

Probabilistic method

A *sample space* is a set of elements that we call *outcomes*. An *event* is a subset $A \subseteq \Omega$. The *event space* $\mathcal{F}$ is a sigma-algebra: a collection of subsets $\mathcal{F} \subseteq 2^\Omega$ that includes Ω , is closed under taking complements and countable unions. For our purposes it is fine to assume $\mathcal{F} = 2^\Omega$. A *probability measure* is a function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$, such that $\mathbb{P}(\Omega) = 1$, and $\mathbb{P}$ is countable additive, meaning that $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$, for disjoint events $A_1, A_2, \dots$. Now, a *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$.

A *random variable* is a measurable function $X : \Omega \rightarrow S$, where S is the *state space* of X . We will always deal with $S \subseteq \mathbb{R}$, and most of the time we will deal with discrete nonnegative random variables, such that $S \subseteq \mathbb{N}_0$. We use the notation $\mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\})$. (Note that in that expression $B \subseteq \mathbb{R}$.) We let I_A be an *indicator random variable* for the event A . This means that

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise.} \end{cases}$$

In the case of a discrete random variable, we define the *expectation* of X as

$$\mathbb{E}(X) = \sum_{x \in S} x \mathbb{P}(X = x).$$

We say that two events A_1 and A_2 are *independent* if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

In order to understand this idea more intuitively, we need the following definition. We let

$$\mathbb{P}(A_1|A_2) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)}.$$

The *conditional probability* $\mathbb{P}(A_1|A_2)$ indicates the probability of event A_1 under the assumption of A_2 . Another way of thinking about conditional probabilities is the following: we assume that $\mathbb{P}(A_2) = 1$ and therefore $\mathbb{P}(\omega) = 0$ for any $\omega \notin A_2$. We then obtain a new probability measure by scaling $\mathbb{P}(\omega)$ by $1/\mathbb{P}(A_2)$ for each $\omega \in A_2$ so that our total probability remains 1. Now, we see that two events A_1 and A_2 are independent if and only if

$$\mathbb{P}(A_1|A_2) = \mathbb{P}(A_1).$$

In other words, the events A_1 and A_2 are independent if knowledge of one event does not yield information about the probability of the other.

Similarly, for random variables X, Y , we say that X and Y are *independent* if

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \text{ for all } x, y.$$

The following lemma is one of the reasons that first-moment method proofs are usually surprisingly straight-forward. We can ignore dependence when working with the expectation of the sum of sets of random variables.

Lemma 1. *We have linearity of expectation:*

$$\mathbb{E}\left(\sum_i X_i\right) = \sum_i \mathbb{E}(X_i),$$

whether the events X_i are dependent or not.

Before we get into more complicated first-moment lemmas, we make a simple observation:

Proposition 2. *We have*

$$\mathbb{P}(X \leq \mathbb{E}(X)) > 0,$$

$$\mathbb{P}(X \geq \mathbb{E}(X)) > 0.$$

This simple observation gives us a powerful method of proving the existence of structures without finding explicit constructions. If $X(G)$ is an invariant of, for example, graphs, and we sample graphs from some distribution, then we immediately know that graphs with $X \geq \mathbb{E}(X)$ as well as graphs with $X \leq \mathbb{E}(X)$ exist.

Lemma 3 (Markov's inequality). *For a non-negative random variable, we have*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

The following is obtained from Markov by letting $t = 1$.

Lemma 4 (First moment method). *For a non-negative, integer-valued random variable, we have*

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X).$$

Let $G = (V, E)$ be a graph, and let $n = |V|$ be the number of vertices in G . A k -edge-coloring of a graph is a coloring that uses k colors. We wish to find the largest n s.t. there exists a 2-edge coloring such that G contains no monochromatic K_s .

Exercise 1. *Show that $R(s, s) > n$ if $\binom{n}{s} 2^{1-\binom{s}{2}} > 1$. (Can you replace the $>$ by $\geq$ in that last inequality?)*

Optional Exercise. *Show that $R(s, s) > n - \binom{n}{s} 2^{1-\binom{s}{2}}$ for any n .*

Applications: cuts and dominating sets

We prove the following proposition in class.

