3 Cluster expansion

In this lecture we focus on one very important tool from statistical physics: the cluster expansion. At a very high level, the cluster expansion provides a way to show that a collection of random variables is 'close' to independent. When it applies, you can deduce almost any probabilistic information you can deduce for collections of independent random variables. We will see that it is very useful for combinatorial enumeration.

3.1 Ideal gas laws

An *equation of state* for a gas relates the quantities that determine the state of the gas: volume, pressure, and temperature.

The *ideal gas* is a gas of non-interacting particles. In the continuum we can model the ideal gas as a Poisson process of intensity λ on a region Λ of finite volume in $\mathbb{R}^d$. The partition function of the ideal gas is given by

$$Z_\Lambda(\lambda) = \sum_{k\geq 0} \int_{\Lambda^k} \frac{\lambda^k}{k!}\, dx_1 \dots dx_k = e^{\lambda|\Lambda|}\,.$$

The density of the ideal gas is simply λ, while the pressure is the normalized logarithm of the partition function: $p_\Lambda(\lambda) = \frac{1}{|\Lambda|}\log e^{\lambda|\Lambda|} = \lambda$. So for the ideal gas the equation of state is

$$p = \rho$$

where p is the infinite volume pressure and ρ is the infinite volume density. In physics this is stated with various constants (and can be derived from the equation $PV = nRT$ which you may be familiar with).

What happens when we deviate from the ideal gas and particles interact with each other? The equation of state must change to account for the interaction, but we can hope that when the density is small enough or the temperature large enough, the new equation of state will be a small perturbation of the ideal gas law.

The idea of the cluster expansion is to measure deviations from the ideal gas law due to interactions via an infinite series. To begin with the series is just a formal power series, but when the activity λ is small enough it can be shown that the series is convergent. The cluster expansion is also known as the 'Mayer series' or 'Mayer expansion' after the work of Mayer and Mayer and Mayer and Montroll [9, 10]. Rigorous proofs of convergence came later, e.g. [4, 13, 12].

In a discrete setting we can define an ideal gas as well, by taking, for instance, the hard-core model on a graph of n isolated vertices (we could have defined instead a gas with a Poisson number of particles at each site, which would lead to a different discrete ideal gas). Then for a given fugacity λ, the density is $\rho = \frac{\lambda}{1+\lambda}$ while the pressure is $p = \log(1+\lambda)$. Eliminating λ gives an equation of state:

$$p = \log\frac{1}{1-\rho}\,.$$

When we introduce interactions by adding edges to the empty graph, the equation of state will change. But again we can hope that when λ is small enough and interactions are weak enough (in terms of, say, the number of edges of the graph), the new equation of state will be a small perturbation of the ideal law.

3.2 Cluster expansion for low density hard-core model

It will be very useful for what we do later to introduce the cluster expansion in the setting of the multivariate hard-core model on a graph G; that is, each vertex v is assigned its own fugacity $\boldsymbol{\lambda}_v$.

We record some notation before we begin. Let K_k be the complete graph on k vertices and $\mathcal{G}_k$ be the set of all graphs on k vertices; that is, all subgraphs of K_k. Let $\mathcal{C}_k$ be the set of all connected graphs on k vertices. We also define the *Ursell function* (with respect to G): for $v_1, \dots v_k \in V(G)$,

$$\phi(v_1, \dots, v_k) = \frac{1}{k!} \sum_{H \in \mathcal{C}_k} \prod_{(i,j) \in E(H)} \left(-\mathbf{1}_{v_i = v_j \text{ OR } (v_i, v_j) \in E(G)}\right).$$

In particular if $v_1, \dots, v_k$ induce a disconnected subgraph graph of G then $\phi(v_1, \dots, v_k) = 0$.

We begin by writing

$$\begin{aligned} Z_G(\boldsymbol{\lambda}) &= \sum_{I \in \mathcal{I}(G)} \prod_{v \in I} \boldsymbol{\lambda}_v \\ &= \sum_{k \geq 0} \frac{1}{k!} \sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \prod_{1 \leq i < j \leq k} (1 - f(v_i, v_j)) \end{aligned}$$

where $f(v_i, v_j) = 1$ if $v_i = v_j$ or $(v_i, v_j) \in E(G)$. We can expand the last product

$$\prod_{1 \leq i < j \leq k} (1 - f(v_i, v_j)) = \sum_{H \subseteq K_k} \prod_{(i,j) \in E(H)} (-f(v_i, v_j))$$

where H is a subgraph of K_k with edges $E(H)$. We have

$$\begin{aligned} Z_G(\boldsymbol{\lambda}) &= 1 + \sum_{k \geq 1} \frac{1}{k!} \sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \sum_{H \in \mathcal{G}_k} \prod_{(i,j) \in E(H)} (-f(v_i, v_j)) \\ &= 1 + \sum_{k \geq 1} \frac{1}{k!} \sum_{H \in \mathcal{G}_k} \sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \prod_{(i,j) \in E(H)} (-f(v_i, v_j)). \end{aligned}$$

The expression $\sum_{v_1, \dots, v_k \subseteq V(G)} \prod_{j=1}^{k} \boldsymbol{\lambda}_{v_j} \prod_{(i,j) \in E(H)} (-f(v_i, v_j))$ factorizes over the connected components $H_1, \dots H_\ell$ of H, and so we can write

$$Z_G(\boldsymbol{\lambda}) = 1 + \sum_{k \geq 1} \frac{1}{k!} \sum_{\ell=1}^{k} \sum_{\substack{H \in \mathcal{G}_k \\ H = (H_1, \dots, H_\ell)}} \prod_{r=1}^{\ell} \sum_{v_1, \dots, v_{m_r} \subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j) \in E(H_r)} (-f(v_i, v_j))$$

and summing over possible sizes of the connected components and noting that the summand corresponding to H_r only depends on its isomorphism class as a graph,

