

11. Konvergenzsätze für Markov-Ketten; MCMC-Methoden

Notizen

von Timo Schorlepp

Einleitung:

In den letzten beiden Vorträgen haben wir Problemstellungen aus der statistischen Physik kennengelernt, bei denen es trotz endlicher Zustandsräume Z schwierig ist, mittels direkter Simulation Erwartungswerte $E_\mu(f) = \sum_{z \in Z} f(z) \mu_z$ zu berechnen. Beim *hard core model* bestand die Schwierigkeit konkret darin, mit realistischem Aufwand eine Gleichverteilung auf $Z \subset \{0, +1\}^E$ für einen Graphen (E, K) zu simulieren unter weiteren Einschränkungen an die zulässigen Konfigurationen, während beim *Ising-Modell* eine Boltzmann-Verteilung auf $Z = \{-1, +1\}^E$ für eine gegebene Hamiltonfunktion mit Wechselwirkungen nur von nächsten Nachbarn, gegeben durch die Kanten K , zu simulieren war. In beiden Fällen sind naive Ansätze, d.h. die Verwerfungsmethode für das *hard core model* und eine direkte Simulation für das Ising-Modell z.B. mit gleichverteilten Zufallsvariablen auf Z , nicht gangbar, da diese in den konkreten Beispielen auf exponentiell kleine Akzeptanzwahrscheinlichkeiten oder exponentiell große Varianzen führen.

Das entscheidende Konzept, das anschließend eingeführt wurde, um die oben genannten Probleme aufzulösen, sind homogene Markov-Ketten in diskreter Zeit $(X_t)_{t \in \mathbb{N}_0}$. Unser übergeordnetes Ziel ist es dabei, Markov-Ketten zu konstruieren, die einfach zu simulieren sind und deren Verteilung gegen eine gegebene Verteilung aus obigen Problemstellungen konvergiert. Da wir nur endliche Zustandsräume betrachten, sind Markov-Ketten durch Angabe der *Übergangsmatrix* $(Q_{z,z'})_{z,z' \in Z} \in \mathbb{R}^{Z \times Z}$ mit $Q_{z,z'} = P(X_{n+1} = z' | X_n = z)$ (falls $P(X_n = z) \neq 0$ für ein $n \in \mathbb{N}_0$) und der *Startverteilung* $(\mu_z^{(0)})_{z \in Z} \in \mathbb{R}^Z$ mit $X_0 \sim \mu^{(0)}$ eindeutig bestimmt (in Verteilung). Als wichtige Eigenschaften stochastischer Matrizen haben wir Irreduzibilität und Aperiodizität kennengelernt, und die Existenz *stationärer Verteilungen*, d.h. von (positiven, zu Wahrscheinlichkeitsmaßen normierten) Eigenvektoren von Q zum Eigenwert 1, bewiesen.

Im Rahmen dieses Vortrags werden wir nun zuerst beweisen, dass für irreduzible und aperiodische stochastische Matrizen Q die stationäre Verteilung eindeutig ist und die Markov-Kette (X_t) mit Übergangsmatrix Q für beliebige Startverteilungen exponentiell schnell in Verteilung gegen die stationäre Verteilung μ von Q konvergiert. Anschließend werden wir einen Ergodensatz für Markov-Ketten mit irreduzibler und aperiodischer Übergangsmatrix Q beweisen, der die Grundlage für die Markov Chain Monte Carlo-Methode (MCMC) darstellt: $1/n \sum_{i=1}^n f(X_i) \rightarrow E_\mu(f)$ fast sicher. Die verbleibende Aufgabe ist dann klar: Finde eine möglichst allgemeine, für Simulationen konkret nutzbare Konstruktionsvorschrift für irreduzible und aperiodische stochastische Matrizen mit vorgegebener stationärer Verteilung. Wichtige Stichworte hierfür sind die *detailed-balance*-Gleichung und das Festlegen von *Akzeptanzwahrscheinlichkeiten*, die wir beide für das *hard core model* und das Ising-Modell illustrieren werden.

Die Darstellung wird im folgendem naturgemäß vor allem an der Seminarlektüre [1] orientiert sein. Bei Interesse an einer Einführung in die statistische Physik, bspw. für Informationen zum Ising-Modell, sei weiterhin [2] empfohlen, und eine sehr empfehlenswerte, informelle und anschauliche Darstellung der hier formal eingeführten Konzepte, sowie eine Darstellung der Geschichte des Metropolis-Algorithmus, findet sich in [3].

Erinnerung an wichtige Definitionen und Resultate:

Set-Up: (Ω, \mathcal{F}, P) sei W-Raum, Z bezeichne eine endliche Menge mit der Potenzmenge als σ -Algebra, $(X_t)_{t \in \mathbb{N}_0}$ sei zeit-diskreter stochastischer Prozess mit $X_t : \Omega \rightarrow Z$ messbar $\forall t$, und $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$ sei die zugehörige Filtration auf Ω , d.h. $\mathcal{F}_t = \sigma(X_0, \dots, X_t) \subset \mathcal{F}$. Eine Abbildung $T : \Omega \rightarrow \mathbb{N}_0 \cup \{+\infty\} =: \mathbb{N}_\infty$ heißt Stoppzeit bzgl. (X_t) , falls $\{T = t\} \in \mathcal{F}_t \forall t \in \mathbb{N}_\infty$ (mit $\mathcal{F}_\infty = \sigma(X_0, X_1, \dots)$).

(X_t) heißt homogene Markov-Kette, falls $P(X_{n+1} = z_{n+1} | (X_0, \dots, X_n) = (z_0, \dots, z_n)) = P(X_{n+1} = z_{n+1} | X_n = z_n)$ für alle $n \in \mathbb{N}$ und $z_0, \dots, z_{n+1} \in Z$ mit $P((X_0, \dots, X_n) = (z_0, \dots, z_n)) > 0$, sowie $Q_{z,z'} := P(X_{n+1} = z' | X_n = z) \neq Q_{z,z'}(n)$ für $P(X_n = z) > 0$. Existiert $z_0 \in Z$ mit $P(X_n = z_0) = 0$ für alle n , so setze $Q_{z_0,z'} = \delta_{z_0,z'}$. Dann gilt für die so konstruierte Matrix $Q \in \mathbb{R}^{Z \times Z}$: $Q_{z,z'} \geq 0$ für alle $z, z' \in Z$, und für die Zeilensummen gilt $\sum_{z' \in Z} Q_{z,z'} = 1$, womit Q per Definition eine stochastische Matrix ist.

Ist T fast sicher endlich, so ist (X_{T+n}) ebenfalls eine Markov-Kette mit Übergangsmatrix Q . Für die Verteilungen $\mu^{(n)}$ von X_n gilt $\mu^{(n+l)} = \mu^{(n)} \cdot Q^l$. Q heißt irreduzibel, falls $\forall z, z' \in Z : \exists n > 0 : (Q^n)_{z,z'} > 0$. Q heißt aperiodisch, falls $\forall z \in Z : \text{ggT}(\{n > 0 | (Q^n)_{z,z} > 0\}) = 1$. Ist Q irreduzibel und $Q_{z,z} > 0$ für ein $z \in Z$, so ist Q aperiodisch.

Ferner ist Q genau dann irreduzibel und aperiodisch, wenn Q^n positiv ist für ein $n > 0$, d.h. $(Q^n)_{z,z'} > 0 \forall z, z' \in Z$. Für die Stoppzeit $T_z = \inf\{n > 0 | X_n = z\}$ für $z \in Z$ (Eintrittszeit) gilt, falls Q irreduzibel und aperiodisch ist, $ET_z < \infty$ und insbesondere ist T_z fast sicher endlich.

Ein Wahrscheinlichkeitsvektor $\mu \in \mathbb{R}^Z$ (d.h. $\mu_z \geq 0 \forall z \in Z$ und $\sum_{z \in Z} \mu_z = 1$) heißt stationäre Verteilung von Q , falls $\mu = \mu \cdot Q$. Ist Q irreduzibel mit einer stationären Verteilung μ , so ist $\tilde{Q} := (I + Q)/2$ (lazy version von Q) irreduzibel und aperiodisch mit stationärer Verteilung μ . Da μ stationäre Verteilung von Q ist genau dann wenn $\mu_z = \sum_{z' \in Z} \mu_{z'} Q_{z',z}$, folgt sofort, dass, falls μ die detailed-balance-Gleichung $\mu_z Q_{z,z'} = \mu_{z'} Q_{z',z} \forall z, z' \in Z$ erfüllt, μ auch stationäre Verteilung von Q ist. Während für stationäre Startverteilungen eine Markov-Kette mit Übergangsmatrix Q ihre Verteilung nicht ändert und anschaulich gesprochen im Gleichgewicht ist, liefert die detaillierte Balance, dass sogar für beliebige zwei Zustände $z, z' \in Z$ im Mittel gleich viele Übergänge von z nach z' wie von z' nach z stattfinden, d.h. das System ist invariant bzgl. Zeitumkehr und formal gilt $P_{(X_0, \dots, X_n)} = P_{(X_n, \dots, X_0)}$. Abschließend wurde gezeigt, dass für irreduzible und aperiodische Q immer eine positive stationäre Verteilung existiert.

Satz 11.1. (Konvergenz gegen stationäre Verteilungen; Satz 6.28 in [1])

Sei $Q \in \mathbb{R}^{Z \times Z}$ eine irreduzible und aperiodische stochastische Matrix und $\mu \in \mathbb{R}^Z$ eine stationäre Verteilung von Q . Dann existieren (feste) Konstanten $c > 0$ und $\alpha \in (0, 1)$ mit

$$\max_{z \in Z} \left| (\mu^{(0)} \cdot Q^n)_z - \mu_z \right| \leq c \cdot \alpha^n \quad (11.1)$$

für alle Wahrscheinlichkeitsvektoren $\mu^{(0)} \in \mathbb{R}^Z$ und alle $n \in \mathbb{N}$.

