Lecture 13: Branching process approximation and stochastic domination

In the last two lectures, we argued that the initial stages of the exploration process of the component in an Erdős-Rényi random graph should be close to a branching process with Poisson offspring distribution, and, in particular, that the number of explored vertices in the component should be close to the total number of individuals in the branching process. The next lemma quantifies this approximation.

Lemma 13.1. *Let C_v be the size of the component containing vertex v in $G \sim \mathrm{ER}(n, \lambda/n)$, where $\lambda > 0$ is fixed. Let W_λ be the total population of a branching process with* Poisson(λ) *offspring distribution. Then for any $k \geq 0$,*

$$\mathbb{P}(W_\lambda \geq k) - \frac{k^2\lambda(k+\lambda)}{n} \leq \mathbb{P}(C_v \geq k) \leq \mathbb{P}(W_\lambda \geq k).$$

The proof of the lemma requires comparing the Binomial distributions of the Erdős-Rényi exploration process with the Poisson distribution of the branching process. We first discuss these kinds of comparisons in general.

Tool: stochastic domination. We start with a definition.

Definition 13.2. For random variables X and Y, we say $\mathscr{L}(X)$ stochastically dominates $\mathscr{L}(Y)$ if for all $x \in \mathbb{R}$,

$$\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x),$$

and we write $X \geq_{\mathrm{st}} Y$, $\mathscr{L}(X) \geq_{\mathrm{st}} \mathscr{L}(Y)$, $Y \leq_{\mathrm{st}} X$, or $\mathscr{L}(Y) \leq_{\mathrm{st}} \mathscr{L}(X)$.

The definition only depends on X and Y through $\mathscr{L}(X)$ and $\mathscr{L}(Y)$, but suggests that X should be bigger than Y in some sense. But the random variables may not be on the same space, in which case there would be no way to compare them. The next proposition shows what stochastic domination means at the level of random variables, and is useful for showing that one distribution dominates another when the CDFs are not simple.

Proposition 13.3. *For random variables X, Y, $X \geq_{\mathrm{st}} Y$ if and only if there is a coupling $(\widehat{X}, \widehat{Y})$ of X and Y such that*

$$\mathbb{P}(\widehat{X} \geq \widehat{Y}) = 1.$$

Proof. Assume $X \geq_{\mathrm{st}} Y$. We construct a coupling $(\widehat{X}, \widehat{Y})$ with the claimed property. Let F and G be the CDFs of X and Y and define, for $u \in (0,1)$, the generalised inverse

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\},$$

with an analogous definition for G^{-1}. We have the following two facts.

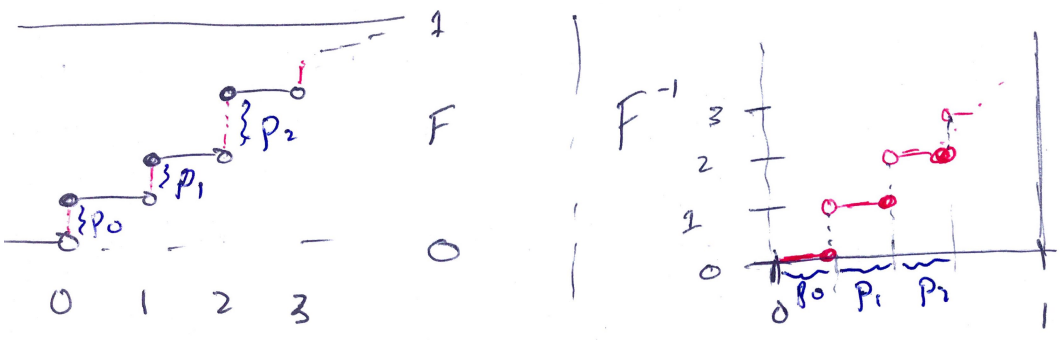
- If $U \sim \mathrm{U}([0,1])$, then
$$(\widehat{X}, \widehat{Y}) := \big(F^{-1}(U), G^{-1}(U)\big)$$
is a coupling of X and Y.
- For all $u \in (0,1)$, $F^{-1}(u) \geq G^{-1}(u)$.

From these two facts, the proposition easily follows. For the first fact, the picture below shows what is going on if X is integer-valued. To prove the fact, we show that for any $u \in (0,1)$, and $y \in \mathbb{R}$,

$$\inf\{x : F(x) \geq u\} \leq y \iff u \leq F(y), \tag{13.1}$$

which implies

$$\mathbb{P}(\widehat{X} \leq y) = \mathbb{P}(F^{-1}(U) \leq y) = \mathbb{P}(\inf\{x : F(x) \geq U\} \leq y) = \mathbb{P}(U \leq F(y)) = F(y).$$



Example CDF F for integer-supported distribution and its generalized inverse F^{-1}.

To see why (13.1) is true, note that if u is such that $z := \inf\{x : F(x) \geq u\} \leq y$, then there is a sequence $z_n \searrow z$ with $F(z_n) \geq u$, which implies that $F(z) \geq u$ by right continuity of F. But since $z \leq y$, $F(z) \leq F(y)$, and so $u \leq F(y)$, as desired. For the other direction, if $u \leq F(y)$, then $y \in \{x : F(x) \geq u\}$, and so $\inf\{x : F(x) \geq u\} \leq y$.