$$= 1 + \sum_{k\geq 1}\sum_{\ell=1}^{k} \frac{1}{\ell!} \sum_{\substack{m_1,\ldots,m_\ell \\ \sum m_r = k}} \frac{k!}{m_1!\cdots m_\ell!} \sum_{H_1\in\mathcal{C}_{m_1},\ldots,H_\ell\in\mathcal{C}_{m_\ell}} \prod_{r=1}^{\ell} \sum_{v_1,\ldots,v_{m_r}\subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1}\sum_{k\geq \ell} \frac{1}{\ell!} \sum_{\substack{m_1,\ldots,m_\ell \\ \sum m_r = k}} \frac{k!}{m_1!\cdots m_\ell!} \sum_{H_1\in\mathcal{C}_{m_1},\ldots,H_\ell\in\mathcal{C}_{m_\ell}} \prod_{r=1}^{\ell} \sum_{v_1,\ldots,v_{m_r}\subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \sum_{m_1,\ldots,m_\ell} \sum_{H_1\in\mathcal{C}_{m_1},\ldots,H_\ell\in\mathcal{C}_{m_\ell}} \prod_{r=1}^{\ell} \frac{1}{m_r!} \sum_{v_1,\ldots,v_{m_r}\subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \sum_{m_1,\ldots,m_\ell} \prod_{r=1}^{\ell} \frac{1}{m_r!} \sum_{H\in\mathcal{C}_{m_r}} \sum_{v_1,\ldots,v_{m_r}\subseteq V(G)} \prod_{j=1}^{m_r} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \prod_{r=1}^{\ell} \sum_{m\geq 1} \frac{1}{m!} \sum_{H\in\mathcal{C}_m} \sum_{v_1,\ldots,v_m\subseteq V(G)} \prod_{j=1}^{m} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j))$$

$$= 1 + \sum_{\ell\geq 1} \frac{1}{\ell!} \left(\sum_{m\geq 1} \frac{1}{m!} \sum_{H\in\mathcal{C}_m} \sum_{v_1,\ldots,v_m\subseteq V(G)} \prod_{j=1}^{m} \boldsymbol{\lambda}_{v_j} \prod_{(i,j)\in E(H_r)} (-f(v_i,v_j)) \right)^{\ell}$$

$$= \exp\left(\sum_{m\geq 1} \sum_{v_1,\ldots v_m\in V(G)} \phi(v_1,\ldots,v_m) \prod_{j=1}^{\ell} \boldsymbol{\lambda}_{v_j} \right)$$

or in other words,

$$\log Z_G(\boldsymbol{\lambda}) = \sum_{m\geq 1} \sum_{v_1,\ldots v_m\in V(G)} \phi(v_1,\ldots,v_m) \prod_{j=1}^{\ell} \boldsymbol{\lambda}_{v_j}\,.$$

There was a step (the second line in the last series of equalities) that was not justified: exchanging the order of summation over k and ℓ. Thus the result is a formal power series for $\log Z_G(\boldsymbol{\lambda})$ which may or may not be convergent. We consider the question of convergence in the next section.

We can rewrite this series in terms of *clusters*. Given the graph G, let $\Gamma = (v_1,\ldots,v_\ell)$ be an ordered tuple of vertices of G (with possible repetitions). The incompatibility graph $H(\Gamma)$ is the graph with vertex set Γ and edges between $v_i, v_j \in \Gamma$ if $v_i = v_j$ or if $(v_i, v_j) \in E(G)$. A *cluster* is an ordered tuple Γ whose incompatibility graph $H(\Gamma)$ is connected. The size of the cluster is the length of the tuple.

We can interpret the Ursell function in terms of the incompatibility graph H:

$$\phi(H) = \frac{1}{|V(H)|!} \sum_{A\subseteq E(H), \text{spanning, connected}} (-1)^{|A|}\,.$$

Then the cluster expansion becomes

$$\log Z_G(\boldsymbol{\lambda}) = \sum_{\Gamma} \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \,,$$

where the sum is over all clusters Γ of G. It will be very important later, for combinatorial, algorithmic, and probabilistic reasons, that the cluster expansion is the sum over *connected* objects.

3.2.1 Examples

1. Suppose G is a single vertex v with fugacity λ. Then for each $\ell \geq 1$ there is a single cluster consisting of ℓ copies of v. The incompatibility graph is the complete graph on ℓ vertices with Ursell function $\frac{(-1)^{\ell+1}}{\ell}$ (exercise: prove this!). Then the cluster expansion is

$$\log Z_G(\lambda) = \sum_{\ell \geq 1} (-1)^{\ell+1} \frac{\lambda^\ell}{\ell}$$

   which is of course the Taylor series for $\log(1+\lambda)$ around 0.

2. Suppose G is a Δ-regular, triangle-free graph on n vertices. Consider the (univariate) hard-core model at fugacity λ. Then we can use symmetry to compute the first few terms of the cluster expansion.

   We list the clusters by size:

   - There are n clusters of size 1 (each a single vertex)
   - There are n clusters of size 2 consisting of two copies of a single vertex; Δn clusters of size 2 consisting of (ordered) edges.
   - There are n clusters of size 3 consisting of 3 copies of the same vertex; $3\Delta n$ clusters of size three with two copies of one vertex and one copy of a vertex joined to it by an edge; $3n\Delta(\Delta-1)$ clusters consisting of a 'v' of three vertices.

