichkeitsmaße  $\nu = \rho^{\otimes \mathbb{N}}$ .  $\square$



## 5 Der abzählbare Phasenraum

In diesem Abschnitt werden wir Zufallsvariablen mit Werten in einem abzählbaren Messraum sowie die große Abweichung deren empirischen Maßes betrachten. Wir ändern also die Voraussetzungen aus (1) und fordern nun

$$\begin{aligned} (X_i)_{i \in \mathbb{N}} \text{ sind i.i.d. Zufallsvariablen } \Omega \rightarrow \mathbb{N}, \\ \text{die Verteilung von } X_1 \text{ besitzt die Zähldichte } (\rho_s)_{s \in \mathbb{N}} \text{ mit } \rho_s > 0 \quad \forall s \in \mathbb{N}. \end{aligned} \quad (8)$$

Analog zu 1.1 definieren wir das empirische Maß

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

und eine Metrik auf  $\mathcal{M}_1(\mathbb{N}) = \{\text{Wahrscheinlichkeitsmaße auf } (\mathbb{N}, \mathcal{P}(\mathbb{N}))\}$

$$d(\nu, \mu) := \frac{1}{2} \sum_{s \in \mathbb{N}} |\nu_s - \mu_s|,$$

wobei die Konvergenz dieser Reihe aus der Dreiecksungleichung folgt:  $\frac{1}{2} \sum_{s \in \mathbb{N}} |\nu_s - \mu_s| \leq \frac{1}{2} \sum_{s \in \mathbb{N}} |\nu_s| + \frac{1}{2} \sum_{s \in \mathbb{N}} |\mu_s| = 1$ .

Wir zeigen die Konvergenzaussage für das empirische Maß  $L_n$ .

**Lemma 5.1.** *Das empirische Maß  $L_n$  konvergiert in der Metrik  $d$  für  $n \rightarrow \infty$  gegen  $\rho = (\rho_s)_{s \in \mathbb{N}}$ .*

*Beweis.* Wie im Beweis von Lemma 1.1 zeigt man die schwache Konvergenz der Folge  $L_n$  gegen  $\rho$  für  $n \rightarrow \infty$ . Es bleibt also zu zeigen, dass die schwache Konvergenz einer Folge  $(\nu^n)_{n \in \mathbb{N}}$  in  $\mathcal{M}_1(\mathbb{N})$  die starke Konvergenz impliziert. Konvergieren also die Zähldichte  $\nu_s^n \rightarrow \nu_s$  punktweise. Dann folgt aus dem Satz über majorisierte Konvergenz:

$$\begin{aligned} d(\nu^n, \nu) &= \frac{1}{2} \sum_{s \in \mathbb{N}} |\nu_s^n - \nu_s| = \frac{1}{2} \sum_{s \in \mathbb{N}} (\nu_s^n - \nu_s + 2 \max\{\nu_s - \nu_s^n, 0\}) \\ &= \sum_{s \in \mathbb{N}} \max\{\nu_s - \nu_s^n, 0\} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

mit der gleichmäßigen Majorante  $\nu_s \geq \max\{\nu_s - \nu_s^n, 0\}$ . □

Um eine Verallgemeinerung von *Satz 1.3* mittels Approximation durch den endlich-dimensionalen Fall zu geben, benötigen wir folgendes Lemma.

**Lemma 5.2.** *Sei  $\pi_N : \mathbb{N} \rightarrow \{1, 2, \dots, N\}$   $\pi_N(s) = \min\{s, N\}$  und  $\pi_N \nu = \nu \circ \pi_N^{-1}$  für  $\nu \in \mathcal{M}_1(\mathbb{N})$ . Dann gilt für  $\rho, \nu \in \mathcal{M}_1(\mathbb{N})$ :*

