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

7.1 Positive Association

Theorem 7.1 (The Four Functions Theorem). *Let $\alpha, \beta, \gamma, \delta : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be non-negative functions defined on subsets of $[n]$. If for any two subsets $A, B \subseteq [n]$ we have*

$$\alpha(A)\beta(B) \leq \gamma(A \cup B)\delta(A \cap B),$$

then, for every two families of subsets $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$, we have

$$\alpha(\mathcal{A})\beta(\mathcal{B}) \leq \gamma(\mathcal{A} \cup \mathcal{B})\delta(\mathcal{A} \cap \mathcal{B}),$$

where $\alpha(\mathcal{A}) = \sum_{A \in \mathcal{A}} \alpha(A)$.

The following theorem, known as the FKG inequality, is a direct consequence of the above theorem:

Definition 7.2 (Log-supermodular probability distributions). *We say a probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ is log-supermodular if for any $A, B \subseteq [n]$, we have*

$$\mu(A)\mu(B) \leq \mu(A \cup B)\mu(A \cap B).$$

This property is also known as Positive Lattice Condition.

For a concrete example, consider the family of Erdos-Reyni $G(n, p)$ random graphs. In such a case, for any set $F \subseteq \binom{n}{2}$ we have

$$\mu(F) = p^{|F|}(1-p)^{\binom{n}{2}-|F|} = (1-p)^{\binom{n}{2}}(\frac{p}{1-p})^{|F|}.$$

We claim that this distribution is log-supermodular. Cancelling out the normalizing constant $(1-p)^{\binom{n}{2}}$, we need to check for any two sets $A, B \subseteq E$, and $q = \frac{p}{1-p}$

$$q^{|A|}q^{|B|} \leq q^{|A \cup B|}q^{|A \cap B|}$$

But this holds simply because $|A| + |B| = |A \cap B| + |A \cup B|$.

Definition 7.3 (Increasing functions). *We say a function $f : 2^{[n]} \to \mathbb{R}_{geq0}$ is increasing if for any $A, B \subseteq [n]$ such that $A \subseteq B$ we have*

$$f(A) \leq f(B).$$

We say f is a decreasing function if the above inequality holds in the reverse direction.

For a concrete example, notice for any $i \in [n]$, $f(A) = \mathbb{I}[i \in A]$ is increasing and $f(A) = \mathbb{I}[i \notin A]$ is decreasing.

But, more generally, consider the domain $\binom{n}{2}$ of the set of all possible edges in a graph with n vertices. Then, for $A \subseteq \binom{n}{2}$,

$$f(A) = \mathbb{I}\left[A \text{ is connected}\right], f(A) = \mathbb{I}\left[A \text{ has a Hamiltonian cycle}\right]$$

are increasing, but,

$$f(A) = \mathbb{I}\left[G(V, A) \text{ is 3-colorable}\right]$$

is decreasing.

Theorem 7.4 (FKG Inequality). *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a log-supermodular probability distribution. Then, for any two increasing functions $f, g : 2^{[n]} \to \mathbb{R}_{\geq 0}$ we have*

$$\mathbb{P}\left[f\right]\mathbb{P}\left[g\right] \leq \mathbb{P}\left[fg\right],$$

i.e., μ is positively associated.

Proof. We use the Four functions theorem with

$$\alpha = \mu \cdot f, \beta = \mu \cdot g, \gamma = \mu \cdot f \cdot g, \text{ and } \delta = \mu.$$

We claim that these four functions satisfy the assumption of the Four functions theorem. In particular, for any $A, B \subseteq [n]$ by log-supermodularity of μ we have

$$\alpha(A)\beta(B) = \mu(A)f(A)\mu(B)g(B) \underset{\text{log-supermodularity}}{\leq} \mu(A \cup B)f(A)g(B)\mu(A \cap B)$$

$$\underset{f,g\text{are increasing}}{\leq} \mu(A \cup B)f(A \cup B)g(A \cup B)\mu(A \cap B) = \gamma(A \cup B)\delta(A \cap B).$$

Therefore, letting $\mathcal{A}, \mathcal{B} = [n]$, we conclude

$$\alpha([n])\beta([n]) = \left(\sum_{A \subset [n]} \mu(A)f(A)\right) \cdot \left(\sum_{B \subseteq [n]} \mu(B)g(B)\right) = \mathbb{E}\left[f\right]\mathbb{E}\left[g\right]$$

On the other hand,

$$\gamma([n])\delta([n]) = \left(\sum_{A \subseteq [n]} \mu(A)f(A)g(A)\right) \left(\sum_{B \subseteq [n]} \mu(B)\right) = \mathbb{E}\left[fg\right] \cdot 1$$

Putting them together proves the theorem.   □

Note that the above inequality also holds if both f, g are decreasing functions. In case that f is increasing and g is decreasing, the inequality holds just in the opposite direction.

