1 Statistical physics foundations

In this lecture we will cover the basic objects and questions of statistical physics and introduce much of the terminology we'll use going forward.

Statistical physics is interested in many different types of physical phenomena and the methods of statistical physics have been applied to many problems outside of physics. In these notes we will be primarily concerned with one of the oldest and most fundamental questions in the field: how do the microscopic interactions between molecules or particles influence the macroscopic behavior of a gas, fluid, magnetic material, or other system. Of particular interest is the phase transition phenomenon, in which small changes in the microscopic interactions lead to qualitative changes in the macroscopic behavior. Examples of a phase transition in nature include a liquid freezing to a solid, a liquid vaporizing to a gas, or a piece of iron exhibiting spontaneous magnetization.

We will see that similar phase transition phenomena occur in the very different setting of combinatorial problems and that tools and intuition from statistical physics can help us understand these problems.

1.1 Gibbs measures and partition functions

While statistical physics encompasses many different types of models, for now we will focus on *spin models on graphs*. This setting is not only rich enough to exhibit many interesting phase transition phenomena, but it is also most relevant to combinatorial applications. We first define a general model then give some important examples.

Fix a finite set of spins Ω. Typical choices include

- $\Omega = \{0, 1\}$
- $\Omega = \{-1, 1\}$
- $\Omega = \{\text{Red, Blue, Green}\}$
- $\Omega = \{1, \dots, q\}$.

For a graph $G = (V, E)$, the set of possible *spin configurations* is Ω^V, all assignments of spins from Ω to the vertices of G.

Next define an *energy function* (or Hamiltonian) from $\Omega^V \to \mathbb{R} \cup \{+\infty\}$ that respects the graph structure (as a sum of functions vertex spins and pairs of spins across edges):

$$H(\sigma) = \sum_{v \in V} f(\sigma_v) + \sum_{(u,v) \in E} g(\sigma_u, \sigma_v)$$

where $f : \Omega \to \mathbb{R}$ and $g : \Omega \times \Omega \to \mathbb{R} \cup \{+\infty\}$ is symmetric. If g takes the value $+\infty$ we say that there is a *hard constraint* in the model.

The *partition function* of the spin model at inverse temperature $\beta \in \mathbb{R}$ is

$$Z_G(\beta) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} . \tag{1}$$

The *Gibbs measure* is the probability distribution on Ω^V defined by

$$\mu_{G,\beta}(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z_G(\beta)}\,, \tag{2}$$

where the partition function plays the role of the normalizing constant ensuring $\mu_{G,\beta}$ is a probability distribution.

Example 1. *The ferromagnetic Ising model with no external field. Let $\Omega = \{+1, -1\}$, $f(\sigma_v) \equiv 0$, $g(\sigma_u, \sigma_v) = -\sigma_u\sigma_v$. Then $\mu_{G,\beta}(\sigma)$ is proportional to $e^{2\beta M(G,\sigma)}$ where $M(G,\sigma)$ is the number of edges of G whose endpoints receive the same spin under σ (monochromatic edges). If $\beta > 0$ then configurations with more monochromatic edges are preferred; this case is the ferromagnetic case.*

To obtain the antiferromagnetic Ising model we take $g(\sigma_u, \sigma_v) = \sigma_u\sigma_v$; then we prefer edges with different spins on their endpoints.

We can add a (uniform) external field by taking $f(\sigma_v) = a\sigma_v$. If $a < 0$ we prefer $+1$ spins.

Consider the ferromagnetic Ising model (with no external field) on $(\mathbb{Z}/n\mathbb{Z})^d$, the d-dimensional discrete torus. If β is large, then we more strongly prefer vertices to have the same spin as their neighbors.

Example 2. *Hard-core model (hard-core lattice gas). $\Omega = \{0, 1\}$, $f(\sigma_v) = -\sigma_v \cdot \log\lambda$, $g(1,1) = +\infty$, $g(0,0) = g(0,1) = g(1,0) = 0$. $\beta = 1$.*

More conveniently, we can associate an independent set I of G (set of vertices that spans no edges) with its indicator vector σ so that $\sigma_v = \mathbf{1}_{v\in I}$. Then

$$\mu_G(I) = \frac{\lambda^{|I|}}{Z_G(\lambda)} \tag{3}$$

where

$$Z_G(\lambda) = \sum_{I\in\mathcal{I}(G)} \lambda^{|I|} = \sum_{k\geq 0} \lambda^k i_k(G)\,. \tag{4}$$

Here $\mathcal{I}(G)$ is the set of all independent sets of G and $i_k(G)$ is the number of independent sets of size k in G. The parameter λ is the fugacity or activity and governs whether we prefer small or large independent sets.

The partition function of the hard-core model, $Z_G(\lambda)$, is also known as the independence polynomial in combinatorics. $Z_G(\lambda)$ encodes a lot of combinatorial information. $Z_G(1) = |\mathcal{I}(G)|$, the number of independent sets of G; the highest order term in $Z_G(\lambda)$ is the independence number of G, $\alpha(G)$, and its coefficient is the number of maximum size independent sets of G.

Take for example, $G = C_4$, the cycle on 4 vertices. Then $Z_G(\lambda) = 1 + 4\lambda + 2\lambda^2$. For any $v \in V(G)$, the probability of choosing $I = \{v\}$ in the hard-core model on G is $\frac{\lambda}{1+4\lambda+2\lambda^2}$.

