1 The Swendsen–Wang method

Several enhancements of Monte Carlo methods are based on a remarkable trick: take a big and difficult problem, and replace it by an even bigger problem that contains the first problem; then solve the bigger problem with methods that were not applicable to the first.

The Swendsen–Wang method takes a spin system such as

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(\sum_{n,n'} J_{nn'} x_n x_{n'} \right), \tag{1}$$

and replaces it by a bigger system that contains both the N original spin variables and M additional 'bond' variables, where M is the number of non-zero couplings $J_{nn'}$ in the original spin system. All the ideas that follow are from Swendsen and Wang (1987).

The extended Ising model

Let's represent the original Ising model by the product of factors

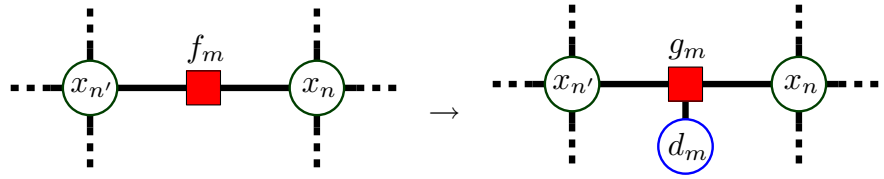
$$P(\mathbf{x}) = \frac{1}{Z} \prod_{m=1}^{M} f_m(\mathbf{x}_m), \tag{2}$$

where each factor $f_m(\mathbf{x}_m)$ depends on just two of the spins, say x_n and $x_{n'}$:

$$f_m(\mathbf{x}_m) = f(x_n, x_{n'}) = \exp(J_{nn'} x_n x_{n'}). \tag{3}$$

In a standard Ising model, all the couplings between neighbouring spins $J_{nn'}$ are equal to a single positive number, J. Later, we will generalize our description to arbitrary spin systems.

We extend this model by introducing M additional 'bond' variables, $\mathbf{d}$.



spins spins and bonds

We define a joint distribution that couples the spins to the bonds,

$$P(\mathbf{x}, \mathbf{d}) = \frac{1}{Z'} \prod_{m=1}^{M} g_m(\mathbf{x}_m, d_m), \tag{4}$$

in such a way that

(a) The marginal distribution of $\mathbf{x}$, $\sum_{\mathbf{d}} P(\mathbf{x}, \mathbf{d})$, is equal to the Ising distribution (2);

(b) the conditional distributions $P(\mathbf{x}|\mathbf{d})$ and $P(\mathbf{d}|\mathbf{x})$ are both simple to sample from.

Each bond variable has two possible states: 1 (also known as 'open' or 'connected') and 0 ('closed' or 'disconnected'). The factor $g_m(\mathbf{x}_m, d_m)$ is defined by the following

eight values:

$$g_m(\mathbf{x}_m, d_m) = \begin{cases} & & d_m=0 & & d_m=1 \\ & & x_{n'}=0 \quad x_{n'}=1 & x_{n'}=0 \quad x_{n'}=1 \\ x_n=0 & & e^{-J} \quad e^{-J} & e^{J}-e^{-J} \quad 0 \\ x_n=1 & & e^{-J} \quad e^{-J} & 0 \quad e^{J}-e^{-J}, \end{cases} \tag{5}$$

The distribution of $\mathbf{x}$ and $\mathbf{d}$ is unchanged if we rescale any factor by any constant; it will be convenient to introduce an alternative notation for the factor g_m, rescaling it by e^J, and defining

$$p \equiv 1 - e^{-2J}. \tag{6}$$

The rescaled factor is:

$$\tilde{g}_m(\mathbf{x}_m, d_m) = \begin{cases} & & d_m=0 & & d_m=1 \\ & & x_{n'}=0 \quad x_{n'}=1 & x_{n'}=0 \quad x_{n'}=1 \\ x_n=0 & & 1-p \quad 1-p & p \quad 0 \\ x_n=1 & & 1-p \quad 1-p & 0 \quad p, \end{cases} \tag{7}$$

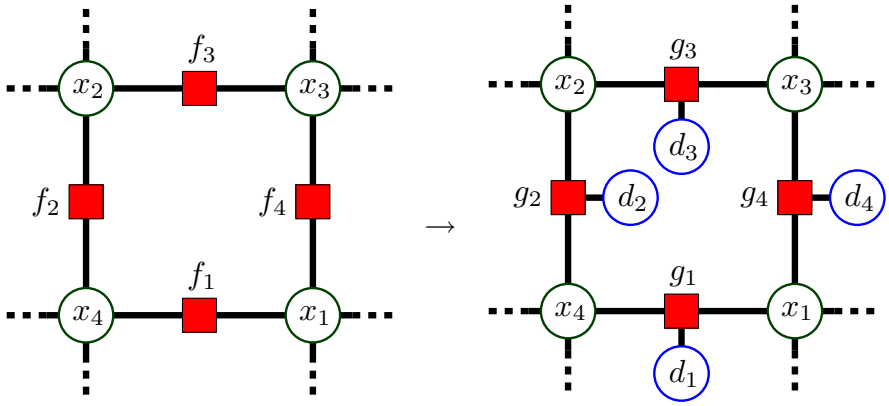
The extended model can thus be written as

