Monte Carlo sampling without Markov chains

Alexandre Chorin, Jakub Kominiarczuk
University of California, Berkeley

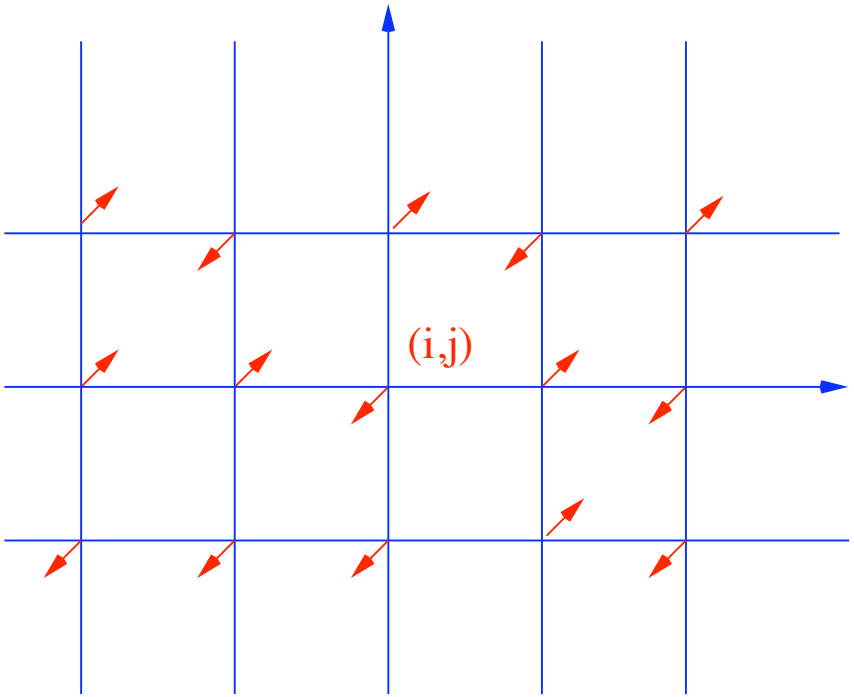
example: Ising model.

At each location of an N by N periodic lattice there is a spin $s_{ij}, 0 \leq i, j \leq N$, $s_{ij} = +1$ or -1. The global state S is the set of all spin values; there are 2^{N^2} states.

The probability of a state is $P(S) = \exp(-\beta H(S))/Z$,

where: $H = -\sum J_{ij,i'j'} s_{ij} s_{i'j'}$, Z is a normalization constant, $J... = 1$ if $ij, i'j'$ are neighbors, 0 otherwise.

$F(S)$ is some function of S. Task: compute $E[F] = \sum F(S)P(S)$.


(i,j)

Suppose N=30; there are $2^{900} \sim 10^{300}$ states, too many to enumerate.

Suppose $\beta = 1$, the ratio P(most likely state)/P(least likely state) $\sim e^{7200}$- no ab-initio sampling is possible.

Markov chain Monte Carlo (MCMC): chain $S_1, S_2, \ldots$, constructed as follows: To go from S_n to S_{n+1}, make a move, accept it with probability $\exp(-\beta(H(new) - H(old)))$. If move rejected, stay in place.

If the move is to a state of higher probability, it is always accepted, a move to a state of lower probability is sometimes accepted.

Favors small moves; cluster algorithms improve matters.

Problem: free energy landscape may have multiple minima.

Goal: non-rejection sampling that samples states with a frequency equal to their probability.