Example 3. *The Potts model. The Potts model is a generalization of the Ising model to $q \geq 2$ spins (or colors); that is, $\Omega = \{1, \ldots, q\}$. Configurations are assignments of q colors to the vertices of a graph. A configuration is chosen with probability $\frac{e^{\beta M(G,\sigma)}}{Z_G(q,\beta)}$ where $M(G,\sigma)$ is the number of monochromatic edges of G under the coloring σ. $\beta \geq 0$ is the ferromagnetic and $\beta \leq 0$ the antiferromagnetic case.*

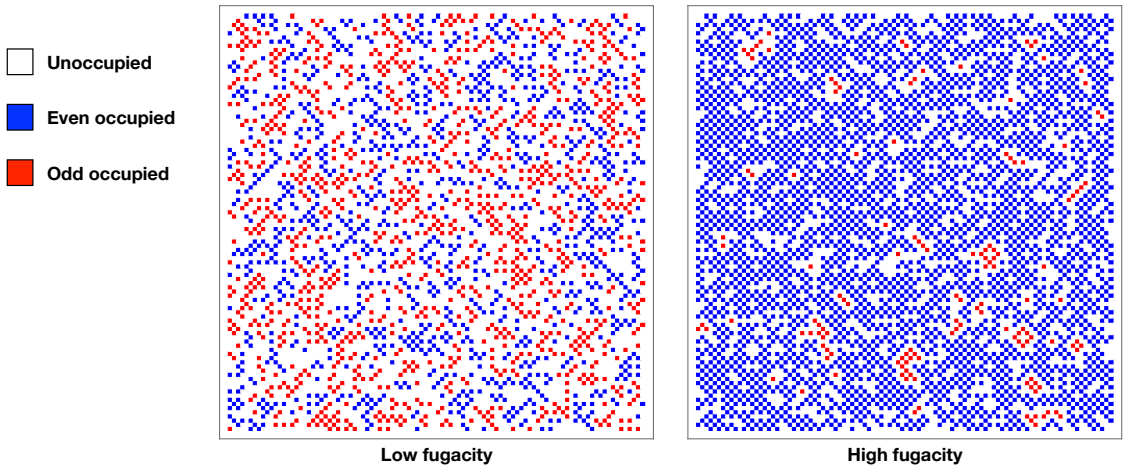

Figure 1: Two instances of the hard-core model on $\mathbb{Z}^2$

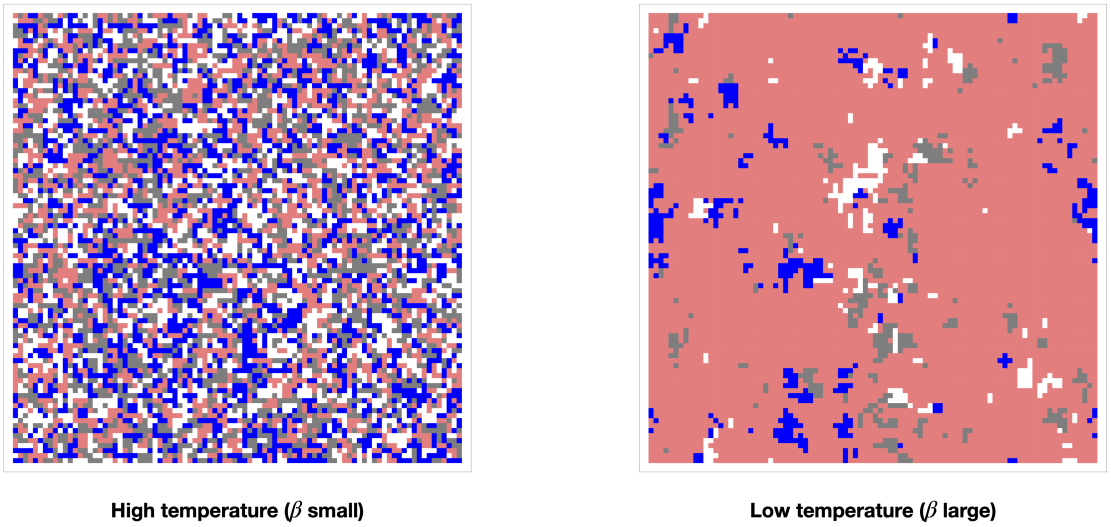

Figure 2: Two instances of the 4-color ferromagnetic Potts model on $\mathbb{Z}^2$

In general, the inverse temperature β controls the strength of the interaction in the model. At high temperatures (small β), local interactions are weak, while at low temperatures (large β), local interactions are strong. In other words,

- At $\beta = 0$ (infinite temperature) the Gibbs measure is simply uniform on Ω^V and so each vertex receives a uniform and independent spin from Ω.
- At $\beta = +\infty$ (zero temperature), the Gibbs measure is uniform over the *ground states* of the model: the configurations σ that minimize the energy $H(\cdot)$. For Gibbs measures on lattices like $\mathbb{Z}^d$, it is often[1] easy to understand the ground states (e.g. all even/all odd occupied configurations for hard-core; monochromatic configurations for Ising/Potts).

[1]But not always! See e.g. the Edwards-Anderson model.

For general graphs though, this need not be the case. In particular, finding a ground state in the hard-core model on a general graph is the max independent set problem, a classic NP-hard problem. Similarly, finding and understanding the ground states of anti-ferromagnetic models on random graphs is a challenging problem, both mathematically and algorithmically.

- Taking β positive and finite interpolates between independence (pure entropy) and optimization (pure energy). Understanding the Gibbs measure and partition function at positive temperature requires balancing energy and entropy.

From the combinatorics perspective, the Gibbs measure interpolates between two objects we study a lot: a purely random object (say a uniformly random cut in a graph) and an extremal object (the max cut or min cut in a graph).

An important theme in statistical physics is that the qualitative properties of the two ends of the interpolation persist at positive, finite temperatures: a weakly interacting system has many of the properties of an independent system, while a strongly interacting system correlates strongly with an extremal object. The switch from one qualitative regime to the other is a phase transition, one of the main topics of statistical physics.