$$P(\mathbf{x}, \mathbf{d}) = \frac{1}{Z} \prod_m g_m(\mathbf{x}_m, d_m) = \frac{1}{\tilde{Z}} \prod_m \tilde{g}_m(\mathbf{x}_m, d_m). \tag{8}$$

The first partition function Z is identical to the partition function of the original Ising model. The second is

$$\tilde{Z} = Z e^{-MJ} \tag{9}$$

In the case where the original spins are connected in a rectangular grid, the factor graph for the extended model looks like this:



spins spins and bonds

1.1 Sampling from the extended model by Gibbs sampling

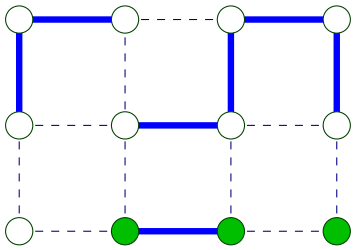
The conditional distributions $P(\mathbf{x}|\mathbf{d})$ and $P(\mathbf{d}|\mathbf{x})$ are as follows:

$P(\mathbf{x}|\mathbf{d})$ – the bonds connect the spins into a number of clusters (connected components); all spins in a cluster must adopt the same state as each other; the states ± 1 are selected with equal probability.

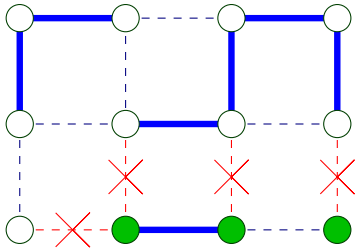
$P(\mathbf{d}|\mathbf{x})$ – conditional on the spins, the bonds are independent. If the two spins surrounding a bond are equal ($x_n = x_{n'}$), set the bond d_m to 1 with probability p (see equation (6)); otherwise set it to zero.

In the case of a rectangular grid, this Gibbs sampling algorithm mixes very rapidly.

The following figures illustrate Gibbs sampling. Spin states up and down are shown by filled and empty circles. Bond states 1 and 0 are shown by thick lines and thin dotted lines. We start from a state with five connected components. (Remember that isolated spins count as connected components, albeit of size 1.)

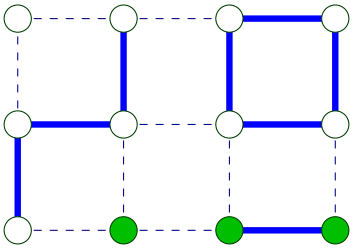


First, let's update the bonds

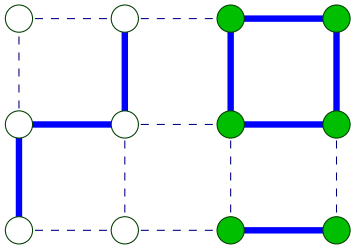


The forbidden bonds are highlighted

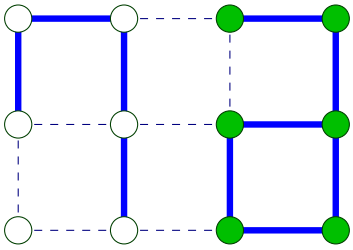
Bonds are forbidden from forming wherever the two adjacent spins are in opposite states. The bonds that are not forbidden are set to the 1 state with probability p.



After updating the bonds



Now we update spins



Update bonds again

1.2 Other properties of the extended model

We already mentioned that the partition function Z is the same as that of the Ising model.

The marginal $P(\mathbf{x})$ is correct, because when we sum the factor g_m over d_m, we get f_m. Summing over d_m is easy because it appears in only one factor.

OK, we've summed out $\mathbf{d}$ and obtained the Ising model. What if we sum out $\mathbf{x}$?

The marginal $P(\mathbf{d})$ is called the *random cluster model*. Summing over $\mathbf{x}$ for given $\mathbf{d}$, all factors are constants. The number of states is $2^{\text{number of clusters}}$. Thus

$$P(\mathbf{d}) = \frac{1}{\tilde{Z}} \prod_m \left(p^{d_m} (1-p)^{1-d_m} \right) 2^{c(\mathbf{d})} \tag{10}$$

where $c(\mathbf{d})$ is the number of connected components in the state $\mathbf{d}$. Isolated spins whose neighbouring bonds are all zero count as single connected components.

