

Simulation of Statistical Mechanical Systems using Stan

Forrest Eli Hurley

July 4, 2019

Abstract

Bayesian statistics is closely coupled with physics. The metropolis algorithm (1953) was developed by scientists working at Los Alamos as a method for thermodynamic simulation of molecular dynamics. Not until the work of W. K. Hastings (1970) was the method generalized to arbitrary probability distributions. Hamiltonian Monte Carlo is even more deeply rooted in physics than the Metropolis-Hastings algorithm. The simulation of states with velocities, energies, and a Hamiltonian describes nothing other than a physical system. It matches a canonical ensemble in that there is not a fixed energy between steps, only an overall fixed temperature. The temperature is usually only implicit, but some tempering methods simulate chains at higher temperatures to smooth the probability distributions. The Ising Model, a proxy for magnetization, is a prevalent introductory model in the study of statistical mechanics. It consists of an N-dimensional grid of spin up or down particles. The energy varies depending on the alignment of spins between nearest neighbors. At low temperatures spins tend to align on a macroscopic scale; at high temperatures they become evenly distributed. We simulate the XY Model, similar to the Ising Model but allowing spins to be represented by unit vectors in two dimensions, using Stan. We create chains at several temperatures to identify the locations of phase transitions in macroscopic properties. Our work shows the applicability of Stan for computation in continuous statistical mechanical problems.

Introduction

This section is a short introduction to statistical mechanics summarizing [source].

The first core assumption of statistical mechanics is that the system of interest exists in one of a discrete set of microstates. The number of microstates can be countably infinite. In fact, the number of microstates is often assumed to be dense enough to be arbitrarily close to any given state. A collection of microstates, often with some defining property, is a macrostate. For example, the collection of microstates with energy 2 is a macrostate. Let $\Omega(E)$ be the size of the macrostate (also called the density of states) with energy E .

In an example system consisting of a single particle on a plane, there are 4 coordinates: 2 positional coordinates and 2 momentum coordinates. The position does not affect energy, which is given by $E = |p|^2/2m$. Every microstate with energy E must have $|p| \propto E^{0.5}$; the momentum coordinates must lie in a circle. The size of the macrostate is proportional to the radius of the circle or $\Omega \propto E^{0.5}$. In higher dimensional space, either corresponding to having more spatial dimensions or to having more particles, this proportionality takes the form $\Omega \propto E^{(D-1)/2}$. Due to the exponential nature of this inequality, the logarithm of the density of states is often used.

Canonical Distribution

Much of the complexity and usefulness of statistical mechanics appears when systems are not entirely isolated. If we take an isolated system with some energy E_T and split it into two systems then $E_T = E_a + E_b + E_{ab}$. If the two portions are sufficiently independent (often by having B be significantly larger than A) then this can be approximated $E_T = E_a + E_b$. While the total energy E_T is conserved, E_a can vary to any value! This collection of two systems which share energy and only energy is called a canonical ensemble. For reasons which can be well justified, but are outside the scope of this paper, we approximate the density of states of B as $\log \Omega_B = \alpha + \beta E_b$. If the density of states of B is independent of A then the density of states $\Omega_a(E_a) = \Omega_b(E_b)$. Combining these various

assumptions together, we find $p(E_a) \propto \omega_A(E_a) \propto e^{-\beta E_a}$. Beta is defined to be $1/kT$. For our purposes, we will use units where $k=1$, so $\beta=1/T$.

In a one dimensional single particle space, $E = p^2/2m$. By substituting into the previous equation we get $\text{prob}(p) \propto e^{-\beta p^2/2m}$. It is simply an unnormalized normal distribution with mean 0 and variance _____. Generally speaking, energy divided by temperature is the negative log likelihood to within a constant. Later we describe how we utilized this correspondence to sample from a statistical mechanical model from within the Bayesian sampling oriented framework of Stan.

XY Model

One particularly simple statistical mechanical model with profound implications on reality is the Ising Model. It consists of an N-dimensional lattice of elements which either have spin up or spin down. The total energy of the system is the sum of interactions between neighboring elements. $\sum -S_i S_j$. If the spins are in the same direction then the energy is lower. In 2 dimensions the Ising Model has two distinct phases. In the ordered phase almost all of the spins are in the same direction and the average spin is not near zero. At higher temperatures, in the unordered phase, the spins are random, and the average spin is almost exactly zero. The transition between the two phases happens suddenly as temperature increases.

