

Markov Chain Monte Carlo, an introduction

Erich Novak

9. Oktober 2010

Zusammenfassung

This lecture was given with the blackboard. Here I only collect a few key words and present some pictures. More details can be found in the forthcoming book on Monte Carlo methods by Müller-Gronbach, Novak and Ritter (Springer, 2011, in German).

Problem

The general problem: Compute an expectation

$$E_{\mu}(f) = \int_D f(x) d\mu(x),$$

where μ is a probability on some space D . For this lecture it is often ok to assume that D is finite.

On Monday I presented some error bounds for Markov Chain Monte Carlo of Daniel Rudolf. In this lecture I discuss the following questions: Why do we need MCMC? Examples? Problems?

Direct Simulation

Here we easily obtain the error bound

$$\sup_{\|f\|_2 \leq 1} E(E_{\mu}(f) - M_n(f))^2 = \frac{1}{n}.$$

Does not depend on the dimension! No curse of dimension!

Problem: We need a random number generator to construct random elements (according to μ) of D .

For SOME problems this works nicely, for other problems it does not work.

EXAMPLE: Problem of Sylvester/Blaschke/Klee: What is the expectation of the volume of a random tetrahedron in a (larger) tetrahedron of volume 1?

This problem was open for more than 100 years. With the help of direct simulation it could be showed (not strictly proved!) that the conjecture of Klee is probably wrong. The problem could be solved exactly only recently.

MCMC

I present a few simple examples where direct simulation does not work since we cannot construct a random number generator.

EXAMPLE: Consider the uniform distribution μ on the set

$$Z = \{A \in \mathbb{R}^{50 \times 50} \mid a_{ij} \in \mathbb{N}_0, \sum_j a_{ij} = \sum_j a_{ji} = 200\}.$$

This is a huge “complicated” set but it is easy to construct a Markov chain that converges to μ .

Result of Rudolf

Here I mention the result of Rudolf (see lecture from Monday) and stress the need of a burn in of the Markov chain. The simplest (asymptotic) error bound is

$$\lim_{n \rightarrow \infty} \sup_{\|f\|_2 \leq 1} n \cdot \mathbb{E}(\mathbb{E}_\mu(f) - M_n(f))^2 = \frac{1 + \beta}{1 - \beta},$$

where β is the second eigenvalue of the transition matrix P and the chain is irreducible, aperiodic and reversible.

Rapidly mixing Markov Chains

If a family (Z_k, μ_k, Q_k) of spaces Z_k and prob. measures μ_k and transition matrices Q_k is given then we call it rapidly mixing if

$$\frac{1}{1 - \beta_k} \leq C \cdot (\log |Z_k|)^\alpha$$

for some $C, \alpha > 0$.

Hard Core Model

Also called “independent sets problem”. A recent result is that the Metropolis algorithm is rapidly mixing in dimension 2 but slowly mixing for larger dimensions.

I present some numerical results.

Ising model

At the end of the lecture I present the Ising model in two dimensions, without an outer magnetic field. Very interesting is the absolute magnetization where we obtain a phase transition.

There is great recent progress concerning fast Monte Carlo simulation of the absolute magnetization and/or the construction of random Ising states.

Discrete cube, need for a burn in

[This is in German, sorry.]

Wir betrachten den diskreten Würfel $Z = \{0, 1\}^d$ mit der Markov-Kette: gehe mit Wahrscheinlichkeit $1/d$ zu einem der d Nachbarn. Die Kette ist irreduzibel, die Gleichverteilung ist stationär. Für $d = 1$ hat man die beiden Eigenwerte 1 und -1 . Für $d > 1$ hat man eine Tensorproduktstruktur und erhält die Eigenwerte

$$\beta_s = \frac{1}{d} \sum_{i=1}^d (-1)^{s_i},$$

die wir hier mit $s \in Z$ indizieren. Der zweitgrößte Eigenwert ist also $\beta_1 = 1 - 2/d$, siehe Bassetti, Diaconis 2006. Allerdings ist diese Kette nicht aperiodisch. Eine aperiodische Kette erhält man wieder durch das „lazy“ machen, dann gilt

$$\beta_1 = 1 - 1/d = 1 - \frac{1}{\log_2 |Z|}.$$

Ein Vergleich mit dem Kreisring zeigt, daß der zweitgrößte Eigenwert β_1 (im Verhältnis zur Größe von Z) jetzt viel langsamer gegen 1 konvergiert. Die Kette ist „schnell mischend“. In LPW09 wird die Kette auf $Z = \{0, 1\}^d$ so beschrieben: Wähle eine zufällige Koordinate und setze dann die entsprechende Koordinate von s zufällig auf 0 oder 1. Betrachte den Integranden

$$f(z) = e^{-|z|} \tag{1}$$

mit $|z| = \sum_i z_i$. Offenbar gilt $E_\mu(f) = (\frac{1}{2} + \frac{1}{2e})^d$. Wir zeigen ein numerisches Experiment mit $N = n + n_0 = 10^6$ und $d = 100$, einmal für $n_0 = 0$ und dann für $n_0 = 200$. Siehe Abbildung 1. Die Kette wird im Ursprung gestartet, hier gilt $f(z) = 1$.

