Lecture 1: Random Graphs and Thresholds

In this section of the course we introduce probability to our investigation of boolean functions $f : \mathbb{F}_2^n \to \{-1, 1\}$.

We take the probability space Ω_n on $\mathbb{F}_2^n$ where each bit is chosen to be 1 independently with probability p (otherwise 0). For any event $\mathcal{A} \subset \mathbb{F}_2^n$ we write $\mu_p(\mathcal{A})$ to be the probability that a randomly chosen $x \in \mathbb{F}_2^n$ lies in the set $\mathcal{A}$. We write $\mu_p(x)$ to denote the probability of the event $\mathcal{A} = \{x\}$, notice

$$\mu_p(x) = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}.$$

As we are dealing with a finite probability space, all expressions for the proabability of an event will be finite weighted sums. We start by introducing/revising some basic notions from probability.

Probability Recap

Lemma 1.1 (Markov's inequality)**.** *If X is a random variable taking only non-negative values and $t > 0$, then $\mathbb{P}(X \geq t) \leq \mathbb{E}[X]/t$.*

Proof. (of Markov's inequality) Let $1_{X \geq t}$ be the indicator function of the event that $X \geq t$. Then always (with probability 1), the random variable X satisfies the relation $X \geq t1_{X \geq t}$. Now take the expectation of both sides to get

$$\mathbb{E}(X) \geq t\mathbb{E}(1_{X \geq t}) = t\mathbb{P}(X \geq t).$$

□

Recall the variance $\mathbb{V}[X]$ of a random variable X is defined by

$$\mathbb{V}[X] = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Lemma 1.2 (Chebyshev's Inequality)**.** *Let X be a random variable and let $t > 0$. Then*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{V}[X]}{t^2}.$$

Proof. This follows from Markov's inequality. We consider the probability of the event that the difference between X and its expectation it at least t. As t is positive,

$$\mathbb{P}\big(|X - \mathbb{E}[X]| \geq t\big) = \mathbb{P}\big((X - \mathbb{E}[X])^2 \geq t^2\big).$$

Then Markov's inequality applies to show this is less than $\mathbb{E}(X - \mathbb{E}(X))^2/t^2$ which is simply $\mathbb{V}[X]/t^2$ and we are done. □

Random Graphs

For this lecture and the next we focus on a particular case of boolean analysis: graphs.

Define a graph $G = (V, E)$ to be a set of labelled vertices $[n] = \{1, 2, \ldots, n\}$ and set of two-subsets of vertices E which we call edges. Write $e(G)$ for the number of edges $|E|$. Technically the edge between vertices i and j should be denoted $\{i, j\}$ but we will use the standard shorthand ij or ji interchangably. We do not allow *loops* which are edges with both end points at the same vertex or *multiple edges* namely each pair of vertices has either zero or one edges between them). (Our graphs are *undirected* but it is possible to define *directed graphs* where each edges has a direction associated with it and $ij \neq ji$.)

Each graph $G = ([n], E)$ can be associated with a boolean vector $x \in \mathbb{F}_2^{\binom{n}{2}}$, identify the $\binom{n}{2}$ positions in the vector x with the set of pairs of vertices in $[n]$ and each edge $e \in E$ is recorded by $x_e = 1$ and each non-edge by $x_e = 0$. For example (where vertex labels are always anticlockwise starting from bottom left e.g. $_1{}^3{}_2$ and the edges listed in lexicographic order e.g. $(12, 13, 23)$ and $(12, 13, 14, 23, 24, 34)$) the graph corresponds to vector $(0, 1, 1)$, likewise to $(1, 0, 0)$ and graph to $(0, 1, 1, 1, 1, 1)$.

Given an integer n and a real number $0 \leq p \leq 1$, the random graph $G(n, p)$ is the graph with vertex set $[n] = 1, 2, ..., n$ in which each possible edge ij, $1 \leq i < j \leq n$, is present with probability p, independently of the others. The notation $G(n, p)$ indicates the probability space of graphs on $[n]$ with the probabilities above. We write $G \sim G(n, p)$ or to mean that G is a random graph with this distribution. For a graph H on n vertices we write $\mu_p(H) = \mu_p(H, n)$ for $\mathbb{P}(G(n, p) = H)$ and for a set of graphs on n vertices, $\mathcal{A}$, write $\mu_p(\mathcal{A}) = \mu_p(\mathcal{A}, n)$ for $\mathbb{P}(G(n, p) \in \mathcal{A})$.
For any given graph H on $[n]$, the probability of H depends only on the number of edges in H,

$$\mathbb{P}(G(n, p) = H) = p^{e(H)}(1 - p)^{\binom{n}{2} - e(H)}.$$

In the special case that $p = 1/2$, then all $\binom{n}{2}$ graphs on vertex set $[n]$ are equally likely.