1D Ising model:

$$P(S) = Z^{-1} \exp(-\beta H),$$
$$H = -\sum s_i s_{i+1},$$

or

$$P(S) = Z^{-1} \exp W(S),$$
$$W(S) = K_0 \sum (s_i s_{i+1}) + A_0,$$

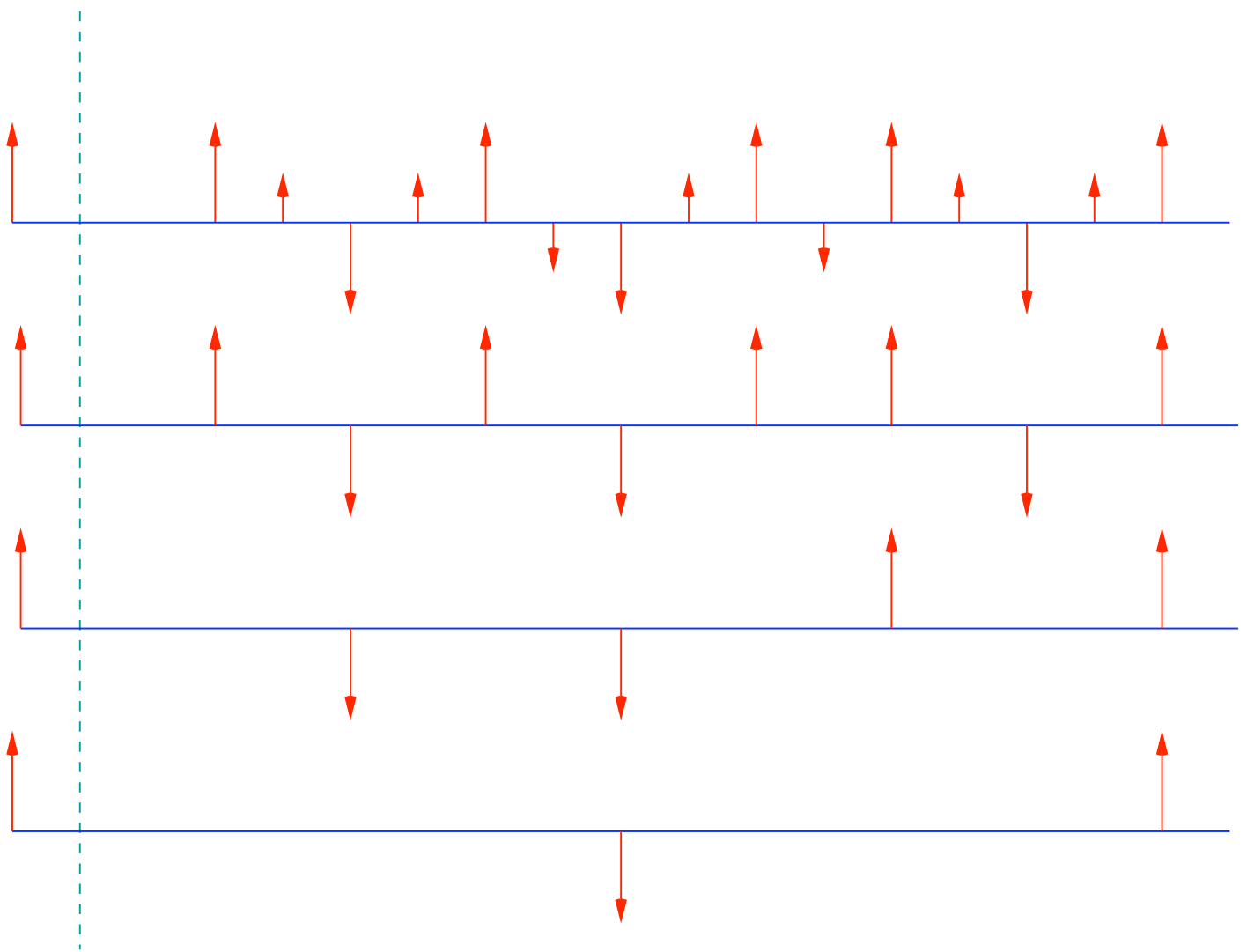
$$A_0 = 0,$$
$$K_0 = \beta.$$

$\widehat{S} = \cdots, s_{-5}, s_{-3}, s_{-1}, s_1, \cdots,$

$\widetilde{S} = \cdots, s_{-2}, s_0, s_2, s_4, \cdots a,\ S = \widehat{S} \cup \widetilde{S}.$

$S = \widehat{S} \cup \widetilde{S}.$

$P(\widehat{S}) = \sum_{\widetilde{S}} Z^{-1} \exp W(\widehat{S}, \widetilde{S})$



Ansatz (Kadanoff renormalization):

$$P(\widehat{S}) = \exp W^{(1)},$$

$$e^{A_1 + K_1 \sum s_i s_{i+2}}/Z = \sum_{\widetilde{s}} e^{A_0 + K_0 \sum' s_i s_{i+1}}/Z,$$

where $\sum'$ is the sum over $s_2 = -1, +1,\ s_4 = -1, +1, \ldots$.