Dieses leicht zu durchschauende Beispiel hat typische Eigenschaften: Der Integrand f ist in weiten Teilen von Z sehr klein, nimmt aber auch große Werte an. Startet man die Kette im Bereich dieser „untypisch großen“ Funktionswerte, so ergibt der Algorithmus M_n viel zu große Werte. Bei diesem Beispiel genügt, weil die Kette schnell mischend ist, der relativ kleine burn in von $n_0 = 200$, um gute Schätzwerte zu erhalten.

Numerical Results for the hardcore model in 2 dimensions

[again in German]

Wir betrachten wieder die Aufgabe, den Erwartungswert $E_\mu(f)$ der Funktion

$$f(z) = \frac{1}{|E|} \sum_{e \in E} z(e)$$

bezüglich der Gleichverteilung μ auf Z auszurechnen. Bei numerischen Experimenten mit dem Standard MCMC Algorithmus M_n und mit Gittergrößen bis zu $m = 128$ hat sich folgendes herausgestellt. Wählt man den Anfangszustand deterministisch, zum Beispiel $z = 0$ oder z gemäß dem Schachbrettmuster besetzt, so benötigt man etwa $50m^2$ Schritte, um die Kette aufzuwärmen, d. h. um typische Zustände zu erzeugen. Diese Aufwärmzeit ist relativ kurz, d. h. die Kette ist schnell mischend.

Die plausible Vermutung, daß für große m ungefähr $E_\mu(f) = 1/4$ gilt, bestätigt sich nicht ganz. Für alle $m > 4$ scheint der Wert kleiner zu sein. Bei diesem Integranden f sind alle Funktionswerte klein: die Funktionswerte sind nichtnegativ und der größte ist nur rund doppelt so groß wie der Erwartungswert. Daher ist das Aufwärmen nicht so wesentlich wie beim Integranden (1).

Abschließend zeigen wir noch einen „typischen“ Zustand beim hard core model mit $m = 32$. Erzeugt wurde das Bild mit der obigen Markov-Kette, wobei 10^7 Schritte gemacht wurden, gestartet wurde mit der Nullfunktion. Schon nach wenigen Zeitschritten, jedenfalls genügen 50 000 Schritte, erhält man Bilder, die qualitativ sehr ähnlich sind.

Numerical Results for the Ising Model

[in German]

Wir haben Standardalgorithmen für das Ising-Modell vorgestellt und man wird sich fragen, ob die entsprechenden Markov-Ketten schnell mischend sind oder nicht. Das folgende tiefliegende Ergebnis wurde vor nicht allzu langer Zeit bewiesen, siehe etwa Martinelli 1999: Es gibt einen kritischen Wert $\beta_c > 0$ so daß für $\beta < \beta_c$ die Kette schnell mischend ist, für $\beta > \beta_c$ jedoch langsam mischend.

Wir wollen uns eine viel schwächere Aussage zumindest plausibel machen. Sei $\beta > 0$ sehr klein, die Temperatur T also sehr groß. Dann unterscheidet sich die Boltzmann-Verteilung nur wenig von der Gleichverteilung und die Irrfahrt auf dem Raum der m^2 Spins unterscheidet sich nur wenig von der einfachen Irrfahrt auf dem diskreten Würfel $\{0, 1\}^{(m^2)}$, von der wir schon wissen, daß sie schnell mischend ist.

Ist $\beta > 0$ sehr groß, dann unterscheidet sich die Boltzmann-Verteilung nur wenig von der Gleichverteilung auf der zweielementigen Menge $\{z = 1, z = -1\}$. Startet man eine Kette beispielsweise im Zustand $z = 1$, so wird man (obwohl die Kette irreduzibel ist) mit großer Wahrscheinlichkeit in einer kleinen Anzahl von Schritten, wobei hier 10^{100} noch als klein gilt, nur solche Zustände \tilde{z} erreichen, die hauptsächlich den Funktionswert 1 annehmen. Die beiden Teilmengen „die meisten Spins sind +1“ und „die meisten Spins sind -1“ sind nur durch einen extrem dünnen Flaschenhals verbunden, dadurch ist die Mischungszeit der Kette bei kleiner Temperatur extrem groß.