Proposition 5. *For any graph G , there exists a subset $S \subseteq V(G)$ such that*

$$|[S, \overline{S}]| \geq \frac{|E|}{2}.$$

Exercise 2. *Can you improve on Proposition 5 by sampling S differently?*

A dominating set in a graph $G = (V, E)$ is a set S such that every vertex is in S or has a neighbor in S . Dominating sets have many applications, for example in communication routing in networks. Often, the goal is to find a smallest possible dominating set. This problem is NP-complete in general.

Let $d(v)$ be the degree of v in G , $\delta(G)$ be the minimum degree in G (and $\Delta(G)$ be the maximum degree in G).

Lemma 6. *Every graph G has a dominating set of at most size $\frac{\log(1+\delta)+1}{1+\delta}n$.*

Suppose $\delta \geq 1$. Create the set X by including each $v \in V$ with probability p i.i.d. Then create $Y = \{v \in V \setminus X : v \text{ has no neighbor in } X\}$. By definition, $X \cup Y$ is a dominating set in G . Since each v is included in X with probability p , the expected size of X is $\mathbb{E}(|X|) = np$. The expected size of Y is

$$\begin{aligned}\mathbb{E}(|Y|) &= \sum_{v \in V} \mathbb{E}(1_{v \in Y}) = \sum_{v \in V} \mathbb{P}(v \in Y) = \sum_{v \in V} \mathbb{P}(v \notin X, v\text{'s neighbors} \notin X) \\ &= \sum_{v \in V} (1-p)^{d(v)+1} \leq \sum_{v \in V} (1-p)^{\delta+1} = n(1-p)^{\delta+1}\end{aligned}$$

Since X and Y are disjoint we have

$$\mathbb{E}(|X \cup Y|) \leq np + n(1-p)^{\delta+1}.$$

Exercise 3. *Show that*

$$\mathbb{E}(|X \cup Y|) \leq np + ne^{-p(\delta+1)}.$$

The following fact is helpful:

$$(1+a)^n \leq e^{an}, \text{ for } a > -1, n \geq 0.$$

Then, find the value of p that minimizes this function, and finish the proof of Lemma 6.

Introduction to variance, Chebyshev, second moment method

The variance of a r.v. X is defined as

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

and can be interpreted as a measure of the spread of X (its expected distance from its mean).

The covariance of two r.v.s X, Y is defined as

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

and is a measure of how X and Y tend to “move together” (both be above (or below) their means).

Lemma 7 (Chebyshev's inequality). *For a random variable X , we have*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{var}(X)}{t^2}.$$

Proof. Follows from Markov's inequality applied to the r.v. $|X - \mathbb{E}(X)|^2$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq t^2) \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|^2)}{t^2} = \frac{\text{var}(X)}{t^2}.$$

□

Lemma 8 (Second moment method). *For a random variable X , we have*

$$\mathbb{P}(X = 0) \leq \frac{\text{var}(X)}{\mathbb{E}(X)^2}.$$

Proof. For a r.v. X we know that $X = 0$ only if X differs from $\mathbb{E}(X)$ by at least $\mathbb{E}(X)$. The second moment method follows from applying Chebyshev's inequality with $t = \mathbb{E}(X)$, giving

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)) \leq \frac{\text{var}(X)}{\mathbb{E}(X)^2}.$$

□

By the first moment method $\mathbb{E}(X) \rightarrow 0$ implies $\mathbb{P}(X = 0) \rightarrow 1$. However, $\mathbb{E}(X) \rightarrow \infty$ does not necessarily imply $\mathbb{P}(X = 0) \rightarrow 0$, as the variance could be large. By the second moment method, if $\text{var}(X) = o(\mathbb{E}(X)^2)$ then $\mathbb{E}(X) \rightarrow \infty$ does imply $\mathbb{P}(X = 0) \rightarrow 0$.

Exercise 4. *Selecting some subset S from $\{1, 2, \dots, n\}$ by choosing each element with probability p , i.i.d. When does S contain a k -arithmetic progression with high probability?*

Threshold for appearance of K_4 in $G(n, p)$

Let $G(n, p)$ be the random graph model where we sample a graph on n vertices by letting each edge appear independently with probability p . When do we first see certain patterns appear in $G(n, p)$, as we increase p from 0 to 1?