The random cluster model can be generalized by replacing the number 2 by a parameter q:

$$P^{(q)}(\mathbf{d}) = \prod_m \left(p^{d_m} (1-p)^{1-d_m} \right) q^{c(\mathbf{d})} \tag{11}$$

The random cluster model can be simulated directly, just as the Ising model can be simulated directly; but the S–W method, augmenting the bonds with spins, is probably the most efficient way to simulate the model. For integer values of q, the appropriate spin system is the 'Potts model', the generalization of the Ising model from 2 spin states to q.

2 Swendsen–Wang for a general spin system

We now include a bias h_n at each spin. And we allow the couplings J to be positive or negative.

Assuming that the original coupling associated with bond m, $J_{nn'}$, is positive, the factor $g_m(\mathbf{x}_m, d_m)$ is defined by the following eight values:

$$g_m(\mathbf{x}_m, d_m) = \begin{cases} & d_m=0 & & d_m=1 \\ & x_{n'}=0 & x_{n'}=1 & x_{n'}=0 & x_{n'}=1 \\ x_n=0 & e^{-J_{nn'}} & e^{-J_{nn'}} & e^{J_{nn'}} - e^{-J_{nn'}} & 0 \\ x_n=1 & e^{-J_{nn'}} & e^{-J_{nn'}} & 0 & e^{J_{nn'}} - e^{-J_{nn'}}, \end{cases} \tag{12}$$

As before we can introduce a parameter $p_m \equiv 1 - e^{-2J_{nn'}}$ and write a rescaled factor:

$$\tilde{g}_m(\mathbf{x}_m, d_m) = \begin{cases} & d_m=0 & & d_m=1 \\ & x_{n'}=0 & x_{n'}=1 & x_{n'}=0 & x_{n'}=1 \\ x_n=0 & 1-p & 1-p & p & 0 \\ x_n=1 & 1-p & 1-p & 0 & p, \end{cases} \tag{13}$$

If $J_{nn'}$ is *negative*, we define the factor g_m thus:

$$g_m(\mathbf{x}_m, d_m) = \begin{cases} & d_m=0 & & d_m=1 \\ & x_{n'}=0 & x_{n'}=1 & x_{n'}=0 & x_{n'}=1 \\ x_n=0 & e^{J_{nn'}} & e^{J_{nn'}} & 0 & e^{-J_{nn'}} - e^{J_{nn'}} \\ x_n=1 & e^{J_{nn'}} & e^{J_{nn'}} & e^{-J_{nn'}} - e^{J_{nn'}} & 0, \end{cases} \tag{14}$$

Two spins surrounding such a bond must be in *opposite* states if the bond is connected ($d_m = 1$).

When each cluster is updated, the total of the associated biases h_n is taken into account when determining the probability of the two states.

2.1 How well does it work?

The Swendsen–Wang method often mixes much better than simple Markov chains such as Gibbs sampling (that is, single-spin updates, also known as the heat bath method).

To be specific, consider the simple two-dimensional rectangular Ising model with positive coupling between neighbours. At high temperatures, spins are roughly independent, so any method works fine. As the temperature is reduced to the critical temperature, typical states have increasingly long correlation-length, and ordinary Gibbs sampling suffers from 'critical slowing-down' – it takes a very long time to obtain independent samples. Below the critical temperature, typical states fall into two sets: mainly-up, and mainly-down. The frequency with which ordinary Gibbs sampling gets between those two sets is exponentially small. In contrast, Swendsen–Wang sampling mixes fast at all temperatures, above, at, or below the critical temperature. Near the critical temperature, most of the spins are in large clusters, so a single flip of sign of one cluster changes a large number of spins in a correlated way. At very low temperatures, almost all spins are in one giant connected cluster.

Swendsen–Wang sampling isn't a cure-all for spin systems, however. It has difficulty with systems that have frustrations. The anti-ferromagnetic hexagonal Ising

model, for example, is full of frustrations, and it has non-zero entropy at absolute zero. If we use the Swendsen–Wang method for negative coupling described above, at very low temperatures, a typical state will have one third of the bonds forbidden and most of the other two thirds present. Those two thirds are likely to connect almost all the spins into a single giant connected component. The only change to the spins that can thus happen is a flip in sign of all spins. Such a move is fine for the very-low-temperature ferromagnet, which only has two states; but it does not do a good job of exploring the frustrated distribution, which has many other equal-energy states accessible to it. Swendsen–Wang does not propose moves that enable rapid exploration.

References

Swendsen, R. H., and Wang, J.-S. (1987) Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters* **58**: 86–88.

David J.C. MacKay (2004).

This is an addition to *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ. Press, 2003), which is available online from

`http://www.inference.phy.cam.ac.uk/mackay/itila/`