Example As an example consider the probability space $G(3, p)$ where the set of possible graphs is $\{ , , , , , , , \}$. (Note that because the graphs are labelled $\neq$.) If we sample a graph $H \sim G(3, p)$ then H is with probability p^3, for each of , , the probability is $p^2(1 - p)$, for each of , , the probablity is $p(1 - p)^2$ and finally for the probability is $(1 - p)^3$.

To study properties of random graphs we need a couple more notions from graph theory. We say that graphs H and G are *isomorphic*, denoted $H \approx G$ if there is a bijective function $\phi : V(H) \to V(G)$ such that $uv \in E(H)$ if and only if $\phi(u)\phi(v) \in E(G)$. For example $\approx$ and $\approx$ $\approx$. Similarly, we say that graph H is a *subgraph* of graph G, denoted $H \subseteq G$ if there is an injective function $\phi : V(H) \to V(G)$ such that if $uv \in E(H)$ then $\phi(u)\phi(v) \in E(G)$. For example $\subseteq$ $\subseteq$ $\subseteq$ but $\not\subseteq$.

We also define some isomorphism classes of graphs. A graph G is a path on n vertices, denoted P_n, if its vertices can be (re)-labelled $v_1, \ldots, v_n$ such that $E(G) = \{v_i v_{i+1} : i \in [n - 1]\}$. For example a path on 2 vertices is and there are three paths on three vertices is , , . A graph G with $n \geq 3$ vertices, denoted C_n, is a cycle if its vertices can be (re)-labelled $v_1, \ldots, v_n$ such that $E(G) = \{v_i v_{i+1} : i \in [n]\}$ where the subscript addition is taken modulo n. For example a cycle on 3 vertices is and there are three cycles on four vertices , , . A graph G is the complete graph on n vertices, denoted K_n, if $uv \in E(G)$ for all $u, v \in V(G)$. A graph G is the empty graph on n vertices, denoted $\overline{K}_n$, if $E(G) = \varnothing$. For example $K_4 =$ and $\overline{K}_4 =$.

Thresholds in Random Graphs

We give an example of the sort of question we will look at. Let $\mathcal{A}_\triangle$ be the set of all graphs which contain $\triangle$ as a subgraph, i.e. all graphs G which contain a set of three vertices $\{u, v, w\} \in V(G)$ such that $uv, uw, vw \in E(G)$. We are interested in how the probability that $G(n, p)$ constains a triangle changes for different values of p. Clearly, for any n, $\mu_0(\mathcal{A}_\triangle, n) = 0$ and for $n \geq 3$, $\mu_1(\mathcal{A}_\triangle, n) = 1$. One can also show that for $p \leq p'$, $\mu_p(\mathcal{A}, n) \leq \mu_{p'}(\mathcal{A}, n)$. For edge probability $p = p(n)$, we investigate the behaviour of $\mu_{p(n)}(\mathcal{A}_\triangle, n)$ as $n \to \infty$. We will find that for $p(n)$ and $p'(n)$ 'not too far apart' that $\mu_{p(n)}(\mathcal{A}_\triangle, n) \to 0$ while $\mu_{p'(n)}(\mathcal{A}_\triangle, n) \to 1$ a sort of a 'phase transition' in the behaviour. These ideas will be made precise in this course as we investigate what are called monotone properties of graphs.

Definition 1.3 (monotone). A set of graphs $\mathcal{A}$ is *monotone* if $H \in \mathcal{A}$ and $H \subseteq G$ implies that $G \in \mathcal{A}$.

A function from $f : \mathbb{F}_2^n \to \mathbb{R}$ is *monotone* if $f(x) \geq f(y)$ whenever $x \geq y$ i.e. for each i $x_i \geq y_i$.

Examples of monotone sets of graphs include the set of graphs containing $\triangle$ as a subgraph, the set of connected graphs (graphs which have a path along edges between any pair of vertices) and the set of all non-planar graphs (i.e. those that can't be drawn in the plane without edges crossing). Non-examples include the set of graphs with an odd number of edges and the set of $\triangle$-free graphs (those graphs *not* containing $\triangle$ as a subgraph).

