CS294 Partition Functions **Fall 2023**

Lecture Note 1

Instructor: Alistair Sinclair

1.1 Introduction

We begin by introducing a framework for studying the type of computational problems covered in this class.

1.1.1 Counting Problems

Our starting point is computational problems of the following type:

1. Given a boolean formula φ in conjunctive normal form (i.e., a constraint satisfaction problem), how many satisfying assignments does ϕ have?

2. Given a graph $G = (V, E)$, how many perfect matchings does G have? When $G = (V_1, V_2, E)$ is bipartite, with $|V_1| = |V_2|$, this is equivalent to computing the *permanent* of the 0-1 adjacency matrix A_G of G.

3. Given a matroid $\mathcal{M}$, how many bases (i.e., maximal independent sets) does $\mathcal{M}$ have?

These problems, and many others like them, are natural generalizations of the more familiar decision problems upon which much of classical complexity theory is based. The decision version of problem #1 is the famous SAT problem, which asks whether φ has a satisfying assignment; this problem is NP-complete. Similarly, the decision version of #2 asks whether G has a perfect matching; this can be solved in polynomial time via a not-so-obvious algorithm (e.g., reduction to network flow). And the decision version of #3 asks whether a matroid has a basis, which is trivially always true. Observe that, viewed in this way, counting problems are strictly more general than decision problems: determining a count is a more challenging task than figuring out if the count is non-zero.

Counting problems have a rich and noble history, dating back to the Greeks, and are arguably more widely studied in mathematics than decision problems. We begin by presenting a more general framework that captures a host of additional important questions and actually sheds more light on the vanilla examples above.

1.1.2 Generating Functions

A *generating function* for an integer sequence $\{a_k\}_{k=0}^{\infty}$ is the formal power series $Z(x) := \sum_i a_k x^k$. In our setting there will be only finitely many non-zero a_i, all of which are non-negative, in which case $Z(x) = a_0 + a_1 x + \cdots + a_d x^d$ becomes a polynomial of degree d with positive integer coefficients.

A canonical example is the so-called *matching polynomial* of a graph $G = (V, E)$, given by

$$Z_G(x) := \sum_{k=0}^{d} m_k x^k, \tag{1.1}$$

where m_k is the number of matchings (independent sets of edges) in G consisting of exactly k edges. Note that always $m_0 = 1$ (the empty matching), and that the degree d can be taken to be $\lfloor \frac{|V|}{2} \rfloor$ (since no matching can have more than this number of edges). Moreover, if $|V| = 2d$ is even then the top coefficient m_d is equal to the number of perfect matchings in G (problem #2 above).

Generating functions for other combinatorial structures may be defined in exactly the same way. For example, for problems #1 and #3 above we can define:

$$\begin{aligned} Z_\varphi(x) &:= \sum_k a_k x^k \quad \text{where } a_k \text{ is the number of assignments that satisfy exactly } k \text{ clauses} \\ Z_{\mathcal{M}}(x) &:= \sum_k b_k x^k \quad \text{where } b_k \text{ is the number of independent sets with exactly } k \text{ elements} \end{aligned}$$

Once again, the top coefficients encode the solutions to the original problems, but the polynomials carry a lot of additional information about the respective combinatorial structures.

1.1.3 Statistical Physics: Spin Systems and Partition Functions

A further generalization is motivated by statistical physics. A *spin system* on an undirected graph $G = (V, E)$ has an allowed set of *configurations*, which are assignments $\sigma : V \to [q]$ of spin values from a finite set $[q] = \{1, 2, \ldots, q\}$ to vertices. The energy of configuration σ is given by a Hamiltonian

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

where f, g are arbitrary real-valued functions. The function f, which determines nearest-neighbor interactions, is usually referred to as an *edge potential*, and the function g as a *vertex potential*. (Edge potentials are assumed to be symmetric, i.e., $f(i, j) = f(j, i)$ for all i, j.) We also allow $f(i, j) = \infty$, which effectively forbids certain pairs of neighboring spins in the configuration; this is referred to as a "hard" constraint.