Beweis. Trotzdem die Aussage nach reiner linearer Algebra aussieht, werden wir zum Beweis Markov-Ketten nutzen bzw. genauer die sogenannte Kopplungsmethode. Sei $\mu^{(0)} \in \mathbb{R}^Z$ beliebiger Wahrscheinlichkeitsvektor und $\mu \in \mathbb{R}^Z$ positive stationäre Verteilung von Q . Seien $(X_n)_{n \in \mathbb{N}_0}$ und $(Y_n)_{n \in \mathbb{N}_0}$ zwei voneinander unabhängige Markov-Ketten mit Zustandsraum Z , Übergangsmatrix Q und Startverteilungen $X_0 \sim \mu^{(0)}$ und $Y_0 \sim \mu$ (d.h. insbesondere $Y_n \sim \mu \forall n \in \mathbb{N}$). Unabhängigkeit der Prozesse ist hier zu verstehen als $P((X_{i_1}, \dots, X_{i_k}) \in A, (Y_{j_1}, \dots, Y_{j_l}) \in B) = P((X_{i_1}, \dots, X_{i_k}) \in A) \cdot P((Y_{j_1}, \dots, Y_{j_l}) \in B)$ für alle $k, l \in \mathbb{N}$, $A \subset Z^k$, $B \subset Z^l$. Definiere den (Produkt-) Prozess $(U_n)_{n \in \mathbb{N}_0}$ durch $U_n : \Omega \rightarrow Z \times Z$, $U_n(\omega) = (X_n(\omega), Y_n(\omega))$. Dann ist auch (U_n) eine Markov-Kette mit irreduzibler und aperiodischer Übergangsmatrix $\tilde{Q} \in \mathbb{R}^{(Z \times Z)^2}$, denn:

$$P(U_{n+1} = (z_{n+1}, z'_{n+1}) | U_0 = (z_0, z'_0), \dots, U_n = (z_n, z'_n)) \quad (11.2)$$

$$= \frac{P(X_{n+1} = z_{n+1}, Y_{n+1} = z'_{n+1}, X_0 = z_0, Y_0 = z'_0, \dots, X_n = z_n, Y_n = z'_n)}{P(X_0 = z_0, Y_0 = z'_0, \dots, X_n = z_n, Y_n = z'_n)} \quad (11.3)$$

$$\stackrel{\perp}{=} \frac{P(X_{n+1} = z_{n+1}, X_0 = z_0, \dots, X_n = z_n) P(Y_{n+1} = z'_{n+1}, Y_0 = z'_0, \dots, Y_n = z'_n)}{P(X_0 = z_0, \dots, X_n = z_n) P(Y_0 = z'_0, \dots, Y_n = z'_n)} \quad (11.4)$$

$$= \frac{P(X_{n+1} = z_{n+1}, X_n = z_n) P(Y_{n+1} = z'_{n+1}, Y_n = z'_n)}{P(X_n = z_n) P(Y_n = z'_n)} \quad (11.5)$$

$$= \frac{P(X_{n+1} = z_{n+1}, Y_{n+1} = z'_{n+1}, X_n = z_n, Y_n = z'_n)}{P(X_n = z_n, Y_n = z'_n)} = P(U_{n+1} = (z_{n+1}, z'_{n+1}) | U_n = (z_n, z'_n)) \quad (11.6)$$

für $P(U_0 = (z_0, z'_0), \dots, U_n = (z_n, z'_n)) > 0$. Insbesondere zeigt obige Rechnung $\tilde{Q}_{(z_1, z'_1), (z_2, z'_2)} = Q_{z_1, z_2} \cdot Q_{z'_1, z'_2}$ (sofern ein $n \in \mathbb{N}$ existiert mit $P(U_n = (z_1, z_2)) = P(X_n = z_1) P(Y_n = z_2) > 0$; da aber Y_n für jedes n ohnehin eine positive Verteilung besitzt und Q irreduzibel ist, existiert ein solches n immer). Per Induktion folgt dann sofort auch $(\tilde{Q}^n)_{(z_1, z'_1), (z_2, z'_2)} = (Q^n)_{z_1, z_2} \cdot (Q^n)_{z'_1, z'_2}$. Da eine stochastische Matrix R genau dann irreduzibel und aperiodisch ist, wenn R^{n_0} positiv ist für ein $n_0 \in \mathbb{N}$, folgt damit, dass \tilde{Q} ebenfalls irreduzibel und aperiodisch ist.

Definiere die Stoppzeit T bzgl. (U_n) durch $T : \Omega \rightarrow \mathbb{N}_\infty$, $T(\omega) = \inf\{n \in \mathbb{N}_0 | X_n(\omega) = Y_n(\omega)\}$ (mit der Konvention $\inf \emptyset = \infty$; es gilt $\{T = n_0\} = \{X_0 \neq Y_0\} \cap \dots \cap \{X_{n_0-1} \neq Y_{n_0-1}\} \cap \{X_{n_0} = Y_{n_0}\}$, d.h. $\{T = n_0\} \in \sigma(U_0, \dots, U_{n_0})$, da bspw. $\{X_0 \neq Y_0\} = \{U_0 \in \Delta^C\}$ für die messbare Teilmenge $\Delta = \{(z, z) \in Z \times Z | z \in Z\} \subset Z \times Z$, also ist T in der Tat eine Stoppzeit bzgl. (U_n)). Definiere nun mithilfe dieser Stoppzeit eine Kopplung zwischen (X_n) und (Y_n) : Sei $(Z_n)_{n \in \mathbb{N}_0}$ der stochastische Prozess mit $Z_n : \Omega \rightarrow Z$, definiert durch

$$Z_n(\omega) = \begin{cases} X_n(\omega), & \text{falls } n \leq T(\omega) \\ Y_n(\omega), & \text{falls } n > T(\omega) \end{cases} \quad (11.7)$$

Mit anderem Worten: Sobald (X_n) und (Y_n) das erste Mal gleich sind, verwende (Y_n) für alle darauf folgenden Zeiten, und vorher (X_n) . Wir bemerken: Da für ein beliebiges $z \in Z$ die Eintrittszeit $T_{(z,z)} = \inf\{n > 0 | U_n = (z, z)\}$ fast sicher endlich ist, ist auch $T \leq T_{(z,z)}$ fast sicher endlich. Der entscheidende Schritt ist es nun zu zeigen, dass der so definierte Prozess (Z_n) ebenfalls eine homogene Markov-Kette mit Übergangsmatrix Q und Startverteilung $\mu^{(0)}$ ist. Haben wir dies gezeigt, so folgt für beliebiges $z \in Z$

$$\left| \mu_z^{(n)} - \mu_z \right| = \left| (\mu^{(0)} \cdot Q^n)_z - \mu_z \right| = |P(Z_n = z) - P(Y_n = z)| \quad (11.8)$$

$$= |P(Z_n = z, T < n) + P(Z_n = z, T \geq n) - P(Y_n = z, T < n) - P(Y_n = z, T \geq n)| \quad (11.9)$$

$$= |P(Y_n = z, T < n) + P(X_n = z, T \geq n) - P(Y_n = z, T < n) - P(Y_n = z, T \geq n)| \quad (11.10)$$

$$= |P(X_n = z, T \geq n) - P(Y_n = z, T \geq n)| \leq P(X_n = z, T \geq n) + P(Y_n = z, T \geq n) \leq 2P(T \geq n), \quad (11.11)$$

sodass $\max_{z \in Z} \left| \mu_z^{(n)} - \mu_z \right| \rightarrow 0$ aus $P(T \geq n) \leq P(T_{(z,z)} \geq n)$ und $ET_{(z,z)} = \sum_{n=1}^{\infty} P(T_{(z,z)} \geq n) < \infty$ folgt. Die konkrete Schranke $c \cdot \alpha^n$ folgt dabei aus dem Beweis zu $ET_{(z,z)} < \infty$ (Satz 6.18 in [1]). Bemerkungen dazu: Die Schranke $2P(T \geq n)$, die hier genutzt wurde, ist schlechter als die in [1] genutzte Schranke $P(T \geq n)$. Das ist für die Aussage aber unerheblich, und mir ist nicht ganz klar, welche in [1] nicht angegebenen Schritte zu der besseren Schranke führen. Eine Möglichkeit hierzu wäre

$$|P(X_n = z, T \geq n) - P(Y_n = z, T \geq n)| \quad (11.12)$$

$$= |P(X_n = z, T = n) + P(X_n = z, T > n) - P(Y_n = z, T = n) - P(Y_n = z, T > n)| \quad (11.13)$$

$$= |P(X_n = z, Y_n = z, T = n) + P(X_n = z, T > n) - P(Y_n = z, X_n = z, T = n) - P(Y_n = z, T > n)| \quad (11.14)$$

$$= |P(X_n = z, T > n) - P(Y_n = z, T > n)| \leq P(X_n = z, T > n) + P(Y_n = z, T > n) \quad (11.15)$$

$$= P(X_n = z, Y_n \neq z, T > n) + P(Y_n = z, X_n \neq z, T > n) \leq P(T > n) \leq P(T \geq n), \quad (11.16)$$

aber die gewünschte Ungleichung lässt sich bestimmt auch schneller zeigen. Dass sich der Erwartungswert der \mathbb{N} -wertigen Zufallsvariablen $T_{(z,z)}$ mit der angegebenen Formel berechnen lässt, folgt aus dem Satz von Fubini:

$$ET_{(z,z)} = \int_{[0,\infty)} x \, dP_{T_{(z,z)}}(x) = \int_{[0,\infty)} \left(\int_0^x 1 \, d\lambda(t) \right) dP_{T_{(z,z)}}(x) = \int_0^\infty \left(\int_{[t,\infty)} 1 \, dP_{T_{(z,z)}}(x) \right) d\lambda(t) \quad (11.17)$$

$$= \int_0^\infty P(T_{(z,z)} \geq t) d\lambda(t) = \sum_{k=1}^{\infty} P(T_{(z,z)} \geq k). \quad (11.18)$$

Die Konstante α ist im Übrigen gegeben als $\alpha = 1 - \delta \in (0, 1)$ mit $\delta = \min_{z, z' \in Z} (Q^n)_{z, z'} > 0$, wobei n so gewählt ist, dass Q^n positiv ist.