Condition:

For all values of s_1, s_3:

$$e^{A_1 + K_1 s_1 s_3} = \sum_{s_2 = \pm 1} (e^{A_0 + K_0 s_1 s_2} + e^{A_0 + K_0 s_2 s_3})$$

Solution:

$$
\begin{aligned}
K_1 &= \frac{1}{2}\log\cosh(2K_0), \\
A_1 &= \frac{1}{2}(\log 2 + 2A_0 + K_1).
\end{aligned}
$$

More dimensions:

$P_0 = Z^{-1} \exp(W^{(0)})$,

$W^{(0)} = \sum s_{i,j} \left(J_{i,j,1} s_{i+1,j} + J_{i,j,2} s_{i,j+1} \right)$.

$P_1 = e^{W^{(1)}}/Z$,

$S = \hat{S} \cup \tilde{S}$ is the set of spins in L_0.

$$P_1(\hat{S}) = e^{W^{(1)}} = \sum_{\tilde{S}} e^{W^{(0)}(S)}$$

or

$$W^{(1)} = \log \sum_{\tilde{S}} e^{W^{(0)}(S)}.$$

Extend the range of values of a spin in $\widehat{S}$:

$s_{i_0,j_0} \to t \in [-1,+1]$.

$$dW/dt = \frac{\sum' \frac{dW^{(0)}}{dt} e^{W^{(0)}(S)}}{\sum' e^{W^{(0)}(S)}},$$

($\sum'$ is a sum over $\widetilde{S}$),

$$\frac{dW^{(1)}}{dt} = E\left[\frac{dW^{(0)}}{dt} \mid \widehat{S}\right].$$

A conditional expectation given $\widehat{S}$ is an orthogonal projection onto the space of functions of $\widehat{S}$, and we approximate it by projecting onto the span of a finite basis of functions of $\widehat{S}$.

Expand W in successive linkages (Kadanoff):

$W = W_1 + W_2 + \ldots$

$s_{i_o,j_0} \to t \in [-1, +1]$,

$dW/dt =$ a local expansion.

Example:

$W = \sum s_i s_{i+1} = 1/2(\cdots + s_1 t + t s_3 + s_3 s_4 + \ldots)$,

$dW/dt = \frac{1}{2}(s_1 + s_3)$.

Heuristics:

$P(a,b) \sim P(a)P(b),$

$logP(a,b) \sim logP(a) + logP(b),$

$dP/dt \sim (d/dt)logP(a).$

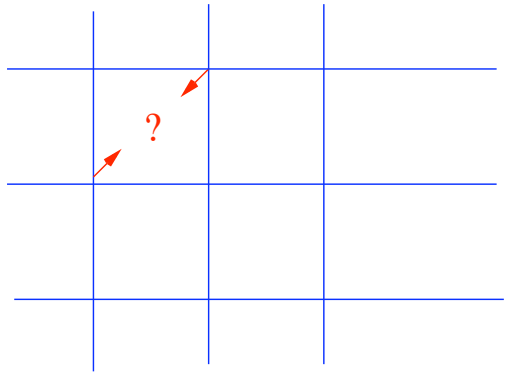
Choice of basis: in 1D, the basis $(\tanh(\beta(s_{+2}+t)),$ $\tanh(\beta(s_{-2}+t)))$ is exact.

dW/dt is not a linear function of t.

In 2D, $\tanh(B(s_{i'j'}+t))$, B determined in the projection process.