Not all Gibbs measures are spin models on graphs. Some other important examples include the following, some of which also have particular importance in combinatorics.

1. The monomer-dimer model. Allowed configurations are matchings in G, with $\mu_{G,\lambda}(M) = \frac{\lambda^{|M|}}{Z_G^{\text{match}}(\lambda)}$. 'Dimers' are edges in the matching while 'monomers' are unmatched vertices. The monomer-dimer model is the hard-core model on the line graph of G. This is an example of an *edge coloring model* (see e.g. [9]).

2. Spin models on hypergraphs. Here the energy function H is a sum of functions on vertices and functions on hyperedges. For example, we can consider the hard-core model on a hypergraph $G = (V, E)$. Configurations are subsets S of vertices that contain no hyperedge, weighted by $\lambda^{|S|}$. An interaction on a hyperedge of size > 2 is called a *multibody interaction*.

3. Gibbs point processes (continuum models). The hard sphere model [5] is a continuum model of a gas and perhaps the original model in statistical mechanics. This is a probability distribution over packings of equal-sized spheres in Euclidean space.

Gibbs measures arise in many other contexts beyond statistical physics, including machine learning, Bayesian statistics, mathematical biology, and many others. They are sometimes called probabilistic graphical models, Markov random fields, log linear models, exponential families, or Boltzmann distributions.

1.2 Markov random fields

A Gibbs measure defined as a spin model on a graph (with interactions across edges) is a *Markov random field* with respect to the graph. This means that it satisfies the following spatial Markov property. Let A, B, S be disjoint subsets of vertices of G so that S separates

A and B: any path from a vertex $a \in A$ to a vertex $b \in B$ must pass through S. Then with respect to the Gibbs measure μ_G, if we condition on the spins in S, $\sigma_S = \tau_S$, the spins σ_A in A and σ_B in B are independent.

A special case of the spatial Markov property is that the spin at v, σ_v, is independent of the other spins in the graph conditioned on the spins of its neighbors $\sigma_{N(v)}$. We can write down a formula for the distribution of σ_v given that $\sigma_{N(v)} = \tau_{N(v)}$:

$$\mu_G(\sigma_v = \omega | \sigma_{N(v)} = \tau_{N(v)}) = \frac{\exp\left[-\beta\left(f(\omega) + \sum_{u \in N(v)} g(\omega, \tau_u)\right)\right]}{\sum_{\omega' \in \Omega} \exp\left[-\beta\left(f(\omega') + \sum_{u \in N(v)} g(\omega', \tau_u)\right)\right]}. \tag{5}$$

For the hard-core model, this formula simplifies considerably. Say v is blocked with respect to an independent set I if $N(v) \cap I \neq \emptyset$ and unblocked otherwise. In particular, v can only be in I if it is unblocked. Then we have $\mu_{G,\lambda}(v \in I | v \text{ unblocked}) = \frac{\lambda}{1+\lambda}$.

This spatial Markov property makes it easy to compute the partition function of a spin model on a tree.

1.3 Moments, cumulants, and derivatives of the log partition function

The energy $H(\cdot)$ is a *local function*, with respect to the geometry of the underlying graph: it is a sum of functions on vertices and edges. As a random variable, $H(\sigma)$ is a locally computable statistic or *observable* of the model (it can be computed from σ by summing over vertices and edges). For instance in the hard-core model $H(\sigma)$ counts the size of an independent set while in the Ising and Potts models $H(\sigma)$ counts the number of monochromatic edges (or equivalently the number of crossing edges of a cut).

As we will see, understanding the behavior of the random variable $H(\sigma)$ for large underlying graphs can tell us a lot about the behavior of the model and any phase transitions that might occur as parameters are varied.

To begin to understand the random variable $H(\sigma)$ we'd like to know its expectation, variance, and then perhaps higher moments.

We can write down the expectation:

$$\begin{aligned}
\mathbb{E}[H] &= \sum_{\sigma \in \Omega^V} H(\sigma) \mu_G(\sigma) \\
&= \sum_{\sigma \in \Omega^V} H(\sigma) \frac{e^{-\beta H(\sigma)}}{Z_G(\beta)} \\
&= \frac{\sum_{\sigma \in \Omega^V} H(\sigma) e^{-\beta H(\sigma)}}{Z_G(\beta)} \\
&= \frac{-\frac{\partial}{\partial \beta} Z_G(\beta)}{Z_G(\beta)} \\
&= -\frac{\partial}{\partial \beta} \log Z_G(\beta). \qquad (6)
\end{aligned}$$

In the case of the hard-core model (due to the slightly different form of the distribution), we have

$$\begin{aligned}
\mathbb{E}_{G,\lambda}|\mathbf{I}| &= \sum_{I\in\mathcal{I}(G)} |I|\mu_{G,\lambda}(I) \\
&= \sum_{I\in\mathcal{I}(G)} |I|\frac{\lambda^{|I|}}{Z_G(\lambda)} \\
&= \frac{1}{Z_G(\lambda)} \sum_{I\in\mathcal{I}(G)} |I|\lambda^{|I|} \\
&= \frac{\lambda\frac{\partial}{\partial\lambda}Z_G(\lambda)}{Z_G(\lambda)} \\
&= \lambda\frac{\partial}{\partial\lambda}\log Z_G(\lambda)\,. \qquad (7)
\end{aligned}$$

This is an important equation that we will use throughout this lecture series. We now see the same fact in more generality.

For a random variable X, its moment generating function is $M_X(t) = \mathbb{E}e^{tX}$. Its cumulant generating function is its logarithm of the moment generating function $K_X(t) = \log \mathbb{E}e^{tX}$. The *cumulants* of X are the coefficients in the Taylor series around 0:

$$K_X(t) = \sum_{n=1}^{\infty} \kappa_n(X)\frac{t^n}{n!}\,. \qquad (8)$$

Or in other words, $\kappa_n(X) = K_X^{(n)}(0)$.