There are a variety of interesting and efficient ways of sampling from the distribution of states of the Ising Model, however due to it only being defined by discrete variables Stan is not one of them. For this work we instead focus on the XY Model. In the XY model the spins are not up or down. Instead the spins are unit vectors in the XY plane. The energy due to the interaction of spins between two neighbors is $-S_i \cdot S_j$. While the XY model does not have a phase transition of the same sort as the Ising model, it does have similar behaviours in more complex variables. The spins orient into larger scale cyclic structures called vortices. At higher temperatures there is suddenly a drastic increase in the density of vortices.

The XY Model is similar to the Ising Model, but with an extra degree of freedom for the spins. The energy of any given state is related to the dot products of the spins of nearest neighbors on the lattice.

$$E = -J \sum_{\langle i,j \rangle} \hat{\sigma}_i \cdot \hat{\sigma}_j$$

Each spin is a unit vector in the XY plane rather than the binary up or down seen in the Ising Model. Unsurprisingly, this change leads to a drastic shift in the dynamics. The Ising Model has a second order phase transition with changing temperature. It also has a first order phase transition with changing external magnetic field. However, the XY Model on a 2 dimensional lattice has no first or second order phase transitions. It does however exhibit a Kosterlitz-Thouless transition. The KT transition relates to the appearance of vortices on the lattice (Tobochnik and Chester 1979). A vortex is a swirling pattern centered on a lattice tile. At the critical temperature T_{KT} vortex pairs unbind and move apart. Previous work has shown that for the XY Model on a 2D lattice, $T_{TK} \approx 0.893$ (Jensen, n.d.). When we refer to the XY Model in the rest of this work, we will mean the 2D case.

```
functions {
  real xy_energy(vector[,] spins) {

    real energy = 0;
    int spin_dim[4] = dims(spins);

    for (dir in 1:2) {

      //calculate the offset for nearest neighbor calculations
      int dir_x = dir % 2 + spin_dim[1];
      int dir_y = dir / 2 + spin_dim[2];
```

```

    for (x in 1:spin_dim[1]) {
      for (y in 1:spin_dim[2]) {
        //calculate the x,y index of the nearest neighbor
        int x_idx = ( (x + dir_x - 1) % spin_dim[1] ) + 1;
        int y_idx = ( (y + dir_y - 1) % spin_dim[2] ) + 1;

        energy += dot_product( spins[x_idx, y_idx], spins[x, y] );
      }
    }
  }

  return energy;
}

//given the temperature and energy, calculate the likelihood
real canonical_ensemble_lpdf(real temperature, real energy) {
  return energy / temperature;
}
}

```

Methods

Parameters and Iterations

For most experiments we ran 5000 NUTS iterations each on 4 parallel chains with a goal rejection rate of 0.7 and a maximum tree depth of 15. The effective sample sizes were always at least a tenth of the number of iterations except on two of the tests around a temperature of 1 on the 24x24 and 32x32 experiments with $j = -1$. The aforementioned 24x24 experiment was run with 1000 NUTS iterations. All calculations were performed on an Intel Core i7 8th generation processor. The total computation time was on the order of 5 days.

We ran experiments on 4x4, 8x8, 16x16, 24x24, and 32x32 lattices. In plots N refers to the side length of the lattice: a $N \times N$ lattice. Energy refers to the energy per lattice site.

Working with Stan

Stan is a Bayesian statistics package for sampling and summarizing distributions with minimal parameter tuning. It primarily uses the NUTS algorithm with an adaptation period during warm up for setting the simulation step size. All calculations are automatically differentiated using an operation tree for the Hamiltonian simulation. Due to the continuous nature of HMC and NUTS, Stan does not support discrete parameters (Team 2018). While small numbers of discrete parameters can be marginalized out manually, this procedure is complex for the large number of parameters in which we are interested.