The second fact follows because $F(x) \leq G(x)$ for all x (this follows from the definition of stochastic domination and that $F(x) = \lim_{y \searrow x} \mathbb{P}(X < y)$), and so for all $u \in (0, 1)$,

$$\{x : F(x) \geq u\} \subseteq \{x : G(x) \geq u\},$$

and so the infimum on the right hand side is smaller than that of the left hand side.

Now assume that there is a coupling $(\widehat{X}, \widehat{Y})$ of X and Y with $\mathbb{P}(\widehat{X} \geq \widehat{Y}) = 1$. Then $\widehat{Y} \geq x$ implies $\widehat{X} \geq x$ and so

$$\mathbb{P}(Y \geq x) = \mathbb{P}(\widehat{Y} \geq x) \leq \mathbb{P}(\widehat{X} \geq x) = \mathbb{P}(X \geq x). \qquad \square$$

The particular coupling used in the proof is rarely useful in practice, but the coupling characterisation of stochastic domination is useful for showing stochastic domination when a simple expression for the CDF is not available.

Example 13.4. Let $X \sim \mathrm{Bi}(n, p)$ and $Y \sim \mathrm{Bi}(m, p)$ with $m \leq n$. We show $X \geq_{\mathrm{st}} Y$ by creating a coupling. let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{Be}(p)$ and $\widehat{X} = \sum_{i=1}^{n} X_i$ and $\widehat{Y} = \sum_{i=1}^{m} X_i$. Then clearly $(\widehat{X}, \widehat{Y})$ is a coupling of X and Y with $\mathbb{P}(\widehat{X} \geq \widehat{Y}) = 1$.

Example 13.5. Let $X \sim \mathrm{Ge}_1(p)$ and $Y \sim \mathrm{Ge}_1(q)$ with $p \leq q$. We can couple X and Y together by coupling i.i.d. Bernoulli (p and q) trials and comparing the number to the first success. More precisely, let $(\widehat{X}_i, \widehat{Y}_i)_{i \geq 1}$ be i.i.d. with

$$\mathbb{P}(\widehat{X}_1 = x, \widehat{Y}_1 = y) = \begin{cases} 1 - q, & (x, y) = (0, 0), \\ q - p, & (x, y) = (0, 1), \\ p, & (x, y) = (1, 1). \end{cases}$$

Then $\mathbb{P}(\widehat{X}_i \leq \widehat{Y}_i) = 1$, and we have the coupling of X and Y

$$\widehat{X} := \inf\{i : \widehat{X}_i = 1\} \geq \inf\{i : \widehat{Y}_i = 1\} =: \widehat{Y},$$

and Proposition 13.3 implies $X \geq_{\mathrm{st}} Y$.

The following proposition provides another important example of using coupling to show stochastic domination.

Proposition 13.6. *If $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^n$ are two sequences of independent random variables such that $X_i \geq_{\text{st}} Y_i$ for all $i = 1, \ldots, n$, then*

$$\sum_{i=1}^n X_i \geq_{\text{st}} \sum_{i=1}^n Y_i.$$

Proof. Since $X_i \geq_{\text{st}} Y_i$ for each $i = 1, \ldots, n$, there are couplings $(\widehat{X}_i, \widehat{Y}_i)$ with $\mathbb{P}(\widehat{X}_i \geq \widehat{Y}_i) = 1$. Letting these collections of pairs be independent, we have the coupling of X and Y

$$\widehat{X} := \sum_{i=1}^n \widehat{X}_i \geq \sum_{i=1}^n \widehat{Y}_i =: \widehat{Y},$$

and Proposition 13.3 implies $X \geq_{\text{st}} Y$. □

We want to compare Binomial (Erdős-Rényi exploration) and Poisson (BP exploration) distributions. We first compare Bernoulli and Poisson distributions, and then apply the previous proposition.

Proposition 13.7. *If $X \sim \text{Po}(\lambda)$, $Y \sim \text{Bi}(n, p)$, and $\lambda \geq -n \log(1-p)$, then $X \geq_{\text{st}} Y$.*

Proof. We can write $X = \sum_{i=1}^n X_i$ for $(X_i)_{i=1}^n$ i.i.d. distributed Poisson with mean λ/n and $Y = \sum_{i=1}^n Y_i$ for $(Y_i)_{i=1}^n$ i.i.d. distributed Bernoulli with parameter p, and so by Proposition 13.6, it's enough to show that $X_1 \geq_{\text{st}} Y_1$. But clearly $\mathbb{P}(X_1 \geq k) \geq \mathbb{P}(Y_1 \geq k)$ for $k = 0, 2, 3, \ldots$, so we only need to check that

$$\mathbb{P}(X_1 \geq 1) = 1 - e^{-\lambda/n} \geq p = \mathbb{P}(Y_1 \geq 1),$$

which follows from the condition $\lambda \geq -n \log(1-p)$. □

Note that the condition of the proposition holds for $p = \lambda/n$. In the next lecture we use this proposition to prove Lemma 13.1.