Counting and Sampling **Fall 2017**

Lecture 20: Log Concavity in Counting and Stability Perserverss

Lecturer: Shayan Oveis Gharan *December 8th*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In the last lecture we discussed applications of log concavity of $\mathbb{H}$-stable polynomials in optimization. Recall that a polynomial $p(z_1, \ldots, z_n)$ is $\mathbb{H}$ -stable if $p(z_1, \ldots, z_n) \neq 0$ whenever $\Im z_i > 0$ for all i.

In the last lecture we used the following lemma.

Lemma 20.1. *Any $\mathbb{H}$ stable polynomial $p \in \mathbb{R}_+[z_1, \dot{;} z_n]$ with nonnegative coefficients is log concave in its variables.*

In this lecture and the next one we will see applications of the above lemma in optimization and counting.

20.1 Log Concavity and Counting

Let $p(z_1, \ldots, z_n) = \sum_S a_S z^S$ be a homogeneous multilinear $\mathbb{H}$-stable polynomial with non-negative coefficients. The main result of this section is to design an algorithm to approximate the sum of all coefficients of p.

We can treat p as a probability distribution μ, where for any $S \subseteq [n]$, $\mu(S) \propto a_S$. These probability distributions are called *strongly Rayleigh* and they are extensively studied in [**?**] and employed to design approximation algorithms [**?**, **?**]. For example, for any graph G the uniform distribution over all spanning trees of G is a strongly Rayleigh distribution.

For a probability distribution $\mu : 2^{[n]} \to \mathbb{R}_+$, the marginal probability of an element $i \in [n]$ is defined as

$$\mu_i := \mathbb{P}_{S \sim \mu}[i \in S].$$

Observe that if μ is the distribution corresponding to a p then

$$\mu_i = \frac{\partial_{z_i} p}{p} = \partial_{z_i} \log p,$$

where we use ∂_{z_i} to denote the partial differential operator $\frac{\partial}{\partial z_i}$. Recall that the entropy of μ is defined as

$$\mathcal{H}(\mu) = \sum_S \mu(S) \log \mu(S).$$

In this section we prove the following theorem which gives an approximation to the entropy μ given the marginal probability of the underlying elements. In the next section we see applications of this theorem in counting.

Theorem 20.2. *For any homogeneous multilinear $\mathbb{H}$-stable polynomial with non-negative coefficients p and the corresponding strongly Rayleigh distribution μ, we have*

$$\max\left\{\sum_{i=1}^n \mu_i \log \frac{1}{\mu_i}, \sum_{i=1}^n (1-\mu_i) \log \frac{1}{1-\mu_i}\right\} \leq \mathcal{H}(\mu) \leq \sum_{i=1}^n \mu_i \log \frac{1}{\mu_i} + (1-\mu_i) \log \frac{1}{1-\mu_i}.$$

Observe that the right inequality simply follows by the sub-additivity of entropy and it holds for *any* probability distribution μ. In fact the distribution with maximum possible entropy is an independent set of Bernoulli random variables where the i-th one has prior μ_i. So, the main non-trivial part is the LHS. Furthermore, observe that the LHS is within a factor 2 of the RHS. So, the above theorem asserts that the strongly Rayleigh distributions are not very far from independent Bernoulli distributions.

Proof. The proof simply follows by log-concavity and the Jensen's inequality. First, let us recall Jensen's inequality: Let μ be a probability distribution on $2^{[n]}$, $x : 2^{[n]} \to \mathbb{R}^n$ be a function that associates a vector to every set $S \subseteq [n]$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a concave function, then

$$\sum_S \mu(S) f(z_S) \leq f\left(\sum_S \mu(S) z_S\right). \tag{20.1}$$

To use the above inequality we let μ be the given strongly Rayleigh distribution. So, we just need to define $z_S = (z_{S1}, \ldots, z_{Sn})$ for any set S. For any set $S \subseteq [n]$ let

$$z_{Si} = \begin{cases} \frac{1}{\mu_i} & \text{if } i \in S \\ 0 & \text{otherwise.} \end{cases}$$

Also, for a vector $z = (z_1, \ldots, z_n)$ define

$$f(z) = \log p(z) = \log \sum_S \mu(S) z^S.$$

Firstly, we need to evaluate each side of (20.1). For a set $T \subseteq [n]$ observe that

$$\begin{aligned} f(x_T) &= \log \sum_S \mu(S) (z_T)^S \\ &= \log \sum_S \mu(S) \prod_{i \in S} \frac{1}{\mu_i} \mathbb{I}[i \in T] \\ &= \log \mu(S) \prod_{i \in S} \frac{1}{\mu_i}. \end{aligned}$$