Stan uses prior probabilities and likelihood functions in its simulation of the posterior. The default prior is the uninformative improper uniform distribution on the interval of possible values. We must still specify a likelihood function which captures the distribution we wish to sample. Due to floating point precision issues, Stan uses log likelihood; this prevents near 0 probabilities from disappearing (Stan Development Team 2018). In a canonical ensemble the log probability of any state is $-\beta E + c$ where E is the energy of the state and c is an arbitrary constant. We define the likelihood as $p(\beta|E) \sim e^{-\beta E}$ and sample states with various E .

Currently Stan only supports sampling from continuous parameters. It does support constraining of those parameters to specific ranges. A value known to be a probability could be limited between 0 and 1. More relevant for the XY Model, vectors can be constrained to have $\|\vec{x}\|_2 = 1$ (Stan Development Team 2018).

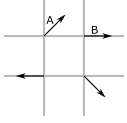


Figure 1: An example configuration of spins on a lattice square. Because the shortest angles between each spin sum to -2π , it is an anti-vortex.

Our sampling parameters are a multidimensional array of unit vectors. Every iteration we calculate the energy based on nearest neighbor interactions and update the likelihood accordingly.

Potential Scale Reduction Statistic

The potential scale reduction statistic is a measure of convergence. It requires multiple independent chains. It compares the variance of the individual chains to the pooled variance to determine if they are at equilibrium in the same distribution. It is also used for checking for stationarity of a single chain by splitting the chain in half and comparing the beginning and the end. If W is the average within chain variance and $\widehat{\text{var}}^+(\theta|y)$ is the best estimate of overall variance given the within chain and between chain variances,

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\theta|y)}{W}}$$

\hat{R} should be between 0.9 and 1.1 if chains have converged properly. In all of our experiments, the final \hat{R} for the energy was between 1.05 and 1.00.

Effective Sample Size

Simulating 4 independent chains also allows us to calculate the effective sample size of our simulation. As the samples from the chains are correlated, effective sample size is the equivalent number of independent samples for estimates of posterior parameters. If ρ_t is the autocorrelation of the Markov chain with lag t then the effective sample size is

$$N_{eff} = \frac{N}{\sum_{t=-\infty}^{\infty} \rho_t} = \frac{N}{1 + 2 \sum_{t=1}^{\infty} \rho_t}$$

Using the same chain for calculating mean and variance (required for autocorrelation) as for calculating the autocorrelations leads to a biased estimate.

Vorticity

The KT transition arises from an appearance of large vortices. We identify vortices by looking for lattice cells where the corner spin vectors tend to point clockwise or counterclockwise. Specifically we calculate the shortest difference in angles in the counterclockwise direction and sum. In Figure 1, the angle between vectors A and B would be -45 degrees. This sum can either be 2π , 0, or -2π (technically it can also be 4π and -4π but those cases are exceedingly unlikely). On a lattice with periodic boundary conditions, there will be the same number of vortices (2π) and anti-vortices (-2π) (Imriška 2009). We generally calculate the density of vortices on 100 random samples from spin configurations at each temperature.

```
data {
  real<lower=0> temp;
  int<lower=1> dim_x;
```

```

    int<lower=1> dim_y;
}
parameters {
    unit_vector[2] spin[dim_x, dim_y];
}

transformed parameters{
    real energy = xy_energy(spin);
    real energy_per_spin = energy / dim_x / dim_y;
}
model {
    temp ~ canonical_ensemble(energy);
}

```

```

#calculate vorticity given a single set of spins
def vorticity_calc(spin_array):
    rolled = [spin_array,
              np.roll(spin_array,1,0),
              np.roll(spin_array,1,(0,1)),
              np.roll(spin_array,1,1)]

    #calculate the dot product for the shortest angle
    dotted = []
    for i in range(len(rolled)):
        dotted.append(np.sum(np.multiply(rolled[i-1],rolled[i]),axis=-1))

    #calculate the cross product for the direction of the shortest angle
    crossed = []
    for i in range(len(rolled)):
        crossed.append(np.cross(rolled[i-1],rolled[i],axis=-1))

    angles = [np.arccos(dot) * np.sign(cross) for dot, cross in zip(dotted,crossed)]

    #sum the angles around each lattice square
    around_angle = np.sum(angles,axis=0)

    #find vortices
    vortex = np.where(around_angle>1,1,0)

    #count the number of vortices
    return np.sum(vortex,axis=(-1,-2))

```

Results

```

model = stan_model('xy_model.stan')

