CSE 525: Randomized Algorithms **Spring 2023**

Lecture 8: Strongly Rayleigh Distributions

Lecturer: Shayan Oveis Gharan *04/20/2023*

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

The main goal of this lecture is to define a systematic way to study negative correlation and negative association. We will study a recently developed theory of negative dependence using the field of geometry of polynomials.

Definition 8.1 (d-homogeneous)**.** *A polynomial $p \in \mathbb{R}[z_1, \ldots, z_n]$ is d-homogeneous if $p(\lambda z_1, \ldots, \lambda z_n) = \lambda^d p(z_1, \ldots, z_n)$ for any $\lambda \in \mathbb{R}$.*

A multivariate polynomial $p \in \mathbb{R}[z_1, \ldots, z_n]$ is $\mathcal{H}$*-stable* (or stable for short) if $p(z_1, \ldots, z_n) \neq 0$ whenever $(z_1, \ldots, z_n) \in \mathcal{H}^n$ where

$$\mathcal{H} = \{c \in \mathbb{C} : \Im(c) > 0\}$$

is the upper-half of the n-dimensional complex plane. We say p is *real* stable if all coefficients of p are real. Unless otherwise specified, all polynomials that we work with in this course have real coefficients.

Fact 8.2. *A univariate polynomial $p \in \mathbb{R}[t]$ is real rooted iff it is real stable.*

This simply follows from the fact that the roots of p come in conjugate pairs. So, if p has a root t with $\Im(t) < 0$, we have $\bar{t}$ is also a root with $\Im(\bar{t}) > 0$.

So, it turns out that this definition is the right generalization of the real rootedness to multivariate polynomials.

Let us discuss several examples of real stable polynomials

Linear Functions: A linear polynomial $p = a_1 z_1 + \cdots + a_n z_n$ is real stable iff $a_1, \ldots, a_n \geq 0$. To see this note that if all z_i have positive imaginary value then any positive combination also has a positive imaginary value and thus is non-zero.

Elementary Symmetric Polynomial: For any n and any k the elementary symmetric polynomial $e_k(z_1, \ldots, z_n) = \sum_{S \in \binom{n}{k}} \prod_{i \in S} z_i$ is real stable. I leave this as an exercise.

Non-example The polynomial $z_1^2 + z_2^2$ is not real stable; for example let $z_1 = e^{\pi i/4}$ and $z_2 = e^{3\pi i/4}$. Similarly, the polynomial $z_1 z_2 + z_3 z_4$ is not real stable (recall that this was an example of a generating polynomial that satisfies negative lattice condition but it is not negatively correlated).

8.1 Closure Properties

The following theorem of Hurwitz simply following from the fact that the roots of a polynomial are continuous functions of its coefficients.

Lemma 8.3 (Hurwitz [**Hur95**]). *Let $\{p_k\}_{k\geq 0}$ be a sequence of Ω-stable polynomials over $z_1, \ldots, z_n$ for a connected and open set $\Omega \subseteq \mathbb{C}^n$ that uniformly converge to p over compact subsets of Ω. Then, p is Ω-stable.*

One of the fundamental properties of real stable polynomials are their closure properties. A real stable polynomial is closed under the following operations:

Product If p, q are real stable so is $p \cdot q$.

Symmetrization If $p(z_1, z_2, \ldots, z_n)$ is real stable then so is $p(z_1, z_1, z_3, \ldots, z_n)$.

Specialization If $p(z_1, \ldots, z_n)$ is real stable then so is $p(a, z_2, \ldots, z_n)$ for any $a \in \mathbb{R}$. First, note that for any integer k, $p_k = p(a + i2^{-k}, z_2, \ldots, z_n)$ is a stable polynomial (note that p_k may have complex coefficients). Therefore by Hurwitz theorem 8.3, the limit of $\{p_k\}_{k\geq 0}$ is a stable polynomial, so $p(a, z_2, \ldots, z_n)$ is stable.

External Field If $p(z_1, \ldots, z_n)$ is real stable then so is $q(z_1, \ldots, z_n) = p(\lambda_1 \cdot z_1, \ldots, \lambda_n \cdot z_n)$ for any positive vector $w \in \mathbb{R}^n_{\geq 0}$. If $q(z_1, \ldots, z_n)$ has a root $(z_1, \ldots, z_n) \in \mathcal{H}^n$ then $(\lambda_1 z_1, \ldots, \lambda_n z_n) \in \mathcal{H}^n$ is a root of p so p is not real stable.

