http://www.math.nyu.edu/faculty/goodman/teaching/MonteCarlo20/

Class notes: Monte Carlo methods
Week 8, Collective Modes and Samplers
Jonathan Goodman
October 28, 2020

1 Introduction

A *collective mode* is a function of the state $u(x)$ that has a large range in the target distribution ρ but moves little in one MCMC step. It's not important to quantify this qualitative idea in detail, but you could take the ratio of the mean square change in one MCMC step divided by the static variance as a measure of slowness

$$S(u) = \frac{\mathrm{E}_\rho\Big[\big(u(X_1) - X_0\big)^2\Big]}{\mathrm{var}_\rho(u(X))} \ .$$

If u makes this $S(u)$ small, then it takes many MCMC steps to move u from one end of its range to the other. Such slow functions u are often *collective modes* in that they depend on all the components of x in some way. Exercise 3 of Week 7 illustrated that local potential wells can lead to slow modes. Many collective mode phenomena are more subtle than this.

This class will give some examples of high dimensional distributions where common MCMC methods have slow collective modes. These are just examples because whatever theory there is beyond these examples is both deep and model specific (not general). Nevertheless, the examples motivate some fancier "moves" (MCMC transformations) that reduce auto-correlation times by moving collective motes more.

The examples come from lattice models in statistical physics. These models are easy to describe and get some intuitions about. It is an active part of research in MCMC to adapt ideas and methods that work on these simple models to other problems in Bayesian statistics and other more modern MCMC applications.

2 Lattice models

A *lattice* is a periodic regular network of points in $\mathbb{R}^n$. Since we're mostly interested in intuition, we will stick to the simple *integer lattice*, which is all $j = (j_1, \cdots, j_n) \in \mathbb{Z}^n$. In words, this is points in $\mathbb{R}^n$ with all integer coordinates. Such a j is a *lattice site*. At every lattice site there is a variable or small collection of variables, which we will call the *field* variable or *lattice* variable, X_j. The collection of all lattice variables is the *lattice field*, X:

$$X = \{ \, X_j \ , \ \ j \in \text{ the lattice } \} \ .$$

We are interested in physical problems, so the dimension of the lattice usually will be $n = 1, 2, 3$. A four dimensional lattice (the fourth dimension representing

something related to time) is important in *lattice QCD*, which is how particle physicists compute predictions of fundamental theories of elementary particles.

For computing, we make the lattice finite with *lattice size* L. This means that we have separate lattice variables only for lattice sites with coordinates in the range from 0 to $L-1$, which is written $0 \le j_k < L$ for $k = 1, \cdots, n$. The set of all lattice sites in this box is our *finite size* lattice, which we call

$$\Lambda = \{0, \cdots, L-1\}^n \ .$$

We would say there are L lattice sites in each direction, so the total number of lattice sites is $|\Lambda| = L^n$. If the lattice variable X_j has r components, then the lattice field has d components with

$$d = r\,|\Lambda| = rL^n \ .$$

The lattice field is

$$X = \{\, X_j \mid j \in \Lambda \} \ .$$

The probability distribution is determined by the temperature and the lattice energy $\phi(x)$. The probability density is

$$\rho(x) = \frac{1}{Z(\beta)} e^{-\beta\phi(x)} \ .$$

Many questions in statistical physics involve understanding how $Z(\beta)$ and ρ depend on β.

One class of models is the *gaussian free field* model. This model is considered trivial in statistical physics because it is easily solvable and the solution doesn't do much by statistical physics standards. That said, it is a good place to start when thinking about lattice field models and MCMC methods for them. For this model, $r = 1$ and X_j is a single real number. The lattice energy may be written (explanations below)

$$\phi(x) = \frac{1}{2}\left[\sum_{\langle ij \rangle} (x_i - x_j)^2 + m \sum_{j \in \Lambda} x_j^2\right] \ . \tag{1}$$

The notation $\langle ij \rangle$ refers to *bonds* between *nearest neighbors* in Λ. Lattice sites i and j are nearest neighbors if they differ only in one coordinate and that coordinate differs only by ± 1. Each nearest neighbor pair is a *bond*. The first sum on the right is a sum over bonds and the second is a sum over sites. The constant m in the second sum is called "mass" because it represents mass in some quantum mechanical application.

