## **Chapter 3.** The Probabilistic Method

The main reference for this section is Alon & Spencer The Probabilistic Method. (scanned pages on Moodle under "Lectures")

The probabilistic method is a non-constructive existence proof method, pioneered by Paul Erdős in the 1950s.

Idea: To prove that some element of a set $\Omega$ has a desired property, define a probability distribution on $\Omega$ and show that a random element of $\Omega$ satisfies the desired property with positive probability.

First, a quick **revision of discrete probability**.

$\pi(x) \geq 0$ for all $x$ and $\sum_{x \in \Omega} \pi(x) = 1$.

Let $\Omega$ be a finite set and let $\pi : \Omega \to [0,1]$ be a probability distribution on $\Omega$. Then $(\Omega, \pi)$ is a probability space on $\Omega$.

Often we work with the uniform distribution defined by $\pi(z) = 1/|\Omega|$ for all $z \in \Omega$.

for each $x \neq y$, decide whether $xy \in E(G)$

**Example.** Let $\Omega$ be the set of all graphs on the vertex set $\{1, 2, \ldots, n\}$. The uniform distribution on $\Omega$ is given by

$$\pi(G) = 1/|\Omega| = 2^{-\binom{n}{2}} \quad \text{since } |\Omega| = 2^{\binom{n}{2}}$$

for all $G \in \Omega$. This is the uniform model of random graphs.

We often just write $\Pr(\cdot)$ instead of $\pi(\cdot)$.

$\mathbb{P}(\cdot)$ or $\mathbb{P}(\cdot)$

An event is a subset of $\Omega$. For any event $A \subseteq \Omega$, we define

$$\Pr(A) = \sum_{z \in A} \Pr(z).$$

If $\pi$ is the uniform distribution on $\Omega$ then

$$\Pr(A) = \frac{|A|}{|\Omega|}$$

**Fact:** If $A$ and $B$ are events in $\Omega$ then

$$\Pr(A \cup B) = \frac{|A \cup B|}{|\Omega|} = \frac{|A| + |B| - |A \cap B|}{|\Omega|}$$

"A or B"

$\Pr(A) + \Pr(B) - \Pr(A \cap B)$ = "A and B"

So if $A$ and $B$ are disjoint events then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Two events $A$, $B$ are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

A random variable on $\Omega$ is a function $X : \Omega \to \mathbb{R}$.
Given any $T \subseteq \mathbb{R}$, the statement "$X \in T$" defines an event $A = \{z \in \Omega : X(z) \in T\}$, and hence

$$\Pr(X \in T) = \Pr(A).$$

The expected value of the random variable $X : \Omega \to \mathbb{R}$, denoted $\mathbb{E}X$ or $\mathbb{E}[X]$, is defined by

$$\mathbb{E}X = \sum_{z \in \Omega} \Pr(z)\, X(z).$$

**Fact:** If $\mathbb{E}X = \mu$ then there exists $z, w \in \Omega$ with $X(z) \leq \mu$ and $X(w) \geq \mu$.

Can we be sure that there is some $z, w \in \Omega$ with $X(z) < \mu$ and $X(w) > \mu$?

Binomial Theorem: $(a+b)^r = \sum_{j=0}^{r} \binom{r}{j} a^j b^{r-j}$

**Lemma**

The expected number of edges in a uniformly chosen graph on the vertex set $\{1, 2, \ldots, n\}$ is $\frac{1}{2}\binom{n}{2}$.

*Proof.* ("~~Dumb~~ LONG" way: from definition)

For $0 \le m \le \binom{n}{2} = N$, there are exactly $\binom{N}{m}$ of graphs on vertex set $\{1, \ldots, n\}$ with $m$ edges

Let $X$ be the number of edges in the random graph.

Then $\mathbb{E}X = \sum_{m=0}^{N} \Pr(X=m) \cdot m = \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N} \cdot m$

$= \frac{N}{2^N} \sum_{m=1}^{N} \frac{(N-1)!}{(m-1)!(N-m)!}$ [Check!]

$[j = m-1] = \frac{N}{2^N} \sum_{j=0}^{N-1} \binom{N-1}{j} = \frac{N}{2^N} \cdot 2^{N-1}$ by the binomial theorem

$= \frac{N}{2} = \frac{1}{2}\binom{n}{2}$. ☐ ✓

(For an easier proof, see Problem Sheet 3.)

[Start at 11:05]

Let $A \subseteq \Omega$ be an event. The indicator variable $\mathcal{I}_A$ for $A \subseteq \Omega$ is

$$\mathcal{I}_A(z) = \begin{cases} 1 & \text{if } z \in A, \\ 0 & \text{otherwise.} \end{cases}$$

(It is also called the characteristic function of $A$, when $A$ is viewed as a set.)