Die angegebene Kette ist also im allgemeinen *nicht* zur Berechnung von Erwartungswerten geeignet. Allerdings gibt es Ausnahmen: Ist der Integrand symmetrisch im Sinne von $f(z) = f(-z)$, so spielt der „Flaschenhals“ keine Rolle und man kann, wenn man mit dem Zustand $z = 1$ startet, doch relativ schnell beispielsweise den Erwartungswert der spontanen Magnetisierung berechnen. Auf diese Weise kann man die Abbildung 4 erzeugen, bei der die spontane Magnetisierung für verschiedene Gittergrößen berechnet wurde. Im Vergleich dazu ist auch der Erwartungswert der absoluten Magnetisierung $E_{\mu\beta}(f)$ für den thermodynamischen Limes, d. h. $m \rightarrow \infty$, angegeben, die entsprechende Formel lautet $(1 - \sinh^{-4}(2\beta))^{1/8}$ für $\beta > \beta_c$ und 0 für $\beta < \beta_c$. Das kritische β_c beträgt

$$\beta_c = \frac{1}{2} \log(1 + \sqrt{2}) \approx 0.44069.$$

Diesen Grenzwert hat Onsager im Jahr 1944 ausgerechnet, damit wurde das Phänomen Phasenübergang erstmals für ein halbwegs realistisches Modell mathematisch verstanden. Phasenübergang heißt hier folgendes: Bei kleinen Temperaturen, also $T < T_c$, richten sich die Spins „spontan“, d. h. ohne äußeres Magnetfeld aus,

der Ferromagnet wirkt als Magnet. Diese spontane Magnetisierung verschwindet bei $T > T_c$.

Zur Veranschaulichung zeigen wir noch typische Spinzustände bei einem 250×250 Gitter mit $\beta = 0.9\beta_c$, $\beta = \beta_c$ und $\beta = 1.05\beta_c$.

Jetzt sollten andere, bessere, Algorithmen für das Ising-Modell beschrieben werden, aber die Zeit war kurz!

Danke an Mario Ullrich, der alle Bilder produziert hat!