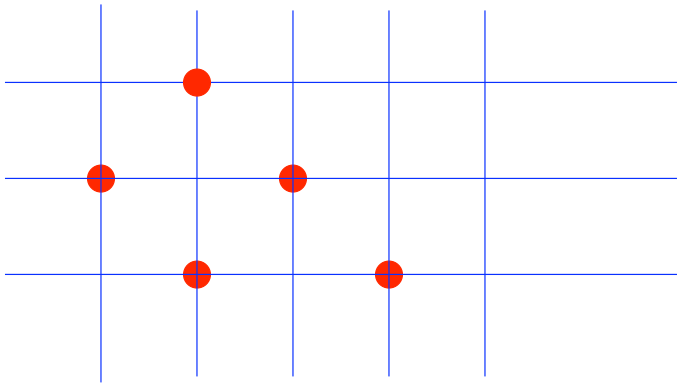
Once dW/dt has been determined, one obtains W by integration.

Warning: one cannot expect $(\partial^2 W)/(\partial t_1 \partial t_2)$ to be symmetric in i, j, but in the probability density, one can use the symmetrized matrix.



Nested lattices:

2D is easy:



in 3D, divide a set of nodes into interior and boundary nodes, interior nodes connected only to boundary nodes, small number of boundary nodes.

As one samples, one computes the probability of the sample:

Given the neighboring nodes, $P_+ = \exp W_+/(\exp W_+ + \exp W_-)$, $P_- = \exp W_-/(\exp W_+ + \exp W_-)$ are known. Accumulate their logs.

$\log P(sample) = \sum \log(P\pm)$.

$P(outcome) = \exp W^{(0)}$.

$\int g(x)f(x) = E[g(\eta)] = \int g(x)(\frac{f(x)}{f_0(x)})f_0(x) = E[g(\eta')Q(\eta')]$, where $Q(\eta') = f(\eta')/f_0(\eta')$ are integration weights.

Here $Q = P(outcome)/P(sample)$.

Q reduces the number of effective samples.

Calculating the expansion coefficients:

At each level, project dW/dt on a basis of functions of $\widehat{S}$, so that $dW/dt = c_1\psi(\widehat{S}) + c_2\psi_2(\widehat{S}) + \ldots$. (important: always project from the finest level !). The ψ_i are not assumed to be orthonormal.

Sample (assuming you know the sampling weights!) and accumulate $b_k = E[dW/dt\psi_k], \quad a_{ij} = E[\psi_i\psi_j]$.

Solve for the vector c of coefficients: $Ac = b$. (accumulation of c produces bias).

Near T_c, α is nearly independent of level.

Balance n. of samples vs. n. of polynomials.

Once you have coefficients the samples are cheap.

Bootstrapping the sampling:

Start with a guess of the expansion coefficients for each level: $c^n_{j,\ell}$, n=iteration, ℓ=level, j=coefficient.

Initial guess $A^0 = I$,
$b^0 = c^0$=the assumed coefficients.

Given c^n, accumulate values of A, b; every now and then set $A^{n+1} = (1/2)(A^n + A)$, $\quad b^{n+1} = (1/2)(B^n + b)$, then compute $c^{n+1} = (A^{n+1})^{-1} b^{n+1}$; because of bias, do not accumulate c.

Example 1: Ising model in 2D. Competitive with Metropolis, not competitive with a cluster algorithm. Interesting connections with the renormalization group.

Example 2: 3D Anderson-Edwards spin glass: H as before, $J_{ijk\ell} = N(0,1)$ random variables.

(We got the same bad results as everyone else but much faster).

Define: $< \cdot >_T$= average over one realization of the J...,
$[\cdot]_{Av}$=average over the J's.