Cumulants are related to moments but are often more convenient to work with in statistical physics. For example, the cumulants of a Gaussian $N(\mu,\sigma^2)$ are $\kappa_1 = \mu, \kappa_2 = \sigma^2$, $\kappa_k = 0$ for $k \geq 3$ (and the vanishing of the higher cumulants characterizes the Gaussian distribution). The cumulants of a Poisson(λ) random variable are all λ.

The partition function (1) looks similar to a moment generating function, and in fact we can write the moment generating function of the random variable $H(\sigma)$ as a ratio of partition functions:

$$\begin{aligned}
\mathbb{E}e^{tH(\sigma)} &= \frac{1}{Z_G(\beta)}\sum_{\sigma} e^{tH(\sigma)}e^{-\beta H(\sigma)} \\
&= \frac{Z_G(\beta - t)}{Z_G(\beta)}\,, \qquad (9)
\end{aligned}$$

which gives

$$K_H(t) = \log Z_G(\beta - t) - \log Z_G(\beta)\,. \qquad (10)$$

Thus taking derivatives of $\log Z(\beta)$ in β we obtain the cumulants of the random variable $H(\sigma)$. Above we computed

$$\frac{d}{d\beta}\log Z(\beta) = -\mathbb{E}H(\sigma) = -\kappa_1(H)\,.$$

We can do a similar calculation with the second derivative:

$$\begin{aligned}\frac{d^2}{d\beta^2}\log Z(\beta) &= \frac{\frac{d^2}{d\beta^2}Z(\beta)}{Z(\beta)} - \left(\frac{\frac{d}{d\beta}Z(\beta)}{Z(\beta)}\right)^2 \\ &= \mathbb{E}[H(\sigma)^2] - (\mathbb{E}H(\sigma))^2 \\ &= \operatorname{var}(H(\sigma)) \\ &= \kappa_2(G)\,.\end{aligned}$$

The higher derivatives recover the higher cumulants of the energy:

$$\frac{d^k}{d\beta^k}\log Z(\beta) = (-1)^k \kappa_k(H)\,. \tag{11}$$

1.4 Multivariate partition functions

It is often useful and interesting to generalize partition functions or graph polynomials from a single variable to multiple variables (for a great discussion of this, see [8]).

The multivariate hard-core partition function (or multi-variate independence polynomial) of a graph G is

$$Z_G(\boldsymbol{\lambda}) = \sum_{I\in\mathcal{I}(G)}\prod_{v\in I}\lambda_v\,, \tag{12}$$

where $\boldsymbol{\lambda} = (\lambda_v)_{v\in V(G)}$ is a vector of activities indexed by the vertices of G. This is a generalization since we can obtain $Z_G(\lambda)$ by taking $\boldsymbol{\lambda}\equiv\lambda$. $Z_G(\boldsymbol{\lambda})$ is a multi-linear polynomial.

We can do the same for a general Gibbs measure. If our original model has partition function

$$Z_G(\beta) = \sum_{\sigma\in\{\pm1\}^V} e^{-\beta H(\sigma)}$$

we can add *non-uniform external fields* $\mathbf{t} = (t_v)_{v\in V(G)}$ and set

$$Z_G(\beta,\mathbf{t}) = \sum_{\sigma\in\{\pm1\}^V} e^{-\beta H(\sigma)}\prod_{v\in V(G)} e^{t_v\sigma_v}\,. \tag{13}$$

We recover the original partition function by taking $\mathbf{t}\equiv 0$.

More generally if we have a q-spin model; that is, $|\Omega| = q$, we can pick one distinguished spin ω, and put non-uniform external fields in the direction of ω. This gives the partition function

$$Z_G(\beta,\mathbf{t}) = \sum_{\sigma\in\Omega^V} e^{-\beta H(\sigma)}\prod_{v\in V(G)} e^{t_v\mathbf{1}_{\sigma_v=\omega}} \tag{14}$$

and Gibbs measure

$$\mu_{G,\beta,\mathbf{t}}(\sigma) = \frac{e^{-\beta H(\sigma)}\prod_{v\in V(G)} e^{t_v\mathbf{1}_{\sigma_v=\omega}}}{Z_G(\beta,\mathbf{t})}\,. \tag{15}$$

External fields introduce a bias in favor (or against if the external field is negative) of a certain spin. The (log of the) fugacity λ in the hard-core model is an external field favoring (or penalizing) occupied vertices.

An external field is uniform if it is the same at every vertex. A non-uniform external field is *consistent* if the sign of the external field is the same at every vertex.

1.5 Marginals and correlations

Central to the statistical physics point of view is considering how *correlations* in a given model behave and how this behavior depends on the parameters. All of the discussion below pertains to general graphs, but for intuition it is helpful to keep in mind a graph like $\mathbb{Z}^2$ or $\mathbb{Z}^d$ with very natural geometry.

We will also focus here mostly on two-spin models, like Ising or hard-core where a probability distribution on the spin set Ω can be specified by its expectation.

We will use the following notation: $\sigma_v \in \Omega$ is the spin at vertex v in the configuration σ. For a subset of vertices $S \subseteq V(G)$, $\sigma_S \in \Omega^S$ is the spin assignment to S given by σ (in other words it is the restriction of σ to the coordinates given by S).

The *marginal* or *occupation probability* of a vertex v is $\mu_v = \mathbb{E}[\sigma_v]$; for instance, in the hard-core model $\mu_v = P(v \in \mathbf{I})$. (For a q-spin model like Potts the marginal would be a probability distribution on $\Omega = [q]$, or we could specify a single spin ω and ask for the marginal probability of ω: $P(\sigma_v = \omega)$).