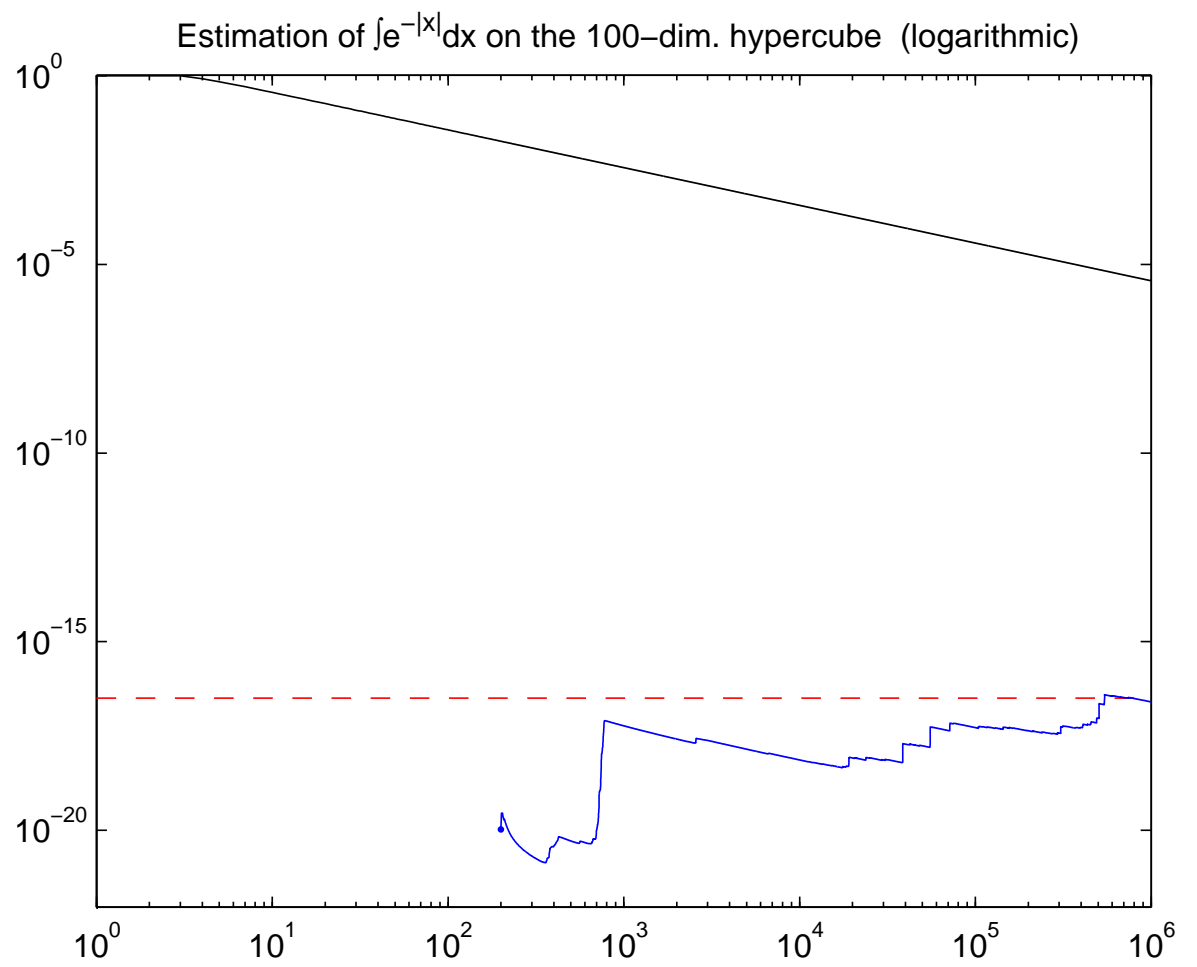


Abbildung 1: Wirkung vom burn in

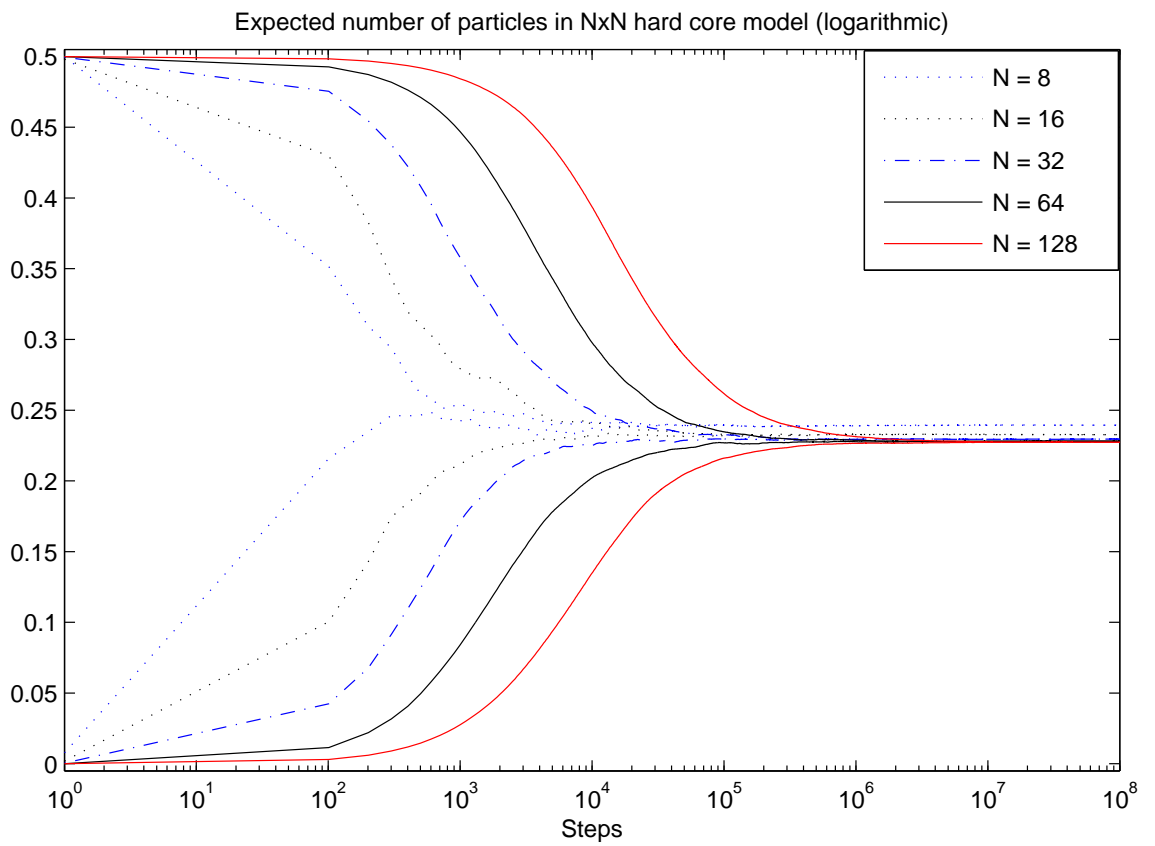
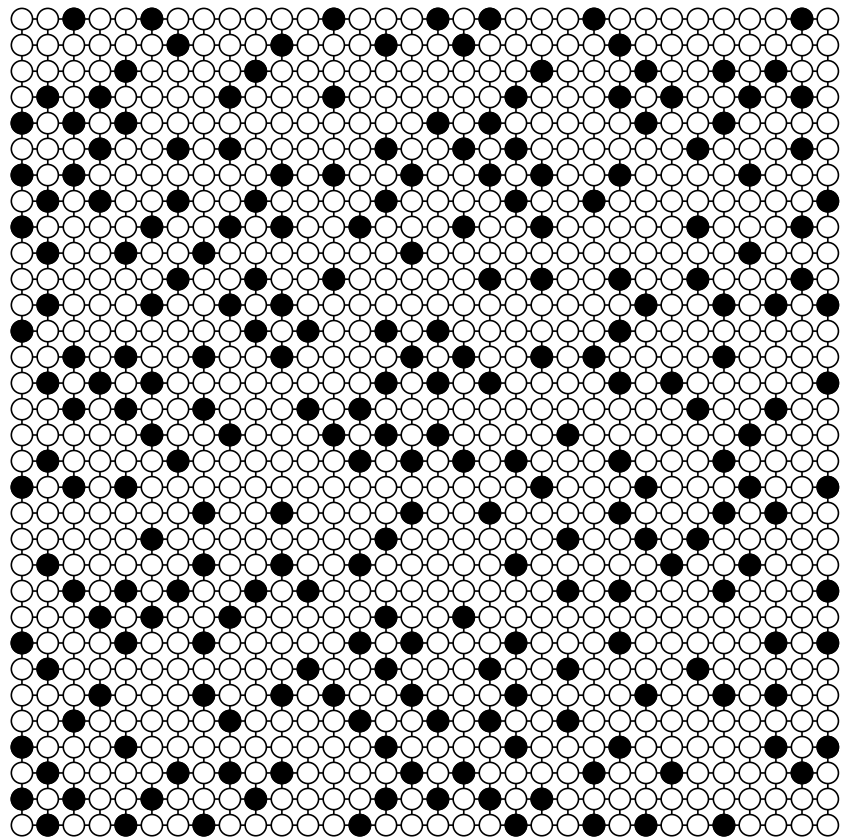