Therefore, we can rewrite the LHS of (20.1) as follows:

$$\begin{aligned} \sum_S \mu(S) f(z_S) &= \sum_S \mu(S) \log \mu(S) + \sum_S \mu(S) \sum_{i \in S} \log \frac{1}{\mu_i} \\ &= -\mathcal{H}(\mu) + \sum_{i=1}^n \left(\log \frac{1}{\mu_i}\right) \sum_{S: i \in S} \mu(S) \\ &= \sum_i \mu_i \log \frac{1}{\mu_i} - \mathcal{H}(\mu). \end{aligned}$$

To finish the proof of the theorem it is enough to show that the RHS (20.1) is 0. First, let us evaluate $\sum_S \mu(S) z_S$. For a coordinate i

$$\sum_S \mu(S) z_{Si} = \sum_{S: i \in S} \mu(S) \cdot \frac{1}{\mu_i} = 1.$$

Therefore, the RHS of (20.1) is exactly

$$f(1, 1, \ldots, 1) = \log \sum_S \mu(S) = \log 1 = 0$$

as desired. □

20.2 Counting the number of Bases of a Matroid

Note that the proof that we discussed in the last lecture works out for any log concave polynomial with nonnegative coefficients. In this section we use that theorem to count the number of bases of a matroid.

Definition 20.3 (Matroid)**.** *Let E be a set of elements and $\mathcal{I} \subseteq 2^E$ be a family of subsets of elements called em independent sets. We say $\mathcal{M}(E, \mathcal{I})$ is a matroid if it satisfies the following properties:*

- *For any $S \in \mathcal{I}$ and any $T \subseteq S$, we have $T \in \mathcal{I}$.*
- *For any $S, T \in \mathcal{I}$ such that $|S| < |T|$ there exists an element $e \in T \setminus S$ such that $S \cup \{e\} \in \mathcal{I}$.*

We say a set $B \in \mathcal{I}$ is a *base* of a matroid if it is a maximal independent set. The *rank* of a matroid is the number of elements of its bases.

Matroids in the first place are studied to generalize the notion of linear independence of a set vectors. They are extensively studied in the field of combinatorial optimization. For example, we can maximize/minimize any linear function over a matroid. The convex hull of all bases of a matroid called *matroid polytope*, $P(\mathcal{M})$ is well understood and we can test if any given point x belongs to $P(\mathcal{M})$ in polynomial time.

One of the fundamental open problems in the field of counting is counting the number of bases of a general matroid. Note that in some special cases we can do this: For example if $\mathcal{M}$ is a *graphic* matroid where for a given graph $G = (V, E)$, E is the set of elements of $\mathcal{M}$ and a set $S \subseteq E$ is independent if does not have any cycle. In this case the bases correspond to the set of spanning trees of G, and it turns out that we can exactly count the number of spanning tree of any given graph by computing the determinant of the Laplacian matrix of G.

It has been conjectured that the natural Metropolis on the rule bases of a matroid, a.k.a., the bases exchange random walk, mixes in polynomial-time for any matroid $\mathcal{M}$. This conjecture was proved by Feder and Mihail [**?**] for a family of matroid called Balanced matroid which satisfy a form of negative correlation. But to this date we are not aware of any other classes of matroid for which we can count the number of bases even to an exponential factor.

In this section we prove the following theorem.

Theorem 20.4 (AOV17)**.** *There is a deterministic polynomial time algorithm that for any matroid $\mathcal{M}(E, \mathcal{I})$ given oracle access to $\mathcal{I}$ gives a e^r approximation to the number of bases of $\mathcal{M}$ where r is the rank of $\mathcal{M}$.*

For a matroid $\mathcal{M}$ with elements $z_1, \ldots, z_n$, one can associate a polynomial

$$p_{\mathcal{M}}(z_1, \ldots, z_n) = \sum_{S \in \mathcal{I}} z^S$$

At the heart of the above theorem we prove that for any matroid $\mathcal{M}$, $p_{\mathcal{M}}(z)$ is log-concave. The proof of this fact builds on a long line of work in algebraic geometry on the proof of the log-concavity of the sequence of the coefficients of $p_{\mathcal{M}}(z, \ldots, z)$.