Noch zu erbringen ist also jetzt der Beweis, dass (Z_n) eine homogene Markov-Kette mit Übergangsmatrix Q und Startverteilung $\mu^{(0)}$ ist. Dass die Startverteilung von (Z_n) durch $\mu^{(0)}$ gegeben ist, ist klar, da $T \geq 0$ und folglich $Z_0(\omega) = X_0(\omega) \forall \omega \in \Omega$. Um nun die Aussage zu zeigen, genügt es, für alle $n \in \mathbb{N}$ und $z_0, \dots, z_n \in Z$ mit $P((Z_0, \dots, Z_{n-1}) = (z_0, \dots, z_{n-1})) > 0$ nachzuweisen, dass

$$P((Z_0, \dots, Z_n) = (z_0, \dots, z_n)) = \mu_{z_0}^{(n)} Q_{z_0, z_1} \dots Q_{z_{n-1}, z_n} \quad (11.19)$$

gilt. Um einzusehen, dass aus Glg. (11.19) in der Tat die Markov-Eigenschaft folgt und dass die Übergangsmatrix durch Q gegeben ist, sei $n \in \mathbb{N}$ beliebig und $z_0, \dots, z_{n+1} \in Z$ mit $P((Z_0, \dots, Z_n) = (z_0, \dots, z_n)) > 0$. Dann gilt

$$P(Z_{n+1} = z_{n+1} | Z_0 = z_0, \dots, Z_n = z_n) = \frac{P(Z_{n+1} = z_{n+1}, Z_0 = z_0, \dots, Z_n = z_n)}{P(Z_0 = z_0, \dots, Z_n = z_n)} \quad (11.20)$$

$$= \frac{\mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{n-1}, n} Q_{z_n, z_{n+1}}}{\mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{n-1}, n}} = Q_{z_n, z_{n+1}} \quad (11.21)$$

und andererseits

$$P(Z_{n+1} = z_{n+1} | Z_n = z_n) = \frac{P(Z_{n+1} = z_{n+1}, Z_n = z_n)}{P(Z_n = z_n)} \quad (11.22)$$

$$= \frac{\sum_{z_0, \dots, z_{n-1} \in Z} P(Z_{n+1} = z_{n+1}, Z_n = z_n, Z_0 = z_0, \dots, Z_{n-1} = z_{n-1})}{\sum_{z_0, \dots, z_{n-1} \in Z} P(Z_n = z_n, Z_0 = z_0, \dots, Z_{n-1} = z_{n-1})} \quad (11.23)$$

$$= \frac{\sum_{z_0, \dots, z_{n-1} \in Z} \mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{n-1}, n} Q_{z_n, z_{n+1}}}{\sum_{z_0, \dots, z_{n-1} \in Z} \mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{n-1}, n}} = Q_{z_n, z_{n+1}} \quad (11.24)$$

Insbesondere hängt das Ergebnis nur von den Zuständen z_n und z_{n+1} ab und nicht von dem Zufallsvariablenindex, sodass Homogenität mit Übergangsmatrix Q gezeigt ist. Es bleibt nun also letztendlich nur noch Glg. (11.19) zu zeigen. Dazu schreiben wir

$$P((Z_0, \dots, Z_n) = (z_0, \dots, z_n)) = \sum_{m=0}^n [P((Z_0, \dots, Z_n) = (z_0, \dots, z_n), T = m)] + P((Z_0, \dots, Z_n) = (z_0, \dots, z_n), T > n) \quad (11.25)$$

und berechnen die einzelnen Summanden wie folgt:

$$P((Z_0, \dots, Z_n) = (z_0, \dots, z_n), T > n) = P(X_0 = z_0, \dots, X_n = z_n, T > n) \quad (11.26)$$

$$= P(X_0 = z_0, \dots, X_n = z_n, Y_0 \neq z_0, \dots, Y_n \neq z_n) \quad (11.27)$$

$$\stackrel{=}{=} P(X_0 = z_0, \dots, X_n = z_n) P(Y_0 \neq z_0, \dots, Y_n \neq z_n) = \mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{n-1}, z_n} P(Y_0 \neq z_0, \dots, Y_n \neq z_n), \quad (11.28)$$

wobei in der letzten Zeile Eigenschaft (11.19) für die Markov-Kette (X_n) genutzt wurde (gehe dazu induktiv vor: $P(X_0 = z_0, \dots, X_n = z_n) = P(X_n = z_n | X_0 = z_0, \dots, X_{n-1} = z_{n-1}) P(X_0 = z_0, \dots, X_{n-1} = z_{n-1}) = P(X_n = z_n | X_{n-1} = z_{n-1}) P(X_0 = z_0, \dots, X_{n-1} = z_{n-1}) = Q_{z_{n-1}, z_n} P(X_0 = z_0, \dots, X_{n-1} = z_{n-1}) = \dots$). Analog folgt für die anderen Summanden mit $0 \leq m \leq n$:

$$P(Z_0 = z_0, \dots, Z_n = z_n, T = m) \quad (11.29)$$

$$= P(X_0 = z_0, \dots, X_m = z_m, Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m, \dots, Y_n = z_n) \quad (11.30)$$

$$\stackrel{=}{=} P(X_0 = z_0, \dots, X_m = z_m) P(Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m, \dots, Y_n = z_n) \quad (11.31)$$

$$= \mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{m-1}, z_m} P(Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m, \dots, Y_n = z_n) \quad (11.32)$$

$$= \mu_{z_0}^{(0)} Q_{z_0, z_1} \dots Q_{z_{m-1}, z_m} P(Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m) Q_{z_m, z_{m+1}} \dots Q_{z_{n-1}, z_n}. \quad (11.33)$$

In der letzten Zeile haben wir dabei eine allgemeinere Version der Markov-Eigenschaft genutzt, s. Lemma 6.7 in [1]. Um den entsprechenden Schritt einzusehen, schreibe

$$P(Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m, \dots, Y_n = z_n) \quad (11.34)$$

$$= \sum_{\tilde{z}_0 \neq z_0} \dots \sum_{\tilde{z}_{m-1} \neq z_{m-1}} P(Y_0 = \tilde{z}_0, \dots, Y_{m-1} = \tilde{z}_{m-1}, Y_m = z_m, \dots, Y_n = z_n) \quad (11.35)$$

$$\stackrel{(11.19)}{=} \sum_{\tilde{z}_0 \neq z_0} \dots \sum_{\tilde{z}_{m-1} \neq z_{m-1}} \mu_{\tilde{z}_0} Q_{\tilde{z}_0, \tilde{z}_1} \dots Q_{\tilde{z}_{m-1}, z_m} Q_{z_m, z_{m+1}} \dots Q_{z_{n-1}, z_n} \quad (11.36)$$

$$\stackrel{(11.19)}{=} \sum_{\tilde{z}_0 \neq z_0} \dots \sum_{\tilde{z}_{m-1} \neq z_{m-1}} P(Y_0 = \tilde{z}_0, \dots, Y_{m-1} = \tilde{z}_{m-1}, Y_m = z_m) Q_{z_m, z_{m+1}} \dots Q_{z_{n-1}, z_n} \quad (11.37)$$

$$= P(Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m) Q_{z_m, z_{m+1}} \dots Q_{z_{n-1}, z_n}. \quad (11.38)$$

Damit folgt nun insgesamt

$$P((Z_0, \dots, Z_n) = (z_0, \dots, z_n)) = \sum_{m=0}^n [P((Z_0, \dots, Z_n) = (z_0, \dots, z_n), T = m)] + P((Z_0, \dots, Z_n) = (z_0, \dots, z_n), T > n) \quad (11.39)$$

$$= \mu_{z_0}^{(0)} Q_{z_0, z_1} \cdots Q_{z_{n-1}, z_n} \underbrace{\left(\sum_{m=0}^n P(Y_0 \neq z_0, \dots, Y_{m-1} \neq z_{m-1}, Y_m = z_m) + P(Y_0 \neq z_0, \dots, Y_n \neq z_n) \right)}_{=1} \quad (11.40)$$

$$= \mu_{z_0}^{(0)} Q_{z_0, z_1} \cdots Q_{z_{n-1}, z_n}, \quad (11.41)$$

was den Beweis vollendet. \square

Bemerkung 11.2.

(a) Aus dem Satz folgt sofort die Eindeutigkeit der stationären Verteilung von Q , deren Existenz im letzten Vortrag gezeigt wurde und die somit stets positiv ist (angenommen es existieren zwei stationäre Verteilungen μ_1 und μ_2 von Q , die sich in mindestens einem Eintrag z_0 unterscheiden, dann gilt $\max_{z \in Z} |(\mu_1 \cdot Q^n)_z - (\mu_2)_z| = \max_{z \in Z} |(\mu_1)_z - (\mu_2)_z| \geq |(\mu_1)_{z_0} - (\mu_2)_{z_0}| > 0$ für alle $n \in \mathbb{N}$, im Widerspruch zu $\max_{z \in Z} |(\mu_1 \cdot Q^n)_z - (\mu_2)_z| \rightarrow 0$ für $n \rightarrow \infty$ laut obigem Satz).

(b) Ist (X_n) eine homogene Markov-Kette mit Übergangsmatrix Q , die den Voraussetzungen des Satzes genügt, und beliebiger Startverteilung $\mu^{(0)}$, so zeigt der Satz $X_n \xrightarrow{\mathcal{D}} \mu$ für $n \rightarrow \infty$, denn für beliebige $f : Z \rightarrow \mathbb{R}$ (eigentlich $f \in C_b(Z)$, aber alle Funktionen auf endlichen Mengen mit der diskreten Topologie sind stetig und beschränkt) gilt:

$$|E_{\mu^{(n)}}(f) - E_\mu(f)| \leq \sum_{z \in Z} |f(z)| \left| \mu_z^{(n)} - \mu_z \right| \leq |Z| \max_{z \in Z} |f(z)| \max_{z \in Z} \left| \mu_z^{(n)} - \mu_z \right| \quad (11.42)$$

$$= |Z| \max_{z \in Z} |f(z)| \max_{z \in Z} \left| (\mu^{(0)} \cdot Q^n)_z - \mu_z \right| \xrightarrow{n \rightarrow \infty} 0. \quad (11.43)$$

Satz 11.3. (Ergodensatz für homogene Markov-Ketten; Satz 6.30 in [1])

Sei $(X_n)_{n \in \mathbb{N}_0}$ eine homogene Markov-Kette mit endlichem Zustandsraum Z , irreduzibler und aperiodischer Übergangsmatrix Q und stationärer Verteilung $\mu \in \mathbb{R}^Z$. Dann gilt für jede Abbildung $f : Z \rightarrow \mathbb{R}$

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow[n \rightarrow \infty]{f.s.} E_\mu(f). \quad (11.44)$$

Bemerkung 11.4. Die Aussage sieht formal wie das starke Gesetz der großen Zahlen aus, mit dem entscheidenden Unterschied, dass die einzelnen Zufallsvariablen in der Summe hier im Allgemeinen weder unabhängig noch identisch verteilt sind. Die Summe auf der linken Seite wird dabei in der MCMC-Methode die bisher verwendete direkte Simulation D_n ersetzen. Anschaulich sagt der Satz, dass der „zeitliche“ Mittelwert auf der linken Seite (mit der Interpretation der Indices i als diskrete Zeitpunkte) gegen den „echten“ Mittelwert, d.h. den Erwartungswert, der stationären Verteilung der Markov-Kette konvergiert.

Beweis. Behauptung: Es genügt, den Fall zu betrachten, dass (X_n) deterministisch in einem $z_* \in Z$ startet, das heißt, dass die Verteilung von X_0 gegeben ist durch $\mu^{(0)} \in \mathbb{R}^Z$ mit $\mu_z^{(0)} = \delta_{z, z_*}$.