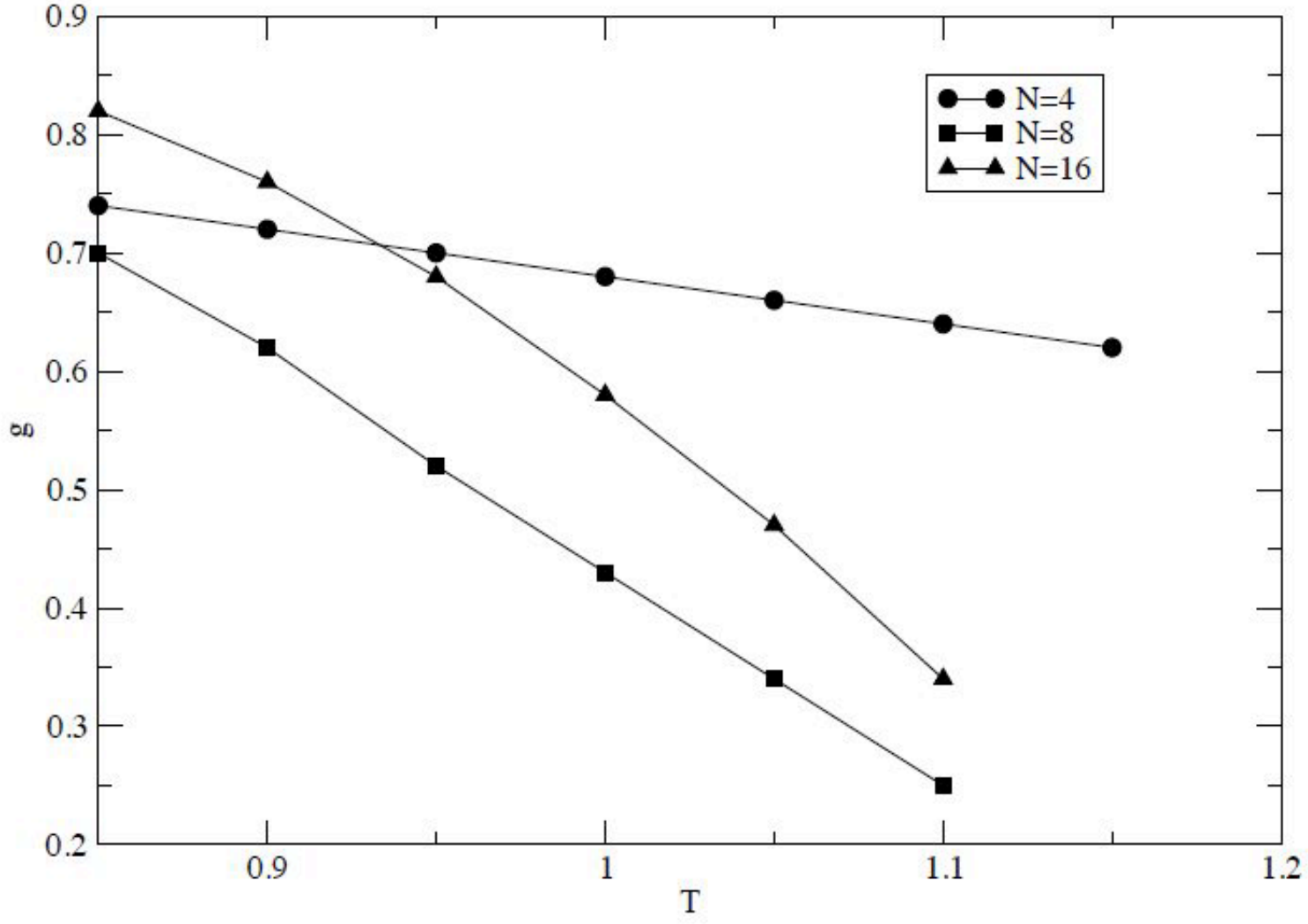
For two sets of spins $S_{1,2} = \{s_{ijk}^{1,2}\}$, overlap q is $q = N^{-3} \sum s_{i,j,k}^{1} s_{i,j,k}^{2}$.

The Binder cumulant is: $g = 0.5(3 - [< q^4 >]_{Av}/[< q^2 >_T]_{Av}$.