For a pair of vertices u, v, the joint marginal is $\mu_{u,v} = \mathbb{E}[\sigma_u \sigma_v]$. In the hard-core model, this is $\mu_{uv} = \Pr_{G,\lambda}[u, v \in \mathbf{I}]$. (For a q-spin model, the joint marginal would be described by a $q \times q$ matrix).

For a subset $S \subseteq V$, the joint marginal is $\mu_S = \mathbb{E}[\prod_{v \in S} \sigma_v]$. If $|S| = k$, then μ_S is also called the *k-point correlation function.*

We are often interested in how strong correlations between spins are, as a function of the parameters of the model and the graph distance between vertices. A natural way to measure the correlation between the spins at vertices u and v is to compute a covariance:

$$\kappa(u, v) = \mu_{uv} - \mu_u \mu_v \, .$$

If σ_u and σ_v were independent then $\kappa(u, v)$ would be 0; if $\kappa(u, v)$ is small in absolute value then we can say σ_u and σ_v are weakly correlated. The quantity $\kappa(u, v)$ is called the *truncated 2-point correlation function.*

1.5.1 Exponential decay of correlations

In probability and combinatorics we are very happy to work with independent random variables. We can compute variances, prove Chernoff bounds, prove Central Limit Theorems. In many or most interesting situations, however, we are not working with independent random variables. Figuring out how to generalize tools from the independent case is an important topic in probability theory. (For example, Martingales and Azuma's inequality generalize Chernoff bounds).

An important heuristic in statistical physics is that a weakly interacting system is well approximated by a system of independent spins or particles. We will see several precise

characterizations of this heuristic, including several notions of 'weakly interacting' and several notions of 'well approximated by'.

One notion of weakly interacting is that of decay of correlations.

Definition 1.1. *Let $\mathcal{G}$ be a family of graphs. We say a family of Gibbs measures μ_G, $G \in \mathcal{G}$, exhibits exponential decay of correlations if there exists constants $A, B > 0$ so that for all $G \in \mathcal{G}$, and all $u, v \in V(G)$,*

$$|\kappa(u, v)| \leq Ae^{-B\text{dist}(u,v)}, \tag{16}$$

where $\text{dist}(u, v)$ is the graph distance from u to v in G.

Note that exponential decay of correlations only makes sense for an infinite family of graphs (or for an infinite graph G if we define an infinite volume Gibbs measure, see Section 1.8). For any finite graph G (or any finite collection of finite graphs) we can always find A, B satisfying (16); what is important in the definition is that these constants are uniform over the collection of graphs $\mathcal{G}$.

Exercise 1. *Let P_n be the path graph on n vertices. Fix $\lambda > 0$. Show that hard-core model on the family of path graphs $\{P_n\}_{n\geq 1}$ exhibits exponential decay of correlations. Hint: write down a recursive equation for $Z_{P_n}(\lambda)$.*

1.6 Joint cumulants and truncated k-point correlation functions

We can also define truncated k-point correlation functions. To do that it will be helpful to define the joint cumulants of a collection of random variables.

The moment generating function for a collection of random variables $\vec{X} = (X_1, \ldots, X_k)$ defined on the same probability space is

$$M_{X_1,\ldots,X_k}(\mathbf{t}) = \mathbb{E}e^{\sum_{j=1}^k t_j X_j},$$

a function from $\mathbb{R}^k \to \mathbb{R}$. The joint cumulant generating function is

$$K_{X_1,\ldots,X_k}(\mathbf{t}) = \log \mathbb{E}e^{\sum_{j=1}^k t_j X_j}.$$

The *joint cumulants* of $\vec{X}$ are the coefficients of the multivariate Taylor series for $K_{X_1,\ldots,X_k}(\mathbf{t})$ around $\mathbf{t} = \vec{0}$. More precisely, for non-negative integers $\ell_1, \ldots, \ell_k$, we define the joint cumulant

$$\kappa(X_1^{(\ell_1)}, \ldots, X_k^{(\ell_k)}) = \frac{\partial^{\sum_{j=1}^k \ell_j}}{\prod_j \partial t_j^{\ell_j}} K_{X_1,\ldots,X_k}(\mathbf{t})\big|_{\mathbf{t}=\vec{0}}.$$

Specializing to our multivariate spin model with partition function

$$Z_G(\beta, \mathbf{t}) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} \prod_{v \in V(G)} e^{t_v \sigma_v},$$

we can define the *truncated k-point correlation function*

$$\kappa(u_1, \ldots, u_k) = \frac{\partial^k}{\partial t_{u_1} \cdots \partial t_{u_k}} \log Z_G(\beta, \mathbf{t})\big|_{\mathbf{t}=\vec{0}},$$

where $u_1, \ldots, u_k \in V(G)$.

Exercise 2. *Prove that*

$$\mu_v = \frac{\partial}{\partial t_v} \log Z_G(\beta, \mathbf{t})\big|_{\mathbf{t}=\vec{0}} \tag{17}$$

and

$$\mu_{uv} - \mu_u \mu_v = \kappa(u, v) = \frac{\partial^2}{\partial t_u \partial t_v} \log Z_G(\beta, \mathbf{t})\big|_{\mathbf{t}=\vec{0}}. \tag{18}$$

1.7 Dynamics

So far we have discussed the *equilibrium measure* μ_G. It is also very natural from several different perspectives (physical, computational, mathematical) to consider dynamics on the space of configurations Ω^V; that is a stochastic process on Ω^V whose behavior at large time scales looks like μ_G, and whose behavior might mimic the physical fluctuations of a gas or a magnet. Specifically, we can consider Markov chains whose stationary distribution is μ_G. One could take an entire graduate course in math or computer science on Markov chains (see, e.g., [4]), so we will be brief here.