Beweis der Behauptung: Ist $\mu^{(0)}$ nicht von der angegeben Form, so setze

$$Z_{>} := \{z \in Z | P(X_0 = z) > 0\} \quad (11.45)$$

und

$$C = \left\{ (z_n)_{n \in \mathbb{N}} \in Z^{\mathbb{N}} \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(z_i) = E_\mu(f) \right\} \quad (11.46)$$

als die Menge der (deterministischen) Folgen in Z , für die die zu zeigende Aussage gilt. Dann gilt

$$P((X_n)_{n \in \mathbb{N}} \in C) = \sum_{z \in Z_{>}} \underbrace{P((X_n)_{n \in \mathbb{N}} \in C | X_0 = z)}_{=1} P(X_0 = z) = 1, \quad (11.47)$$

da $P((X_n)_{n \in \mathbb{N}} \in C | X_0 = z) = P((Y_n^z)_{n \in \mathbb{N}} \in C) = 1$ für eine deterministisch in $z \in Z$ gestartete Markov-Kette $(Y_n^z)_{n \in \mathbb{N}}$ mit Übergangsmatrix Q .

Sei nun also $z_* \in Z$ der deterministische Startpunkt der Markov-Kette (X_n) . Definiere für $n \in \mathbb{N}_0$ die Stoppzeiten $T^{(n)} : \Omega \rightarrow \mathbb{N}_\infty$ durch $T^{(0)}(\omega) \equiv 0$ und

$$T^{(n)}(\omega) = \inf\{k \in \mathbb{N} | k > T^{(n-1)}(\omega), X_k(\omega) = z_*\} \quad (11.48)$$

für $n > 0$ (d.h. die $T^{(n)}$ stellen die Zeitpunkte dar, zu denen die Markov-Kette zum n -ten Mal in den Anfangszustand z_* zurückkehrt). Diese Stoppzeiten sind alle fast sicher endlich (folgt per Induktion; Eintrittszeiten sind nach einem Satz aus dem vorigen Vortrag endlich, und für die Markov-Kette $(X_{n+T^{(k)}})$ ist dann $T^{(k+1)}$ wieder eine „gewöhnliche“ Eintrittszeit ohne weitere Bedingungen). Unterteile die $f(X_i)$ nun in disjunkte Partialsummen gemäß

$$Z^{(n)} := \sum_{i=T^{(n-1)}}^{T^{(n)}-1} f(X_i) \quad (11.49)$$

(außerhalb der Nullmenge, wo mindestens ein $T^{(k)} = \infty$ ist). Behauptung: Die reellwertigen ZV $Z^{(n)}$ sind unabhängig und identisch verteilt (!). Diese zunächst vielleicht erstaunliche Aussage lässt sich zumindest plausibel machen, wenn man realisiert, dass nach Satz 6.12 in [1] (starke Markov-Eigenschaft) für jede fast sicher endliche Stoppzeit T der Prozess $(X_{T+n})_{n \in \mathbb{N}_0}$ wieder eine Markov-Kette mit Übergangsmatrix Q ist. Da per Konstruktion $P(X_{T^{(k)}} = z_*) = 1$, ist für alle $k \in \mathbb{N}$ also $(X_{T^{(k)}+n})_{n \in \mathbb{N}_0}$ wieder eine Markov-Kette mit der gleichen deterministischen Startverteilung δ_{z_*} und Übergangsmatrix Q . Unabhängigkeit und identische Verteilung der $Z^{(k)}$ folgt dann anschaulich daraus, dass das deterministische Neustarten der Kette alle Abhängigkeiten zu vorigen Zeitpunkten eliminiert; **TODO: ein formaler Beweis mit Satz 6.12 sollte hier trotzdem noch eingefügt werden!** [1] ist dafür allerdings keine große Hilfe, [4] auch nicht..

Damit gilt nun nach dem starken Gesetz der großen Zahlen fast sicher

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m Z^{(i)} = EZ^{(1)}. \quad (11.50)$$

Zur Berechnung von $EZ^{(1)}$: Für alle $z \in Z$ gilt

$$E \left(\sum_{i=0}^{T^{(1)}-1} 1_{\{z\}}(X_i) \right) = \mu_z E T^{(1)} \quad (11.51)$$

(s. Beweis von Satz 6.27 in [1], auf diese Weise wurde gerade die stationäre Verteilung μ_z konstruiert), sodass sich

$$EZ^{(1)} = E \left(\sum_{i=0}^{T^{(1)}-1} f(X_i) \right) = E \left(\sum_{i=0}^{T^{(1)}-1} \left(\sum_{z \in Z} f(z) 1_{\{z\}}(X_i) \right) \right) \quad (11.52)$$

$$= \sum_{z \in Z} f(z) E \left(\sum_{i=0}^{T^{(1)}-1} 1_{\{z\}}(X_i) \right) = E T^{(1)} E_\mu(f) \quad (11.53)$$

ergibt. Zur weiteren Berechnung der linken Seite im GGZ bezeichne mit $R(m)$ für $m \in \mathbb{N}$ die Anzahl der *Eintritte* der Markov-Kette in den Zustand z_* bis (einschließlich) zum Zeitpunkt m , d.h.

$$R(m) = \sum_{i=1}^m 1_{\{z_*\}}(X_i). \quad (11.54)$$

Offensichtlich ist die Folge $(R(m)(\omega))_{m \in \mathbb{N}}$ monoton wachsend für alle $\omega \in \Omega$, und es gilt f. s. $R(m) \xrightarrow{n \rightarrow \infty} \infty$ aufgrund der Abschätzung

$$T^{(R(m))} \leq m < T^{(R(m)+1)}, \quad (11.55)$$

da sonst, falls $R(m)(\omega) \xrightarrow{n \rightarrow \infty} R_\infty(\omega) < \infty$ auf einer Menge $A \subset \Omega$ positiven Maßes, d.h., da die ZV diskrete Werte in \mathbb{N} annimmt, $R(m) = R_\infty$ ab einem $m_0 = m_0(\omega)$, $T^{(R_\infty(\omega)+1)}(\omega) = \infty$ für alle $\omega \in A$. Sei $\{R_0, R_1, R_2, \dots\} \subset \mathbb{N}$ die Menge der in A auftretenden Grenzwerte $R_\infty < \infty$. Da die Mengen $\{\omega \in A | R(m)(\omega) \rightarrow R_i\}$ eine abzählbare Partition von A darstellen, muss wegen $P(A) > 0$ mindestens ein $i_0 \in \mathbb{N}$ existieren mit $P(R(m) \rightarrow R_{i_0}) > 0$. Auf dieser Menge gilt dann aber $T^{(R_{i_0}+1)} = \infty$, im Widerspruch zur fast sicheren Endlichkeit aller Eintrittszeiten $T^{(k)}$. Die oben angegebenen Ungleichungen folgen dabei im Prinzip direkt aus der Definition der Eintrittszeiten und von $R(m)$: Für jedes $\omega \in \Omega$ ist $R(m)$ die Anzahl der Eintritte der Kette (X_n) in den Zustand z_* innerhalb der Indexmenge $\{1, 2, \dots, m\}$. $T^{(R(m))}$ ist dann der *Zeitpunkt*, zudem der $(R(m))$ -te Eintritt geschieht, und dies muss folglich per Definition zum Zeitpunkt m oder früher geschehen. Andererseits muss $m < T^{(R(m)+1)}$ gelten, da es ansonsten ja $(R(m) + 1)$ Eintritte in den Zustand z_* innerhalb $\{1, 2, \dots, m\}$ gäbe, was der Definition von $R(m)$ widerspräche.

Nehme im Folgenden ohne Einschränkung $f \geq 0$ an (sonst zerlege $f = f^* - f^-$ und verfähre für $f^+, f^- \geq 0$ analog). Dann gilt wegen Gl. (11.55):

$$\frac{1}{R(m)} \sum_{i=1}^{R(m)} Z^{(i)} = \frac{1}{R(m)} \sum_{i=0}^{T^{(R(m))}-1} f(X_i) \leq \frac{1}{R(m)} \sum_{i=0}^m f(X_i) \leq \frac{1}{R(m)} \sum_{i=0}^{T^{(R(m)+1)}-1} f(X_i) \quad (11.56)$$

$$= \frac{R(m)+1}{R(m)} \frac{1}{R(m)+1} \sum_{i=1}^{R(m)+1} Z^{(i)}, \quad (11.57)$$

sodass mithilfe des GGZ, da wegen $R(m) \xrightarrow{f.s.} \infty$

$$\frac{1}{R(m)} \sum_{i=1}^{R(m)} Z^{(i)} \xrightarrow{f.s.} EZ^{(1)} ; \quad \frac{1}{R(m)+1} \sum_{i=1}^{R(m)+1} Z^{(i)} \xrightarrow{f.s.} EZ^{(1)}, \quad (11.58)$$

per Sandwich-Kriterium dann auch

$$\frac{1}{R(m)} \sum_{i=0}^m f(X_i) \xrightarrow{f.s.} EZ^{(1)} = ET^{(1)} \cdot E_\mu(f) \quad (11.59)$$

folgt (die Summe kann auch bei $i = 1$ begonnen werden, das ist für den Grenzwert egal und entspricht dann der Form aus der Formulierung des Satzes). Betrachtet man dann abschließend den Spezialfall $f \equiv 1$, so ergibt sich also

$m/R(m) \xrightarrow{f.s.} ET^{(1)}$, d.h.

$$\frac{1}{m} \sum_{i=1}^m f(X_i) = \frac{R(m)}{m} \frac{1}{R(m)} \sum_{i=1}^m f(X_i) \xrightarrow{f.s.} E_\mu(f). \quad (11.60)$$

□

Bemerkung 11.5. Wie konstruiert man nun im Allgemeinen eine irreduzible und aperiodische stochastische Matrix zu einer gegebenen Verteilung $\mu \in \mathbb{R}^Z$, sodass die zugehörige Markov-Kette mit möglichst wenig Aufwand zu simulieren ist? Starte dazu mit einer *beliebigen* irreduziblen und symmetrischen stochastischen Matrix $\tilde{Q} \in \mathbb{R}^{Z \times Z}$. Wähle dann für $z \neq z'$ sogenannte Akzeptanzwahrscheinlichkeiten $\alpha_{z,z'} \in (0, 1]$, die bezüglich der stationären Verteilung μ die detailed-balance-artigen Gleichungen

$$\mu_z \cdot \alpha_{z,z'} = \mu_{z'} \cdot \alpha_{z',z} \quad (11.61)$$

erfüllen. Dann ist die Matrix Q , definiert durch $Q_{z,z'} = \tilde{Q}_{z,z'} \cdot \alpha_{z,z'}$ für $z \neq z'$ und $Q_{z,z} = 1 - \sum_{z' \neq z} Q_{z,z'}$, ebenfalls eine stochastische irreduzible Matrix mit stationärer Verteilung μ . Ist \tilde{Q} aperiodisch, so auch Q .