- (a)  $0 \leq d(\nu, \rho) - d_N(\pi_N \nu, \pi_N \rho) \leq \sum_{s=N}^{\infty} \rho_s$ .
- (b) Für  $I_\rho(\nu) = \sum_{s \in \mathbb{N}} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right)$  gilt, dass  $I_\rho(\nu) \in [0, \infty]$ , also konvergiert die Reihe. Die Abbildung  $N \mapsto I_{\pi_N \rho}^N(\pi_N \nu)$  ist monoton steigend und  $I_{\pi_N \rho}^N(\pi_N \nu) \leq I_\rho(\nu)$ .
- (c) Für  $J(a) = \inf_{\nu \in B_a^c(\rho)} I_\rho(\nu)$  mit  $a > 0$  und  $B_a^c(\rho) = \{\nu \in \mathcal{M}_1(\mathbb{N}) : d(\nu, \rho) > a\}$  gilt, dass die Abbildung  $a \mapsto J(a)$  monoton steigt und rechtsseitig stetig ist.

*Beweis.* Um die untere Schranke aus (a) zu bekommen rechnen wir

$$\begin{aligned} d_N(\pi_N \nu, \pi_N \rho) &= \frac{1}{2} \sum_{n=1}^N |(\pi_N \nu)_n - (\pi_N \rho)_n| = \frac{1}{2} \left( \sum_{n=1}^{N-1} |\nu_n - \rho_n| + \left| \sum_{n=N}^{\infty} (\nu_n - \rho_n) \right| \right) \\ &\leq \frac{1}{2} \sum_{n \in \mathbb{N}} |\nu_n - \rho_n| = d(\nu, \rho). \end{aligned}$$

Für die obere Schranke setzten wir (vgl. [3, S. 125])  $K = \{s \geq N : \nu_s \geq \rho_s\}$  und  $A = \frac{1}{2} \sum_{s \in K} (\nu_s - \rho_s) \geq 0$  und  $B = -\frac{1}{2} \sum_{s \notin K, s \geq N} (\nu_s - \rho_s) \geq 0$ . Dann gilt

$$\begin{aligned} d(\nu, \rho) - d_N(\pi_N \nu, \pi_N \rho) &= \frac{1}{2} \left( \sum_{n=N}^{\infty} |\nu_n - \rho_n| - \left| \sum_{n=N}^{\infty} (\nu_n - \rho_n) \right| \right) \\ &= A + B - \frac{1}{2} \left| \sum_{n=N}^{\infty} (\nu_n - \rho_n) \right| = A + B - |A - B| \\ &\stackrel{A-B \leq |A-B|}{\leq} 2B = \sum_{s \notin K, s \geq N} (\rho_n - \nu_n) \leq \sum_{s \geq N} \rho_s. \end{aligned}$$

Zu Teil (b): wir setzen  $h(x) = x \log(x)$ . Dann ist  $h$  konvex auf  $[0, \infty)$  und wir rechnen mit Hilfe der Jensen'schen Ungleichung

$$I_\rho(\nu) = \sum_{s \in \mathbb{N}} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) = \sum_{s \in \mathbb{N}} \rho_s h\left(\frac{\nu_s}{\rho_s}\right) \geq h\left(\sum_{s \in \mathbb{N}} \rho_s \frac{\nu_s}{\rho_s}\right) = h(1) = 0,$$



wobei die absolute Konvergenz bzw. die bestimmte Divergenz der Reihe aus der maßtheoretischen Variante der Jensen'schen Ungleichung folgt. Wir zeigen nun die Monotonie von  $N \mapsto I_{\pi_N \rho}^N(\pi_N \nu)$ . Dazu setzen wir  $\nu_{[N]} = \sum_{s>N} \nu_s$  (vgl. [3, S. 125]) und rechnen