The probability of finding the system in configuration σ is given by the *Gibbs (or Boltzmann) distribution*

$$\pi(\sigma) := \frac{1}{Z(\beta)} \exp(-\beta H(\sigma)), \tag{1.2}$$

where β is a non-negative parameter that is inversely proportional to the temperature, and the normalizing factor is the *partition function*

$$Z(\beta) := \sum_{\sigma} \exp(-\beta H(\sigma)). \tag{1.3}$$

If, as is often the case, H is integer-valued and non-negative (as can be easily ensured by adding a constant to the energy function) then by the change of variable $\lambda = \exp(-\beta)$ we can rewrite Z as a polynomial

$$Z(\lambda) = \sum_k a_k \lambda^k, \tag{1.4}$$

where a_k is the number of configurations with energy exactly k and $\lambda \in (0, 1]$.

The form of the Gibbs distribution and partition function in (1.2) and (1.3) is quite general, and applies to any statistical mechanical system in equilibrium: higher weight is assigned to lower energy configurations,

and this bias increases with the parameter β (decreases with temperature): when β is small the Gibbs distribution is almost uniform over σ; when it is large, most of the weight is on ground state (minimum energy) configurations. The exponential form of the distribution arises from the principle that it should *maximize entropy* subject to the constraint that the mean energy is fixed. The temperature-dependent scaling factor β arises from the quantitative relationship between entropy, energy and temperature.

We now give a few canonical examples of spin systems that will recur throughout the course.

1. Ferromagnetic Ising model. The first significant spin system to be studied (and still among the most important) was the Ising model of ferromagnetism [Isi25, Len20]. Configurations of the Ising model on graph $G = (V, E)$ are assignments $\sigma : V \to \{\pm 1\}$ of atomic spins to vertices. (Conventionally the spin values $\{\pm 1\}$ or $\{\uparrow, \downarrow\}$ are used here.) The edge and vertex potentials are

$$f(i,j) = 1 - \delta_{i,j} = \begin{cases} 1 & \text{if } i \neq j; \\ 0 & \text{otherwise} \end{cases} \qquad g(i) = \begin{cases} -h & \text{if } i = +1; \\ 0 & \text{otherwise,} \end{cases}$$

where $h \geq 0$ is an external magnetic field. The energy function is therefore

$$H(\sigma) := \#\mathrm{Dis}(\sigma) - h\#\mathrm{Pos}(\sigma), \tag{1.5}$$

where $\#\mathrm{Dis}(\sigma)$ is the number of edges in σ with unequal spins on their endpoints, and $\#\mathrm{Pos}(\sigma)$ is the number of $+1$ spins in σ. Note that the lowest energy (highest weight) configurations are those with many neighboring spins aligned (few disagreements) and many spins aligned with the external field. The partition function is

$$Z_G(\beta, h) := \sum_{\sigma} \exp(-\beta H(\sigma)), \tag{1.6}$$

which by the change of variables $\lambda := \exp(-\beta)$, $\mu := \exp(\beta h)$ can be written as the polynomial

$$Z_G(\lambda, \mu) = \sum_{\sigma} \lambda^{\#\mathrm{Dis}(\sigma)} \mu^{\#\mathrm{Pos}(\sigma)} = \sum_{j,k} a_{j,k} \lambda^j \mu^k, \tag{1.7}$$

where $a_{j,k}$ is the number of configurations σ with k "+1" spins and j "misaligned" edges. Note that $\lambda \in (0, 1]$ and $\mu \geq 1$; the case $\mu = 1$ corresponds to zero external field ($h = 0$).

