Chapter 13. The Probabilistic Method

Note. Some of the material in this chapter overlaps with topics covered in other probability and/or statistics classes you may have taken. In particular, much of the material overlaps with Mathematical Statistics 1 (MATH 4047/5047); see my online notes for Mathematical Statistics 1. I also have some online notes for an Intermediate Probability and Statistics class (not a formal class at ETSU). A thorough class on probability theory requires a background in measure theory and functional analysis. I have some online notes for a Measure Theory Based Probability class (not a formal ETSU class). However, in graph theory we are mostly concerned with finite graphs and this greatly simplifies the probability spaces and the probability functions with which we deal. We start from the basics here.

Section 13.1. Random Graphs

Note. We first define a probability space. We only consider finite sample spaces and finite additivity of the probability function, so we avoid the many problems involved in more complicated settings. For example, in Mathematical Statistics 1 you would consider an arbitrary set as the sample space and the collection of events would form a "σ-field" or a "σ-algebra" of sets (see my online Mathematical Statistics 1 notes on 1.3. The Probability Set Function or my online notes on Real Analysis 1 [MATH 5210] on 1.4. Borel sets).

Definition. A (*finite*) *probability space* (Ω, P) consists of a finite set Ω, called the *sample space*, and a *probability function* $P : \Omega \to [0, 1]$ satisfying $\sum_{\omega \in \Omega} P(\omega) = 1$.

Definition. With the set $\mathcal{G}_n$, the set of all labeled simple graphs on n vertices (or, equivalently, the set of all spanning subgraphs of a labeled K_n), as the sample space of a finite space $(\mathcal{G}_n, P)$, the result of selecting an element G of this sample space according to the probability function P is a *random graph* (of order n).

Example. There are $2^N = 2^{\binom{n}{2}}$ simple graphs of order n (see equation $(*)$ in Section 1.2. Isomorphisms and Automorphisms). So if P is a uniform probability function (that is, each graph has the same probability of being chosen), the probability space $(\mathcal{G}_n, P)$ satisfies $P(G) = 1/2^N$ for all $G \in \mathcal{G}_n$ where $N = \binom{n}{2}$. Notice we could also view P in terms of choosing each of the $\binom{n}{2}$ edges of K_n for inclusion in graph G with probability $1/2$ each.

Example. Similar to the previous example, if we were to choose edges of K_n with the same probability p for each, then for a graph G with $m = |E(G)|$ edges, we have $P(G) = p^m(1-p)^{N-m}$ where $N = \binom{n}{2}$. Notice that there are $\binom{\binom{n}{2}}{m} = \binom{N}{m}$ labeled simple graphs on n vertices with m edges, so

$$\begin{aligned}\sum_{G \in \mathcal{G}_n} P(G) &= \sum_{m=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} = \sum_{m=0}^{N} \binom{N}{m} p^m (1-p)^{N-m} \\ &= (p + (1-p))^N \text{ by the Binomial Theorem} \\ &= 1.\end{aligned}$$

So $(\Omega, P) = (\mathcal{G}_n, P)$ is a probability space. This space is denoted $\mathcal{G}_{n,p}$. Figure 13.1 gives the probability space $\mathcal{G}_{3,p}$:

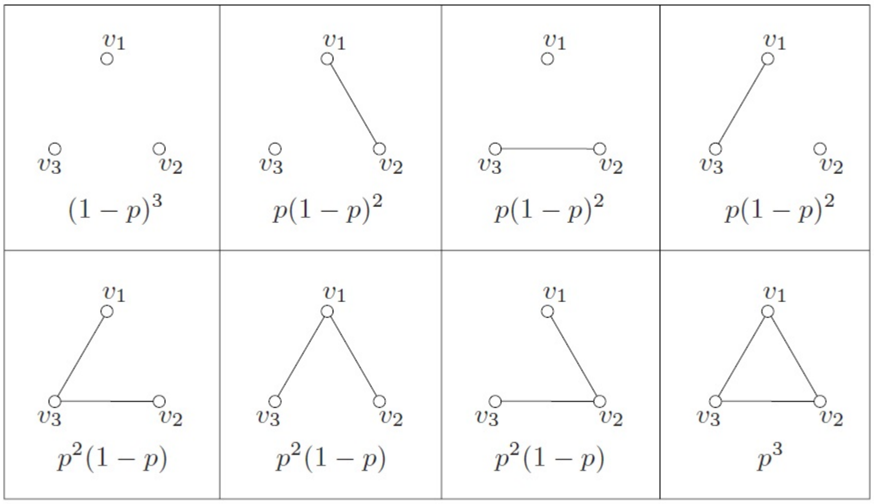

Figure 13.1. The probability space $\mathcal{G}_{3,p}$.