**Fact:** $\mathbb{E}I_A = \Pr(A)$.

**Linearity of expectation:**

Let $X_1, \ldots, X_k$ be random variables on $\Omega$ and let $c_1, \ldots, c_k \in \mathbb{R}$. Define the random variable $X = c_1 X_1 + \cdots + c_k X_k$. Then

$$\mathbb{E}[X] = c_1 \mathbb{E}[X_1] + c_2 \mathbb{E}[X_2] + \cdots + c_k \mathbb{E}[X_k].$$

(This holds whether the $X_i$ are independent or not!)

**Markov's inequality**

Suppose that $X : \Omega \to [0, \infty)$ is a nonnegative random variable on $\Omega$ and let $k > 0$. Then

$$\Pr(X \geq k) \leq \frac{\mathbb{E}[X]}{k}.$$

In particular, if $X$ is a nonnegative integer-valued random variable then

$$\Pr(X \neq 0) \leq \mathbb{E}[X].$$

Let $k \geq 2$ be an integer. Events $A_1, \ldots, A_k$ in $\Omega$ are mutually independent if for all $j, \ell_1, \ldots, \ell_j$ with $2 \leq j \leq k$ and $1 \leq \ell_1 < \ell_2 < \cdots < \ell_j \leq k$,

$$\Pr\left(\bigcap_{i=1}^{j} A_{\ell_i}\right) = \prod_{i=1}^{j} \Pr(A_{\ell_i}).$$

Sometimes we just say "independent" rather than "mutually independent".

Roughly speaking, events are independent if they have no effect on each other.

---

See the worksheet "Revision of discrete probability" (available on Moodle in the Lectures folder) for some revision exercises.

---

$|\Omega| = 2^n$

**Lemma**

Let $\Omega$ be the set of all subsets of some given set $S$, where $|S| = n$.

Define a random set $X \subseteq S$ by setting $\Pr(x \in X) = \frac{1}{2}$, independently for each $x \in S$.

[You can picture this as an experiment where you flip a fair coin for each $x \in S$, independently, and put $x$ into $X$ if and only if the coin flip for $x$ comes up heads.]

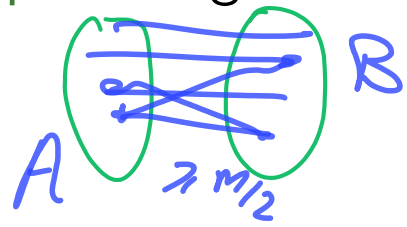
Then $\Pr(X = A) = 2^{-n}$ for all $A \subseteq S$, so this gives the uniform probability space on $\Omega$.

*Proof.* Fix $A \subseteq \Omega$. Then

$$\Pr(X = A) = \prod_{x \in A} \Pr(\text{heads}) \cdot \prod_{x \notin A} \Pr(\text{tails})$$ using independence

$$= \left(\frac{1}{2}\right)^{|A|} \cdot \left(\frac{1}{2}\right)^{n-|A|} = \left(\frac{1}{2}\right)^n = 2^{-n},$$ as claimed. □

Exercise: Describe the uniform model of random graphs using independent coin flips. (See Problem Sheet 3.)




**Theorem** (Alon & Spencer, Theorem 2.2.1)
Let $G$ be a graph with $n$ vertices and $m$ edges. Then $G$ contains a bipartite subgraph with at least $m/2$ edges.

*Proof.* Let $\Omega$ be the set of all subsets of $V(G)$. Then $|\Omega| = 2^n$. Consider the uniform probability space on $\Omega$ [ie flip fair coin, independently for each $v \in V$]. Let $A \subseteq V$ be a randomly chosen element of $\Omega$, and define $B = V - A$. Call $xy \in E(G)$ a crossing edge if exactly one of $x, y$ belongs to $A$. Let $X$ be the number of crossing edges.

Finally, for each edge $e \in E(G)$ define the indicator variable $X_e = \begin{cases} 1 & \text{if } e \text{ is a crossing edge,} \\ 0 & \text{otherwise} \end{cases}$

Then $X = \sum_{e \in E(G)} X_e$. For any $e = xy \in E(G)$,

we have

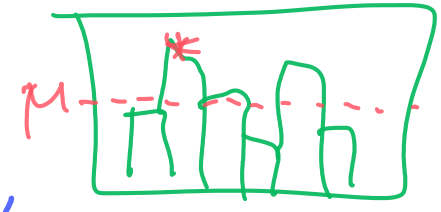
$$\Pr(x \in A \text{ and } y \notin A) = \Pr(x \in A)\Pr(y \notin A) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

using independence!