The terminology is related to the atomic structure of simple crystals. We imagine that there is an atom at every lattice site and a chemical bond that connects the atom at each site to each of its nearest neighbors. Our integer lattice would be called a *cubic lattice*, because it consists of a regular array of cubes with lattice sites at the corners. The lattices of most common crystals are

not not cubic. For example, diamond is a crystal made up of carbon atoms. Each carbon atom (you may remember from elementary chemistry) makes bonds with four other carbon atoms. But a site in a cubic lattice has six nearest neighbors, two in each coordinate direction. The actual arrangement of carbon atoms in diamond is called *diamond cubic*, and can be found at Wikipedia. Each site in a diamond cubic lattice has four nearest neighbors.

A finite lattice Λ has *boundary sites*, which are sites that have at least one nearest neighbor that is not in Λ. *Boundary conditions* say what to do with the terms $\langle ij \rangle$ when $i \in \Lambda$ and $j \notin \Lambda$. Here are three commonly used boundary conditions.

- *Free* boundary conditions. The first sum in (1) includes terms $\langle ij \rangle$ with both $i \in \Lambda$ and $j \in \Lambda$. The lattice field X is "free" because it is not influenced by anything outside the lattice.

- *Dirichlet* boundary conditions. Take $X_i = 0$ for any $i \notin \Lambda$. This sets the field to zero at the boundary.

- *Periodic* boundary conditions. This makes the field X_j a periodic function of j with period L in each direction. Let m be one of the directions $1 \leq m \leq n$ and e_m the corresponding unit vector in the m direction. X is periodic if the value at $j + e_m$ is the same as the value at j for each m. You can implement periodic boundary conditions by getting the value X_i if $i \notin \Lambda$ by copying from the unique image of i that is in Λ.

The effects of boundaries and boundary conditions are called *finite size effects*. These describe the difference between a lattice and an ideal infinite lattice without boundaries. A real crystal is not infinite, but is likely have more sites (atoms) than will fit in a simulation. Many simulations choose periodic boundary conditions to reduce finite size effects, as the periodic field has no boundary. The Gaussian free field with energy (1) has finite variance with $m = 0$ only for Dirichlet boundary conditions. Otherwise, a constant field $x_j = c$ for all $j \in \Lambda$ has $\phi(x) = 0$. This gives the $x = c$ mode infinite variance.

The field energy (1) is a model of lattice deformations at finite temperature. The field variable X_j might represent the displacement of the atom at site j from it's resting place exactly at site j. If $m = 0$, then then the forces on atom j are only from the neighbors. If the $j + e_1$ atom is displaced to the right, this pulls the atom at j to the right also. The "mass" part of the energy, which is the second sum in (1) represents a force that makes an atom "want" to be at it's equilibrium position regardless of where its neighbors are. You might wonder where a force like that would come from. The model with $m = 0$ makes sense only with Dirichlet boundary conditions.

Another class of models is *spin models*, where X_j represents the orientation of the "spin" of atom j. In this context, "spin" refers to the fact that an atom can act like a small magnet, in part because of the electrons going around the nucleus making a small current. [Feel free to look up "spin" in Wikipedia for a more accurate version of this explanation.] In this model, a "spin" at site j is a

unit vector $|X_j| = 1$. Different spin models have spins in different dimensions. The *Ising* model has a one dimensional spin variable. A one dimensional variable with length 1 has two possible values $X_j = \pm 1$. The state $X_j = +1$ is called spin *up*, and $X_j = -1$ is spin *down*. The *XY model* has a two dimensional spin variable, which is the same as giving an angle: $X_j = (\cos(\theta_j), \sin(\theta_j))$. The *Heisenberg model* has a three dimensional spin variable. Normally, spin orientations would be random, so the net magnetic field from the crystal would be tiny. *Ferromagnetic* materials (iron is Fe(something) in Latin) have the surprising property that neighboring spins "like to" line up. [This is a quantum mechanical effect called *exchange energy*.] The energy for a spin field is given by (1), with $m = 0$. The mass is irrelevant because $|X_j|^2 = 1$ always.