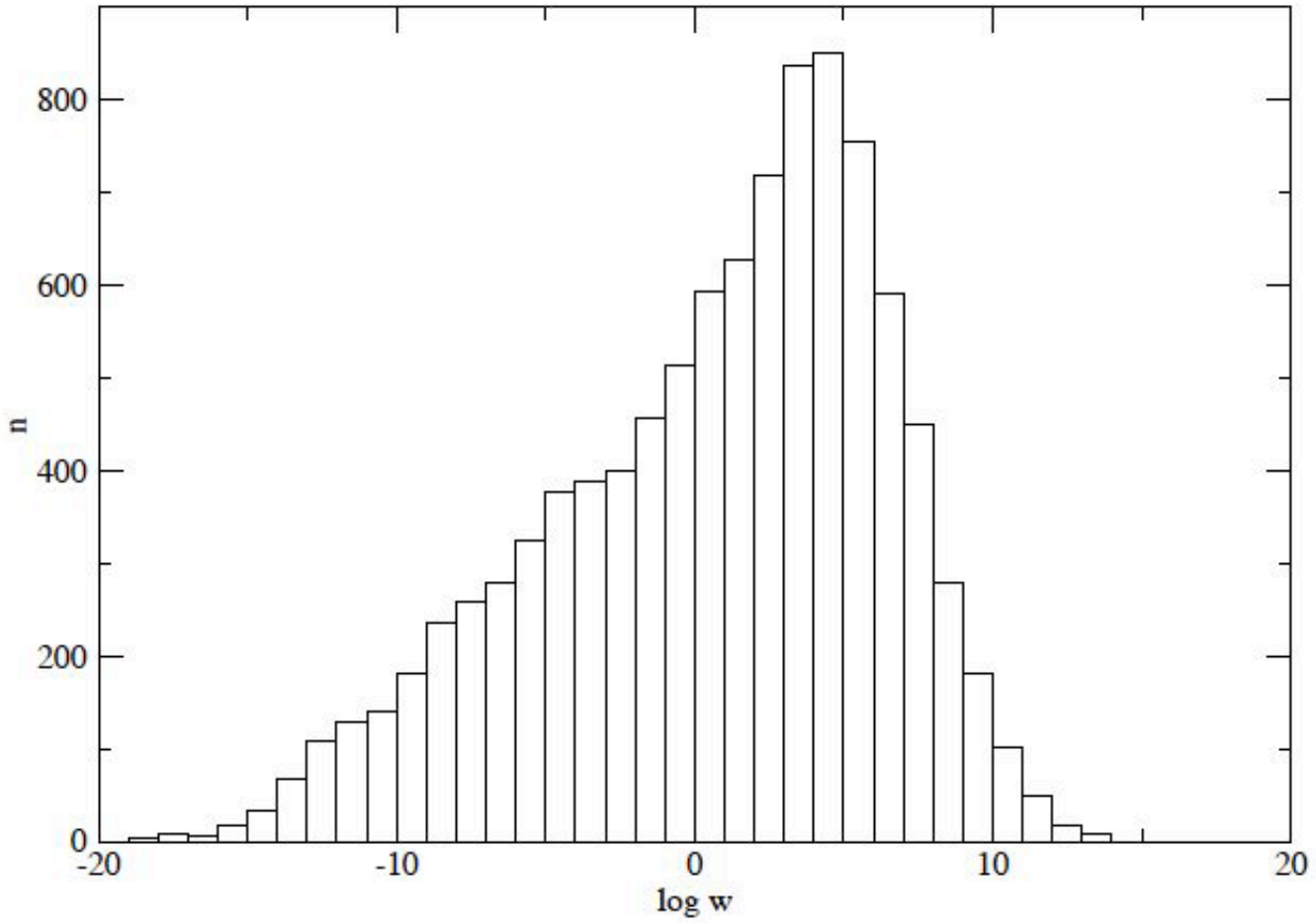
$g = g(T)$ is "universal", curves cross at $T = T_c$.

Old calculation: $\log Q$ goes up to 25; new calculation: $\log Q$ goes up to 4.

The Binder cumulant g as a function of the temperature T in the three-dimensional AE model



Histogram of weights for the Ising model, $N = 32$, 10^4 samples



Another example: sampling a Brownian bridge (Weare), application to filtering.

Variant: Parallel marginalization (Weare).