Abbildung 2: hard core model

32x32 hard core model



Step = 10000000

Abbildung 3: hard core model

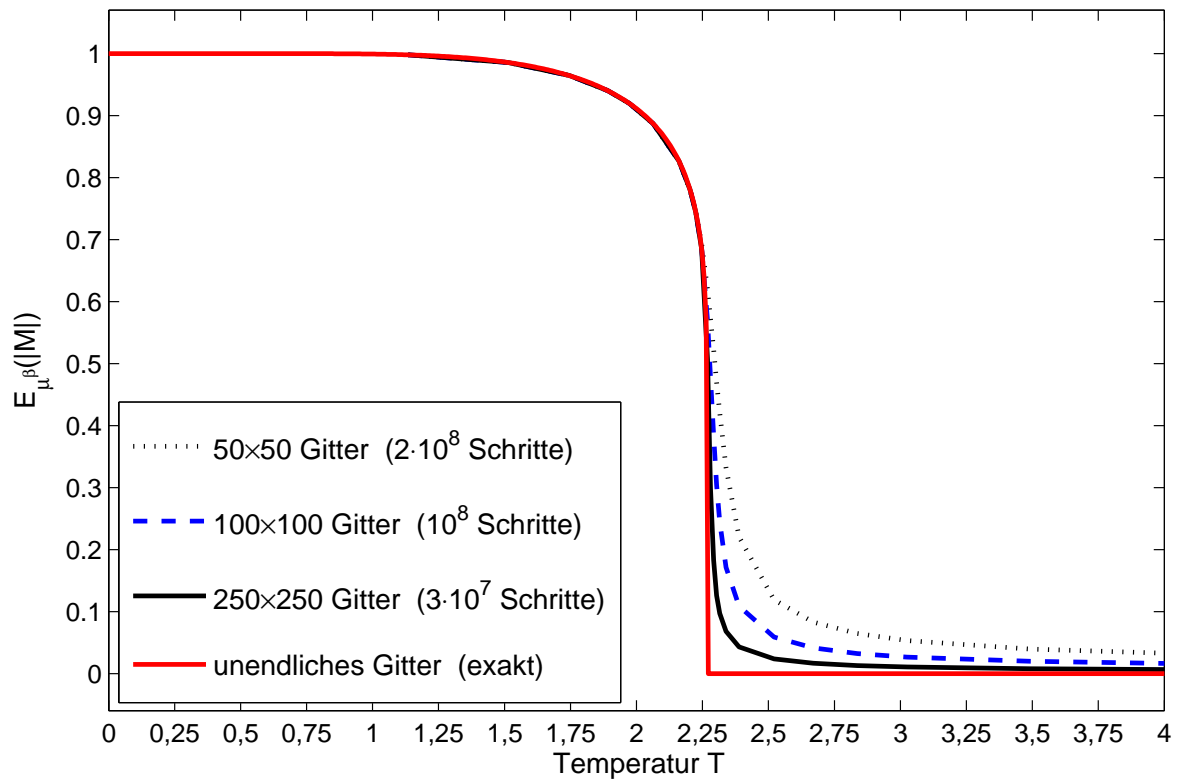


Abbildung 4: Erwartungswert der absoluten Magnetisierung bei verschiedenen Gittern; im Vergleich dazu das theoretische Ergebnis von Onsager im thermodynamischen Limes, $m \rightarrow \infty$.