Cluster	Size	Count	Ursell function
1 vertex	1	n	1
2 copies of 1 vertex	2	n	$-1/2$
Ordered edge	2	Δn	$-1/2$
3 copies of 1 vertex	3	n	$1/3$
An edge with a repeated vertex	3	$3\Delta n$	$1/3$
A path of three vertices	3	$3n\Delta(\Delta-1)$	$1/6$

Table 1: A list of clusters up to size 3 in a regular, triangle-free graph

   Putting this together we get

$$\frac{1}{n} \log Z_G(\lambda) = \lambda - \frac{\lambda^2}{2} + \frac{\lambda^3}{3} - \frac{\Delta}{2}\lambda^2 + \frac{2\Delta + \Delta(\Delta-1)}{2}\lambda^3 + \dots$$

We can also measure the deviation from the ideal gas by removing the terms corresponding to clusters consisting of k copies of a single vertex:

$$\frac{1}{n}\log\frac{Z_G(\lambda)}{(1+\lambda)^n} = \frac{1}{n}\log Z_G(\lambda) - \log(1+\lambda) = -\frac{\Delta}{2}\lambda^2 + \frac{2\Delta+\Delta(\Delta-1)}{2}\lambda^3 + \dots$$

The expression $\frac{Z_G(\lambda)}{(1+\lambda)^n}$ has a nice interpretation: it is the probability that when picking a subset $S \subset V(G)$ by including each vertex independently with probability $\lambda/(1+\lambda)$ that the set S is an independent set. Janson's Inequality bounds probabilities such as this, and the first term of the expansion is exactly the term in the exponent in Janson's Inequality. The cluster expansion gives a way to get arbitrarily many terms in the expansion (but of course we need to ask about convergence). A generalization of Janson's Inequality due to Mousset, Noever, Panagiotou, and Samotij [11] deals with the probability of obtaining an independent set in a hypergraph when picking vertices independently at random.

3.3 Convergence criteria

The cluster expansion will only be a useful tool if we can show that it converges (and even more, bound the truncation error). There is a very large body of literature in mathematical physics devoted to finding convergence criteria for the cluster expansion. This criteria seek to balance three qualities: how sharp the bound is, how general its applicability, and how easy it is to check and apply. In this course we will give two such criteria: one that is very general and easy to check; the other that is sharp in a specific situation and has connections to combinatorics.

While we are usually interested in non-negative valued fugacities $\boldsymbol{\lambda}$, the convergence criteria will often be written in terms of complex-valued $\boldsymbol{\lambda}$. This is for several reasons. The cluster expansion is a power series and so will be convergence in a (poly-)disk in the complex plane. Showing convergence in complex domain allows one to deduce analyticity properties of the pressure and prove absence of phase transition in the Lee–Yang sense. Finally, we will see that evaluating the independence polynomial at negative fugacities is closely related to the Lovász Local Lemma. For much more on this last connection see the paper of Scott and Sokal [14].

The following result of Shearer gives the optimal zero-free (poly)disk for graphs of maximum degree Δ

Theorem 3.1 ([15])**.** *Let G be a graph of maximum degree Δ and suppose that for all $v \in V(G)$,*

$$|\boldsymbol{\lambda}_v| \le \lambda_s(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{\Delta^\Delta}\,. \tag{1}$$

The the cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges absolutely, and in particular, $Z_G(\boldsymbol{\lambda}) \neq 0$.

Note that $\lambda_s(\Delta) = \frac{1+o(1)}{e\Delta}$ as $\Delta \to \infty$, so it has the same asymptotic performance as the Kotecký–Preiss bound. Shearer's bound, however, is tight: by taking finite truncations of the infinite Δ-regular tree, we can find zeroes of the independence polynomials of max degree Δ

graphs that approach $-\lambda_s(\Delta)$ from below on the negative real axis. In fact, Groeneveld [4] (and Scott–Sokal [14]) show that the coefficients of the cluster expansion for $\log Z_G(\lambda)$ alternate in sign and so the closest complex zero of Z_G must lie on the negative real axis, and this zero determines the radius of convergence of the cluster expansion.

We can derive bounds on the rate of convergence using Theorem 3.1 an idea of Barvinok [1]. Define the kth order truncation of the cluster expansion as

$$T_k(G, \boldsymbol{\lambda}) = \sum_{\Gamma : |\Gamma| \le k} \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \,.$$

Theorem 3.2 ([15, 1])**.** *Suppose G is a graph of maximum degree Δ on n vertices and for $\eta \in (0,1)$ suppose $|\boldsymbol{\lambda}_v| \le \eta \lambda_s(\Delta)$ for all $v \in V(G)$. Then*

$$|\log Z_G(\boldsymbol{\lambda}) - T_k(G, \boldsymbol{\lambda})| \le \frac{n\eta^k}{1-\eta} \,.$$

Proof. Let $\hat{Z}(t) = Z(t\boldsymbol{\lambda})$. $\hat{Z}$ is a polynomial in t of degree at most $N \le n$ and $Z(t) \neq 0$ for any complex t satisfying $|t| \le \eta^{-1}$. We are interested in evaluating $\log \hat{Z}(1)$. By the fundamental theorem of algebra we can write

$$\hat{Z}(t) = \prod_{j=1}^{N} \left(1 - \frac{t}{r_j}\right)$$

where $r_1, \ldots, r_N$ are the complex roots of $\hat{Z}(t)$. We write

$$\log \hat{Z}(t) = \sum_{j=1}^{N} \log \left(1 - \frac{t}{r_j}\right)$$

and then Taylor expand $\log\left(1 - \frac{t}{r_j}\right)$ at $t = 1$ around $t = 0$ using the fact that $|1/r_j| \le \eta$. This gives

$$\begin{aligned} |\log Z_G(\boldsymbol{\lambda}) - T_k(G, \boldsymbol{\lambda})| &\le \sum_{j=1}^{N} \sum_{i=k+1}^{\infty} \frac{\eta^i}{i} \\ &\le \frac{N\eta^k}{1-\eta} \\ &\le \frac{n\eta^k}{1-\eta} \,. \end{aligned}$$

□

The following is a specialization of the Kotecký–Preiss condition. We will see the general condition shortly.