Inversion If $p(z_1, \ldots, z_n)$ is real stable and degree of z_1 is d_1 then $p(-1/z_1, z_2, \ldots, z_n)z_1^{d_1}$ is real stable. This is because the map $z_1 \mapsto -1/z_1$ is a bijection between $\mathcal{H}$ and itself.

Differentiation If $p(z_1, \ldots, z_n)$ is real stable then so is $q = \partial p/\partial z_1$. This follows from Gauss-Lucas theorem. If $q(z_1, \ldots, z_n)$ is not real stable it has a root $(z_1^*, \ldots, z_n^*)$. Define $f(z_1) = p(z_1, z_2^*, \ldots, z_n^*)$. Then, $f'(z_1)$ has a root in $\mathcal{H}$. But the roots of $f'(z_1)$ are in the convex hull of the roots of $f(z_1)$ we get a contradiction because the complement of $\mathcal{H}$ is convex.

8.2 Strongly Rayleigh Distributions

In the previous lecture we observed that the generating polynomial of a family of n independent Bernoullis can be written as follows:

$$\prod_{i=1}^{n}(p_i z_i + 1 - p_i).$$

Note that this polynomial is real stable simply because it is a product of linear (real stable) polynomials.

Borcea Branden and Liggett [**BBL**] in an influential paper introduced a new class of probability distributions, namely Strongly Rayleigh distributions. These are all probability measures $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ such that g_μ is real stable.

One of the nice properties of SR distributions is that the closure properties of real stable polynomials nicely translate into closure properties of SR distributions.

Product Product of two SR distributions is SR.

Conditioning if μ is SR then for any $i \in [n]$, $\mu|i$ and $\mu|\bar{i}$ are SR. This because

$$g_{\mu|\bar{i}}(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n) =\propto g_\mu(z_1, \ldots, z_{i-1}, z_i = 0, z_{i+1}, \ldots, z_n), \text{and}$$

$$g_{\mu|\bar{i}}(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n) \propto \partial_{z_i} g_\mu(z_1, \ldots, z_n).$$

Projection Given a probability distribution μ, and a set $S \subseteq [n]$, the projection of μ onto S is the distribution the trace outs every variable not in S, or only looks at elements in S,

$$g_{\mu_S}(\{z_i\}_{i \in S}) = g_\mu(\{z_i = 1\}_{i \notin S}, \{z_j\}_{j \in S})$$

External Field Given any $\lambda \in \mathbb{R}^n_{\geq 0}$, suppose we put λ_i copies of i for all i, the resulting distribution is still SR since $g_\mu(\lambda_1 z_1, \ldots, \lambda_n z_n)$ is real stable.

Truncation If μ is SR then for any integer $k > 0$, the measure μ_k that is μ restricted to subsets of size k is SR, i.e., the measure where for for every $S \in \binom{n}{k}$,

$$\mu_k(S) = \frac{\mu(S)}{\sum_{T \in \binom{n}{k}} \mu(T)}.$$

One of the most important family of real-stable polynomials is the determinant polynomial.

Lemma 8.4. *Given PSD matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ and a symmetric matrix $B \in \mathbb{R}^{d \times d}$, the polynomial*

$$p(z) = \det\left(B + \sum_{i=1}^{n} z_i A_i\right)$$

is real stable.

We don't prove this lemma here, but instead explain two applications:

Uniform Spanning Trees Given a graph $G = (V, E)$ the generating polynomial of all spanning trees of G is SR, this is because

$$g_\mu(\{z_e\}_{e \in E}) = \sum_{T \text{ spanning tree}} z^T = \det(\frac{1}{n}\mathbf{1}\mathbf{1}^T + \sum_{e \in E} z_e L_e),$$

where L_e is the Laplacian of edge e is a PSD matrix. This identity is called the matrix tree theorem that we don't prove here.

k-DPPs Given n vectors $v_1, \ldots, v_n \in \mathbb{R}^d$, consider the distribution over $S \in \binom{n}{k}$ where the $\mathbb{P}[S] = \det(\sum_{i \in S} v_i v_i^T)$, a.k.a., the volume of the parallelepiped defined by vectors in S. This distribution has an abundance of applications in ML in addressing diversity and it is SR.

8.3 Negative Correlation Property of SR distributions

Theorem 8.5. *A multilinear polynomial $p \in \mathbb{R}[z_1, \ldots, z_n]$ is real stable iff for any i, j*

$$\partial_{z_i} p(x) \cdot \partial_{z_j} p(x) \geq p(x) \cdot \partial_{z_i} \partial_{z_j} p(x), \tag{8.1}$$

where $x \in \mathbb{R}^n$.