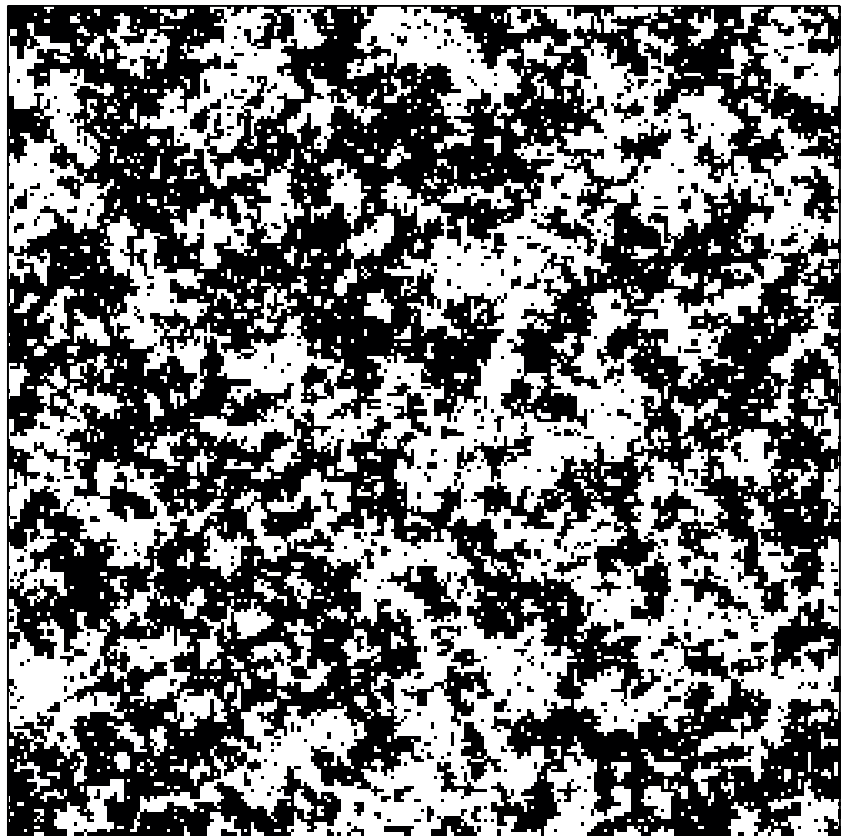


Abbildung 5: Spinzustand bei einem 250×250 Gitter mit $\beta = 0.9\beta_c$.

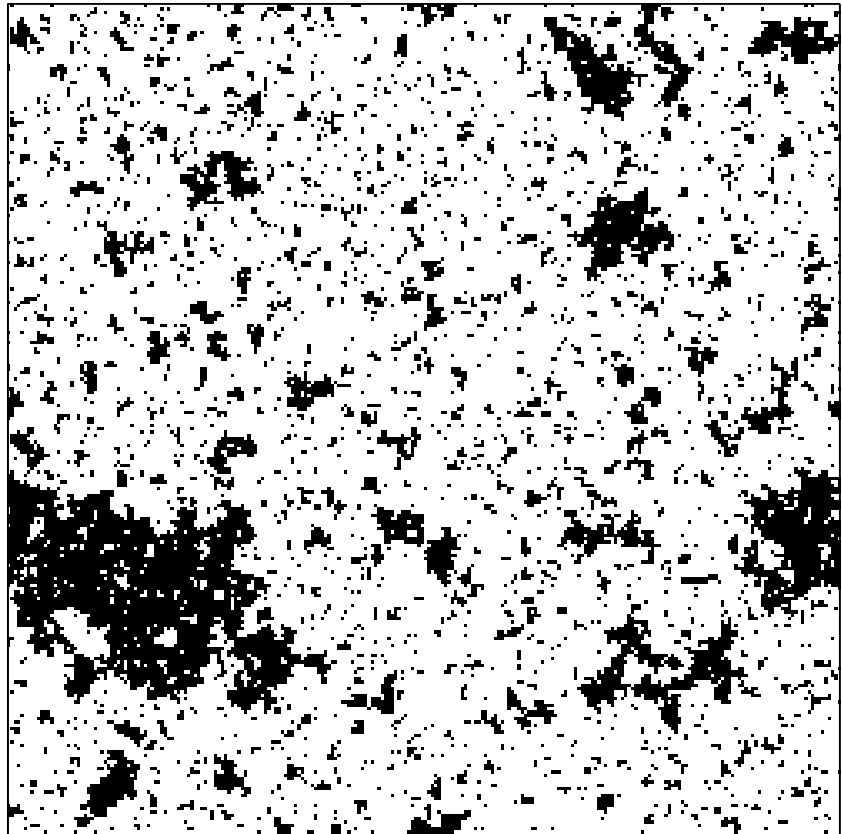


Abbildung 6: Spinzustand bei einem 250×250 Gitter mit $\beta = \beta_c$.

