Harris-FKG Inequalities

Consider a random red/blue coloring of the edges of K_n. Intuitively, the event that we have a red K_3 and the event that we have a red C_5 should be positively correlated, since they both occur when there are more red edges. It is not obvious how to make such intuition precise, however, since one is not a subgraph of the other. This is exactly what the Harris-FKG inequalities are for.

Suppose that $\Omega = \{0,1\}^n$, and each coordinate is assigned independently. In other words, we have a set of independent Bernoulli random variables $X_1, \dots, X_2$. For example, this occurs when we apply a uniform independent random 2-coloring of vertices or edges, or when we assign edges independently in $G(n,p)$.

Define a partial order on the elements of Ω as follows: $(x_1, \dots, x_n) \geq (y_1, \dots, y_n)$ if and only if $x_i \geq y_i$ for all $1 \leq i \leq n$.

We say that an event $A \subseteq \Omega$ is *increasing* if $x \in A$ and $y \geq x$ implies that $y \in A$. We say that an event $A \subseteq \Omega$ is *decreasing* if $x \in A$ and $y \leq x$ implies that $y \in A$. Then, Harris' Inequality says that events are positively correlated if they go in the same direction (are both increasing or both decreasing).

Theorem 1 (Harris)**.** *In the above setting, if A and B are increasing events, then*

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A)\mathbb{P}(B).$$

The above setting will be the most commonly used in discrete problems, but we will state and prove a more general version of Theorem 1. This is sometimes attributed to Fortuin, Kasteleyn and Ginibre, who independently proved this result in a more general setting. Hence, it is commonly referred to as the Harris-FKG inequality.

Let $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ be our sample space, and suppose that we have a linear ordering on each Ω_i. Again, for $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \Omega$, we say that $x \geq y$ if and only if $x_i \geq y_i$ for all $1 \leq i \leq n$. We say that a function $f : \Omega \to \mathbb{R}$ is increasing if $x \geq y$ implies $f(x) \geq f(y)$.

Theorem 2 (Harris)**.** *If f and g are increasing functions of independent random variables, then*

$$\mathbb{E}(fg) \geq \mathbb{E}(f)\mathbb{E}(g).$$

Proof. We will use induction on n. Let $n = 1$. Note that for any $x, y \in \Omega_1$, we have that $f(x) - f(y)$ and $g(x) - g(y)$ have the same sign. Therefore,

$$\mathbb{E}((f(x) - f(y))(g(x) - g(y))) = 2\mathbb{E}(fg) - 2\mathbb{E}(f)\mathbb{E}(g) \geq 0,$$

which implies the result.

Exercise 1. *Show the the first equality above.* 1

Suppose that $n \geq 2$, and define

$$\begin{aligned} f_1(y_1) &= \mathbb{E}(f \mid x_1 = y_1) = \mathbb{E}(f(y_1, x_2, \dots, x_n)) \\ g_1(y_1) &= \mathbb{E}(g \mid x_1 = y_1) = \mathbb{E}(g(y_1, x_2, \dots, x_n)) \\ (fg)_1(y_1) &= \mathbb{E}(fg \mid x_1 = y_1) = \mathbb{E}((fg)(y_1, x_2, \dots, x_n)). \end{aligned}$$

For a fixed y_1, the functions $f(y_1, x_2, \ldots, x_n)$, $g(y_1, x_2, \ldots, x_n)$ and $(fg)(y_1, x_2, \ldots, x_n)$ are increasing functions in $n-1$ variables, so by the inductive hypothesis, we have

$$(fg)_1(y_1) \geq f_1(y_1)g_1(y_1).$$

Also note that f_1 and g_1 are increasing functions in one variable, and therefore

$$\mathbb{E}(f_1 g_1) \geq \mathbb{E}(f_1)\mathbb{E}(g_1)$$

Therefore,

$$\mathbb{E}(fg) = \mathbb{E}((fg)_1) \geq \mathbb{E}(f_1 g_1) \geq \mathbb{E}(f_1)\mathbb{E}(g_1) = \mathbb{E}(f)\mathbb{E}(g).$$

□

Exercise 2. *Use the Harris-FKG bound to bound (from below) the probability that $G(n,p)$ is triangle-free. One way to bound this probability is as follows:* 3

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) \geq \mathbb{P}(G(n,p) \text{ is empty}) = (1-p)^{\binom{n}{2}}.$$

Compare this bound to your Harris-FKG bound.

Janson's Inequalities

Assume a probability space where a random subset is chosen from $[n]$, with each element included independently of other elements. (They need not have the same probability of being included.) Let $S_1, \ldots, S_t \subseteq [n]$ be a set of subsets, let A_i be the event that all the elements of S_i are chosen and let X_i be the indicator random variable of A_i. Let $X = \sum_{i=1}^t X_i$. Then we have the following upper bound on the probability that none of the events happen.

Theorem 3 (Janson)**.** *With the definition of X given above, we have*

$$\mathbb{P}(X = 0) \leq e^{-\mu + \Delta/2},$$

where $\mu = \mathbb{E}(X)$ and

$$\Delta = \sum_{i \neq j \mid S_i \cap S_j \neq \emptyset} \mathbb{P}(A_i \cap A_j)$$

Proof. (This proof follows Yufei Zhao's notes.) First we let

$$p_i = \mathbb{P}(A_i \mid \overline{A_1} \cap \ldots \overline{A_{i-1}}).$$

Exercise 3. *Show that* 1

$$\mathbb{P}(X = 0) \leq e^{-\sum_{i=1}^t p_i}.$$

First, we will show that

$$p_i \geq \mathbb{P}(A_i) - \sum_{j<i \mid S_i \cap S_j \neq \emptyset} \mathbb{P}(A_i \cap A_j).$$

Let

$$B = \bigcap_{j<i \mid S_i \cap S_j = \emptyset}$$

and

$$C = \bigcap_{j<i \mid S_i \cap S_j \neq \emptyset} .$$

Then

$$\begin{aligned}
p_i &= \mathbb{P}(A_i \mid B \cap C) \\
&= \frac{\mathbb{P}(A_i \cap B \cap C)}{\mathbb{P}(B \cap C)} \\
&\geq \frac{\mathbb{P}(A_i \cap B \cap C)}{\mathbb{P}(B)} \\
&= \mathbb{P}(A_i \cap C \mid B) \\
&= \mathbb{P}(A_i \mid B) - \mathbb{P}(A_i \cap \overline{C} \mid B).
\end{aligned}$$

Exercise 4. *Finish the proof.* 2

□