Counting and Sampling	**Winter 2020**

Lecture 11: Log Concavity and Convex Programming

Lecturer: Shayan Oveis Gharan	*Feb 21st*

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

In this lecture we see applications of (completely) log-concave polynomials in convex optimization and approximate counting.

11.1 Log Concavity and Counting

We say a probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ is (completely) log-concave if the generating polynomial g_μ is (completely) log-concave.

Recall that the *entropy* of μ is defined as

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

The following theorem is the main result of this section:

Theorem 11.1. $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ *be a homogeneous log-concave probability distribution. For* $1 \leq i \leq n$, *let* $\mu_i := \mathbb{P}_{S \sim \mu} [i \in S]$. *Then,*

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

Observe that the right inequality simply follows by the sub-additivity of entropy and it holds for *any* probability distribution μ: Say $X_1, \ldots, X_n$ be random variables where X_i is the indicator random variable of i. Then,

$$
\begin{aligned}
\mathcal{H}(\mu) &= \mathcal{H}(X_1) + \mathcal{H}(X_2|X_1) + \cdots + \mathcal{H}(X_n|X_1, \ldots, X_{n-1}) \\
&\leq \mathcal{H}(X_1) + \mathcal{H}(X_2) + \cdots + \mathcal{H}(X_n) = \sum_{i=1}^n \mu \log \frac{1}{\mu_i} + (1 - \mu_i) \log \frac{1}{1 - \mu_i}.
\end{aligned}
$$

So, we get an equality above if μ is an independent Bernoulli distribution where the success probability of the i-th Bernoulli is μ_i.

So, the main non-trivial part is the proof of the left inequality.

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

$$\sum_S \mu(S) f(x_S) \leq f\left(\sum_S \mu(S) x_S \right). \tag{11.1}$$

To use the above inequality we let μ be the given log-concave probability distribution. So, we just need to define $x_S = (x_{S1}, \ldots, x_{Sn})$ for any set S. For any set $S \subseteq [n]$ let

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

Also, for a vector $x = (x_1, \ldots, x_n)$ define

$$
f(x) = \log g_\mu(x) = \log \sum_S \mu(S) x^S.
$$

Firstly, we need to evaluate each side of (11.1). For a set $T \subseteq [n]$ observe that

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

Therefore, we can rewrite the LHS of (11.1) as follows:

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

To finish the proof of the theorem it is enough to show that the RHS (11.1) is 0. First, let us evaluate $\sum_S \mu(S) x_S$. For a coordinate i

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

Therefore, the RHS of (11.1) is exactly

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

as desired. $\square$

Note that if μ is the uniform distribution over the bases of a matroid $M = ([n], \mathcal{I})$ we can also show

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

The idea is to take work with the dual of M. Given a matroid $M = ([n], \mathcal{I})$, the dual, M^*, is a matroid on the same ground set $[n]$ where B is a base of M^* iff $\overline{B}$ is a base of M. So, if M has rank r, then M^* has rank $n - r$. Furthermore, we can write

$$
g_{M^*} = z_1 \ldots z_n g_M(1/z_1, \ldots, 1/z_n).
$$

It is well-know that for any matroid M, its dual is also a matroid, so g_{M^*} is also a completely log-concave polynomial. It then follows that the marginal of distribution μ^* uniform over the bases of M^* are $1 - \mu_1, \ldots, 1 - \mu_n$ but $\mathcal{H}(\mu^*) = \mathcal{H}(\mu)$.

Using these two facts, it follows that $\mathcal{H}(\mu)$ can be approximated within a multiplicative factor 2 by $\sum_{i=1}^{n} \mu_i \log \frac{1}{\mu_i} + \sum_{i=1}^{n} (1 - \mu_i) \log \frac{1}{1-\mu_i}$. This shows that uniform distributions over bases of a matroid are "too far" from independent distributions.

11.2   An Application in Combinatorics

Theorem 11.2. *Let $M = ([n], \mathcal{I})$ be a matroid of rank r. Suppose that M has k disjoint bases $B_1, \ldots, B_k$, i.e., for any i, j, $B_i \cap B_j = \emptyset$. Then, M has at least k^r many bases.*

Proof. Let μ be the uniform distribution over the bases of M. Then, $\mathcal{H}(\mu) = \log \#B$. So, we just need to prove $\mathcal{H}(\mu) \geq r \log k$.