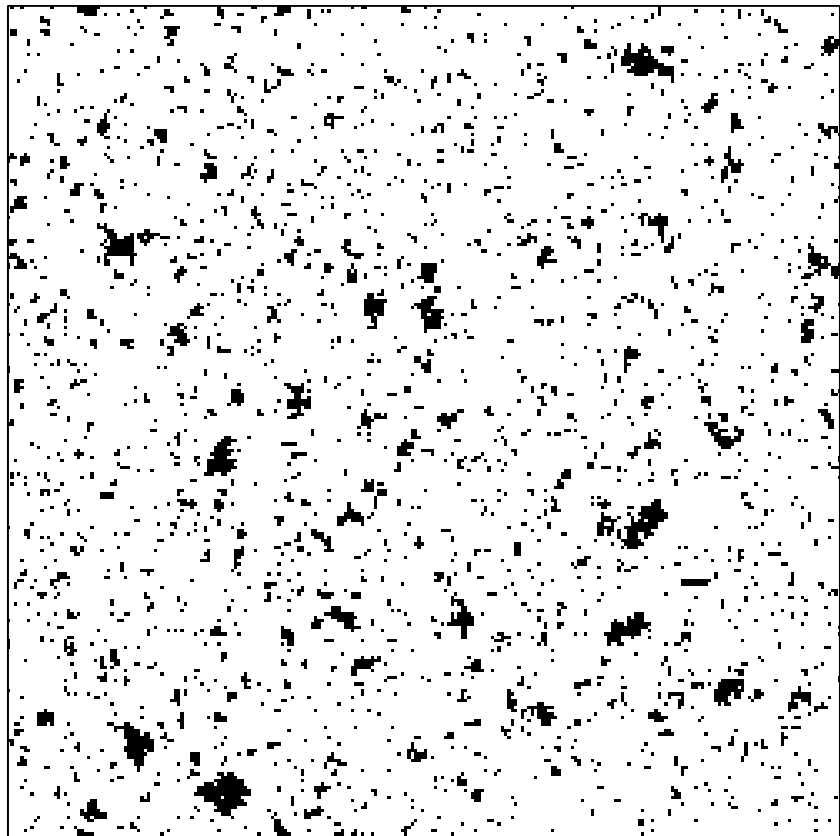


Abbildung 7: Spinzustand bei einem 250×250 Gitter mit $\beta = 1.05\beta_c$.

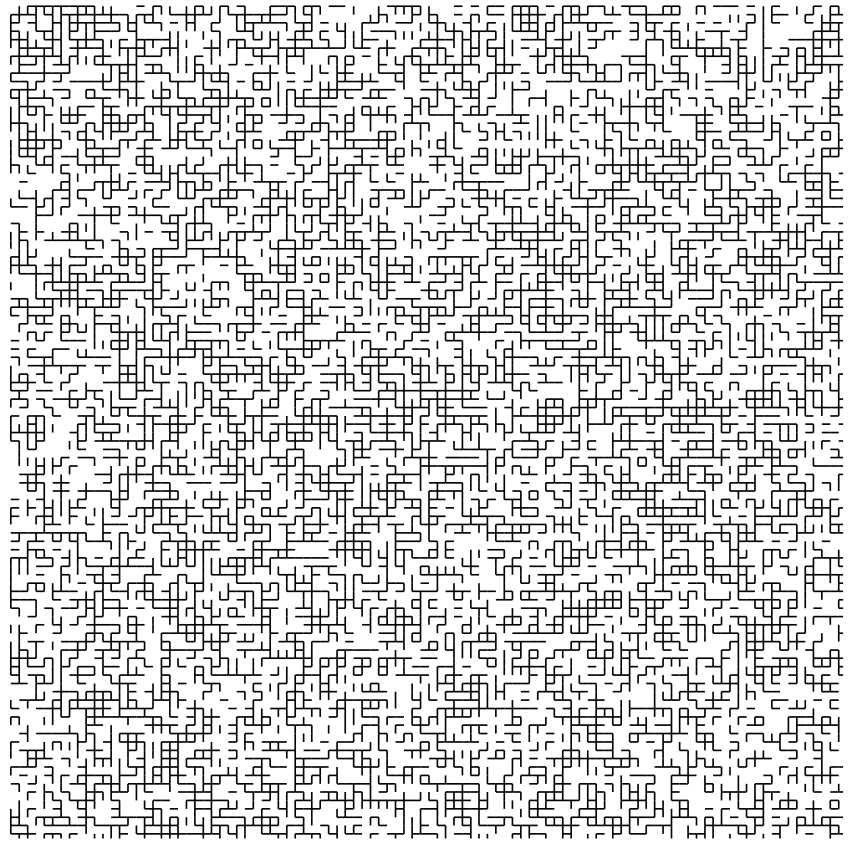


Abbildung 8: random cluster Zustand, 100×100 Gitter mit p_c .