$$\begin{aligned}
& I_{\pi_{N+1} \rho}^{N+1}(\pi_{N+1} \nu) - I_{\pi_N \rho}^N(\pi_N \nu) \\
&= \sum_{s=1}^{N+1} (\pi_{N+1} \nu)_s \log \left( \frac{(\pi_{N+1} \nu)_s}{(\pi_{N+1} \rho)_s} \right) - \sum_{s=1}^N (\pi_N \nu)_s \log \left( \frac{(\pi_N \nu)_s}{(\pi_N \rho)_s} \right) \\
&= \sum_{s=N}^{N+1} (\pi_{N+1} \rho)_s h \left( \frac{(\pi_{N+1} \nu)_s}{(\pi_{N+1} \rho)_s} \right) - (\pi_N \rho)_N h \left( \frac{(\pi_N \nu)_N}{(\pi_N \rho)_N} \right) \\
&= \rho_{[N]} h \left( \frac{\nu_{[N]}}{\rho_{[N]}} \right) + \rho_N h \left( \frac{\nu_N}{\rho_N} \right) - (\rho_{[N]} + \rho_N) h \left( \frac{\nu_{[N]} + \nu_N}{\rho_{[N]} + \rho_N} \right) \\
&\geq 0.
\end{aligned}$$

Dazu haben wir die Konvexität von  $h$  benutzt.

Teil (c) beweist man analog zu *Lemma 4.3 (b)*. □

Wir formulieren nun den Satz von Sanov aus [3, II.7 *Theorem II.36*] für den abzählbaren Phasenraum.

**Satz 5.3.** *Seien  $(X_i)_{i \in \mathbb{N}}$  wie in (8),  $a > 0$ ,  $B_a^c(\rho)$  und  $J$  wie in Lemma 5.2, dann gilt für das empirische Maß  $L_n$ :*

- (a)  $\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) \geq -J(a)$
- (b)  $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) \leq -J(a-)$

mit  $J(a-) = \lim_{\delta \downarrow 0} J(a - \delta)$ .

*Beweis.* Den Beweis entnehmen wir aus [3, II.7 *Theorem II.36*]. Um Teil (a) zu zeigen, folgern wir aus *Lemma 5.2 (a)*

$$\mathbb{P}(d(L_n, \rho) > a) \geq \mathbb{P}(d_N(\pi_N L_n, \pi_N \rho) > a) \quad \forall N \in \mathbb{N}.$$

Weil  $\pi_N L_n = \frac{1}{n} \sum_{i=1}^n \delta_{\pi_N X_i}$  und  $(\pi_N X_i)_{i \in \mathbb{N}}$  i. i. d.  $\pi_N \rho$ -verteilte Zufallsvariablen sind, folgern wir mit Anwendung von *Satz 1.3*, dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d_N(\pi_N L_n, \pi_N \rho) > a) = - \inf_{\nu \in \mathcal{M}(\mathbb{N}), d_N(\pi_N \nu, \pi_N \rho) > a} I_{\pi_N \rho}^N(\pi_N \nu) \quad \forall N \in \mathbb{N},$$

da die Abbildung  $\pi_N : \mathcal{M}_1(\mathbb{N}) \rightarrow \mathcal{M}_1(\{1, 2, \dots, N\})$  surjektiv ist. Aus  $\lim_{N \rightarrow \infty} \sum_{s=N}^{\infty} \rho_s = 0$  und *Lemma 5.2 (a)* folgt, dass für jedes  $\delta > 0$  und  $\nu \in \mathcal{M}_1(\mathbb{N})$  ein  $N_0 \in \mathbb{N}$  existiert, so dass

$$\{\nu \in \mathcal{M}_1(\mathbb{N}) : d_N(\pi_N \nu, \pi_N \rho) > a\} \supset \{\nu \in \mathcal{M}_1(\mathbb{N}) : d(\nu, \rho) > a + \delta\} \quad \forall N \geq N_0.$$

Mit  $I_{\pi_N \rho}^N(\pi_N \nu) \leq I_\rho(\nu)$  aus *Lemma 5.2 (b)* folgt nun die Abschätzung

$$\inf_{\nu \in \mathcal{M}_1(\mathbb{N}), d_N(\pi_N \nu, \pi_N \rho) > a} I_{\pi_N \rho}^N(\pi_N \nu) \leq \inf_{\nu \in \mathcal{M}_1(\mathbb{N}), d(\nu, \rho) > a + \delta} I_\rho(\nu) = J(a + \delta)$$

und daher

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) \geq -J(a + \delta).$$

Mit  $\delta \downarrow 0$  und mit der rechtsseitigen Stetigkeit von  $J$  aus *Lemma 5.2 (c)* folgt die untere Schranke aus dem Satz.