Note. We are interested in computing (or estimating) the probability that a random graph has a particular property, such as connectedness. For example, in $\mathcal{G}_{3,p}$ there are four connected graphs and the sum of the probabilities of choosing one of these graphs is: $3 \times p^2(1-p) + p^3 = 3p^2 - 2p^3$. In $\mathcal{G}_{3,p}$ there are seven bipartite graphs and one nonbipartite graph and the sum of the seven probabilities of choosing one of the bipartite graphs is the sum of the probabilities: $(1-p)^3 + 3(1-p)^2p + 3(1-p)p^2 = 1 - p^3$.

Definition. In a (finite) probability space (Ω, P), any subset A of Ω is referred to as an *event* and the *probability* of event A is defined as $P(A) = \sum_{\omega \in A} P(\omega)$.

Definition. Events A and B in a probability space (Ω, P) are *independent* if $P(A \cap B) = P(A)P(B)$; otherwise events A and B are *dependent*. More generally, events A_i for $i \in I$ are *(mutually) independent* if, for any subset S of I, $P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$.

Note. This is exactly the same definition of independent and mutually independent events which you see in Mathematical Statistics 1 (see 1.4. Conditional Probability and Independence).

Example. Let A be the event "G is connected" and let B be the event "G is bipartite" in probability space $\mathcal{G}_{3,p}$. Then as shown above, $P(A) = 3p^2 - 2p^3$ and $P(B) = 1 - p^3$. Now from Figure 13.1, we see that the event $A \cap B$, "G is connected and bipartite," has three elements and each has probability $p^2(1-p)$ so that $P(A \cap B) = 3p^2(1-p)$. Then for $p \notin \{0, 1\}$ we have

$$P(A)P(B) = (3p^2 - 2p^3)(1 - p^3) = 3p^2 - 3p^5 - 2p^3 + 2p^6$$

$$= p^2(3 - 2p - 3p^3 + 2p^4) \neq 3p^2(1-p) = P(A \cap B),$$

so events A and B are dependent.

Note. It is to be shown in Exercise 13.1.2 that a set of events may be pairwise independent but that the set of events may not be mutually independent. The exercise uses random vertex colourings.

Definition. A *random variable* on a probability space (Ω, P) is any real-valued function defined on the sample space Ω.

Note. This is the same as the definition of a random variable which you see in Mathematical Statistics 1; see my online notes on 1.5. Random Variables (notice Definition 1.5.1).

Note. Some random variables on probability space $\mathcal{G}_{n,p}$ are the connectivity κ (see Section 9.1. Vertex Connectivity), the clique number ω (see Section 8.3. $\mathcal{NP}$-Complete Problems), and the stability number α (also in Section 8.3).

Example. Let S be a set of vertices of a random graph $G \in \mathcal{G}_{n,p}$. We can define a random variable X_S associated with set S as

$$X_S(G) = \begin{cases} 1 & \text{is } S \text{ is a stable set of } G \\ 0 & \text{otherwise.} \end{cases}$$

This is an example of an "indicator random variable."

Definition. In probability space (Ω, P), an *indicator random variable* X_A is defined for event $A \subset \Omega$ as

$$X_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise.} \end{cases}$$

Note. For each event $A \subset \Omega$ we can associate a random variable (namely the indicator random variable). Conversely, we can associate with each random variable X and each $t \in \mathbb{R}$ the event $\{\omega \in \Omega \mid X(\omega) = t\} \subset \Omega$. We denote this event as "$X = t$." We can similarly define the events $\{\omega \in \Omega \mid X(\omega) < t\}$ (denoted $X < t$), $\{\omega \in \Omega \mid X(\omega) \leq t\}$ (denoted $X \leq t$), $\{\omega \in \Omega \mid X(\omega) > t\}$ (denoted $X > t$), and $\{\omega \in \Omega \mid X(\omega) \geq t\}$ (denoted $X \geq t$). We are interested in the probabilities of such events. For example, if X is the number of components of $G \in \mathcal{G}_{3,p}$, the event $X \geq 2$ consists of the first four graphs in Figure 13.1 and the probability of this event is $(1-p)^3 + 3p(1-p)^2 = (1-p)^2(1-p+3p) = (1-p)^2(1+2p)$.

Definition. Random variables X_i for $i \in I$ are (*mutually*) *independent* if the events $X_i = t_i$ for $i \in I$ are independent for all $t_i \in \mathbb{R}$. Random variables are *dependent* if they are not independent.

Revised: 2/24/2021