Therefore

$$\begin{aligned} \mathbb{E}X_e &= \Pr((x \in A \text{ and } y \notin A) \text{ or } (x \notin A \text{ and } y \in A)) \\ &= \Pr(x \in A \text{ and } y \notin A) + \Pr(x \notin A \text{ and } y \in A) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

since these events are disjoint



Hence, by linearity of expectation,

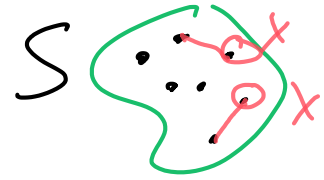
$$\mathbb{E}X = \sum_{e \in E(G)} \mathbb{E}X_e = \frac{m}{2}.$$

Thus there exists a fixed set $A_0 \subseteq V(G)$ which has at least $m/2$ crossing edges. The corresponding bipartition $(A_0, V(G) - A_0)$ defines a bipartite subgraph consisting of the $\geq \frac{m}{2}$ crossing edges, as required. □

or "stable set"

An independent set in a graph $G$ is a subset $U \subseteq V$ such that if $v, w \in U$ then $vw \notin E(G)$.

Let $\alpha(G)$ be the size (cardinality) of a maximum (that is, largest) independent set in $G$, called the independence number.




**Theorem** (Alon & Spencer, Theorem 3.2.1)
Let $G$ have $n$ vertices and $nd/2$ edges, where $d \geq 1$.
Then $\alpha(G) \geq \frac{n}{2d}$. (Note, $d$ is the average degree of $G$.)

*Proof.* Define the random subset $S \subseteq V(G)$ by $\Pr(v \in S) = p$, independently for all $v \in V$.
[Like flipping a biased coin which has $\Pr[\text{heads}] = p$.]
Here $p \in [0,1]$ which we will fix later.
Let $X = |S|$ and let $Y$ be the number of edges of $G$ with both endvertices in $S$. ("bad edges").
Then $\mathbb{E}X = pn$ [Exercise! Write $X$ as a sum of indicator variables (or recognise that $X \sim \text{Bin}(n,p)$).]

For $e \in E(G)$ let $Y_e$ be the indicator variable for the event "$e \subseteq S$". Then for $e = xy \in E(G)$,

$$\begin{aligned} \mathbb{E}\, Y_e &= \Pr(x \in S \text{ and } y \in S) \\ &= \Pr(x \in S) \cdot \Pr(y \in S) \quad \text{by independence} \\ &= p^2. \end{aligned}$$

Therefore, by linearity of expectation and the fact that $Y = \sum_{e \in E(G)} Y_e$, we have

$$\mathbb{E}\, Y = \sum_{e \in E(G)} \mathbb{E}\, Y_e = \frac{nd}{2} \cdot p^2.$$

By linearity of expectation,

$$\mathbb{E}(X - Y) = \mathbb{E}X - \mathbb{E}Y = pn - p^2 \frac{nd}{2}.$$

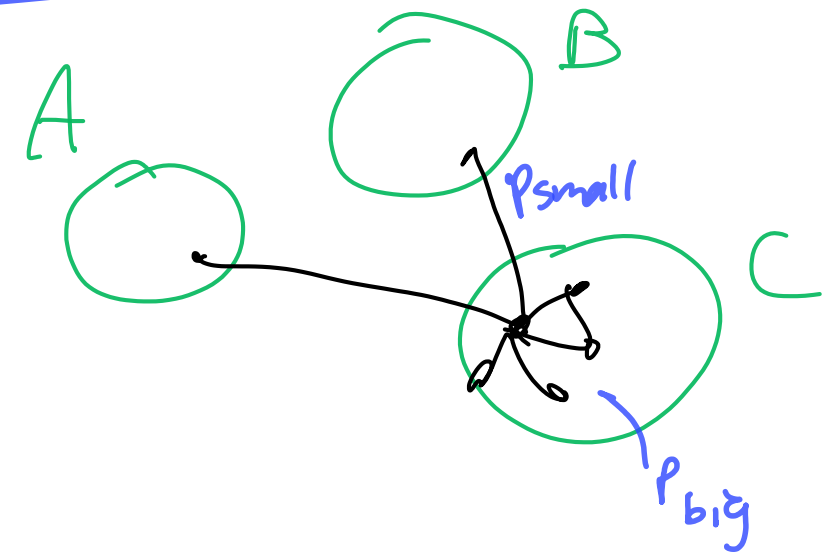
Want to choose $p$ to maximise this.

Check: $p = \frac{1}{d}$ maximises this expression, and $p \in [0, 1]$.

Substituting gives $\mathbb{E}(X - Y) = \frac{n}{2d}$. [check!!]

Hence there exists a fixed set $S_0 \subseteq V(G)$ with $|S_0| - (\#\text{edges in } S_0) \geq \frac{n}{2d}$.

Delete one vertex from each edge within $S_0$ to give a set $S^*$ of at least $\frac{n}{2d}$ vertices which is an independent set. □



□

[End of Chapter 3. Try Problem Sheet 3.]