Spin models models were created to model ordinary iron magnets. A piece of iron can hold a magnetic field after it is magnetized. If the metal is heated above the *Curie temperature* (about 700°C for iron), it looses its magnetism and stops being ferromagnetic. This is explained by the probabilities

$$\rho(x) = \frac{1}{Z} \exp\left[-\beta \sum_{\langle ij \rangle} |X_i - X_j|^2 \right]$$

A bond $\langle ij \rangle$ is *aligned* if $X_i = X_j$ (both up or both down) and misaligned if one is up and the other down. The energy of an aligned bond is zero and a misaligned bond is 2. Therefore, spin fields with more aligned bonds are more likely. How strong this effect is depends on the temperature parameter β. This leads to a dramatic prediction called *phase transition*, which is the mathematical explanation of the Curie temperature. There is a critical (inverse) temperature β_c so that if $\beta > \beta_c$ (low temperature) there is a net mean magnetization (defined below), while if $\beta > \beta_c$, there is none. This is a *phase transition* from the magnetized phase (low temperature) to the unmagnetized phase (high temperature). The transition happens as β goes from $\beta > \beta_c$ to $\beta < \beta_c$. The behavior near $\beta = \beta_c$ is *critical behavior*, which is observed in experiments and computations but hard to understand theoretically. Computational studies of critical behavior motivated much research on MCMC methods, including the "Multigrid Monte Carlo" method I was involved with.

Phase transitions have are about the behavior of observables that are bulk quantities. "Bulk" means an average of some local quantity over the lattice. The *net magnetization* is an example

$$M = \frac{1}{|\Lambda|} \sum_{j \in \Lambda} X_j \ .$$

The mean of this is zero, by symmetry ("flipping" every spin doesn't change probabilities). But the mean square is not zero

$$m^2(\beta, L) = \mathrm{E}_\beta\big[\, M^2 \big] \ .$$

This depends on the temperature and the lattice size. The *infinite volume limit* is the limit as the lattice becomes infinite. I hope it is plausible that the infinite

volume limit of square mean magnetization exists

$$m(\beta) = \lim_{L \to \infty} m^2(\beta, L) \; .$$

MCMC studies of spin models use large lattices because people are interested in the infinite volume limit. [Physical crystals in the lab are bigger than computational crystals, even with the best supercomputers.] The Ising model magnetic phase transition is that in more than one dimension ($n > 1$) $m^2(\beta) = 0$ for $\beta < \beta_c$ (above critical temperature) and $m^2(\beta) > 0$ if $\beta > \beta_c$ (low temperature).

The non-zero net mean magnetization is a consequence of long range correlations between spins. The distance k correlation (actually, covariance) is

$$C_k(L) = \operatorname{cov}(X_j, X_{j+k}) \; .$$

For any fixed offset k, the infinite volume limit exists:

$$C_k = \lim_{L \to \infty} C_k(L) \; .$$

At high temperature, we have only "short range" correlations, which means that $C_k \to 0$ as $|k| \to \infty$. It can be shown that the rate is exponential in the high temperature phase. There is a *correlation length* $\xi(\beta) > 0$ so that

$$C_k(\beta) \sim e^{-\frac{|k|}{\xi(\beta)}} \; .$$

The correlation length diverges to infinity as you approach the critical temperature, by a power law (it turns out)

$$\xi(\beta) \sim (\beta_c - \beta)^{-\nu} \quad \text{as } \beta \to \beta_c \, , \beta < \beta_c \; .$$

For $\beta > \beta_c$,

$$\lim_{|k| \to \infty} C_k(\beta) > 0 \; .$$

Distant spins remain positively correlated at any distance. This makes the fraction of up-spins different from the fraction of down-spins and leads to net mean square magnetization. The power ν is a *critical exponent*. Critical phenomena and critical exponents are studied by Monte Carlo. It takes large lattices to estimate ν accurately.