```

```

## DIAGNOSTIC(S) FROM PARSER:
## Info: integer division implicitly rounds to integer. Found int division: dir / 2
## Positive values rounded down, negative values rounded up or down in platform-dependent way.

## recompiling to avoid crashing R session

```

```

#basic parameters
chains <- 4
vorticity_samples <- 1000
iterations <- 10000
dim_x <- 4
dim_y <- 4
temp <- 2.

#initialize the data dictionary
data <- list(dim_x=dim_x, dim_y=dim_y, temp=temp)

#simulate the distribution
fit <- sampling(model,
  data=data,chains=chains,
  iter=iterations,
  verbose=FALSE,
  control = list(max_treedepth=15))

```

Warning: There were 12 divergent transitions after warmup. Increasing adapt_delta above 0.8 may help
 ## <http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup>

Warning: Examine the pairs() plot to diagnose sampling problems

```
print(summary(fit,pars="energy_per_spin")$summary)
```

```
##               mean      se_mean      sd      2.5%      25%
## energy_per_spin 0.606715 0.003873833 0.3250774 0.03313525 0.3786287
##               50%      75%      97.5%      n_eff      Rhat
## energy_per_spin 0.57943 0.8181796 1.304858 7041.928 1.000068
```

```

energy = mean(extract(fit,pars='energy_per_spin')$energy_per_spin)
energy_var = var(extract(fit,pars='energy_per_spin')$energy_per_spin)
specific_heat = var(extract(fit,pars='energy')$energy) / dim_x / dim_y / temp / temp

spin_data <- extract(fit,pars="spin")$spin

```

```

crossvec <- function(x,y){
  cv <- x[1]*y[2]-x[2]*y[1]
  return(invisible(cv))
}

```

```

vorticity_calc <- function(spin_array) {
  spin_dim = dim(spin_array)[1:2]
  angles = array(0,c(4,spin_dim))
  for (dir in 1:4) {
    dir_x <- dir %% 2 + spin_dim[1]
    dir_y <- dir %/% 2 + spin_dim[2]
    for (x in spin_dim[1]) {
      for (y in spin_dim[2]) {

        x_idx <- ( (x + dir_x - 1) %% spin_dim[1] ) + 1;

```

```

        y_idx <- ( (y + dir_y - 1) %% spin_dim[2] ) + 1;

        dotted <- crossprod(
            spin_array[x,y,1:2], spin_array[x_idx,y_idx,1:2])
        crossed <- crossvec(
            spin_array[x,y,1:2], spin_array[x_idx,y_idx,1:2])
        angles[dir,x,y] <- acos(dotted) * sign(crossed)
    }
}
}

around_angles <- apply(angles,c(2,3),sum)
vortex <- sum(around_angles > 1)
return(vortex)
}

apply(spin_data[1:2,,,],1,vorticity_calc)

## [1] 1 1

#Calculate and print various statistics related to the distribution
vorticity = np.mean([vorticity_calc(x) for x in
    r.spin_data[:int(r.vorticity_samples)]])
vortex_density = vorticity / float(r.dim_x * r.dim_y)
print("Energy:",r.energy,"Specific heat:",r.specific_heat,"Vortex Density:",vortex_density)

## Energy: 0.6067150450396009 Specific heat: 0.4227011521006108 Vortex Density: 0.087375

```

KT Phase Transition

We found that the energy vs temperature curves had converged by $N=8$. Figure 2 indicates that the energy of the system increases fairly linearly with temperature before the KT transition. Figure 3 shows the specific heat of the various lattice sizes and temperatures. It starts around $c = 0.5$ at a temperature of 0. There is no extreme discontinuity of the specific heat for any lattice size up to $N=24$. The peak specific heat occurs around a temperature of 1.1. While vortices start appearing in large numbers around a temperature of 0.9, Figure 4 does not show any particularly noticeable characteristics around the peak specific heat. The initial rate of vortex pair splitting and vortex density increase does not differ between any of the lattice sizes. The density of vortices is perfectly linearly related to the density of energy in the system. As seen in Figure 5, once the energy is greater than about -1.5, the vortex density for all of the lattice sizes lie on a line with slope approximately 0.1.