A very natural Markov chain for a spin model on a graph is the *Glauber dynamics*. One step of the chain is easy to perform:

- Select a vertex $v \in V(G)$ uniformly at random;
- Resample the spin at v, σ_v, conditioned on the rest of the configuration. By the Markov property, this conditional distribution just depends on the current spins of the neighbors of v.

For example, if μ_G is the hard-core model on G, then given the current independent set I, we pick $v \in V(G)$ uniformly at random. If any neighbor of v is in I, then we do nothing and leave v unoccupied. If no neighbors of v are in I, then with probability $\frac{\lambda}{1+\lambda}$ we include v in the independent set and with probability $\frac{1}{1+\lambda}$ we remove v from the independent set.

It is not hard to check that the Glauber dynamics are a reversible Markov chain with unique stationary distribution μ_G, and so as the number of steps of the chain tends to ∞, the distribution converges to μ_G. How quickly it converges is a very interesting and practically relevant question. One way to measure the speed of convergence is via the *mixing time*.

1.8 Phase transitions

A central concept in statistical physics is that of a *phase transition*. There are several different mathematical definitions of a phase transition, and they do not always coincide. We will give three definitions here. One thing to remember is that a phase transition does not occur on a fixed, finite graph; instead it is a phenomenon of infinite graphs or of infinite sequences of finite graphs. Three examples of graphs to keep in mind are:

- An infinite lattice like $\mathbb{Z}^d$; this is the classical setting of statistical physics. Often the most interesting and physically relevant cases are $d = 2, 3$.

- A sequence of boxes in $\mathbb{Z}^d$ with increasing sidelengths or a sequence of discrete d-dimensional torii increasing sidelengths.
- The infinite Δ-regular tree $\mathbb{T}_\Delta$. Like $\mathbb{Z}^d$, this is also an infinite graph of bounded degree but with some important differences, including 1) the graph is non-amenable (meaning that boundaries of finite sets are comparable to the size of the sets) 2) trees have a recursive structure that can facilitate computations.
- Sequences of random graphs, like the random Δ-regular graph on n vertices or the random Δ-regular bipartite graph.

1.8.1 Boundary conditions and infinite volume limits

Suppose we want to understand a Gibbs measure (and the phase transition phenomenon) on an infinite graph like $\mathbb{Z}^d$ or $\mathbb{T}_\Delta$. One immediate difficulty is that because there are now an infinite number of possible configurations, one cannot define μ_G as above. Luckily there is a natural way to define infinite volume Gibbs measures via boundary conditions and conditional probability measures.

We start by describing how to impose *boundary conditions* on a Gibbs measure: that is, fixing the spins on a subsets of vertices (the boundary) and studying the resulting conditional Gibbs measure. We will define and study the notion of boundary conditions for arbitrary graphs (with arbitrary boundaries) but it is useful to keep in mind a box in $\mathbb{Z}^d$ with its natural geometric boundary.

Specify a spin model with spin set Ω and energy function $H(\cdot)$. Let $G = (V, E)$ be a graph and let $S \subseteq V$ be a specified set of boundary vertices. Let $\tau_S \in \Omega^S$ be a fixed assignment of spins to the vertices in S. Then the Gibbs measure with boundary conditions τ_S is the probability distribution $\mu_G^{\tau_S}$ on Ω^V defined by

$$\mu_G^{\tau_S}(\sigma) = \frac{e^{-\beta H(\sigma)} \mathbf{1}_{\sigma_S = \tau_S}}{Z_G^{\tau_S}(\beta)}$$

where

$$Z_G^{\tau_S}(\beta) = \sum_{\sigma \in \Omega^V} e^{-\beta H(\sigma)} \mathbf{1}_{\sigma_S = \tau_S}.$$

In words, $\mu_G^{\tau_S}$ is the distribution of μ_G conditioned on the event $\{\sigma_S = \tau_S\}$.

In the case of the hard-core model, boundary conditions take a particularly simple form. If we specify that a vertex u is not in I, then this is the same as removing u from $V(G)$. If we specify that u is in I, then this is the same as removing $N(u)$ from $V(G)$ and keeping u as an occupied isolated vertex (which will have no effect on the other vertices of G).

How can we define a Gibbs measure on an infinite graph like $\mathbb{Z}^d$ or the infinite Δ-regular tree? We cannot define it via (1) and (2) since the number of vertices is infinite.

One approach to make sense of a Gibbs measure on an infinite graph is that of Dobrushin, Lanford, and Ruelle based on conditional probabilities and boundary conditions.

A *specification* of a spin model is a fixed choice of parameters, say fixing λ in the hard-core model, or fixing the function f, g in the general formulation of a spin model on a graph.

For a given specification and a countably infinite graph G, we say a probability measure μ on $\Omega^{V(G)}$ is a Gibbs measure with the given specification if for every finite set $\Lambda \subset V(G)$ and every configuration $\tau \in \Omega^{V(G)}$, we have the following equality between conditional probabilities and finite volume Gibbs measures with boundary conditions:

$$\mu(\sigma_\Lambda = \cdot \big| \sigma_{\Lambda^c} = \tau_{\Lambda^c}) = \mu_\Lambda^{\tau_{\partial\Lambda}}(\cdot)\,, \tag{19}$$

where $\partial\Lambda$ is the external neighborhood of Λ, vertices in Λ^c that are joined to some vertex in Λ. In words, this condition says that the conditional distributions of μ induced on any finite set by a configuration on its complement is exactly the finite volume Gibbs measure with the same specification and the given boundary condition. For more details on the construction of infinite volume Gibbs measures see [7, 3, 2].