Theorem 3.3 ([8]). *Consider the multivariate hard-core model on a graph G with complex-valued fugacity vector $\boldsymbol{\lambda}$. Suppose that for some $b \geq 0$ and each $v \in V(G)$,*

$$\sum_{u \in N(v) \cup \{v\}} |\boldsymbol{\lambda}_u| e^{1+b} \leq 1 \,. \tag{2}$$

Then the cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges absolutely, and moreover, for each $v \in V(G)$ and each $t \geq 0$,

$$\sum_{\substack{\Gamma \nsim v \\ |\Gamma| \geq t}} \left| \phi(H(\Gamma)) \prod_{u \in \Gamma} \boldsymbol{\lambda}_u \right| \leq e^{-bt} \,. \tag{3}$$

In particular, $Z_G(\boldsymbol{\lambda}) \neq 0$.

Here we write $v \nsim \Gamma$ if there is some $u \in \Gamma$ so that $u = v$ or $u \in N(v)$; that is, if $\Gamma \cup \{v\}$ is a cluster.

We can apply this condition to a Δ-regular graph with uniform fugacity λ, taking $b = 0$. In this case, the condition is

$$|\lambda| \leq \frac{1}{e(\Delta + 1)} \,.$$

Taking $b > 0$ allows us to deduce strong bounds on the tail of the cluster expansion and correlation decay properties. We will see this in Section 3.4.

Connections to Lovász Local Lemma

Shearer proved Theorem 3.1 in the context of the Lovász Local Lemma. Consider a set of events $E_1, \ldots, E_n$ on a probability space, and a *dependency graph* G on the vertex set $[n]$ with the property that the event E_i is independent of the events $\{E_j : j \notin \{i \cup N(i)\}\}$. Let $p_j = P(E_j)$. The Lovász Local Lemma gives conditions in terms of G and $\{p_j\}$ to ensure that $\Pr\left(\bigcap_{j=1}^n \overline{E}_j\right) > 0$. Shearer showed a very nice connection between such a condition and the independence polynomial of the dependency graph G.

Theorem 3.4 ([15]). *Given a set of events $E_1, \ldots, E_n$ with probabilities $p_1, \ldots, p_n$ and dependency graph G, suppose that*

$$Z_G(-\boldsymbol{q}) \neq 0$$

for all $\boldsymbol{q} \leq \boldsymbol{p}$ coordinatewise. Then

$$\Pr\left(\bigcap_{j=1}^n \overline{E}_j\right) > 0 \,.$$

Combining the two theorems gives an optimal condition for the Lovász Local Lemma.

Corollary 3.5. *Given a set of events $E_1, \ldots, E_n$ with probabilities $p_1, \ldots, p_n$ and dependency graph G of maximum degree Δ. Then if $p_j \leq \frac{(\Delta-1)^{\Delta-1}}{\Delta^\Delta}$ for all j. Then $\Pr\left(\bigcap_{j=1}^n \overline{E}_j\right) > 0$.*

3.4 Cluster expansion and (joint) cumulants

An important application of a convergent cluster expansion is to proving strong correlation decay properties.

Recall from Lecture 1 that we defined the truncated k-point correlation functions in terms of the partial derivatives of a log partition function with external fields. We restate the definition here in the setting of the multivariate hard-core model.

With $\boldsymbol{\lambda}$ and $\mathbf{t}$ both vectors of non-negative numbers indexed by $V(G)$, let

$$Z_G(\boldsymbol{\lambda}, \mathbf{t}) = \sum_{I \in \mathcal{I}(G)} \prod_{v \in V(G)} \boldsymbol{\lambda}_v e^{t_v} .$$

For $u_1, \ldots, u_k \in V(G)$, the truncated k-point correlation function is

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

In particular, the marginal of v is $\mu_v = \frac{\partial}{\partial t_v} \log Z_G(\boldsymbol{\lambda}, \mathbf{t})$ and the covariance of spins at u and v is $\kappa(u, v) = \mu_{uv} - \mu_u \mu_v = \frac{\partial^2}{\partial t_u \partial t_v} \log Z_G(\boldsymbol{\lambda}, \mathbf{t})$.

When the cluster expansion for $\log Z_G(\boldsymbol{\lambda})$ converges, we can differentiate it term-by-term and get a convergent expansion for the truncated correlation functions. Moreover, because the cluster expansion is a sum over connected objects, we can deduce correlation decay properties. (See [3] or [2] for more details).

Lemma 3.6. *Consider the multivariate hard-core model on G with fugacity vector $\boldsymbol{\lambda}$. Suppose that for some $b > 0$ the Kotecký–Preiss condition (2) holds. Then the model exhibits exponential decay of correlations, i.e. for all $u, v \in V(G)$,*

$$|\kappa(u, v)| \leq C(b) e^{-b \cdot \mathrm{dist}(u,v)} ,$$

where the constant C only depends on the constant b. More generally, for any $k > 0$ and $u_1, \ldots, u_k \in V(G)$, we have

$$|\kappa(u_1, \ldots, u_k)| \leq C(b, k) e^{-b \cdot \mathrm{MST}(u_1, \ldots, u_k)} ,$$

where $\mathrm{MST}(u_1, \ldots, u_k)$ is the minimum number of edges of G in a connected subgraph containing $u_1, \ldots, u_k$.

Recall that joint cumulants of independent random variables are 0, and so upper bounds on the magnitude of joint cumulants show a kind of approximate independence. It is not difficult to convert these bounds into bounds on the total variation distance between the joint distribution of spins at $u_1, \ldots, u_k$ and independent spins with marginals $\mu_{u_1}, \ldots, \mu_{u_k}$.