Beweis. Offensichtlich gilt für $z \neq z'$: $\tilde{Q}_{z,z'} > 0 \Leftrightarrow Q_{z,z'} > 0$, sowie $\tilde{Q}_{z,z} > 0 \Rightarrow Q_{z,z} > 0$, sodass, falls \tilde{Q} irreduzibel oder aperiodisch ist, sofort per Definition dieser Eigenschaften auch Q irreduzibel bzw. aperiodisch ist (Q hat mindestens überall dort positive Einträge, wo \tilde{Q} positive Einträge hat, woraus sofort $(\tilde{Q}^m)_{z,z'} > 0 \Rightarrow (Q^m)_{z,z'} > 0$ für beliebige $m \in \mathbb{N}$ und $z, z' \in Z$ folgt). Außerdem ist Q reversibel bezüglich μ , d.h.

$$\mu_z Q_{z,z'} = \mu_z \alpha_{z,z'} \tilde{Q}_{z,z'} = \mu_{z'} \alpha_{z',z} \tilde{Q}_{z',z} = \mu_{z'} Q_{z',z} \quad (11.62)$$

für $z \neq z'$, wobei insbesondere die Symmetrie von \tilde{Q} benutzt wurde. Aus Reversibilität/detailed balance folgt aber Stationarität, denn

$$\mu_z = \mu_z \sum_{z' \in Z} Q_{z,z'} = \sum_{z' \in Z} \mu_{z'} Q_{z',z} = (\mu \cdot Q)_z. \quad (11.63)$$

□

Bemerkung 11.6. Zwei spezielle Wahlen von Akzeptanzwahrscheinlichkeiten, die für beliebige μ die detailed-balance-Gleichung erfüllen, sind der Metropolis-Algorithmus [5] mit

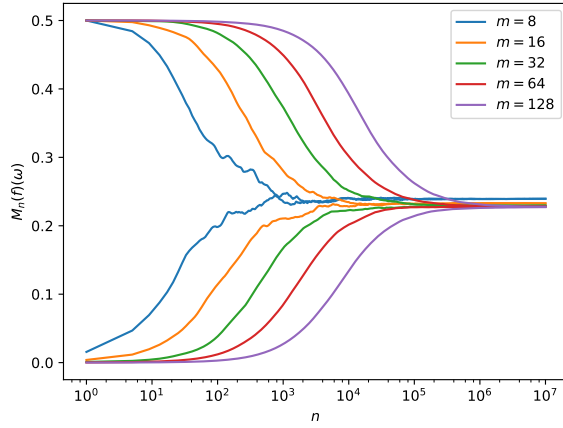
$$\alpha_{z,z'} = \min \left(1, \frac{\mu_{z'}}{\mu_z} \right) \quad (11.64)$$

und der Gibbs-Sampler

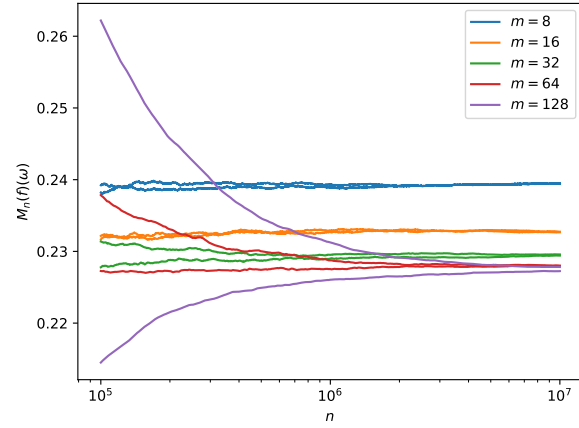
$$\alpha_{z,z'} = \frac{\mu_{z'}}{\mu_z + \mu_{z'}}. \quad (11.65)$$

Zum Vergleich verschiedener Akzeptanzwahrscheinlichkeitswahlen und entsprechenden Gütekriterien siehe [3]. Wichtig ist hier insbesondere, dass in beiden Fällen nur Verhältnisse von Wahrscheinlichkeiten benötigt werden, der Normierungsfaktor von μ muss also nicht bekannt sein.

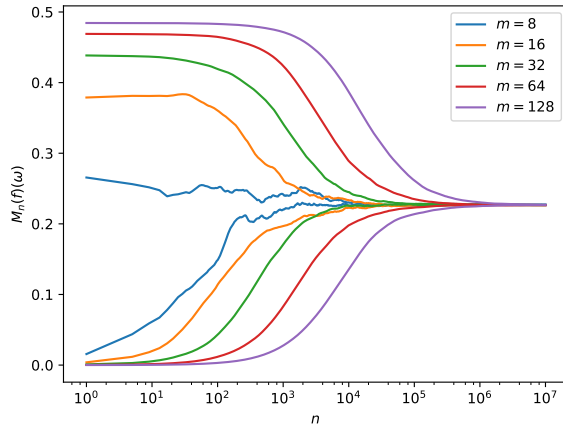
Beispiel 11.7. Wir kehren nun zunächst zum hard core model zurück. Zur Erinnerung: Wir betrachten einen ungerichteten Graphen (E, K) mit $|E| < \infty$, $K \neq \emptyset$ und $Z \subset \tilde{Z} = \{0, 1\}^E$, sodass für $z \in Z$ gilt, dass $z(e) = z(e') = 1 \Rightarrow \{e, e'\} \notin K$. Die gesuchte stationäre Verteilung sei nun die Gleichverteilung auf Z , und eine natürlich Frage ist zum Beispiel: Wie groß ist die mittlere Besetzungszahl, d.h. was ist $E_\mu(f)$ für $f : Z \rightarrow \mathbb{R}$, $z \mapsto 1/|E| \sum_{e \in E} z(e)$? Die Wahl der Übergangswahrscheinlichkeiten ist für den Spezialfall einer Gleichverteilung natürlich trivial: $\alpha_{z,z'} = 1$ für alle $z, z' \in Z$. Wir benötigen also nur eine symmetrische, irreduzible und aperiodische Matrix $Q \in \mathbb{R}^{Z \times Z}$, die zu einer einfach zu simulierenden Markov-Kette korrespondiert. Wir setzen dazu für jedes $z \in Z$: $Q_{z,z'} = 1/|E|$, falls $z' \in Z$ ein benachbarter Zustand zu z ist, d.h. wenn genau ein $e \in E$ existiert mit $z(e) \neq z'(e)$. Ansonsten setze $Q_{z,z'} = 0$ für alle nicht benachbarten Zustände $z' \neq z$, und zur korrekten Normierung $Q_{z,z} = 1 - \sum_{z' \in Z} Q_{z,z'}$ (da jeder Zustand höchstens $|E|$ Nachbarn hat, ist $Q_{z,z} \geq 0$). Die so erhaltene Matrix ist offenbar symmetrisch, da für $z \neq z'$ stets $Q_{z,z'} = Q_{z',z}$ gilt, denn die Definition benachbarter Zustände ist symmetrisch. Zur Irreduzibilität: Da sich offensichtlich von jedem beliebigen Zustand z aus mit positiver Wahrscheinlichkeit in $\sum_{e \in E} z(e)$ Schritten die Nullkonfiguration $z \equiv 0$ erreichen lässt und entsprechend dann in $\sum_{e \in E} z'(e)$ Schritten jeder andere beliebige Zustand $z' \in Z$, ist Q per Definition irreduzibel und insbesondere gibt es, da $K \neq \emptyset$ und somit Zustände mit weniger als $|E|$ zulässigen Nachbarzuständen existieren, mindestens einen Zustand z mit $Q_{z,z} > 0$, sodass Q auch aperiodisch ist nach Lemma 6.15 in [1]. Die Simulation der zu Q gehörenden Markov-Kette ist einfach: Ist die Kette im Zustand z , so wähle gleichverteilt ein $e \in E$. Falls sich durch Ändern von $z(e)$ ein zulässiger Zustand ergibt, so ändere $z(e)$, andernfalls bleibe im Zustand z . Hierbei ist es also insbesondere nicht notwendig, die Anzahl an benachbarten Zuständen oder die „Verweilwahrscheinlichkeit“ $Q_{z,z}$ explizit auszurechnen! Als konkretes Beispiel betrachten wir als Graphen wie in [1] ein zweidimensionales, kartesisches und nicht-periodisches $m \times m$ -Gitter $E = \{1, 2, \dots, m\}^2$ mit Kanten zwischen Punkten mit $|e_i - e'_i| = 1$ für genau ein $i \in \{1, 2\}$ und $e_j = e'_j$ für $j \neq i$, d.h. $|e_1 - e'_1| + |e_2 - e'_2| = 1$ (Periodizität wäre hier aber unter anderen deshalb wünschenswert, weil so für kleine Gittergrößen Randeffekte keine Rolle spielen, weshalb auch diese Randbedingungen implementiert wurden, s. folgende Ergebnisse). Es wurden folgende numerische Experimente durchgeführt: Für $N = 10^7$ Schritte und Gittergrößen $m \in \{8, 16, 32, 64, 128\}$ wurde sowohl für periodische als auch nicht-periodische Gitter je eine MCMC-Simulation gestartet, um die mittlere relative Besetzungszahl $E_\mu(f)$ mittels $M_n(f) = \sum_{i=1}^n f(X_i)$ für $n = 1, \dots, N$ zu schätzen. Die Ergebnisse sind in Abbildung 11.1 und Tabelle 11.1 zu finden. Beachte, dass wegen der fehlenden Unabhängigkeit



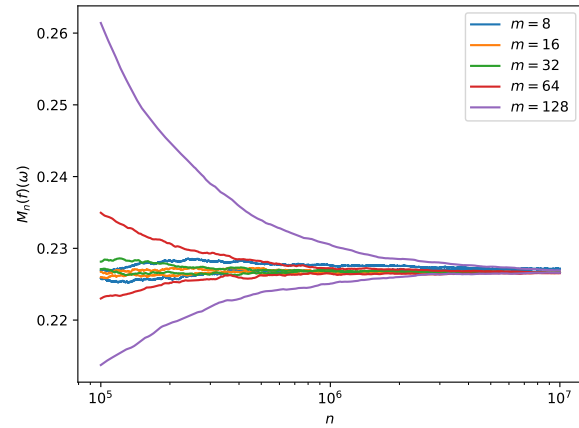
(a) Nicht-periodische Randbedingungen



(b) Zoom in (a)



(c) Periodische Randbedingungen



(d) Zoom in (c)

Abbildung 11.1: Verhalten des Schätzers $M_n(f)$ für die mittlere relative Besetzungszahl im hard core model in Abhängigkeit der Anzahl an generierten Zuständen n für verschiedene Gittergrößen $m \times m$ und Randbedingungen.

der einzelnen ZV hier keine einfache Möglichkeit besteht, (asymptotische) Konfidenzintervalle anzugeben (für die empirische Standardabweichung als erwartungstreuen Schätzer bräuchte man Unabhängigkeit; tatsächlich lässt sich aber mithilfe einer Modifikation dieses Schätzers durch die sogenannte Korrelationszeit der MCMC-Simulation ein geeigneter Schätzer für die Varianz gewinnen [3]). Die Ergebnisse deuten darauf hin, dass der Erwartungswert der mittleren relativen Besetzungszahl für große m etwa bei 0.227 liegt (und nicht etwa bei 0.25, wie man naiv vermuten könnte).