To get a more intuitive form for this polynomial, note that we can identify σ with the set $S = S(\sigma) := \{v \in V : \sigma(v) = +1\}$; the corresponding set of misaligned edges $E(S, \bar{S}) := \{uv \in E : \sigma(u) \neq \sigma(v)\}$ is then just the *cut* in G separating S from $\bar{S}$. Now we can rewrite (1.7) as

$$Z_G(\lambda, \mu) = \sum_{S \subseteq V} \lambda^{|E(S,\bar{S})|} \mu^{|S|}. \tag{1.8}$$

The Ising model partition function in (1.8) is a bivariate polynomial with non-negative combinatorial coefficients that generalizes the types of counting problems we have seen earlier. It is natural to study it as a function of either λ or μ, or both. For example, if we fix $\mu = 1$ (zero field) we get

$$Z_G(\lambda) = \sum_{j} c_j \lambda^j, \tag{1.9}$$

where c_j is the number of cuts $(S, \bar{S})$ in G with exactly j edges; this is the *cut generating polynomial* of G. In the ferromagnetic regime $\lambda \in (0, 1]$, this polynomial puts largest weight on *minimum* cuts.

Finally, we note that this model, as well as the other examples below, can be generalized by introducing separate parameters λ_e for each edge, and μ_v for each vertex (corresponding to introducing edge-dependent

and vertex-dependent weights into the Hamiltonian (1.5)). The partition function in (1.8) then becomes a function of all these parameters. In addition to being a natural extension of the model, this will sometimes be mathematically convenient.

2. Antiferromagnetic Ising model. As in the ferromagnetic Ising model, configurations are assignments $\sigma : V \to \{\pm 1\}$, but now the edge potential is $f(i,j) = \delta_{i,j}$ so that the energy function becomes

$$H(\sigma) := \#\mathrm{Agr}(\sigma) - h\#\mathrm{Pos}(\sigma) = |E| - \#\mathrm{Dis}(\sigma) - h\#\mathrm{Pos}(\sigma),$$

where $\#\mathrm{Agr}(\sigma)$ is the number of edges whose spins are aligned. Notice that low energy configurations are now those with *more* misaligned edges; i.e., in contrast to the ferrogmagnetic model, neighboring spins now prefer to be different. Setting $\lambda := \exp(\beta)$ and (as before) $\mu := \exp(-\beta h)$, the partition function becomes

$$Z_G(\lambda,\mu) = \sum_\sigma \lambda^{-|E|+\#\mathrm{Dis}(\sigma)} \mu^{\#\mathrm{Pos}(\sigma)} = \lambda^{-|E|} \sum_{j,k} a_{j,k} \lambda^j \mu^k,$$

which is the same as (1.7) except that now we are in the regime $\lambda \geq 1$ (i.e., favoring larger cuts). Note that the factor $\lambda^{-|E|}$ is a constant independent of σ and hence can be omitted.

3. Hard-core model. Here configurations are assignments $\sigma : V \to \{0,1\}$, with the hard constraint that no two "1" spins may be adjacent. Equivalently, a configuration σ corresponds to an *independent set* $S = \{v \in V : \sigma(v) = 1\}$ in G. This model represents an *excluded volume* constraint: if a site v is occupied, then none of its neighbors is allowed to be occupied. This is achieved by setting the edge potential $f(1,1) = \infty$ and $f(i,j) = 0$ for other values of i,j. We also set the vertex potential to $g(1) = 1$ and $g(0) = 0$. Thus the only configurations with positive weight in the Gibbs distribution correspond to independent sets S, and the corresponding Gibbs weight is $\exp(-\beta|S|)$.

Under the change of variable $\lambda := \exp(-\beta)$, the partition function can be written as

$$Z_G(\lambda) = \sum_S \lambda^{|S|} = \sum_k a_k \lambda^k,$$

where a_k is the number of independent sets of size k in G. As written we are in the regime $\lambda \in (0,1]$ (favoring small independent sets); however, the model makes equal sense in the complementary regime $\lambda \geq 1$.