Für Teil (b) wählen wir ein  $\delta > 0$ . Aus *Lemma 5.1 (a)* folgt nun, dass ein Index  $n_0 \in \mathbb{N}$  existiert, so dass

$$\mathbb{P}(d(L_n, \rho) > a) \leq \mathbb{P}(d_N(\pi_N L_n, \pi_N \rho) > a - \delta) \quad \forall n \geq n_0.$$

Wie bei der unteren Schranke wenden wir *Satz 1.3* auf  $\pi_N L_n$  an und folgern

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) \leq - \inf_{\nu \in \mathcal{M}_1(\mathbb{N}), d_N(\pi_N \nu, \pi_N \rho) > a - \delta} I_{\pi_N \rho}^N(\pi_N \nu) \quad \forall N \in \mathbb{N}.$$

Es bleibt nur noch die Ungleichung  $J(a - \delta) \leq \inf_{\nu \in \mathcal{M}_1(\mathbb{N}), d_N(\pi_N \nu, \pi_N \rho) > a - \delta} I_{\pi_N \rho}^N(\pi_N \nu)$  für ein beliebiges  $N \in \mathbb{N}$  zu zeigen. Sei  $\nu \in \mathcal{M}_1(\mathbb{N})$  mit  $d_N(\pi_N \nu, \pi_N \rho) > a - \delta$ . Wir definieren ein  $\tilde{\nu} \in \mathcal{M}_1(\mathbb{N})$  durch

$$\tilde{\nu}_s := \begin{cases} \nu_s, & \text{falls } s < N \\ \frac{\nu_{[N]}}{\rho_{[N]}} \rho_s, & \text{falls } s \geq N \end{cases}$$

mit  $\nu_{[N]} = \sum_{s=N}^{\infty} \nu_s$  und  $\rho_{[N]} = \sum_{s=N}^{\infty} \rho_s$ . Dann gilt  $\pi_N \tilde{\nu} = \pi_N \nu$  und

$$\begin{aligned}
I_\rho(\tilde{\nu}) &= \sum_{s \in \mathbb{N}} \tilde{\nu}_s \log\left(\frac{\tilde{\nu}_s}{\rho_s}\right) = \sum_{s=1}^{N-1} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) + \sum_{s=N}^{\infty} \frac{\nu_{[N]}}{\rho_{[N]}} \rho_s \log\left(\frac{\nu_{[N]}}{\rho_{[N]}}\right) \\
&= \sum_{s=1}^{N-1} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) + \nu_{[N]} \log\left(\frac{\nu_{[N]}}{\rho_{[N]}}\right) = I_{\pi_N \rho}^N(\pi_N \nu).
\end{aligned}$$

Weiterhin gilt, dass  $d(\tilde{\nu}, \rho) = d_N(\pi_N \tilde{\nu}, \pi_N \rho)$ , denn

$$\begin{aligned}
d(\tilde{\nu}, \rho) &= \frac{1}{2} \sum_{s=1}^{N-1} |\nu_s - \rho_s| + \frac{1}{2} \sum_{s=N}^{\infty} \left| \frac{\nu_{[N]}}{\rho_{[N]}} \rho_s - \rho_s \right| \\
&= \frac{1}{2} \sum_{s=1}^{N-1} |\nu_s - \rho_s| + \frac{1}{2} |\nu_{[N]} - \rho_{[N]}| = d_N(\pi_N \tilde{\nu}, \pi_N \rho).
\end{aligned}$$

Da  $\nu \in \mathcal{M}_1(\mathbb{N})$  beliebig war, folgt mit der Definition von  $J$  die gewünschte Ungleichung  $J(a - \delta) \leq \inf_{\nu \in \mathcal{M}(\mathbb{N}), d_N(\pi_N \nu, \pi_N \rho) > a - \delta} I_{\pi_N \rho}^N(\pi_N \nu)$  und damit auch die Behauptung des Satzes.  $\square$

Wir formulieren einige Eigenschaften der Funktion  $I_\rho$  aus *Lemma 5.2*.