Beispiel 11.8. Als zweites Beispiel behandeln wir das zweidimensionale Ising-Modell für einen Ferromagneten ohne äußeres Magnetfeld. Allgemein betrachten wir wieder einen ungerichteten Graphen (E, K) mit $|E| < \infty$, $K \neq \emptyset$ und $Z = \{-1, 1\}^E$. Das Ziel ist es, Erwartungswerte bezüglich einer Boltzmann-Verteilung μ^β auf Z mit $\beta > 0$ zu berechnen, d.h. $\mu_z^\beta = 1/C_\beta \exp(-\beta H(z))$ für $z \in Z$ mit der Zustandssumme C_β als Normierungsfaktor und der Hamilton-Funktion $H : Z \rightarrow \mathbb{R}$ von der allgemeinen Form $H(z) = \sum_{e \in E} h_e(z(e)) + \sum_{\{e, e'\} \in K} h_{\{e, e'\}}(z(e), z(e'))$, wobei die Funktionen $h_e : \{-1, 1\} \rightarrow \mathbb{R}$ bspw. ein äußeres, ortsabhängiges Magnetfeld modellieren, und die Funktionen $h_{\{e, e'\}} : \{-1, 1\}^2 \rightarrow \mathbb{R}$ Wechselwirkungen benachbarter Spins modellieren. Wir beschränken uns wieder auf ein zweidimensionales, kartesisches, periodisches Gitter $E = \{1, \dots, m\}^2$ sowie auf den Fall, dass $h_e = 0$ für alle $e \in E$ und dass $h_{\{e, e'\}}(\lambda_1, \lambda_2) = -\lambda_1 \lambda_2$ (d.h. parallele Ausrichtungen benachbarter Spins liefern niedrigere Energien, also

Tabelle 11.1: Mittelwert der mittleren Besetzungszahl $M_n(f)$ für $n = 9 \cdot 10^6, \dots, 1 \cdot 10^7$ im hard core model für verschiedene Randbedingungen („p“: periodisch, „np“: nicht-periodisch) und Startkonfigurationen („zero“: starte mit $z \equiv 0$, „full“: starte mit maximaler Besetzung, d.h. Schachbrettmuster). Die empirische Standardabweichung lag dabei in allen Fällen in der Größenordnung 10^{-5} .

m	np, zero	np, full	p, zero	p, full
8	0.23944	0.23947	0.22715	0.22710
16	0.23279	0.23267	0.22656	0.22667
32	0.22942	0.22958	0.22660	0.22660
64	0.22791	0.22804	0.22654	0.22657
128	0.22724	0.22783	0.22660	0.22694

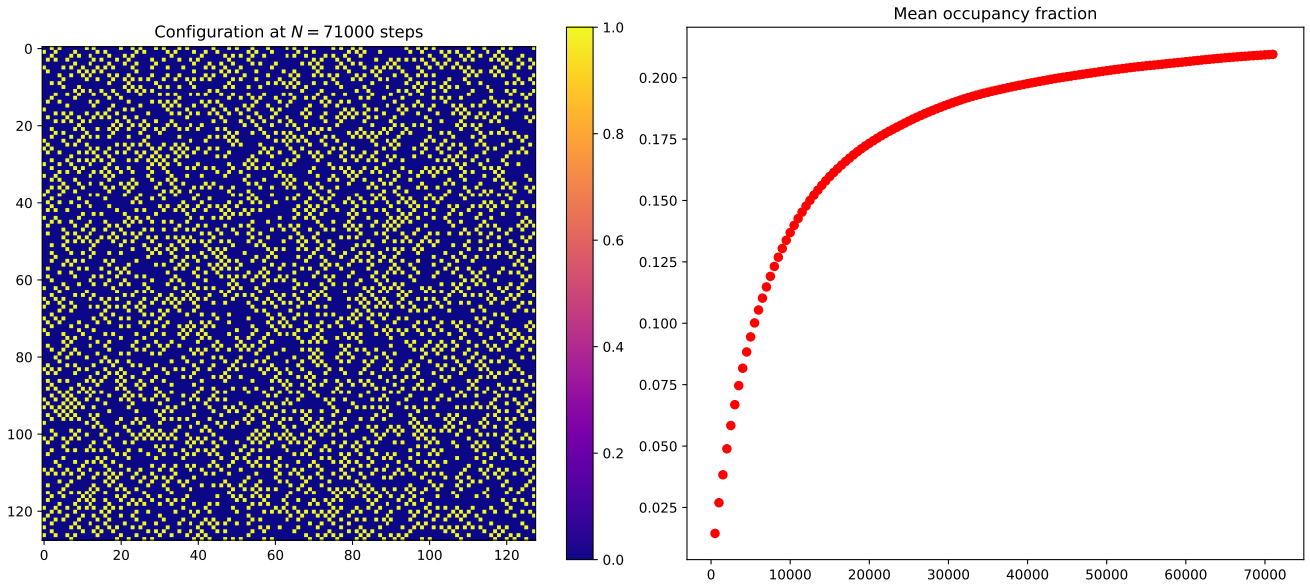


Abbildung 11.2: Typische Konfiguration im hard core model für $m = 128$ mit periodischen Randbedingungen nach 71000 Schritten der Markov-Kette. Rechts ist die mittlere Besetzungszahl $M_n(f)$ in Abhängigkeit der Anzahl an Schritten n dargestellt.

Zustände mit höherer Wahrscheinlichkeit). Für $\beta \rightarrow 0$, d.h. mit steigender Temperatur, geht μ^β punktweise gegen die Gleichverteilung auf Z , während für $\beta \rightarrow \infty$, d.h. mit fallender Temperatur, nur die beiden Minima von H , d.h. $z \equiv 1$ und $z \equiv -1$, mit Wahrscheinlichkeit $1/2$ „überleben“. Zur Simulation der Boltzmann-Verteilung auf E bezeichnen wir zwei Zustände $z, z' \in Z$ als benachbart, falls sie sich in genau einer Spinausrichtung unterscheiden, und setzen für jedes $z \in Z$: $\tilde{Q}_{z,z'} = 1/|E|$, falls z und z' benachbart sind, und $\tilde{Q}_{z,z'} = 0$ sonst. Offensichtlich ist die so erhaltene Matrix irreduzibel (gleiche Argumentation wie für hard core), aber in diesem Fall nicht (!) aperiodisch, da nur nach einer geraden Anzahl an Schritten in einen gegebene Zustand z zurückgekehrt werden kann. Da für den Gibbs-Sampler aber wegen $\mu_z > 0$ für alle z stets $\alpha_{z,z'} < 1$ gilt, hat die Matrix Q also positive Diagonalelemente und ist somit aperiodisch. Für den Metropolis-Algorithmus ist analog ebenfalls, da H nicht invariant ist unter Flippen eines Spins, $\alpha_{z,z'} < 1$ für einige benachbarte Zustände und somit Q aperiodisch. Zur Implementation dieser Übergangsmatrix Q gehe nun wie folgt vor: Befindet sich die Kette im Zustand z , so wähle gleichverteilt ein $e \in E$ und flippe den erhaltenen Spin mit Wahrscheinlichkeit $\alpha_{z,z'}$ in den Zustand z' . Für den Metropolis-Algorithmus heißt da konkret: Falls der Zustand z' , der sich durch Flippen des Spines an der Position e ergibt, eine niedrigere oder gleiche Energie aufweist, so akzeptiere die Änderung immer. Falls der geflippte Zustand eine höhere Energie aufweist, so ändere den Zustand mit der Wahrscheinlichkeit $\mu_{z'}/\mu_z$ und bleibe ansonsten im Zustand z (zur Implementation: generiere u gleichverteilt in $[0, 1]$ und akzeptiere, falls $u \leq \mu_{z'}/\mu_z$). Beim Gibbs-Sampler entfällt diese Fallunterscheidung und es wird in jedem Fall mit Wahrscheinlichkeit $\mu_{z'}/(\mu_z + \mu_{z'})$ akzeptiert. Mithilfe diese MCMC-Methode könnte man nun den Erwartungswert der (absoluten) Magnetisierung $f(z) = \frac{1}{|E|} |\sum_{e \in E} z(e)|$ in Abhängigkeit der Temperatur numerisch bestimmen. Für $m \rightarrow \infty$ berechnete Onsager 1944 [6], dass der Erwartungswert gegeben ist als

$$E_{\mu^\beta}(f) = \begin{cases} (1 - \sinh^{-4}(2\beta))^{1/8}, & \beta > \beta_c = \frac{1}{2} \log(1 + \sqrt{2}) \approx 0.44069 \\ 0, & \beta \leq \beta_c \end{cases}, \quad (11.66)$$

siehe Abbildung 11.3. Bei $\beta = \beta_c$ findet also ein Phasenübergang statt, d.h. dass nur für $\beta > \beta_c$ spontane Magnetisierung auftritt. Will man dies nun allerdings mit der vorgestellten MCMC-Methode numerisch verifizieren, stößt man auf mehrere Probleme: Oberhalb von β_c ist unsere MCMC-Methode nicht mehr schnell mischend und benötigt exponentiell viele Schritte in m , um den Totalvariationsabstand zur stationären Verteilung unter gegebene Schranken zu bringen; es ist nicht klar, wie groß der burn-in zu wählen ist; und Fehlerschranken/Konfidenzintervalle für die MCMC-Methode müssten bestimmt werden, um für verschiedene Gittergrößen vergleichbare Ergebnisse zu erhalten. Aus diesen Gründen kann ich hier leider keine Ergebnisse analog zu Abbildung 6.12 in [1] zeigen. In Zukunft sollte dann, falls ich dazu irgendwann einmal Zeit finde, hierfür ein Cluster-Algorithmus oder eine perfekte Simulation implementiert werden. Um die Schwierigkeiten bei großen β zu illustrieren, zeigt Abbildung 11.4 die Ergebnisse unserer MCMC-Methode auf einem 128×128 -Gitter für die Magnetisierung $f(z) = \frac{1}{|E|} \sum_{e \in E} z(e)$ ohne Betragsstriche, für die der exakte Erwartungswert aus Symmetriegründen immer 0 sein muss für beliebige β . Die Markovkette wurde für verschiedene β zum Einen im vollständig magnetisierten Zustand $z = 1$ gestartet und zum anderen im unmagnetisierten Zustand, in dem die einzelnen Spins Bernoulli- $1/2$ -verteilt sind. Offensichtlich konvergiert die Kette für höhere β wesentlich langsamer gegen den Erwartungswert 0, da das Akzeptieren von Spin-Flips aus dem Zustand $z = 1$ heraus, der ein Minimum von H darstellt, mit steigendem β immer unwahrscheinlicher wird. Schließlich zeigen noch Abbildungen 11.5 bis 11.8 einige typische Zustände der Markov-Kette des Metropolisalgorithmus, ausgehend von einer unmagnetisierten Startverteilung, für verschiedene β .