Proof. To start, we give a formula for the truncated correlation functions.

$$\kappa(u_1, \ldots, u_k) = \sum_{\Gamma \supseteq \{u_1, \ldots, u_k\}} \prod_{i=1}^{k} Y_{u_i}(\Gamma) \cdot \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \tag{4}$$

where $Y_u(\Gamma)$ is the number of occurrences of v in the cluster Γ. In particular,

$$\mu_v = \sum_{\Gamma \ni v} Y_v(\Gamma)\phi(H(\Gamma)) \prod_{u \in \Gamma} \boldsymbol{\lambda}_u \,.$$

Observe that

$$\frac{\partial^k}{\partial t_{u_1} \cdots \partial t_{u_k}} \prod_{v \in \Gamma} \boldsymbol{\lambda}_v e^{\mathbf{t}_v}\big|_{\mathbf{t}=\vec{0}} = \prod_{i=1}^{k} Y_{u_i}(\Gamma) \cdot \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \,,$$

and then apply this term-by-term to the cluster expansion to obtain (4). Differentiating the power series term-by-term is justified since we are inside the radius of convergence.

Now we can apply the pinned tail estimate (3) from Theorem 3.3. Each cluster that contains all vertices $u_1, \ldots, u_k$ has size at least $\mathrm{MST}(u_1, \ldots, u_k)$ since its incompatibility graph must be connected. Moreover, each additional copy of a vertex from $u_1, \ldots, u_k$ adds at least 1 to the size of a cluster. Therefor, by applying (3), for each $r_1, \ldots r_k \geq 0$, we have

$$\begin{aligned}\sum_{\Gamma : Y_{u_j}(\Gamma) = r_j + 1, 1 \leq j \leq k} \left| \prod_{i=1}^{k} Y_{u_i}(\Gamma) \cdot \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \right| &\leq e^{-b \cdot (\mathrm{MST}(u_1, \ldots, u_k) + \sum_j r_j - 1)} \prod_{j=1}^{k} (1 + r_j) \\ &\leq C(b,k) e^{-b \cdot \mathrm{MST}(u_1, \ldots, u_k)} \,.\end{aligned}$$

The result for $\kappa(u, v)$ follows immediately since $\mathrm{MST}(u, v) = \mathrm{dist}(u, v)$.

□

Note that in the setting of the univariate hard-core model on graphs of maximum degree Δ, Weitz's result proves exponential decay of correlations (in fact, strong spatial mixing) for a wider range of parameters than we obtain by cluster expansion convergence. The power of the cluster expansion method will become evident later when we apply it to models with very non-uniform fugacities.

3.5 Limit theorems and large deviations

A general principle is that whenever we have a convergent cluster expansion, we can, without too much work, deduce any probabilistic information we want to know, and in particular, any qualitative result that holds for independent random variables will likely hold in our setting. Convergence of the cluster expansion tells us that the perturbation from independent spins is small enough that we can control it.

We can use the cluster expansion to prove convergence of random variables. Recall that a random variable X has a $N(\mu, \sigma^2)$ distribution if and only if $\kappa_1(X) = \mu$, $\kappa_2(X) = \sigma^2$ and $\kappa_k(X) = 0$ for $k \geq 3$. Since the Normal distribution is determined by its moments, we can show that a sequence of random variables X_n converges to $N(\mu, \sigma^2)$ in distribution if $\kappa_1(X_n) \to \mu$, $\kappa_2(X_n) \to \sigma^2$, and $\kappa_k(X_n) \to 0$ for each fixed $k \geq 3$.

Consider the multivariate hard-core model, and assume the Kotecký–Preiss condition holds with some $b > 0$. Let $X = |\mathbf{I}|$ be the size of the random independent set drawn from the model (X implicitly depends on G and we can take the size of G to ∞). We can compute

the first two cumulants of X by taking $\mathbf{t} = t$ above and taking derivatives of $\log Z_G(\boldsymbol{\lambda}, t)$ in t:

$$\kappa_1(X) = \sum_{\Gamma} |\Gamma| \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v$$
$$\kappa_2(X) = \sum_{\Gamma} |\Gamma|^2 \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v$$

Now let us assume that $\kappa_2(X) = \mathrm{var}(X) = \Theta(n)$ (which we can prove in many cases, assuming, say, that $\boldsymbol{\lambda}$ is bounded above and below and G has an independent set of linear size). Let $\tilde{X} = (X - \mathbb{E}X)/\sqrt{\mathrm{var}(X)}$ be the centered and scaled version of X. Then to show that $\tilde{X} \Rightarrow N(0,1)$ as $n \to \infty$, it is enough to show that $\kappa_k(\tilde{X}) \to 0$ for each $k \geq 3$. Since centering by the mean does not change higher cumulants, and since $\mathrm{var}(X) = \Theta(n)$ it is enough to show that $\kappa_k(X/\sqrt{n}) \to 0$ for $k \geq 3$. We can compute

$$\begin{aligned}
\kappa_k(X/\sqrt{n}) &= n^{-k/2} \sum_{\Gamma} |\Gamma|^k \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \\
&\leq n^{-k/2} \sum_{t \geq 1} t^k \sum_{|\Gamma| = t} \left| \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \right| \\
&\leq n^{-k/2} n \sum_{t \geq 1} t^k e^{-bt} \\
&= O(n^{1-k/2}) \\
&= o(1)\,,
\end{aligned}$$

where we used (3) and summed over all vertices in the second inequality.

Next we can ask about large deviations for X; that is, prove an upper bound on probabilities like $\Pr(X \geq (1+\varepsilon)\mathbb{E}X)$ for $\varepsilon > 0$ small and fixed. If we assume $\mathbb{E}X = \Theta(n)$ and $\mathrm{var}(X) = \Theta(n)$ we would like to say that the large deviation probability is exponentially small in n. We will be able to prove this using the exponential Markov inequality if we can bound the moment generating function $\mathbb{E}e^{tX}$ for some $t > 0$. With Z_t denoting the partition function with weights $\boldsymbol{\lambda} e^t$ and Z denoting the partition function with weights $\boldsymbol{\lambda}$, we can compute

$$\begin{aligned}
\log \mathbb{E}e^{tX} &= \log Z_t - \log Z \\
&= \sum_{\Gamma} (e^{t|\Gamma|} - 1) \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v
\end{aligned}$$