Under very general conditions, there always exists an infinite volume Gibbs measure with a given specification on an infinite graph G. The main question is whether there is only one: is the infinite volume Gibbs measure unique or not? (Note that for every finite graph G there is only one Gibbs measure).

For some infinite graphs (like $\mathbb{Z}^d$) we can define a number (the 'pressure') that describes the exponential growth rate of the partition function of finite subgraphs.

For a Gibbs measure defined on $\mathbb{Z}^d$, the (infinite volume) *pressure* at inverse temperature β is

$$p(\beta) = \lim_{\Lambda_n \uparrow \mathbb{Z}^d} \frac{1}{|\Lambda_n|} \log Z_{\Lambda_n}(\beta)\,. \tag{20}$$

The notation $\Lambda_n \uparrow \mathbb{Z}^d$ means that the sequence of subsets Λ_n is increasing, their limit is $\mathbb{Z}^d$, and

$$\frac{|\partial\Lambda_n|}{|\Lambda_n|} \to 0$$

(called a limit in the sense of van Hove). For most models this limit exists and is independent of the sequence Λ_n (and one may additionally impose arbitrary boundary conditions on each Λ_n without changing the limit). To define the pressure it is very important that the infinite graph be amenable, otherwise boundary conditions have a significant effect on $\log Z_{\Lambda_n}$ and the limit need not exist (or make any sense).

1.8.2 Notions of phase transition

1. **Uniqueness**. If G is an infinite graph, then we say a phase transition occurs at β_c if for all ε small enough there exists a unique infinite volume Gibbs measure on G for $\beta = \beta_c - \varepsilon$ and multiple infinite volume Gibbs measures for $\beta = \beta_c + \varepsilon$ for arbitrarily small $\varepsilon > 0$.

2. **Analyticity**. We say a phase transition occurs at β_c if the infinite-volume pressure $p(\beta)$ is non-analytic at β_c. This definition makes sense if we can define $p(\beta)$, so G should be an infinite, amenable graph.

3. **Correlation decay**. We say a phase transition occurs at β_c if for all ε small enough correlations decay exponentially fast when $\beta = \beta_c - \varepsilon$ and do not decay exponentially

fast when $\beta = \beta_c + \varepsilon$. This definition makes sense for an infinite graph (in which case the correlation decay should hold for the infinite volume Gibbs measure). This also makes sense for an infinite sequence of finite graphs (where the exponential decay of correlation should be uniform over the sequence).

Often (but not always) these notions of phase transition agree or are expected to agree, but there is not in general an equivalence. E.g. there are models in which there is a transition from exponential decay of correlations to polynomial decay of correlations yet the Gibbs measure remains unique.

Here we list some approaches to proving absence of phase transition or uniqueness of Gibbs measure for a given specification. We will see some of these methods in more details in upcoming lectures.

- Cluster expansion
- Dobrushin uniqueness
- Disagreement percolation
- Markov chain mixing
- Computational trees

Conversely, how does one show that multiple Gibbs measures exist for a given specification? There are a number of different techniques, depending on the model and the type of graph. For infinite trees, one can set up a recursion for marginal probabilities and show that there exist multiple fixed points. For lattices like $\mathbb{Z}^d$, one usually uses a Peierls argument, bounding the number and 'cost' of contours separating the origin from infinity; a much more sophisticated form of this technique is Pirogov-Sinai theory.

Background reading

The textbook of Friedli and Velenik [2] covers many fundamentals of statistical physics models on lattices from a mathematical point of view. Ruelle's textbook [7] is a classic reference for both lattice and continuum systems. The lecture notes of Adams [1] cover the mathematical fundamentals of statistical physics. It is also well worth reading about statistical physics from a physics perspective; the lecture notes of Tong are a good place to start [10]. Finally, Okounkov's recent overview of the work of Hugo Duminil–Copin [6] gives a nice feeling for some exciting questions and recent results in statistical physics.

1.9 Exercises and examples

1.9.1 Solved examples

1. Let $Z_G(\lambda)$ be the hard-core partition function on a graph G. Let $\mathbf{I}$ be a random independent set from G distributed as $\mu_{G,\lambda}$, the hard-core model at fugacity λ. Let $\alpha_G(\lambda) = \mathbb{E}|\mathbf{I}|$.

(a) Write an expression for $\alpha_G(\lambda)$ in terms of $\log Z_G(\lambda)$.

(b) Prove that $\alpha_G(\lambda)$ is strictly increasing as a function of λ (assume the graph G has at least one vertex).

(c) Write an expression for the number of maximum independent sets of G in terms of a limit involving $\log Z_G(\lambda)$.

(d) Suppose G has n vertices and maximum degree Δ. Show that

$$\mathrm{var}(|\mathbf{I}|) \geq cn$$

where c is a constant that depends on Δ and λ. Hint(s): use the law of total variance; use the fact that G has an independent set of size at least $n/(\Delta+1)$.

Solution:

(a) We can write

$$\begin{aligned}
\alpha_G(\lambda) &= \sum_{I \in \mathcal{I}(G)} |I| \mu_G(I) \\
&= \frac{\sum_{I \in \mathcal{I}(G)} |I| \lambda^{|I|}}{Z_G(\lambda)} \\
&= \lambda \frac{\sum_{I \in \mathcal{I}(G)} |I| \lambda^{|I|-1}}{Z_G(\lambda)} \\
&= \lambda \frac{Z_G'(\lambda)}{Z_G(\lambda)} \\
&= \lambda (\log Z_G(\lambda))' .
\end{aligned}$$

(b) We take the derivative of $\alpha_G(\lambda)$ with respect to λ:

$$\begin{aligned}
\alpha_G'(\lambda) &= \frac{\lambda Z_G(\lambda) Z_G''(\lambda) + Z_G(\lambda) Z_G'(\lambda) - \lambda (Z_G'(\lambda))^2}{Z_G(\lambda)^2} \\
&= \frac{\mathrm{var}(|\mathbf{I}|)}{\lambda} > 0
\end{aligned}$$

since $\{|\mathbf{I}| = 0\}$ and $\{|\mathbf{I}| = 1\}$ both have positive probability when $\lambda > 0$.