4. Potts model. This is a generalization of the Ising model to a larger number of spin values [Pot52]. Configurations are assignments $\sigma : V \to [q] := \{1,2,\ldots,q\}$ of spins, and the edge potential is again given by $f(1,j) = 1 - \delta_{i,j}$ in the ferromagnetic case, and $f(i,j) = \delta_{i,j}$ in the antiferromagnetic case. (We could also introduce an external field of the form $-h_i \sum \delta_{\sigma(v),i}$ for each spin $i \in [q]$, but for simplicity we omit it.) The case $q = 2$ is equivalent to the Ising model (with zero field).

Under the transformation $\lambda := \exp(\pm\beta)$ (according to whether we are in the ferromagnetic or antiferromagnetic case), the partition function becomes

$$Z_G(\lambda) = \sum_{S_1,\ldots,S_q} \lambda^{|E(S_1,\ldots,S_q)|} = \sum_k a_k \lambda^k,$$

where the sum is over all *q-way cuts* in G (i.e., partitions of V into q parts $S_1,\ldots,S_q$), $E(S_1,\ldots,S_q)$ denotes the set of edges in the cut, and a_k is the number of q-way cuts with exactly k edges, and we have omitted a prefactor of $\lambda^{-|E|}$ in the antiferromagnetic case. As with the Ising model, the ferromagnetic and antiferromagnetic cases correspond to the regimes $\lambda \in (0,1]$ and $\lambda \geq 1$, respectively.

Configurations of the antiferromagnetic Potts model ($\lambda \geq 1$) correspond to vertex colorings with q colors, with larger weight assigned to colorings with few *monochromatic* edges; in the limit as $\lambda \to \infty$, the Gibbs distribution is just the uniform distribution on *proper* q-colorings of G, a widely studied object in combinatorics.

5. Monomer-dimer model. This is not strictly a spin system, as assignments are to edges rather than vertices. Configurations are assignments $\sigma : E \to \{0,1\}$, with the constraint that no two adjacent edges may receive a 1 spin. Thus, a configuration σ corresponds to a *matching* $M = \{e \in E : \sigma(e) = 1\}$ in G. This model represents (among other things) a mixture of monatomic and diatomic molecules (monomers and dimers), which occupy one and two adjacent sites respectively. The energy is $H(\sigma) = |M|$, and analogous to #3 above the partition function becomes

$$Z_G(\lambda) := \sum_k m_k \lambda^k,$$

where $\lambda = \exp(-\beta)$ and m_k is the number of matchings of size k in G. This is just the matching polynomial, as in (1.1). As for the hard-core model, this model also makes sense for $\lambda \geq 1$.

1.1.4 Markov Random Fields

A *Markov random field (MRF)* on a graph $G = (V, E)$ is a collection of random variables $(X_v)_{v \in V}$ associated with the vertices of G whose joint distribution satisfies the (global) *Markov property*: for any subsets $A, B, S \subseteq V$ such that S *separates* A and B (i.e., removing S from G disconnects A from B), X_A is independent of X_B conditional on X_S. (Here X_A denotes the collection $\{X_v : v \in A\}$.) MRFs (also known as *undirected graphical models*) are widely used in machine learning to represent dependencies between sets of random variables.

It should be clear from their definition that spin systems are MRFs, whose configurations are collections of values of the random variables, $\{X_v = \sigma(v)\}_{v \in V}$. [**Exercise:** Why?] Surprisingly, it turns out that, under a fairly mild condition, any MRF can be represented as a (slightly generalized) spin system! This fact is known as the *Hammersley-Clifford Theorem*, though Hammersley and Clifford never formally published their proof. We sketch below a proof due to Besag [Bes74].