Let's look for the first appearance of K_4 , the complete graph on 4 vertices. This graph can appear in $\binom{n}{4}$ positions in $G(n, p)$ (with overlap). Let X_i for $1 \leq i \leq \binom{n}{4}$ be the indicator random variable for the presence of a subgraph isomorphic to K_4 on a given set of 4 vertices S_i . We have $\mathbb{P}(X_i = 1) = p^6$ for each i , since 6 edges need to be present. Therefore,

$$\mathbb{E}(X_i) = p^6, \quad 1 \leq i \leq \binom{n}{4}.$$

Now, we let X count the number of K_4 s in $G(n, p)$: $X = \sum_{i=1}^{\binom{n}{4}} X_i$. Then, by linearity of expectation

$$\mathbb{E}(X) = \sum_{i=1}^{\binom{n}{4}} \mathbb{E}(X_i) = \binom{n}{4} p^6 = \Theta(n^4 p^6).$$

When $p \ll n^{-4/6}$, then $\mathbb{E}(X) \rightarrow 0$ as $n \rightarrow \infty$. This implies that $\mathbb{P}(X > 0) \rightarrow 0$ by the first moment method. So, with high probability, $G(n, p)$ does not have a K_4 when $p \ll n^{-4/6}$.

What happens when $p \gg n^{-4/6}$? Then, $\mathbb{E}(X) \rightarrow \infty$. Be careful! This does not mean that $\mathbb{P}(X > 0) \rightarrow 1$. For that, we need to show that X is concentrated enough around its mean, using the second moment method. We need the variance of X , given by

$$\text{var}(X) = \text{var}\left(\sum_{i=1}^{\binom{n}{4}} X_i\right) = \sum_i \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j).$$

We have $\text{var}(X_i) = p^6(1 - p^6) = O(p^6)$, for all i . As for the $\text{cov}(X_i, X_j)$, this depends on the cardinality of $S_i \cap S_j$. We separate the following cases:

- $S_i \cap S_j \in \{0, 1\}$: in this case, X_i and X_j are independent, and we have $\text{cov}(X_i, X_j) = 0$,
- $S_i \cap S_j = 2$: in this case, we have $\mathbb{E}(X_i X_j) = p^{11}$, since this asks for the appearance of 11 edges. The number of such pairs is $\binom{n}{6} \cdot \binom{6}{2} \cdot \binom{4}{2} = O(n^6)$.
- $S_i \cap S_j = 3$: in this case, we have $\mathbb{E}(X_i X_j) = p^9$. The number of such pairs is $\binom{n}{5} \cdot \binom{5}{2} \cdot \binom{2}{1} = O(n^5)$.

Putting this together, we get

$$\text{var}(X) = O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9) = O(n^4 p^6).$$

By Chebyshev's inequality, we have

$$\mathbb{P}(X = 0) \leq \frac{\text{var}(X)}{\mathbb{E}(X)^2} = \frac{O(n^4 p^6)}{\Theta(n^4 p^6)^2} = O\left(\frac{1}{\mathbb{E}(X)}\right) \rightarrow 0, \text{ as } n \rightarrow \infty,$$

since $\mathbb{E}(X) \rightarrow \infty$ as $n \rightarrow \infty$.

Exercise 5. Use the previous method to find the threshold of appearance for the path P_4 in $G(n, p)$.

Exercise 6. Use the previous method to find the threshold of appearance for the path graph H in $G(n, p)$, where H is a triangle with a leaf.

General threshold for the appearance of subgraph H in $G(n, p)$

For a graph H , we define $m(H) = \frac{|E(H)|}{|V(H)|}$ as a form of edge density of H . We also define a *balanced* graph H as a graph such that $m(H') \leq m(H)$ for all subgraphs $H' \subseteq H$. Let $m^*(H) = \max_{H' \subseteq H} m(H')$, be the highest density “inside of H ”. Intuitively, in $G(n, p)$, it takes longer to see denser structure (as p increases from 0 to 1). The threshold of appearance turns out to depend only on the densest part of the graph.

Theorem 9. For a graph H , the threshold for appearance of H as a subgraph of $G(n, p)$ is $p^* = n^{-m^*(H)^{-1}}$. If H is balanced, then the threshold is $p^* = n^{-|V(H)|/|E(H)|}$.