Consequently, FKG theorem implies that any pair of elements i, j are positively correlated in a log-supermodular probability distribution,

$$\mathbb{P}\left[i\right]\mathbb{P}\left[j\right] \leq \mathbb{P}\left[i, j\right] \Leftrightarrow \mathbb{P}\left[i|j\right] \geq \mathbb{P}\left[i\right]$$

More interestingly, we can use it to prove the following fact about $G(n, p)$ graphs:

Fact 7.5. *For any $0 \leq p \leq 1$, let G be a random Erdos-Reyni graph with parameter p.*

$$\mathbb{P}\left[G \text{ has a Hamiltonian cycle}|G \text{ is 3-colorable}\right] \leq \mathbb{P}\left[G \text{ has a Hamiltonian cycle}\right].$$

7.2   Negatively Correlated Random Variables

We say that a collection $\{X_1, \ldots, X_n\}$ of random variables are negatively correlated if it holds that for any subset $S \subseteq [n]$:

$$\mathbb{E}\left[\prod_{i \in S} X_i\right] \leq \prod_{i \in S} \mathbb{E}\left[X_i\right].$$

Note that if $\{X_1, \ldots, X_n\}$ are independent, then this holds with equality.

Furthermore, we say $X_1, \ldots, X_n$ are pairwise negatively correlated if for all $1 \leq i < j \leq n$,

$$\mathbb{E}\left[X_i X_j\right] \leq \mathbb{E}\left[X_i\right] \mathbb{E}\left[X_j\right].$$

Theorem 7.6 (Chernoff for negatively correlated random variables). *Suppose $X_1, \ldots, X_n$ are negatively correlated Bernoulli random variables (instead of independent), then the conclusion of the multiplicative Chernoff bound still holds.*

Proof. To see this, note that the one place we used independence in the proof of the Chernoff bound is in the calculation: When $X = X_1 + \cdots + X_n$,

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t \sum_i X_i}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right].$$

The main observation is that the above statement still holds except the last identity will be an inequality. So the rest of the proof of the Chernoff bound follows. In particular, when $X_1, \ldots, X_n$ are negatively correlated we show

$$\mathbb{E}\left[e^{tX}\right] \leq \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right].$$

Let $\{\tilde{X}_1, \ldots, \tilde{X}_n\}$ be independent Bernoulli random variables with $\mathbb{E}\left[\tilde{X}_i\right] = \mathbb{E}\left[X_i\right]$ for each $i \in \{1, \ldots, n\}$ and define $\tilde{X} := \tilde{X}_1 + \cdots + \tilde{X}_n$. For any nonnegative integer k,

$$\mathbb{E}\left[X^k\right] = \sum_{\alpha} \mathbb{E}\left[X_1^{\alpha_1} X_2^{\alpha_2} \ldots X_n^{\alpha_n}\right]$$

$$\underset{X_i \in \{0,1\}}{=} \sum_{\alpha} \mathbb{E}\left[\prod_{i=1}^{n} X_i^{\mathbb{I}[\alpha_i \equiv 1 \bmod 2]}\right]$$

$$\underset{\text{Negative Correlation}}{\leq} \sum_{\alpha} \prod_{i=1}^{n} \mathbb{E}\left[X_i^{\mathbb{I}[\alpha_i \equiv 1 \bmod 2]}\right]$$

$$= \sum_{\alpha} \prod_{i=1}^{n} \mathbb{E}\left[X_i^{\alpha_i}\right] \underset{\mathbb{E}[X_i] = \mathbb{E}[\tilde{X}_i]}{=} \sum_{\alpha} \prod_{i=1}^{n} \mathbb{E}\left[\tilde{X}_i^{\alpha_i}\right]$$

where the sum is over all non-negative integer vectors α such that $\sum_i \alpha_i = k$.