**Lemma 5.4.** *Die Funktion  $I_\rho$  aus Lemma 5.2 (b) hat folgende Eigenschaften.*

- (a)  $I_\rho$  ist unterhalbstetig und streng konvex auf  $\mathcal{M}_1(\mathbb{N})$ .
- (b) Die Niveaumengen  $(\{I_\rho \leq c\})_{c \in \mathbb{R}}$  sind kompakt in  $\mathcal{M}_1(\mathbb{N})$ .
- (c)  $I_\rho(\nu) \geq 0$  und Gleichheit gilt genau dann, wenn  $\nu = \rho$ .

*Bemerkung:* Mit diesem Lemma folgt insbesondere, dass  $I_\rho$  eine Ratenfunktion ist.

*Beweis.* Zu Teil (a): wir zeigen, dass die Abbildung aus *Lemma 5.2 (b)* für  $N \rightarrow \infty$  gegen die Ratenfunktion  $I_\rho$  konvergiert, also  $\lim_{N \rightarrow \infty} I_{\pi_N \rho}^N(\pi_N \nu) = I_\rho(\nu)$ . Es gilt:

$$\begin{aligned}
I_{\pi_N \rho}^N(\pi_N \nu) &= \sum_{s=1}^{N-1} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) + \nu_{[N]} \log\left(\frac{\nu_{[N]}}{\rho_{[N]}}\right) \\
&= \underbrace{\sum_{s=1}^{N-1} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right)}_{\xrightarrow{N \rightarrow \infty} I_\rho(\nu)} + \underbrace{\nu_{[N]} \log(\nu_{[N]})}_{\xrightarrow{N \rightarrow \infty} 0} - \underbrace{\nu_{[N]} \log(\rho_{[N]})}_{> 0}
\end{aligned}$$

mit  $\nu_{[N]} = \sum_{s=N}^{\infty} \nu_s$  und analog für  $\rho$ . Mit  $I_\rho(\nu) \geq I_{\pi_N \rho}^N(\pi_N \nu)$  aus *Lemma 5.2 (b)* folgern wir, dass  $\lim_{N \rightarrow \infty} I_{\pi_N \rho}^N(\pi_N \nu) = \lim_{N \rightarrow \infty} \sum_{s=1}^N \nu_s \log \left( \frac{\nu_s}{\rho_s} \right) = I_\rho(\nu)$ .

Damit ist  $I_\rho$  der Grenzwert aufsteigender stetiger Funktionen. Daher ist es unterhalbstetig, was man wie im Beweis von *Lemma 4.6 (a)* zeigt. Die strenge Konvexität folgt aus der strengen Konvexität von  $h$ . Damit ist Teil (a) gezeigt.

Um Teil (b) zu zeigen brauchen wir den *Satz von Prochorow*. Dazu zunächst eine Definition.

**Definition.** Eine Menge  $\mathcal{Q}$  von Borel-Wahrscheinlichkeitsmaßen auf einem metrischen Raum  $(M, d)$  heißt *straff*, wenn es für jedes  $\epsilon > 0$  eine kompakte Menge  $K \subset M$  gibt mit  $\mu(K) \geq 1 - \epsilon$  für alle  $\mu \in \mathcal{Q}$ .

**Satz 5.5** (Satz von Prochorow).

$$\mathcal{Q} \text{ ist straff} \implies \mathcal{Q} \text{ ist relativ schwach folgenkompakt.}$$

Einen Beweis findet man zum Beispiel in [4, *Theorem 8.9*].