Theorem 1.1 (Hammersley-Clifford Theorem)**.** *Let $\mathcal{M}$ be a Markov random field on $G = (V, E)$ with discrete variables $(X_v)_{v \in V}$. If $\mathcal{M}$ assigns non-zero probability to all possible configurations $x = \{x_v : v \in V\}$, then $\mathcal{M}$ is equivalent to a generalized spin system on G with Hamiltonian $H(x) = \sum_C \Psi_C(x_C)$, where the sum ranges only over* cliques *C in G and Ψ_C is a potential associated with C.*

Proof. For ease of notation we'll assume that the value 0 is in the support of each random variable X_v. Also, we'll write $x = (x_v)$ for a set of values of the X_v, and $\pi(x)$ for their joint probability. Define the real-valued function

$$g(x) := \log \frac{\pi(x)}{\pi(\mathbf{0})},$$

where $\mathbf{0}$ denotes the valuation $x_v = 0 \; \forall v$; so $g(\mathbf{0}) = 0$. Now note that for all $x \neq \mathbf{0}$ we can uniquely write

$$g(x) = \sum_{U \subseteq V} \Psi_U(x_U) \tag{1.10}$$

for suitable real-valued functions Ψ_U depending only on x_U and satisfying $\Psi_U(x_U) = 0$ unless $x_U > \mathbf{0}$. [**Exercise:** Show how to do this, starting by defining $\Psi_{\{v\}}$ for singletons $\{v\}$ and building up from there.] Our goal is to show that necessarily the function Ψ_U is identically zero unless the vertices U form a clique in G.

To this end, let u, v be two vertices in U that are not connected by an edge. For any valuation x, let x' denote the same valuation with x'_u set to 0. Then by the spatial Markov property the difference

$$g(x) - g(x') = \log \frac{\pi(x)}{\pi(x')} = \log \frac{\pi(x_u, x_{V \setminus \{u\}})}{\pi(0, x_{V \setminus \{u\}})} = \log \frac{\pi(x_u \mid x_{V \setminus \{u\}})}{\pi(0 \mid x_{V \setminus \{u\}})}$$

is independent of x_v for all x. On the other hand, from (1.10) we can write

$$g(x) - g(x') = \sum_{U \ni u} \Psi_U(x_U).$$

Now, setting $x_w = 0$ for all $w \notin \{u, v\}$ and fixing x_u to an arbitrary non-zero value tells us that the expression

$$\Psi_{\{u\}}(x_u) + \Psi_{\{u,v\}}(x_u, x_v)$$

is independent of x_v, which implies that $\Psi_{\{u,v\}}$ must be identically zero. (Why?) Proceeding in similar fashion through subsets $U \ni \{u, v\}$ of increasing sizes, and setting $x_w = 0$ for all $w \notin U$, we can see that Ψ_U is identically zero for all such U. [**Exercise:** Carry through this argument explicitly for the case $U = \{u, v, y\}$, and hence convince yourself that it works for all $U \ni \{u, v\}$.] Since $\{u, v\}$ was an arbitrary pair of non-adjacent vertices, this ensures that Ψ_U is non-zero only when U is a clique. □

Note that the above proof makes use of the requirement that $\pi(x) \neq 0$ for all possible valuations x. Indeed, the requirement that $\mathcal{M}$ has full support is real: the theorem fails without this assumption (which, in our earlier language, means in the presence of hard constraints) [Mou74]. As far as I know, the question of how to formulate a useful analogous theorem in the presence of hard constraints (possibly with additional conditions) is still open.

Note also that any assignment of (finite) clique potentials Ψ_U for cliques U in (1.10) gives rise to a valid MRF.

Because of the above correspondence, MRFs are in fact often described as spin systems, by specifying as parameters the (vertex, edge, or more generally clique) potentials; unlike the uniform models above, these potentials are typically not uniform over vertices/edges/cliques. Denoting the collection of parameter values by Θ, the partition function and the Gibbs distribution then both become functions of Θ. The Gibbs distribution determines the likelihood of observing any configuration $\{\sigma(v)\}$, under the given values of the parameters Θ.