(c) For $\lambda \gg 2^n$, $Z_G(\lambda) \approx i_{\alpha(G)} \lambda^{\alpha(G)}$; in particular,

$$1 \leq \frac{Z_G(\lambda)}{i_{\alpha(G)} \lambda^{\alpha(G)}} \leq 1 + \frac{2^n}{\lambda} .$$

We can recover $\alpha(G)$:

$$\alpha(G) = \lim_{\lambda \to \infty} \frac{\log Z_G(\lambda)}{\log \lambda} .$$

Then once we know $\alpha(G)$ we can recover $i_{\alpha(G)}$:

$$i_{\alpha(G)} = \lim_{\lambda \to \infty} \frac{Z_G(\lambda)}{\lambda^{\alpha(G)}} .$$

(d) Let $J \in \mathcal{I}(G)$ be an independent set of size at least $n/(\Delta+1)$. Let $X = |\mathbf{I}|$. Let Y be the number of vertices of J 'unblocked' by $\mathbf{I}$; that is,

$$Y = |\{x \in J : N(x) \cap \mathbf{I} = \emptyset\}| .$$

The Law of Total Variance says

$$\mathrm{var}(X) = \mathbb{E}[\mathrm{var}(X|\mathbf{I} \cap J^c)] + \mathrm{var}[\mathbb{E}(X|\mathbf{I} \cap J^c)] .$$

By the spatial Markov property, given $\mathbf{I} \cap J^c$, the unblocked vertices of J appear in $\mathbf{I}$ independently with probability $\lambda/(1+\lambda)$, and so $\mathrm{var}(X|\mathbf{I} \cap J^c) = \frac{\lambda}{(1+\lambda)^2} \cdot Y$. On the other hand, any vertex in a max-degree Δ graph is unblocked with probability at least $\frac{1}{(1+\lambda)^\Delta}$ and so

$$\mathbb{E}Y \geq |J| \frac{1}{(1+\lambda)^\Delta} \geq \frac{n}{(\Delta+1)(1+\lambda)^\Delta} .$$

Putting this together gives

$$\mathrm{var}(X) \geq n \cdot \frac{\lambda}{(\Delta+1)(1+\lambda)^{\Delta+2}} .$$

2. Let G be a graph and $A \subset V(G)$ be a set of d vertices. Suppose you know that $\mu_{G,\lambda}(\mathbf{I} \cap A = \emptyset) = p$. Form the graph G' by adding a single vertex v to $V(G)$ and connecting v to each of the vertices of A.

   (a) Compute $\log Z_{G'}(\lambda) - \log Z_G(\lambda)$.

   (b) Compute $\mu_{G',v,\lambda}$ (the occupation probability of v in the hard-core model on G' at fugacity λ).

   (This is a baby version of the 'cavity method')

   Solution:

   (a) Independent sets of G' can be divided in two sets: those that contain v and those that do not. Independent sets that do not contain v are exactly the independent sets of G. Independent sets that contain v are independent sets of G that contain no neighbor of v, plus v. We can then write

$$Z_{G'}(\lambda) = Z_G(\lambda) + \lambda \sum_{I \in \mathcal{I}: I \cap N(v) = \emptyset} \lambda^{|I|} .$$

Since

$$p = \mu_{G,\lambda}(\mathbf{I} \cap A = \emptyset) = \frac{\sum_{I \in \mathcal{I}: I \cap N(v) = \emptyset} \lambda^{|I|}}{Z_G(\lambda)}$$

we have

$$\frac{Z_{G'}(\lambda)}{Z_G(\lambda)} = 1 + \lambda p$$

and so $\log Z_{G'}(\lambda) - \log Z_G(\lambda) = \log(1 + \lambda p)$.

(b) By a similar calculation,

$$\begin{aligned}\mu_{G',v,\lambda} &= \frac{\lambda \sum_{I \in \mathcal{I}: I \cap N(v) = \emptyset} \lambda^{|I|}}{Z_{G'}(\lambda)} \\ &= \frac{\lambda p Z_G(\lambda)}{(1+\lambda p) Z_G(\lambda)} \\ &= \frac{\lambda p}{1+\lambda p}\,.\end{aligned}$$

(As a check, imagine v has no neighbors in G'; then this formula reduces to $\lambda/(1+\lambda)$, the probability an isolated vertex is occupied).

References

[1] S. Adams. *Lectures on mathematical statistical mechanics.* Number 30. Dublin Institute for Advanced Studies, 2006.

[2] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction.* Cambridge University Press, 2017.

[3] H.-O. Georgii. Gibbs measures and phase transitions. In *Gibbs Measures and Phase Transitions.* de Gruyter, 2011.

[4] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[5] H. Löwen. Fun with hard spheres. In *Statistical physics and spatial statistics*, volume 554, pages 295–331. Springer, 2000.

[6] A. Okounkov. The Ising model in our dimension and our times. *arXiv preprint arXiv:2207.03874*, 2022.

[7] D. Ruelle. *Statistical mechanics: Rigorous results.* World Scientific, 1999.

[8] A. D. Scott and A. D. Sokal. The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *Journal of Statistical Physics*, 118(5-6):1151–1261, 2005.

[9] B. Szegedy. Edge coloring models and reflection positivity. *Journal of the American mathematical Society*, 20(4):969–988, 2007.

[10] D. Tong. Lectures on statistical physics. `https://www.damtp.cam.ac.uk/user/tong/statphys.html`.