when t is sufficiently small that the Kotecký–Preiss condition holds, and so

$$\begin{aligned}
\log \mathbb{E}e^{t(X - \mathbb{E}X)} &= \sum_{\Gamma} (e^{t|\Gamma|} - 1 - t|\Gamma|) \phi(H(\Gamma)) \prod_{v \in \Gamma} \boldsymbol{\lambda}_v \\
&\leq t^2 C n
\end{aligned}$$

for some absolute constant $C = C(b)$. Now applying Markov's inequality we have

$$\Pr(X - \mathbb{E}X \geq \varepsilon n) \leq \frac{\mathbb{E}e^{t(X - \mathbb{E}X)}}{e^{t\varepsilon n}}$$

$$\leq \exp(t^2 Cn - t\varepsilon n)$$
$$\leq \exp\left(-\frac{\varepsilon^2}{4C} n\right),$$

which is our desired exponentially small upper bound. Here we have chosen $t = \varepsilon/(2C)$, which allows for cluster expansion convergence if ε is small enough.

3.6 Polymer models and cluster expansion

An important framework in which to apply the cluster expansion is that of abstract polymer models. This setting is not so far from that of the multivariate hard-core model on a graph, but with an added notion of 'size' that makes polymer models very useful in geometric settings. Gruber and Kunz defined 'subset polymer models' in which polymers were subgraphs of a lattice like $\mathbb{Z}^d$ [5]; Kotecky and Preiss defined a fully abstract notion of a polymer model and gave a very convenient convergence criteria for the cluster expansion [8].

An abstract polymer model consists of a finite or infinite set of polymers $\mathcal{C}$; a (complex-valued) weight function w_γ for each $\gamma \in \mathcal{C}$; and a symmetric, reflexive incompatibility relation '$\nsim$'. We write $\gamma' \nsim \gamma$ if γ and γ' are incompatible and write $\gamma \sim \gamma'$ if they are compatible.

Let Ω be the collection of all sets of mutually compatible polymers from $\mathcal{C}$ (including the empty set). Then the polymer model partition function is

$$Z = \sum_{X \in \Omega} \prod_{\gamma \in X} w_\gamma \,. \tag{5}$$

If the weights are non-negative then there is a natural probability measure on sets of mutually compatible polymers given by

$$\mu(X) = \frac{\prod_{\gamma \in X} w_\gamma}{Z} \,. \tag{6}$$

Example 1. *The canonical example of a polymer model is the multivariate hard-core model on a graph G with fugacities $\lambda_v, v \in V$. We let $\mathcal{C} = V$ and say two vertices u, v are incompatible if $d_G(u, v) \leq 1$. Sets of pairwise compatible polymers are exactly the independent sets of G, that is $X = \mathcal{I}(G)$. Assigning weights functions λ_v to each polymer v, we see that the polymer model partition function equals the multivariate hard-core partition function.*

Example 2. *Consider the ferromagnetic Ising model at inverse temperature β with external field λ on a graph G. That is,*

$$Z_G(\beta, \lambda) = \sum_{\sigma \in \{\pm 1\}^V} \lambda^{N(\sigma,+)} e^{\beta M(G,\sigma)} \,.$$

Now let polymers be connected induced subgraphs of G, with $w_\gamma = \lambda^{-|\gamma|} e^{-\beta|\partial_E \gamma|}$, where $|\gamma|$ is the number of vertices in the polymer γ and $\partial_E \gamma$ is the edge boundary of the subgraph induced by γ. Then we have

$$Z_G(\beta, \lambda) = \lambda^{|V|} e^{\beta|E|} Z \,. \tag{7}$$

If $\lambda > 1$ we expect to see more $+$ spins in a typical configuration and so the polymers represent deviations from the all $+$ ground state.

When the polymer weights w_γ are small we might expect few polymers in a typical configuration X, with weak correlations between the polymers. The cluster expansion gives us a way to measure this, by expanding $\log Z$ around the empty configuration.

As the example of the polymer model representation of the Ising model shows, a polymer will often come with a natural notion of 'size'; in this case, the size might be the number of vertices in the subgraph. The Kotecký–Preiss condition will provide a way to use the size of polymers in balancing the weight functions and the number of polymers a given polymer is incompatible with (in the hard-core setting, the activities and the degrees of vertices).

The following gives the Kotecký–Preiss condition for convergence of cluster expansion for abstract polymer models.

Theorem 3.7 ([8])**.** *Consider a polymer model defined by $(\mathcal{C}, w, \nsim)$. Suppose that there exist functions $a : \mathcal{C} \to [0, \infty)$, $b : \mathcal{C} \to [0, \infty)$ so that for every $\gamma \in \mathcal{C}$,*

$$\sum_{\gamma' \nsim \gamma} |w_{\gamma'}| e^{a(\gamma')+b(\gamma')} \leq a(\gamma)\,. \tag{8}$$

Then for every $\gamma \in \mathcal{C}$ and all $t \geq 0$,

$$\sum_{\substack{\Gamma \nsim \gamma \\ b(\Gamma) \geq t}} \left| \phi(H(\Gamma)) \prod_{\gamma' \in \Gamma} w_{\gamma'} \right| \leq a(\gamma) e^{-bt}\,, \tag{9}$$

where $b(\Gamma) = \sum_{\gamma' \in \Gamma} b(\gamma')$.

Note that if $\mathcal{C}$ is a finite set (and so the partition function Z is finite), then (9) implies convergence of the cluster expansion for $\log Z$ by summing over all $\gamma \in \mathcal{C}$.