Assuming this fact let us prove the above theorem. Consider the uniform distribution μ over all bases of $\mathcal{M}$. By Theorem 20.2 we have that

$$\sum_{i=1}^{n} \mu_i \log \frac{1}{\mu_i} \leq \mathcal{H}(\mu) \leq \sum_{i=1}^{n} \mu_i \log \frac{1}{\mu_i} + \sum_{i=1}^{n} (1 - \mu_i) \log \frac{1}{1 - \mu_i}.$$

Firstly, observe that $\mathcal{H}(\mu) = \log|\mathcal{B}|$ where $\mathcal{B} \subseteq \mathcal{I}$ is the set of bases of $\mathcal{M}$. Therefore, $e^{\sum_{i=1}^n \mu_i \log \frac{1}{\mu_i}}$ gives an approximation of the number of bases up to a multiplicative error $e^{\sum_{i=1}^n (1-\mu_i)\log\frac{1}{1-\mu_i}}$. So, let us upper bound the latter quantity. Observe that

$$e^{\sum_{i=1}^n (1-\mu_i)\log\frac{1}{1-\mu_i}} = \prod_{i=1}^n (1-\mu_i)^{-(1-\mu_i)} \le \prod_{i=1}^n e^{\mu_i} = e^r. \tag{20.2}$$

The inequality follows by the fact that for any $0 \le x \le 1$, $(1-x)^{-(1-x)} \le e^x$ and the last identity follows by linearity of expectation.

Therefore, to prove Theorem 20.4 all we need to do is to compute $\mu_1, \ldots, \mu_n$. But the natural way to compute the marginal probabilities is to compute the partition function. So, it seems that we haven't made any progress. We claim that the following convex program is enough for our purpose:

$$\begin{aligned} \max \quad & \sum_{i=1}^n x_i \log\frac{1}{x_i} + \sum_{i=1}^n (1-x_i)\log\frac{1}{1-x_i} \\ \text{s.t.} \quad & x \in P(\mathcal{M}). \end{aligned} \tag{20.3}$$

Note that the objective function of this convex program is concave so we can solve this program in polynomial time.

Observe that the ideal marginal probabilities $\mu_1, \ldots, \mu_n$ is a feasible solution to the above program. Therefore the optimum is at least

$$\sum_{i=1}^n \mu_i \log\frac{1}{\mu_i} + (1-\mu_i)\log\frac{1}{1-\mu_i} \ge \mathcal{H}(\mu).$$

So, to prove Theorem 20.4 it is enough to show that the optimum of the above program is at most $\mathcal{H}(\mu) + O(r)$.

We first explain the idea: Let x be the optimum solution of the program (20.3). We show that there exists a distribution μ' over the bases of $\mathcal{M}$ such that the polynomial $\sum_{S\in\mathcal{B}} \mu'(S) z^S$ is log-concave and that for all i, $\mu'_i = x_i$. Then, by another application of Theorem 20.2 we have that

$$\mathcal{H}(\mu') \le \sum_{i=1}^n x_i \log\frac{1}{x_i}.$$

On the other hand, since the uniform distribution over all bases has the largest possible entropy we have that $\mathcal{H}(\mu) \ge \mathcal{H}(\mu')$. This will complete the proof of Theorem 20.4.

So, to finish the proof it is enough to show that there exists a distribution μ' over the bases such that for all i, $\mu'_i = x_i$. The *symmetric exclusion of a polynomial $p(z_1, \ldots, z_n)$ with respect to $\lambda_1, \ldots, \lambda_n$ is defined as the polynomial $p(\lambda_1 z_1, \ldots, \lambda_n z_n)$. We use the following fact:*

Fact 20.5. *For any homogeneous log-concave polynomial $p(z_1, \ldots, z_n)$ and any set of $\lambda_1, \ldots, \lambda_n \ge 0$ the symmetric exclusion of p with respect to $\lambda_1, \ldots, \lambda_n$ is log-concave.*

The proof of the above fact simply follows by writing the Hessian of the polynomial and comparing it to the Hessian of the polynomial where $\lambda_i = 1$ for all i.

So, to finish the proof it is enough to show that there exists a symmetric exclusion of p such that the marginal probability of each element i is x_i. This can be done by studying the following maximum entropy *convex program:*

$$\begin{aligned} \max \quad & \sum_{S \in \mathcal{B}} p(S) \log \frac{1}{p(S)} \\ \text{s.t.} \quad & \sum_{S \in \mathcal{B}: i \in S} p(S) = x_i \qquad \forall i \\ & p(S) \geq 0 \qquad \forall S. \end{aligned} \tag{20.4}$$

The following convex program is obviously feasible because $x \in P(\mathcal{M})$. *Furthermore by writing the Lagrangian dual one can show that there exists* $\lambda_1, \ldots, \lambda_n$ *such that for all* $S \in \mathcal{B}$ *we have*

$$p(S) \propto \prod_{i \in S} \lambda_i.$$

The latter fact has nothing to do about matroid. It holds for any homogeneous set system. This completes the proof of Theorem 20.4.