Consider the vector $x = \mathbf{1}_{B_1} + \cdots + \mathbf{1}_{B_k}$. We claim that there is a probability distribution ν with marginals x such that g_ν is log-concave. Since $x \in \text{Newt}(g_\mu)$, by Theorem 5.2, there is an external field $(\lambda_1, \ldots, \lambda_n) \geq 0$ such that the marginals of $\lambda * \mu$ is equal to x. Since g_μ is a log-concave polynomials, and log-concave polynomials are closed under external field, $g_{\lambda*\mu}$ is also log-concave. So, we can let $\nu = \lambda * \mu$.

To be precise, if x is not in the interior of Newton polytope of g_μ we can find a sequence of external field vectors $\lambda^1, \ldots,$ such that the marginals of $\mu * \lambda^t$ converges to x as $t \to \infty$. In such a case, we let ν be $\lim_{t \to \infty} \lambda^t * \mu$. By the closure of log-concave polynomials under taking limits, g_ν is log-concave.

Since g_ν is log-concave, by Theorem 11.1,

$$\mathcal{H}(\nu) \geq \sum_{i=1}^{n} \nu_i \log \frac{1}{\nu_i} = \sum_{i \in B_1 \cup \cdots \cup B_k} \frac{1}{k} \log k = r \log k.$$

Finally, the theorem follows from the fact that μ has the largest possible entropy among all distributions supported over the bases of M. So, $\mathcal{H}(\mu) \geq \mathcal{H}(\nu) \geq r \log k$ as desired. $\square$

As an immediate corollary, a complete graph G with n vertices has at least $n/2 - 1$ disjoint spanning trees. So, the above theorem implies that a complete graph has at least $(n/2 - 1)^{n-1}$ many spanning trees.

11.3   Counting the number of Bases of a Matroid

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 exactly: For example if M is a *graphic* matroid where for a given graph $G = (V, E)$, E is the set of elements of M and a set $S \subseteq E$ is independent if it does not have any cycle, then the bases correspond to the set of spanning trees of G, and we can exactly count the number of spanning trees of any given graph by computing the determinant of the Laplacian matrix of G.

In the next lectures, we will see how to use Markov chains to approximately sample (and count) bases of a matroid.

In this section, we give a deterministic algorithm for this task:

Theorem 11.3 (AOV17). *There is a deterministic polynomial time algorithm that for any matroid $M(E, \mathcal{I})$ given oracle access to $\mathcal{I}$ gives a e^r approximation to the number of bases of M where r is the rank of M.*

Let μ be the uniform distribution over all bases of M. Firstly, observe that $\mathcal{H}(\mu) = \log \#B$. 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

$$e^{\sum_{i=1}^{n}(1-\mu_i)\log \frac{1}{1-\mu_i}} = \prod_{i=1}^{n}(1-\mu_i)^{-(1-\mu_i)} \leq \prod_{i=1}^{n} e^{\mu_i} = e^r. \tag{11.2}$$

The inequality follows by the fact that for any

$$(1-x)^{-(1-x)} \leq e^x, \forall 0 \leq x \leq 1 \tag{11.3}$$

and the last identity follows by linearity of expectation.

Therefore, to prove Theorem 11.3 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 the optimum value of the following convex program gives an additive $O(r)$ approximation to $\mathcal{H}(\mu)$:

$$\max \quad \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i} + \sum_{i=1}^{n}(1-\alpha_i)\log \frac{1}{1-\alpha_i} \tag{11.4}$$
$$\text{s.t.} \quad \alpha \in P(M).$$

Here, $P(M) = \text{Newt}(g_M)$ is the matroid base polytope, i.e., it is the convex hull of the indicator vectors of all bases of M. Note that the objective function of this convex program is concave so we can solve this program in polynomial time. Furthermore, as we discussed

Let α be an optimum solution. Observe that the ideal marginal probabilities $\mu_1, \ldots, \mu_n$ is a feasible solution to the above program. Therefore,

$$\sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i} + (1-\alpha_i)\log \frac{1}{1-\alpha_i} \geq \sum_{i=1}^{n} \mu_i \log \frac{1}{\mu_i} + (1-\mu_i)\log \frac{1}{1-\mu_i} \geq \mathcal{H}(\mu). \tag{11.5}$$

So, to prove Theorem 11.3 it is enough to show that the optimum of the above program is at most $\mathcal{H}(\mu) + O(r)$.