$J=-1$

Many of our early experiments were run with $J = -1$. This may not have been intentional. However, we found that the energy curves of the XY Model with $J = 1$ are identical to those with $J = -1$. Figure 6 shows the difference between the mean energy for chains on the same lattice at the same temperature for the two different values of J . As temperature increases the differences also increase, but the mean difference remains close to zero for all lattice sizes. The lowest energy states for $J = -1$ are the highest energy states for $J = 1$. Flipping half of the spins in a checker pattern by 180 degrees transforms directly between low energy states for the two coupling coefficients.

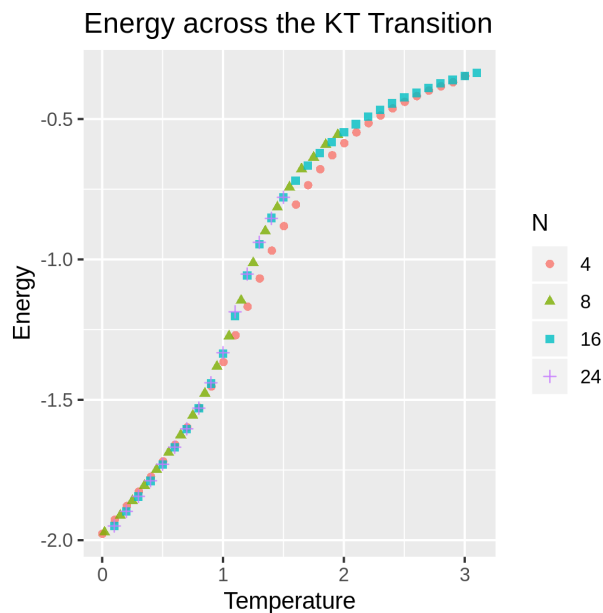


Figure 2: Mean energy of a canonical ensemble of states of the XY Model at several temperatures

Conclusions

The XY Model has a wealth of interesting dynamics. The energy of the system after the KT transition is directly proportional to the number of vortices on average. An interaction strength (J) of negative 1, while it does lead to different patterns of vortex pairing, does not have a significantly different energy at any temperature. The lowest energy states with $J = -1$ are the highest energy states of $J = 1$; there is a strong symmetry between the two extremes of energy. Stan provided fairly sophisticated sampling methods with minimal coding effort. The built in diagnostics helped to validate the resulting chains and gauge the accuracy of their estimates. While its lack of support for discrete parameters limits the models for which it is useful, there are still a wealth of contexts in which it is a good first choice for simulating models.

References

- Imriška, Jakub. 2009. “Thesis "Phase diagram of the modified XY model".” Comenius University in Bratislava. <https://doi.org/10.1088/0022-3719/19/23/007>.
- Jensen, Henrik Jeldtoft. n.d. “The Kosterlitz-Thouless Transition.” Boston: MIT. <http://www.mit.edu/~levitov/8.334/notes>
- Stan Development Team. 2018. *Stan User’s Guide*. <https://mc-stan.org/docs/2.19/stan-users-guide/index.html>.
- Team, Stan Development. 2018. “Stan Reference Manual” 18 (July). <https://mc-stan.org/docs/2.18/reference-manual/convergence.html>.
- Tobochnik, Jan, and G. V. Chester. 1979. “Monte Carlo study of the planar spin model.” *Physical Review B* 20 (9): 3761–9. <https://doi.org/10.1103/PhysRevB.20.3761>.

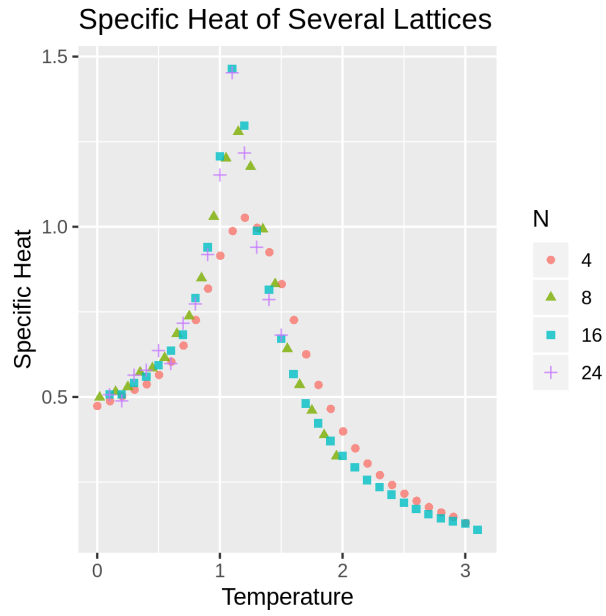


Figure 3: The specific heat of the XY models across temperatures and lattice sizes. The 24x24 lattice only has data points up to a temperature of 1.5.