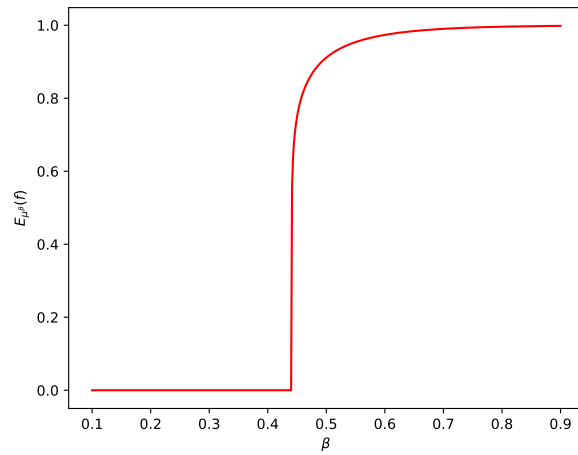


Abbildung 11.3: Erwartungswert der absoluten Magnetisierung im zweidimensionalen Isingmodell ohne äußeres Magnetfeld in Abhängigkeit der inversen Temperatur β nach Onsager.

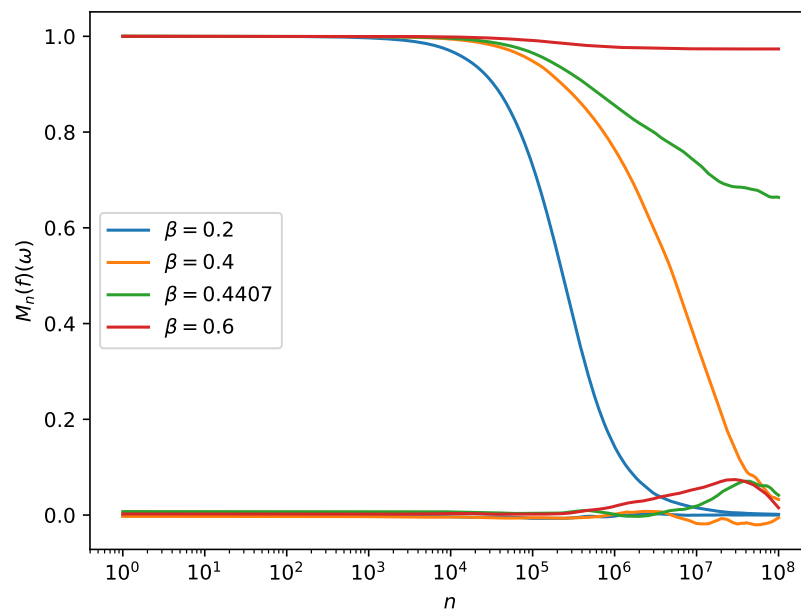


Abbildung 11.4: Numerisch mit dem Metropolis-Algorithmus ermittelter Erwartungswert der Magnetisierung im zweidimensionalen Isingmodell ohne äußeres Magnetfeld (exaktes Ergebnis: 0) in Abhängigkeit der inversen Temperatur β nach 10^8 Schritten der Markov-Kette auf einem 128×128 -Gitter.

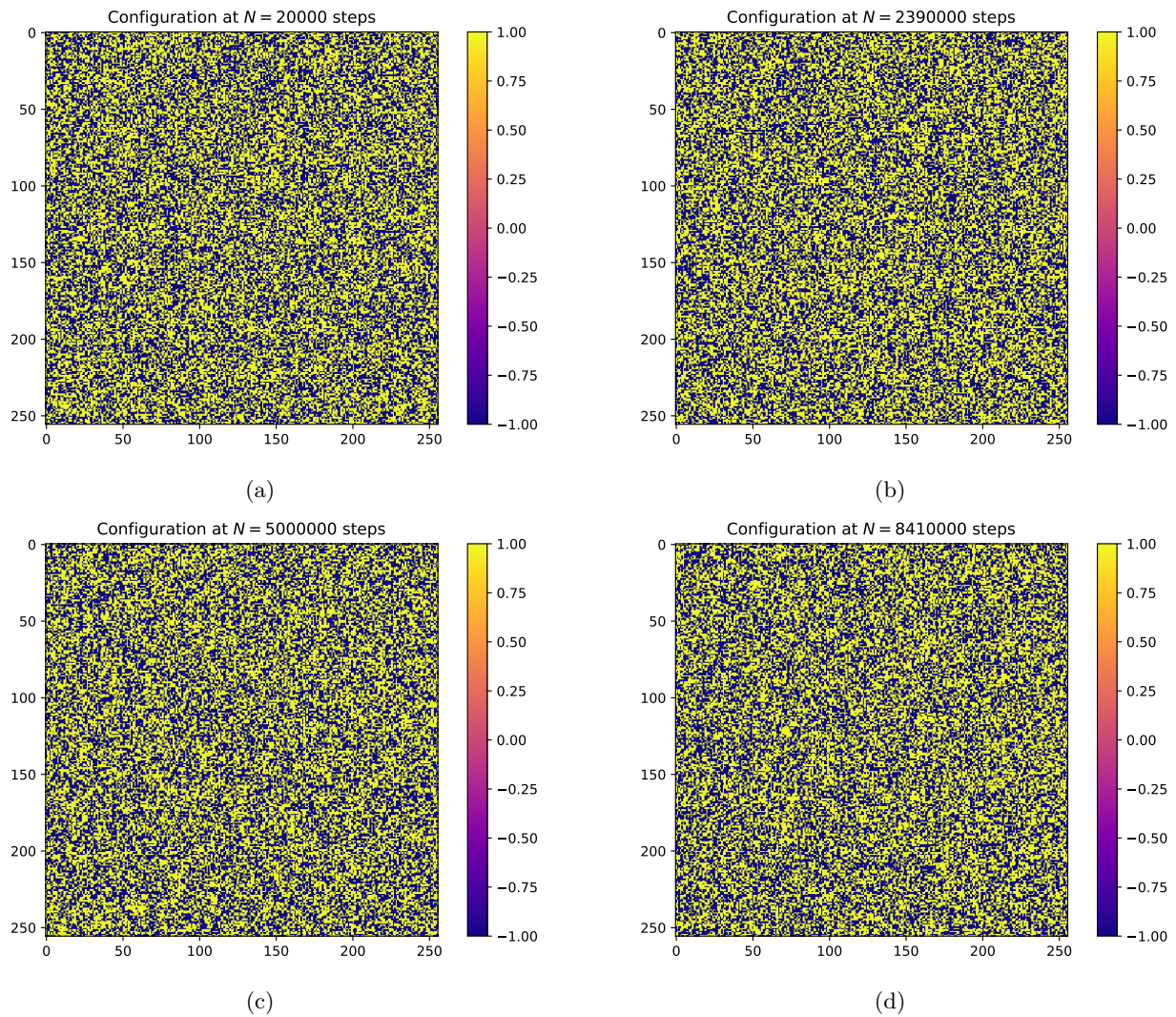


Abbildung 11.5: Beispiel einer Realisierung der Markov-Kette fürs Ising-Modell mit Metropolis-Algorithmus nach verschieden vielen Schritten (auf einem 256×256 -Gitter für $\beta = 0.1\beta_c$ und mit unmagnetisierter Startverteilung)