Here is a some historical irony. What we call the Ising model was invented by Dutch physicist Lenz for his student Ising to study. Ising showed that there is no phase transition in one dimension. It is "easy" to show that $C_k \to 0$ exponentially for $n = 1$ and any $\beta \geq 0$. Ising went on to give an incorrect argument that there is no phase transition in two or three dimensions either. This mistaken conclusion was accepted by the physics community until the German/English physicist Peierls proved was mistaken a decade later, but without a quantitative understanding of critical exponents. The Ising model in two dimensions was essentially solved by Norwegian physicist Onsager in what may be one of

the top ten most surprising and beautiful calculations of all time. In the mean time Heisenberg studied the more realistic Heisenberg model (three dimensional spin orientation) to escape Ising's conclusion. It is a deep theorem ("dynamic mass generation") that the Heisenberg model has exponentially decaying C_k in $n = 2$ dimensions for any β.

3 Critical slowing down

A starting MCMC algorithm for the Ising model would be that one MCMC step is a sweep through the sites in the lattice doing heat-bath re-sampling of each spin. If we are interested in the mean magnetization, we measure the functional that involves the spin variable at site j after k heat bath sweeps through the lattice:

$$M_k = \frac{1}{|\Lambda|} \sum_{j \in \Lambda} X_{j,k} \ .$$

The steady-state lag-t correlation function for this is

$$\rho_t(\beta, L) = \mathrm{corr}_{\rho,\beta}(M_t, M_0) \ .$$

This formula has multiple conflicts of notation: ρ on the left is the time correlation, but in $\mathrm{cor}_{\rho,\beta}$ ρ is the Gibbs Boltzmann distribution for inverse temperature β. The correlation $\mathrm{cor}_{\rho,\beta}$ refers to the correlation at inverse temperature β assuming that the lattice spin field X_0 is already in the target distribution. Here are the sad facts about this correlation function. For $\beta \leq \beta_c$, the limit exists:

$$\rho_t(\beta) = \lim_{L \to \infty} \rho_t(\beta, L) \ .$$

For $\beta < \beta_c$, this correlation decays exponentially

$$\rho_t(\beta) \sim \exp\left[-\frac{t}{\tau_{\exp}(\beta)} \right] \ .$$

The constant $\tau_{\exp}(\beta)$ is the *exponential autocorrelation time*. This is related to the integrated auto-correlation time (what we called correlation time) close to the critical temperature. *Critical slowing down* is the observed fact that $\tau_{\exp}(\beta) \to \infty$ as $\beta \to \beta_c$. This makes Monte Carlo study of critical phenomena very expensive if you use single site heat bath MCMC sampling.

There is an intuitive picture of critical slowing down that guides a search for better MCMC strategies. The picture is that when $\beta < \beta_c$ but close, there is a long correlation length $\xi(\beta)$. Sites closer together than ξ have correlated spins. If $X_j = +1$ (spin up), then if $|i - j| < \xi(\beta)$, then X_i is more likely to be up than down. This suggests that that sites around site j form a "magnetic domain" that is overall spin up on average. The lattice consists of up and down domains of roughly this size. The heat bath resampler is not good at moving or flipping magnetic domains. Suppose site j is in a spin up domain. Then when you resample X_j, its neighbors are more likely to be up than down. Therefore,

the new X_j is more likely to be up than down. Even if the new X_j is down, the spins in its domain are majority spin up. For this reason, it takes many sweeps for the magnetic domain structure to change significantly. To be clear, this picture is heuristic and not completely accurate. These domains are not precisely defined. Even though the spins in a spin up domain are majority up, there will be a positive fraction (but less than half) of down spins. Domains have irregular shapes and sizes.