Theorem 1.4. *For any monotone set of graphs $\mathcal{A}$ and $p' > p$,*

$$\mathbb{P}(G(n, p) \in \mathcal{A}) \leq \mathbb{P}(G(n, p') \in \mathcal{A})$$

Proof. Define $p_1 \in [0, 1]$, by $p + (1 - p)p_1 = p'$. Let $G \sim G(n, p)$ and $G_1 \sim G(n, p_1)$ and define the random graph $G_2 = G \cup G_1$, (this is the graph $([n], E(G) \cup E(G_1))$. Now each edge in G_2 occurs independently with probability $p + (1 - p)p_1 = p'$ and hence $G_2 \sim G(n, p')$. Now because $\mathcal{A}$ is monotone

$$\mathbb{P}(G \in \mathcal{A}) \leq \mathbb{P}(G \cup G_1 \in \mathcal{A}) = \mathbb{P}(G_2 \in \mathcal{A}).$$

□

Definition 1.5 (threshold). The function $p^* = p^*(n)$ is a (coarse)[1] *threshold* for monotone $\mathcal{A}$ if $\mathbb{P}(G(n, p) \notin \mathcal{A}) \to 1$ for $p/p^* \to 0$ and $\mathbb{P}(G(n, p) \in \mathcal{A}) \to 1$ for $p/p^* \to \infty$.

Exercise 1. An *Eulerian circuit* of G is a sequence of vertices $v_1 v_2 \dots v_\ell$ (a vertex may appear more than once) so that every edge $uw \in E(G)$ appears as $v_i v_{i+1}$ for some i in the sequence, and so that $v_1 = v_\ell$. A Eulerian graph is one which has a Eulerian circuit.

A *Hamiltonian cycle* of graph G on atleast three vertices is an sequence $v_1 v_2 \dots v_n$ such that each $u \in V(G)$ appears exactly once, $v_1 = v_2$ and each $v_i v_{i+1} \in E(G)$. A graph is Hamiltonian if it has a Hamiltonian cycle.

(a) Let $\mathcal{A}$ be the set of Eulerian graphs. Show that $\mathcal{A}$ is not monotone.

(b) Let $\mathcal{B}$ be the set of Hamiltonian graphs. Is $\mathcal{B}$ monotone?

[1] The function $p^* = p^*(n)$ is a *sharp threshold* for monotone $\mathcal{A}$ if $\mathbb{P}(G(n, p) \notin \mathcal{A}) \to 1$ for $p < (1 - \varepsilon)p^*$ and $\mathbb{P}(G(n, p) \in \mathcal{A}) \to 1$ for $p > (1 - \varepsilon)p^*$.

Exercise 2. A graph G with $n \geq 3$ vertices, denoted C_n, is a cycle if its vertices can be (re)-labelled $v_1, \ldots, v_n$ such that $E(G) = \{v_i v_{i+1} : i \in [n]\}$ where the subscript addition is taken modulo n. For example a cycle on 3 vertices is  and there are three cycles on four vertices , , .

A *connected graph* is one in which any two vertices uv are connected by a sequence of vertices $v_1 \ldots v_\ell$ so that $u = v_1$, $v = v_\ell$ and each $v_i v_{i+1}$ is an edge. For example  is connected but  is not connected.

(a) A graph with n vertices and n edges must contain a cycle as a subgraph.

(b) A connected graph with n vertices and n edges must contain exactly one cycle.

(c) Give an example to show that the assumption of connectivity is needed for part b.

Exercise 3. Let $\mathcal{A}_\triangle$ be the set of all graphs which contain  as a subgraph.

(a) Show that $\mathbb{P}(G(n, 1/2) \in \mathcal{A}_\triangle) \to 1$.

(b) (optional) Fix a constant $0 < p \leq 1$, and show that $\mathbb{P}(G(n,p) \in \mathcal{A}_\triangle) \to 1$.

Exercise 4. Prove the following:

Let $X_1, X_2, \ldots$ be a sequence of random variables each taking non-negative values. If $\mathbb{E}[X_n] \to 0$ then

$$\mathbb{P}(X_n = 0) \to 1,$$

and if $\mathbb{E}[X_n] > 0$ for each n, and $\mathbb{V}[X_n]/\mathbb{E}[X_n] \to 0$ then

$$\mathbb{P}(X_n = 0) \to 0.$$