On the other hand, since $\tilde{X}_1, \ldots, \tilde{X}_n$ are independent,

$$\sum_{\alpha} \prod_{i=1}^{n} \mathbb{E}\left[\tilde{X}_i^{\alpha_i}\right] = \sum_{\alpha} \mathbb{E}\left[\tilde{X}_1^{\alpha_1} \ldots \tilde{X}_n^{\alpha_n}\right] = \mathbb{E}\left[\tilde{X}\right].$$

Putting these together we obtain, for every $k \geq 0$,

$$\mathbb{E}\left[X^k\right] \leq \mathbb{E}\left[\tilde{X}^k\right] \tag{7.1}$$

Lastly, using the Taylor expansion

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2} + \frac{t^3 X^3}{6} + \dots$$

Applying (7.1) to every monomial above, we get

$$\mathbb{E}\left[e^{tX}\right] \leq \mathbb{E}\left[e^{t\tilde{X}}\right] \underset{\text{independence}}{=} \prod_{i=1}^{n} \mathbb{E}\left[e^{t\tilde{X}_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right]$$

as desired. $\square$

Definition 7.7 (Generating Polynomial)**.** *It is natural to express a probability distribution μ over subsets of $[n]$ by its generating polynomial. To do that we consider n variables, $z_1, \dots, z_n$ and write*

$$g_\mu(z_1, \dots, z_n) = \sum_{S \subseteq [n]} \mu(S) z^S,$$

where $z^S = \prod_{i \in S} z_i$.

For a concrete example, let $B_1, \dots, B_n$ be n independent Bernoulli random variables where B_i has success probability p_i. Then, we can write the corresponding generating polynomial as follows:

$$(p_1 z_1 + 1 - p_1)(p_2 z_2 + 1 - p_2) \dots (p_n z_n + 1 - p_n).$$

The following facts about the generating polynomial are straightforward:

Fact 7.8. *Let μ be a probability distribution over $[n]$ with generating polynomial g_μ, then*

- *$g_\mu(\mathbf{1}) = 1$, i.e., sum of the coefficients of g_μ is 1.*

- *$\partial_i g_\mu(\mathbf{1}) = \mathbb{P}_\mu[i]$, i.e., the marginals can be deduced by take partial derivatives.*

- *i, j are negatively correlated if*

$$g(\mathbf{1}) \partial_i \partial_j g_\mu(\mathbf{1}) = \mathbb{P}[i, j] \geq \mathbb{P}[i] \mathbb{P}[j] = \partial_i g_\mu(\mathbf{1}) \partial_j g_\mu(\mathbf{1}).$$

- *Say we have two probability distributions μ_1, μ_2 over disjoint sets, then the product distribution is the probability distribution with generating polynomial $\mu_1 \cdot \mu_2 = g_{\mu_1} g_{\mu_2}$.*

- *If μ_1, μ_2 are pairwise negatively correlated then so is $\mu_1 \cdot \mu_2$.*

Next, we explain a few examples of negatively correlated random variables:

Example 1: Observe that any probability distribuition over subsets of size (exactly) one among n objects is negatively correlated, namely

$$p_1 z_1 + p_2 z_2 + \dots + p_n z_n$$

where $\sum_i p_i = 1$. Following the above fact, product of these distributions are also negatively correlated.

As an application, recall that in lecture 4, we introduced a probability distribution over paths $P \in \mathcal{P}_i$ connecting the i-th terminal pairs s_i, t_i where we chose one path with its probability y_P and we independently

run the procedure for every i. It follows that the resulting probability distribution over the random variables $Y_P = \mathbb{I}[P \text{ is chosen}]$ is negatively correlated. So, we could have directly apply the Chernoff bound instead of defining a new family of random variables $X_{e,i} = \mathbb{I}[\text{a path of } \mathcal{P}_i \text{ going through } e \text{ is chosen}]$.

Example 2: Edges of a uniform spanning tree One of the most interesting family of negative correlated probability distributions is the distribution of the set of edges of a uniform spanning tree. Namely, let $G = (V, E)$ be a connected undirected graph; assign a variable z_e to every edge $e \in E$, then μ is the distribution with the following generating polynomial,

$$g_\mu(\{z_e\}_{e \in E}) = \sum_{T \text{ spanning tree}} z^T.$$

We will discuss ideas to prove this fact in the next lecture.

7.3 Towards a theory of Negative dependence

One of the ongoing research directions in probability theory is to study under what conditions one can expect negative correlation and and negative association.

Following the above discussion, a natural choice is the reverse of the positive lattice condition, namely negative lattice condition:

$$\mu(A)\mu(B) \geq \mu(A \cup B)\mu(A \cap B), \quad \forall A, B \subseteq [n]$$

Unfortunately, it can be seen that this property does not even imply a pairwise negative correlation property:

Example 7.9. *Consider the distribution μ over $[4]$ with the following generating polynomial,*

$$\frac{1}{2}(z_1 z_2 + z_3 z_4).$$

This distribution satisfies the NLC but it not negative correlated as $\mathbb{P}[1, 2] = 0.5 > 0.25 = \mathbb{P}[1]\mathbb{P}[2]$.

In the next lecture we will introduce strongly Rayleigh distribution as a generic method to study negative dependence.