Large but finite magnetic domains in a near critical Ising model are the *collective phenomenon* responsible for critical slowing down. A better sampler should have a way to flip large numbers of spins at once. In general, "collective mode" MCMC methods try to move collective modes. It turns out to be hard to create collective mode samplers that are correct.

4 Swendsen Wang algorithm

The *Swendsen Wang* algorithm is an MCMC sampler designed for the Ising model. The idea applies to other models, but the range of application is limited. Still, it is an inspiring example of that clever MCMC ideas can accomplish.

The Swendsen Wang algorithm creates *bond variables* $n_{\langle ij\rangle}$ for each nearest neighbor bond $\langle ij\rangle$. These are called *bond occupation* variables. The possible values are $n_{\langle ij\rangle} = 1$ (occupied), and $n_{\langle ij\rangle}$ (unoccupied). Occupied bonds are allowed only between sites with aligned spin. A bond between misaligned spins must be unoccupied:

$$X_i \neq X_j \implies n_{\langle ij\rangle} = 0 \,.$$

A bond between occupied spins may be occupied or not. The probability of being occupied is

$$p = 1 - e^{-2\beta} \,. \tag{2}$$

The algorithm uses connected components of the bond graph. The bond graph has lattice sites as vertices and occupied bonds as edges. Sites i and j are in the same connected component if there is a path from i to j that uses only occupied bonds. These connected components are called *connected clusters*. Every site in a connected cluster has the same spin, because a bond between sites of different spin cannot be occupied. There are relatively simple algorithms to identify the connected clusters, including the path compression algorithm whose analysis made Robert Tarjan famous (as in office hours or look it up). One Swendsen Wang MCMC step starts with a spin field X_k. It first generates a bond field n_k according to the rules above: misaligned (nearest neighbor pairs $\langle ij\rangle$ with $X_{i,k}X_{j,k} = -1$) bonds are unoccupied ($n_{\langle ij\rangle} = 0$) and aligned bonds ($X_{i,k}X_{j,k} = +1$) are occupied with probability (2). All occupation choices are independent. Next, the algorithm decomposes Λ into conected clusters for the bond field n_k with occupation variables $n_{\langle ij\rangle,k}$. Finally, the spins in each cluster are flipped ($+1 \leftrightarrow -1$) with probability $\frac{1}{2}$ with all flips independent. Near the critical point, with long correlation length, there will be large clusters for the algorithm to flip. This leads to faster decorrelation.

Computational experiments with this algorithm show that critical slowing down is a lot less, but is not eliminated. The integrated auto-correlation time $tau(\beta)$ still blows up like a power of $\beta_c - \beta$. But the power is so much lower that you have to do careful experiments to see that the critical exponent is not zero. Several people have claimed to have "multi-level" versions of the Swendsen Wang algorithm without critical slowing down. As far as I can see, and according to people more expert than me, none of these claims is true.

Why the algorithm is correct – see the beautiful paper of my former collaborators Robert Edwards and Alan Sokal, Physical Review Letter, September 1988.

The Swendsen Wang algorithm has been extended to apply to other problems, but its range of application is limited. To illustrate this, here is a problem it cannot be applied to, as far as I know. Here is a model of lattice phase separation that is close to the Ising model. Suppose there are two kinds of atoms A and B that may occupy the sites of a lattice. Suppose A atoms like other A atoms and B atoms like other B atoms more than they like A atoms. An energy function for this would be a sum over nearest neighbor bonds with a $+1$ for every $A - B$ bond and 0 for every $A - A$ or $B - B$ bond. This is the same as the Ising energy, with A for $X_j = -1$ and B for $X_j = +1$. The difference is that atoms can move but they cannot change type (flip). In the language of the Ising model, we restrict to configurations with a specified total magnetization. There is a local MCMC algorithm for this, called *Kawasaki dynamics*. You sweep through the bonds and propose moves that exchange the spins (A or B) on a bond. This "works" but has bad auto-correlation times. There's no method for this problem that is significantly better.

5 Exercises

A few to be added.