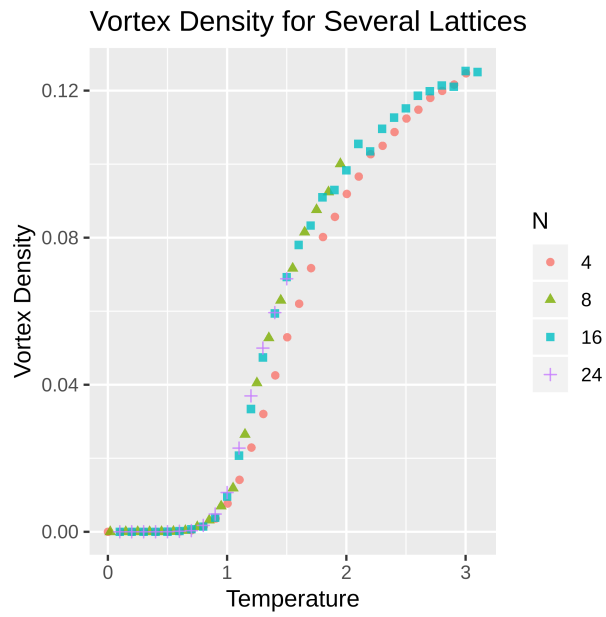


Figure 4: The density of vortices at various temperatures. Below 0.8 it is almost exactly 0.

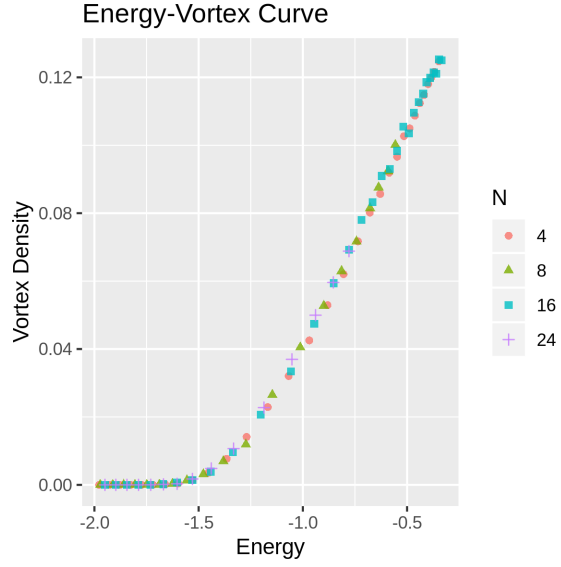


Figure 5: The relationship between energy per spin and vortices per spin for several lattice sizes.

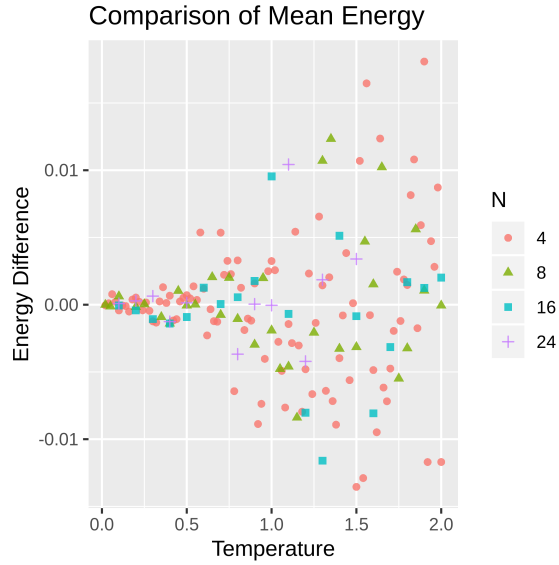


Figure 6: The difference in mean energy between chains generated with $J=1$ and $J=-1$ at several temperatures and lattice sizes.