Mit Hilfe des Satzes zeigen wir, dass die Niveaumengen von  $I_\rho$  kompakt sind. Aus der Jensen'schen Ungleichung folgt  $I_\rho(\nu) \geq I_{\pi_N \rho}^N(\pi_N \nu) \geq R_N(\nu) \geq 0$  mit

$$R_N(\nu) = \nu_{[N]} \log \left( \frac{\nu_{[N]}}{\rho_{[N]}} \right) + (1 - \nu_{[N]}) \log \left( \frac{1 - \nu_{[N]}}{1 - \rho_{[N]}} \right).$$

Wir folgern die Inklusion

$$\{I_\rho \leq c\} \subset \bigcap_{N \in \mathbb{N}} \{R_N \leq c\} = K,$$

wobei die linke Seite wegen der Unterhalbstetigkeit von  $I_\rho$  abgeschlossen ist. Es reicht also zu zeigen, dass  $K$  relativ kompakt ist. Wir zeigen zuerst, dass  $K$  straff ist, damit ist es nach *Satz 5.5* relativ schwach folgenkompakt. Angenommen  $K$  ist nicht straff. Dann existiert ein  $\epsilon > 0$  und für jedes kompakte  $C \subset \mathbb{N}$  ein  $\nu \in K$  mit  $\nu(C^c) > \epsilon$ . Wir wählen kompakte Mengen  $C_N = \{1, 2, \dots, N\}$ . Dann finden wir für jedes  $N \in \mathbb{N}$  ein  $\nu \in K$ , so dass  $\nu(C_N^c) = \nu_{[N]} > \epsilon$ . Damit ist  $\lim_{N \rightarrow \infty} \sup_{\nu \in K} R_N(\nu) = \infty$ , denn

$$\epsilon \log \left( \frac{\epsilon}{\rho_{[N]}} \right) \leq \sup_{\nu \in K} \left\{ \nu_{[N]} \log \left( \frac{\nu_{[N]}}{\rho_{[N]}} \right) \right\}.$$

Wegen  $\rho_{[N]} \downarrow 0$  konvergiert die linke und damit auch die rechte Seite für  $N \rightarrow \infty$  gegen  $\infty$ . Weil  $x \log(x)$  auf  $[0, \infty]$  von unten beschränkt ist, ist auch  $(1 - \nu_{[N]}) \log \left( \frac{1 - \nu_{[N]}}{1 - \rho_{[N]}} \right)$  für alle  $\nu \in K$  und  $N \in \mathbb{N}$  gleichmäßig von unten beschränkt. Damit erhalten wir einen Widerspruch zur Definition von  $K$ . Die Menge  $K$  muss also straff sein.

Wegen der Äquivalenz von Kompaktheit und Folgenkompaktheit in metrischen Räumen, genügt zu zeigen, dass jede Folge in  $K$  eine in  $(\mathcal{M}_1(\mathbb{N}), d)$  konvergente Teilfolge hat. Aus der relativ schwachen Folgenkompaktheit folgern wir, dass jede Folge eine schwach konvergente Teilfolge hat. Wie im Beweis von *Lemma 5.1* zeigt man, dass diese Teilfolge auch in Metrik  $d$  konvergiert.  $\square$



## Literatur

- [1] K. C. BORDER: *Notes on the Kolmogorov Extension Problem*, <http://www.hss.caltech.edu/~kcb/Notes/Kolmogorov.pdf> (1988). Datum des Zugriffs: 2. Juli 2011.
- [2] A. DEMBO AND O. ZEITOUNI: *Large Deviations Techniques and Applications*, 2<sup>nd</sup> edition. Springer, New York (1988).
- [3] F. DEN HOLLANDER: *Large Deviations*, Fields Institute Monographs, Providence, Rhode Island (2000).
- [4] L. B. KARALOV UND Y. G. SINAI: *Theory of Probability and Random Processes*, 2<sup>nd</sup> edition. Springer, Berlin (2007).
- [5] J. NEVEU: *Mathematische Grundlagen der Wahrscheinlichkeitstheorie*, R. Oldenbourg Verlag, München (1969).





## **Erklärung**

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht sind und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

München, 9. August 2011

Unterschrift