In the previous lecture, we observed that if μ is negatively correlated, then g_μ satisfies (8.1) for $x = \mathbf{1}$. A probability distribution is called *Rayleigh* if it satisfies (8.1) for all $x \geq 0$. The notion of Rayleigh distributions was first introduced by Wagner [**Wag06**] with the purpose of generalizing the Rayleigh monotonicity law for effective resistances in graphs (we will see this later in the course). If (8.1) holds for all $x \geq 0$, then it implies that all conditional distributions of μ are negatively correlated.

Proof. We only prove $\Rightarrow$: Assume p is real stable, and let $x \in \mathbb{R}^n$. First, note that for any $a > 0$ and $b \in \mathbb{R}$, by closure properties of real stable polynomials, $p(at+b, z_2, \ldots, z_n)$ is real stable. Then the bivariate restriction

$$g(s,t) = p(x_1, \ldots, x_{i-1}, s + x_i, \ldots, x_{j-1}, t + x_j, x_{j+1}, \ldots, x_n)$$

is real stable. Since p is multilinear, g is multilinear and it can be writen as

$$g(s,t) = a + bs + ct + dst,$$

where $a = p(x)$, $b = \partial_{z_i} p(x)$, $c = \partial_{z_j} p(x)$ and $d = \partial_{z_i} \partial_{z_j} p(x)$. So, to finish the proof it is enough to show that $bc \geq ad$. This is shown in the next lemma. Therefore, (8.1) $\square$

Lemma 8.6. *Let $p \in \mathbb{R}[z_1, z_2] = a + bz_1 + cz_2 + dz_1z_2$ be a real stable polynomial. Prove that $bc \geq ad$.*

Proof. If $d = 0$ we are done since the polynomial is degree 1 so b, c have the same sign. Otherwise we know that $p(t - c/d, t - b/d)$ is real rooted. But this polynomial is equal to $a - \frac{bc}{d} + dt^2$. Therefore the discriminant gives $0 \geq 4(a - \frac{bc}{d})d$, or equivalently $bc \geq ad$ proving the claim. $\square$

So, we immediately get any sum $\sum_{i \in S} X_i$ is strongly concentrated.

8.4 Negative Association Property of SR Distributions

First, notice that unlike, positive association, we cannot hope to prove that any two increasing functions are negatively associated. This is simply because any increasing function f is positively correlated with itself. So, instead we want the functions to be supported on disjoint coordinates:

Definition 8.7. *A probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ is negatively associated if any two increasing functions $f, g : 2^{[n]} \to \mathbb{R}_{\geq 0}$ that depend on disjoint set of coordinates are negatively correlated; namely if f is a function of $S \subseteq [n]$ and g is a function of $T \subseteq [n]$ such that $S \cap T = \emptyset$ then,*

$$\mathbb{E}[f]\,\mathbb{E}[g] \geq \mathbb{E}[fg].$$

Feder and Mihail [**FM92**] proved the following theorem:

Theorem 8.8. *Let S be a class of probability distributions satisfying:*

1) Each $\mu \in S$ is a measure on 2^E, where for some $E \subseteq \{1, 2, \ldots, n\}$;

2) S is closed under conditioning, i.e., for any $\mu \in S$, $\mu|i, \mu|\bar{i} \in S$.

3) For each $\mu \in S$ is pairwise negatively correlated.

4) Each $\mu \in S$ has a homogeneous generating polynomial.

Then all measures in S are negatively associated.

We don't prove this theorem here. Instead, we use to prove that Strongly Rayleigh distributions are negatively associated.

Theorem 8.9 (Borcea-Brändén-Liggett [**BBL08**]). *Any strongly Rayleigh distribution is negatively associated.*

Proof. We sketch the main idea: First we prove that homogeneous SR distributions are negatively associated. This is simply because they are pairwise negatively correlated, and closed under conditioning.

If the given distribution is not homogeneous, there is a generic way to homogenize it while preserving SR property by adding some dummy variables. Then we apply the argument from the previous paragraph. □

For a concrete example, say $f(T)$ is the indicator that the induced graph of a set $T \subseteq E$ on a set $S \subseteq V$ is connected, namely, $T \cap E(S)$ is connected. For any edge $e \notin E(S)$ let $g(T) = \mathbb{I}[e \in T]$. Then, f, g are negatively correlated.

To this date we know many other properties of SR distributions that we don't discuss here such as stochastic dominance, stochastic covering property, and concentration of Lipshitz functions.