In many applications of Theorem 3.7 we will choose the functions a and b to incorporate the 'size' of a polymer (denote it by $|\gamma|$); often we will take $a(\gamma) = a|\gamma|$ and $b(\gamma) = b|\gamma|$ for some constants $a, b > 0$,

Example 3. *Consider the Ising model polymer model as above and suppose the host graph G has maximum degree Δ. For a polymer γ let $a(\gamma) = |\gamma|$, where $|\gamma|$ denotes the number of vertices of the induced subgraph that defines γ. For convenience take $b(\gamma) = 0$, but it is straightforward to extend the analysis below to the case $b(\gamma) = b|\gamma|$ for some $b \geq 0$.*

We need to show that for every γ,

$$\sum_{\gamma' \nsim \gamma} \lambda^{-|\gamma'|} e^{-\beta |\partial_E \gamma'|} e^{|\gamma'|} \leq |\gamma|$$

The number of induced subgraphs of size k containing a given vertex v in a graph of maximum degree Δ is at most $\frac{(e\Delta)^{k-1}}{2}$. In particular, this means that the number of γ' of size k that are incompatible with γ is at most $|\gamma|(\Delta+1)\frac{(e\Delta)^{k-1}}{2}$. Since $\beta \geq 0$ we can bound $w_{\gamma'} \leq \lambda^{-|\gamma'|}$. Then we can bound

$$\sum_{\gamma' \nsim \gamma} \lambda^{-|\gamma'|} e^{-\beta |\partial_E \gamma'|} e^{|\gamma'|} \leq \sum_{k \geq 1} |\gamma|(\Delta+1)\frac{(e\Delta)^{k-1}}{2} e^k \lambda^{-k}$$

$$\leq \frac{|\gamma|}{2} \sum_{k \geq 1} \left(\frac{e^2 \Delta}{\lambda} \right)^k$$
$$\leq |\gamma|$$

if $\lambda \geq 3e^2\Delta/2$, and so the Kotecký–Preiss holds when λ is large enough as a function of Δ. Of course the bound is not tight – there are many points at which we made rough estimates, but we do not expect the condition to be tight itself and so we are usually ok with bounds like this.

3.7 Example: ferromagnetic Potts model on expander graphs

Consider the q-color ferromagnetic Potts model on a graph G and suppose we want to model defects from the all 'red' ground state. Define polymers to be connected induced subgraphs of G with vertices of the subgraph colored by the remaining $q-1$ non-red colors (each different coloring of the same subgraph yields a different polymer). Two polymers γ and γ' are incompatible if their union is connected. The weight of a polymer is

$$w_\gamma = e^{-\beta|\partial_e \gamma| - \beta|E_b(\gamma)|}$$

where $\partial_e \gamma$ is the set of edges from γ to γ^c and $E_b(\gamma)$ are the bichromatic edges of γ. Then we have

$$Z_G(q, \beta) = e^{\beta|E(G)|} \cdot Z$$

where $Z_G(q, \beta)$ is the Potts model partition function and Z is the polymer model partition function. Notice that $e^{\beta|E(G)|}$ is the weight of the ground state (all red configuration), and so Z captures contributions to $Z_G(q, \beta)$ from deviations from the ground state (the empty polymer configuration corresponds to the ground state – no defects). Of course we haven't really gained anything from this representation – the polymer model includes all the configurations that are dominated by blue or by green, etc, while we wanted to capture deviations from the red ground state. We will see below that we can address (in some settings) this by restricting polymers to be 'small'.

We will work in the setting of *expander graphs* (see the excellent survey [6]). The topic of expander graphs is too big to go into here, but expander graphs are extremely useful in mathematics, computer science, and information theory in:

- Derandomizing randomized algorithms
- Building error correcting codes
- Constructing pseudorandom graphs

Random walks mix extremely fast on expander graphs. There are many different constructions of expander graphs including those based on randomness (random regular graphs for example), group theory, number theory, and iterative constructions (the zig-zag product).

There are several different but related definitions of expander graphs: there are spectral expanders, vertex expanders, edge expanders. We will specialize here to Δ-regular graphs to make the connections clearer.

For a vertex subset S, let $\partial_e S$ denote the set of edges with one endpoint in S and one endpoint in S^c.

Definition 3.8. *An n-vertex Δ-regular graph G is an α-edge-expander if for all $S \subset V(G), |S| \leq n/2$, $|\partial_e S| \geq \alpha |S|$.*

Definition 3.9. *An n-vertex Δ-regular graph G is an α-vertex-expander if for all $S \subset V(G), |S| \leq n/2$, $|N(S)| \geq \alpha |S|$.*

Let $A(G)$ be the adjacency matrix of an n-vertex graph G and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of A sorted in decreasing order. Recall that for a Δ-regular graph, $\lambda_1 = \Delta$.

Definition 3.10. *An n-vertex Δ-regular graph G is an ε-spectral-expander if $\lambda_2(G) \leq (1 - \varepsilon)\Delta$*

Later we will use the notion of a bipartite α-expander.

Definition 3.11. *A Δ-regular bipartite graph G with bipartition (L, R) is a bipartite α-expander if for all $S \subset L, |S| \leq |L|/2$, $|N(S)| \geq (1 + \alpha)|S|$ (and likewise for $S \subset R$).*

Returning to the low-temperature Potts model, we will make two assumptions on G:

1. G has maximum degree Δ
2. G is an α-edge-expander for some $\alpha > 0$

For instance, the random Δ-regular graph satisfies these conditions whp.

For large β, we expect configurations to be dominated by one of the q-colors – we expect to see sparse, disordered deviations from one of the q monochromatic ground states. We will control these deviations via polymer models and the cluster expansion.

How do we ensure that our polymer model only captures small deviations from the red ground state and not, for example, small deviations from the green ground state? And what can we do about configurations for which there is no clearly dominant color? We will do this in two steps. In the first we argue that we can divide the state space of the Potts model into q disjoint parts (each associated to a dominant color) so that the complement of these parts has exponentially small probability. We will see later that this is in fact a proof of slow mixing for the Glauber dynamics. In the second step we will show that a polymer model with a convergent cluster expansion can approximate each one of these q parts.

In the polymer model, we will insist that $|\gamma| \leq n/2$ for all polymers γ. Here $|\gamma|$ denotes the number of vertices of γ. This ensures, for example, that the monochromatic green configuration is not represented in the red polymer model. Let Z denote the partition function of this polymer model (note that by symmetry Z does not depend on the ground state color).