Since x is in $\text{Newt}(g_\mu)$, similar to the proof of Theorem 11.2, there a log-concave distribution $\nu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ with marginals x (again by Theorem 5.2, there is an external field $\lambda = (\lambda_1, \ldots, \lambda_n)$ such that $g_{\mu * \lambda}$ has marginals x).

Since g_ν is log-concave and has marginals x, by Theorem 11.1,

$$\mathcal{H}(\nu) \geq \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i}.$$

Since the uniform distribution over all bases has the largest possible entropy we also get $\mathcal{H}(\mu) \geq \mathcal{H}(\nu) \geq \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i}$. Finally, we can write

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

where the last inequality follows by (11.3). This together with (11.5) finishes the proof of Theorem 11.3.

11.4   Counting Common Bases of Two Matroids

Theorem 11.4. *Given two matroids $M_1 = ([n], \mathcal{I}_1), M_2 = ([n], \mathcal{I}_2)$ by independence oracle of rank r. There is a deterministic polynomial time algorithm that counts the number common bases to M_1, M_2 up to a multiplicative factor of c^r, for some constant $c > 1$.*

The algorithm is a natural extension of the algorithm in the previous section:

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i} + \sum_{i=1}^{n} (1 - \alpha_i) \log \frac{1}{1 - \alpha_i} \\
\text{s.t.} \quad & \alpha \in P(M_1) \\
& \alpha \in P(M_2).
\end{aligned}
\tag{11.6}
$$

We will see that the optimum solution of the above program gives an additive $O(r)$ approximation to $\log |B(M_1) \cap B(M_2)|$. As a consequence, it follows that if there are k disjoint bases in the intersection of M_1, M_2, then $|B(M_1) \cap B(M_2)| \geq (k/c)^r$ for some constant $c > 1$ (independent of k, r). This can be seen as a generalization of the van-der-Waerden conjecture to matroid intersection.

Next, we discuss main ideas to prove Theorem 11.4. Let x be an optimum solution of the convex program 11.6. Let μ be the uniform distribution over the bases in $B(M_1) \cap B(M_2)$. Then, obviously $\mu_1, \dots, \mu_n$ is a feasible solution to the above program. So, $\mathcal{H}(\mu) \leq \sum_{i=1}^{n} \mathcal{H}(\alpha_i)$, where $\mathcal{H}(\alpha_i) = \alpha_i \log \frac{1}{\alpha_i} + (1 - \alpha_i) \log \frac{1}{1 - \alpha_i}$. So, the main non-trivial part is to show that

$$
\mathcal{H}(\mu) \geq \sum_{i=1}^{n} \mathcal{H}(\alpha_i) - O(r).
\tag{11.7}
$$

The following theorem, which can be seen as a generalization of Gurvits' machinery to completely log-concave polynomials is the main technical part of the proof:

Theorem 11.5. *Let $p \in \mathbb{R}[y_1, \dots, y_n, z_1, \dots, z_n]$ be a multilinear completely log-concave polynomial. Then, for any $\alpha \in [0, 1]^n$,*

$$
\prod_{i=1}^{n} (\partial_{y_i} + \partial_{z_i}) p|_{y=z=0} \geq \Phi(\alpha) \inf_{y, z > 0} \frac{p(y, z)}{y^{\alpha} z^{1-\alpha}},
$$

where $\Phi(\alpha)$ is defined as follows:

$$
\Phi(\alpha) = (\alpha/e^2)^{\alpha} = \prod_{i=1}^{n} \left(\frac{\alpha_i}{e^2} \right)^{\alpha_i}.
$$

We do not discuss the proof of the above theorem here. The proof is a generalization of Theorem 6.1. We can follow an inductive proof similar to Gurvits' theorem. The main non-trivial part is the base case where p is a bi-variate polynomial. There the proof mainly follows from Problem 4 of HW2, i.e., that a multilinear bivariate polynomial $a + by + cz + dyz$ is completely log-concave iff $2bc \geq ad$.