The spin systems we saw above (generalized to allow vertex- and edge-dependent potentials) are special cases where only the 1- and 2-vertex cliques (vertices and edges) are required to represent the MRF. We will focus on these in the course as they are conceptually simpler while also being general enough to illustrate essentially all the important behaviors.

1.2 Applications of Partition Functions

This course is concerned with computing partition functions of the kind introduced in the previous sections. We conclude by briefly discussing why this is an important topic.

1. Inherent interest. As mentioned earlier, counting combinatorial structures is an important subfield of mathematics with a long history. More recently, the lenses of computational complexity and of efficient approximation algorithms have shed new light on this topic.

2. Miscellaneous applications. Many computational problems can be reduced to counting problems, often in surprising ways. To give just two examples: computing the *effective resistance* of an electrical network can be reduced to counting its spanning trees; and computing the *permanent* of a 0-1 matrix is equivalent to counting perfect matchings in an associated bipartite graph.

3. Statistical physics. The partition function of a spin system encodes essentially full information about the model. As we have seen, the coefficients of this polynomial represent the energy profile of the configurations. Moreover, the partition function is the normalizing factor in the Gibbs distribution, which is usually

not known a priori. (Typically all we know is the Hamiltonian for any given configuration σ; computing the actual probability $\pi(\sigma)$ requires normalizing by Z.) Less obviously, the expectations of many important random variables (or *observables*) under the Gibbs distribution can be expressed as the ratio of two partition functions, or equivalently as a derivative of $\log Z$. To illustrate this point, for a given spin system consider the *mean energy* $\mathrm{E}_\pi(H)$, where the expectation is over the Gibbs distribution π. Following (1.2), we may write

$$\mathrm{E}_\pi(H) = \frac{1}{Z}\sum_\sigma H(\sigma)\exp(-\beta H(\sigma)) = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \log Z}{\partial \beta}. \tag{1.11}$$

Now if we write the partition function as a polynomial $Z(\lambda)$ as in (1.4), where $\lambda = \exp(-\beta)$, then $\partial\lambda = -\lambda\partial\beta$, so (1.11) becomes

$$\mathrm{E}_\pi(H) = \lambda\frac{Z'}{Z},$$

where Z' denotes $\frac{\partial Z}{\partial \lambda}$ and is a combinatorial partition function very similar to Z itself. The expectations of other observables (i.e., simple functions of the potentials) can be calculated similarly.

Exercise: The *heat capacity* or *energy fluctuation* can be defined as the variance $\mathrm{E}((H - \mathrm{E}(H))^2)$. Verify that this quantity is equal to $\frac{Z''}{Z} - (\frac{Z'}{Z})^2$, where Z'' denotes $\frac{\partial^2 Z}{\partial \lambda^2}$.

4. Machine learning. As indicated earlier, an MRF or undirected graphical model is frequently used to represent relationships between random variables. For a given choice of model parameters Θ, the Gibbs distribution determines the probability of any configuration (realization of the random variables) σ. Thus the partition function $Z(\Theta)$, as the normalizing factor in this distribution, is required in order to compute actual probabilities. (The Gibbs weight $\exp(-\beta H(\sigma))$ of any configuration can be computed easily from the potentials, but these weights are unnormalized.) These probabilities are important, e.g., in evaluating the model, as they can be used to compare the probabilities that certain sets of test data are generated by the model under different choices of the parameters Θ: typically we would want to choose Θ to maximize this probability.

References

[Bes74] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.

[Isi25] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31:253–258, 1925.

[Len20] W. Lenz. Beiträge zum Verständnis der magnetischen Eigenschaften in festen Körpern. *Physikalische Zeitschrift*, 21:613–615, 1920.

[Mou74] J. Moussouris. Gibbs and Markov random systems with constraints. *Journal of Statistical Physics*, 10:11–33, 1974.

[Pot52] R.B. Potts. Some generalized order-disorder transformations. *Mathematical Proceedings*, 48:106–109, 1952.