We start by understanding the separation of the q ground states. By using the expansion properties of G we can show that $q \cdot e^{\beta |E(G)|} \cdot Z$ approximates $Z_G(q, \beta)$ to within e^{-n} relative error.

Lemma 3.12 ([7]). *Let G be a n-vertex, α expander of maximum degree at most Δ. For $\beta \geq \frac{4+2\log(q\Delta)}{\alpha}$,*

$$(1-e^{-n})q \cdot e^{\beta|E(G)|}Z \leq Z_G(q,\beta) \leq (1+e^{-n})q \cdot e^{\beta|E(G)|}Z\,.$$

Proof. Proving this takes two steps. The first is to show that when β is large Potts configurations in which no color has a majority have exponentially small relative weight. For $r \in [q]$, let $Z_r(\beta) = \sum_{\sigma:|\sigma^{-1}(r)|>n/2} e^{\beta M(G,\sigma)}$. Then for $\beta \geq 2\log(eq)/\alpha$, we will show that

$$\sum_{r\in[q]} Z_r(\beta) = qZ_r(\beta) \leq Z_G(q,\beta) \leq (1+e^{-n})qZ_r(\beta)\,. \tag{10}$$

The lower bound is immediate since configurations can have at most one majority color. The upper bound is a simple consequence of expansion: when there is no majority there must be many bichromatic edges, and these are penalized heavily for large β. In particular, there must be at least $\frac{n\alpha}{2}$ bichromatic edges, giving a penalty to each configuration of $e^{-n\alpha\beta/2}$ relative to one of the ground state configurations. There are at most q^n configurations with no majority, and taking $\beta \geq$, we have $q^n e^{-n\alpha\beta/2} \leq e^{-n}$. The approximation given by (10) allows us to partition configurations into $q+1$ subsets; one for each color plus an addition error class (no majority) that we can neglect.

The next step is to show that the color r polymer model partition function Z (after scaling by $e^{\beta|E(G)|}$) approximates $Z_r(\beta)$ up to an exponentially small relative error.

$$(1-e^{-n})e^{\beta|E(G)|}Z \leq Z_r(\beta) \leq e^{\beta|E(G)|}Z\,.$$

Every configuration with a majority of vertices colored r is captured by the polymer model since fewer than $n/2$ vertices receive a non-r color; this gives the upper bound. To prove the lower bound we show that configurations in which all non-r connected components are of size at most $n/2$ but which do not have a majority r have small total weight; this also follows from an expansion argument: such configurations must have at least $\alpha n/2$ bichromatic edges and so their total weight is at most $q^n e^{-\beta n/2} \leq e^{-n}$ when $\beta \geq 2\log(eq)/\alpha$, relative to the empty polymer configuration. □

The restriction on polymer sizes will also allow us to show that for β large enough as a functions of q, Δ, α, the Kotecký–Preiss condition is satisfied.

As above we need to bound the number of polymers of a given size and their weight. The number of polymers of size k incompatible with γ is at most $\frac{|\gamma|}{2}(\Delta+1)(q-1)^k(e\Delta)^{k-1}$.

The expansion condition gives us an upper bound on the weight of a polymer:

$$w_\gamma \leq e^{-\alpha\beta|\gamma|}$$

since $|\partial_e\gamma| \geq \alpha|\gamma|$; here we have crucially used the upper bound on the size of γ.

With these two bounds we can verify the Kotecký–Preiss condition. Let $a(\gamma) = b(\gamma) = |\gamma|$. For a given polymer γ,

$$\sum_{\gamma'\nsim\gamma} w_{\gamma'} e^{a(\gamma')+b(\gamma')} \leq \sum_{k\geq 1} (e\Delta(q-1))^k \Delta|\gamma| e^{-\alpha\beta k} e^{2k}$$

$$\leq |\gamma|\Delta \sum_{k\geq 1} \exp\left[k\left(3+\log\Delta+\log(q-1)-\alpha\beta\right)\right]$$

which is at most $a(\gamma) = |\gamma|$ if $\beta \geq \frac{4+2\log(q\Delta)}{\alpha}$.

References

[1] A. Barvinok. *Combinatorics and complexity of partition functions*, volume 9. Springer, 2016.

[2] S. Cannon and W. Perkins. Counting independent sets in unbalanced bipartite graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1456–1466. SIAM, 2020.

[3] R. Dobrushin. Estimates of semi-invariants for the Ising model at low temperatures. *Translations of the American Mathematical Society-Series 2*, 177:59–82, 1996.

[4] J. Groeneveld. Two theorems on classical many-particle systems. *Phys. Letters*, 3, 1962.

[5] C. Gruber and H. Kunz. General properties of polymer systems. *Communications in Mathematical Physics*, 22(2):133–161, 1971.

[6] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[7] M. Jenssen, P. Keevash, and W. Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020.

[8] R. Kotecký and D. Preiss. Cluster expansion for abstract polymer models. *Communications in Mathematical Physics*, 103(3):491–498, 1986.

[9] J. Mayer and M. Mayer. *Statistical Mechanics*. John Wiley, 1940.

[10] J. E. Mayer and E. Montroll. Molecular distribution. *The Journal of Chemical Physics*, 9(1):2–16, 1941.

[11] F. Mousset, A. Noever, K. Panagiotou, W. Samotij, et al. On the probability of nonexistence in binomial subsets. *The Annals of Probability*, 48(1):493–525, 2020.

[12] O. Penrose. Convergence of fugacity expansions for fluids and lattice gases. *Journal of Mathematical Physics*, 4(10):1312–1320, 1963.

[13] D. Ruelle. Correlation functions of classical gases. *Annals of Physics*, 25:109–120, 1963.

[14] 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.

[15] J. B. Shearer. On a problem of Spencer. *Combinatorica*, 5(3):241–245, 1985.