To use the above theorem, let g_{M_1}, g_{M_2} be the generating polynomial of matroids M_1, M_2 respectively. Recall that M_2^* is the *dual* of the matroid M_2.

Having defined the above quantities, we let

$$
p(y_1, \dots, y_n, z_1, \dots, z_n) = g_{M_1}(y_1, \dots, y_n) g_{M_2^*}(z_1, \dots, z_n).
$$

Crucially observe that

$$|B(M_1) \cap B(M_2)| = \prod_{i=1}^{n} (\partial_{y_i} + \partial_{z_i}) p|_{y=z=0}.$$

This is because $\prod_{i=1}^{n} (\partial_{y_i} + \partial_{z_i})$ zeros out any monomial of p where the sum of the degree of y_i, z_i is not 1. Now, to invoke Theorem 11.5 we need to show that p is completely log-concave. This is a non-trivial fact, and it follows from the following lemma:

Lemma 11.6. *For any two homogeneous completely log-concave polynomials $p, q \in \mathbb{R}[z_1, \ldots, z_n]$, $p \cdot q$ is also completely log-concave.*

Following the above discussion, we have

$$
\begin{aligned}
\mathcal{H}(\mu) = \log |B(M_1) \cap B(M_2)| \;\; & \geq \;\; \log \Phi(\alpha) \inf_{y,z>0} \frac{g_{M_1}(y) g_{M_2^*}(z)}{y^\alpha z^{1-\alpha}} \\
& \geq \;\; \Phi(\alpha) + 2 \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i} \\
& = \;\; \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i} - 2 \sum_{i=1}^{n} \alpha_i \\
& = \;\; \sum_{i=1}^{n} \mathcal{H}(\alpha_i) - 2r - \sum_{i=1}^{n} (1 - \alpha_i) \log \frac{1}{1 - \alpha_i} \\
& \geq \;\; \sum_{i=1}^{n} \mathcal{H}(\alpha_i) - 3r.
\end{aligned}
$$

where the first inequality follows by the following fact, and the last inequality follows by (11.3). This proves (11.7) as desired.

Fact 11.7. *For any $\alpha \in \mathrm{Newt}(g_{M_1})$,*

$$\log \inf_{y>0} \frac{g_{M_1}(y)}{y^\alpha} \geq \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i}.$$

Proof. Let μ_1 be the uniform distribution over bases of M_1. Recall that by Lecture 5, $\lambda = \mathrm{argmin}_{y>0} \frac{g_{\mu_1}(y)}{y^\alpha}$ gives a distribution $\lambda * \mu_1$ with generating polynomial $g_{\lambda * \mu_1}(y_1, \ldots, y_n) = g_{\mu_1}(\lambda_1 y_1, \ldots, \lambda_n y_n)$ with marginals $\alpha_1, \ldots, \alpha_n$. But since $g_{\lambda * \mu_1}$ is completely log-concave by Theorem 11.1 $\mathcal{H}(\lambda * \mu_1) \geq \sum_{i=1}^{n} \alpha_i \log \frac{1}{\alpha_i}$. But on the other hand, $\square$

11.5   Future Directions

One natural question is if one can prove an analogue of Theorem 11.4 for matchings in general graphs. The following long-standing open problem was posed by Lovasz and Plummer:

Conjecture 11.8. *Given a k-regular k-edge connected (general) graph $G = (V, E)$ with n vertices. Prove that G has at least $\Omega(k)^n$ many perfect matchings.*

An analogous algorithmic question is if there is a deterministic algorithm c^n-approximation algorithm for counting the number of matchings in general graphs. The questions is if we can prove some form (complete) log-concavity that can be used to approximately count matchings in general graph.