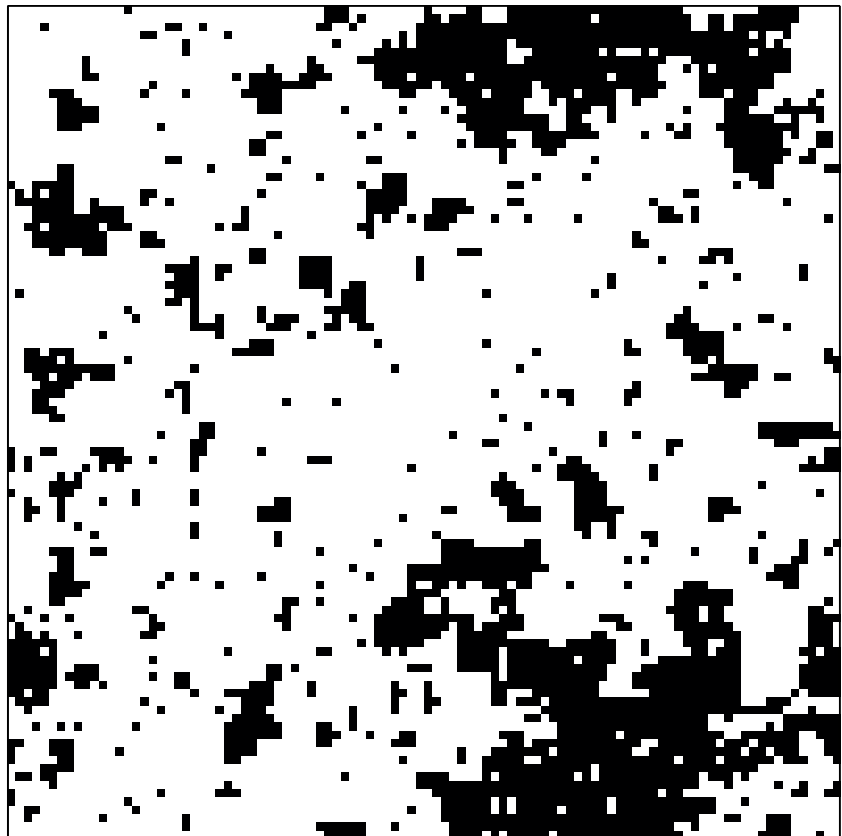


Abbildung 9: Spinzustand erzeugt aus Abb. 8, 100×100 Gitter mit $\beta = \beta_c$.

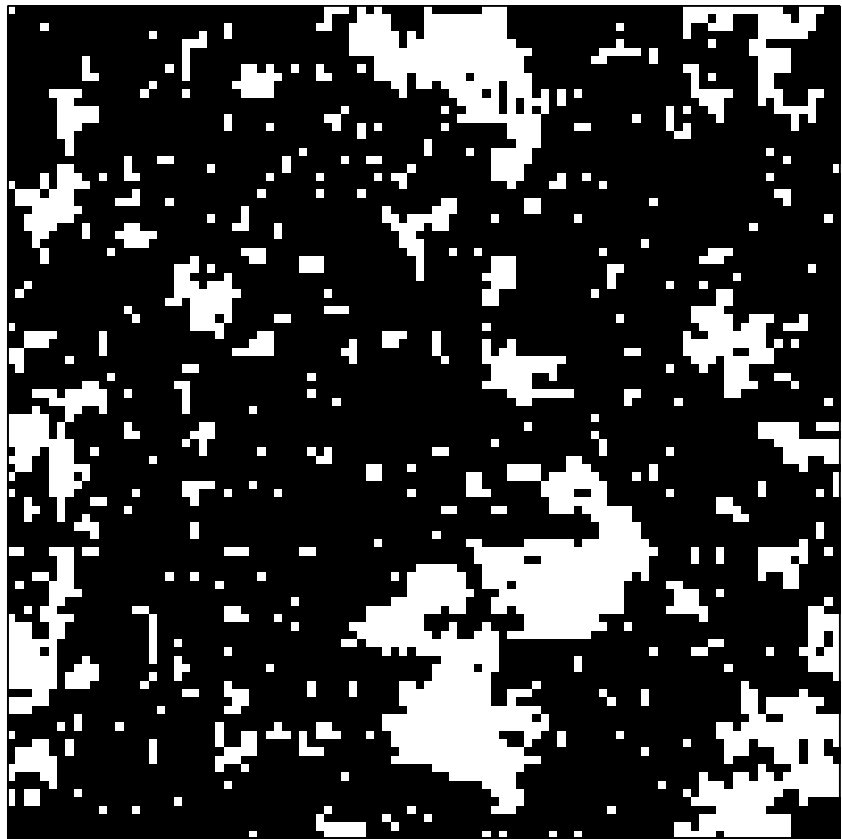


Abbildung 10: Spinzustand erzeugt aus Abb. 8, 100×100 Gitter mit $\beta = \beta_c$.