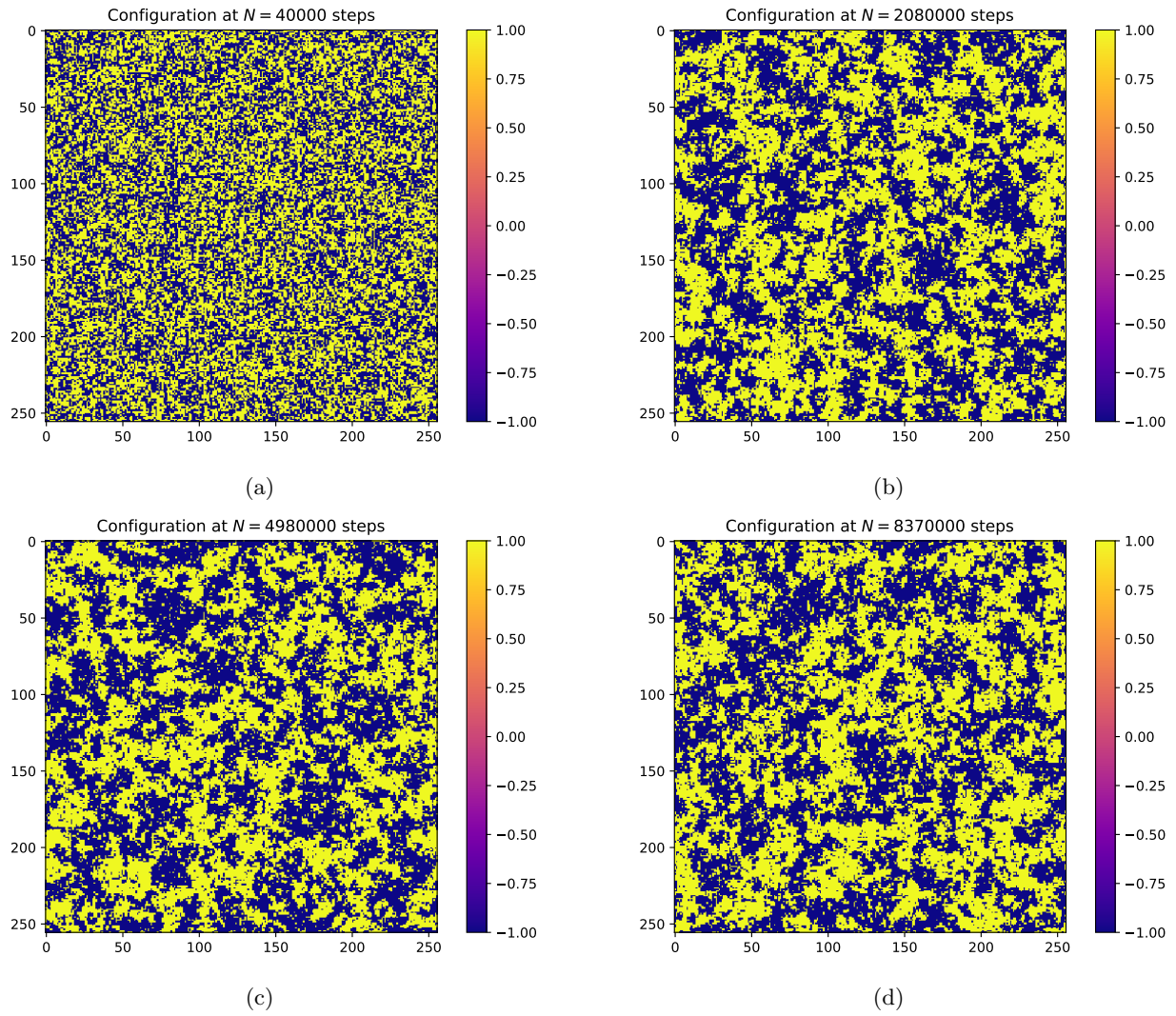


Abbildung 11.6: Beispiel einer Realisierung der Markov-Kette fürs Ising-Modell mit Metropolis-Algorithmus nach verschieden vielen Schritten (auf einem 256×256 -Gitter für $\beta = 0.9\beta_c$ und mit unmagnetisierter Startverteilung)

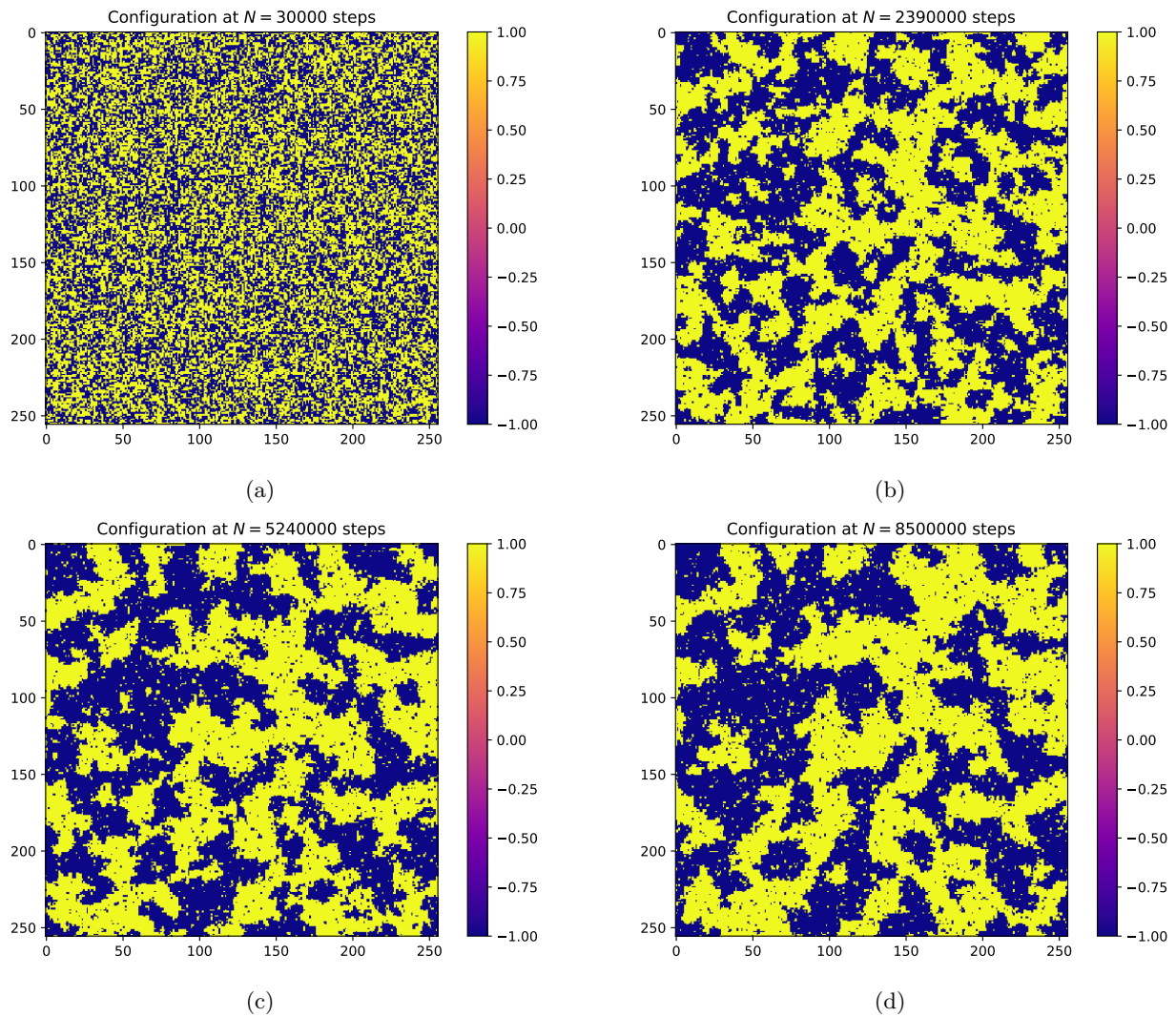


Abbildung 11.7: Beispiel einer Realisierung der Markov-Kette fürs Ising-Modell mit Metropolis-Algorithmus nach verschieden vielen Schritten (auf einem 256×256 -Gitter für $\beta = 1.1\beta_c$ und mit unmagnetisierter Startverteilung)

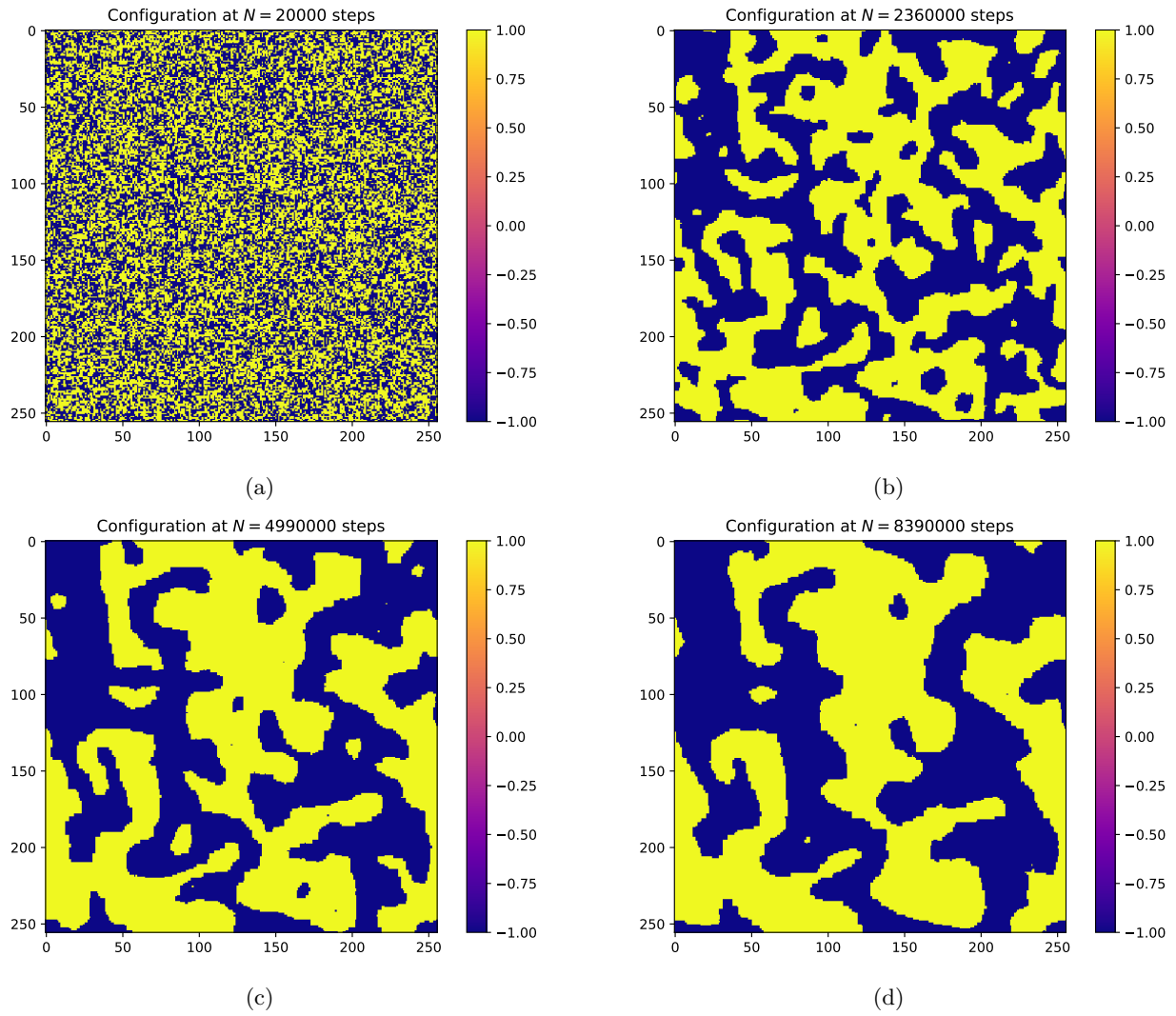


Abbildung 11.8: Beispiel einer Realisierung der Markov-Kette fürs Ising-Modell mit Metropolis-Algorithmus nach verschieden vielen Schritten (auf einem 256×256 -Gitter für $\beta = 3\beta_c$ und mit unmagnetisierter Startverteilung)

Literatur

- [1] T. MÜLLER-GRONBACH, E. NOVAK, K. RITTER, “Monte Carlo-Algorithmen” *Springer-Verlag*, 2012.
- [2] F. SCHWABL, “Statistische Mechanik” *Springer-Verlag*, 2006.
- [3] M. E. J. NEWMAN, G. T. BARKEMA, “Monte Carlo Methods in Statistical Physics” *Oxford University Press*, 2002.
- [4] D. A. LEVIN, Y. PERES, “Markov Chains and Mixing Times” *American Mathematical Soc.*, 2017.
- [5] N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, E. TELLER, “Equation of State Calculations by Fast Computing Machines” *J. Chem. Phys.* *21*, 1087, 1953.
- [6] L. ONSAGER, “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition” *Phys. Rev.* *65*, 117, 1944.