Section 13.2. Expectation

Note. In this section, we define the expectation (or "expected value") of a finite random variable. We give a proof of The Crossing Lemma based on expectation and give two results from combinatorial geometry based on The Crossing Lemma. We also introduce an asymptotic notation related to probability, define the concept of "almost surely" and give a proof of an almost surely upper bound on the stability number of a graph in $\mathcal{G}_{n,p}$ (using Markov's Inequality).

Definition. The average value, or mean, of a random variable X is its *expectation*, denoted $E(X)$ and is $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega)$.

Note. The expectation $E(X)$ is just the weighted mean of X where the values $X(\omega)$ are weighted by the amount $P(\omega)$. As an example, consider the probability space $\mathcal{G}_{3,p}$ of Figure 13.1 again.

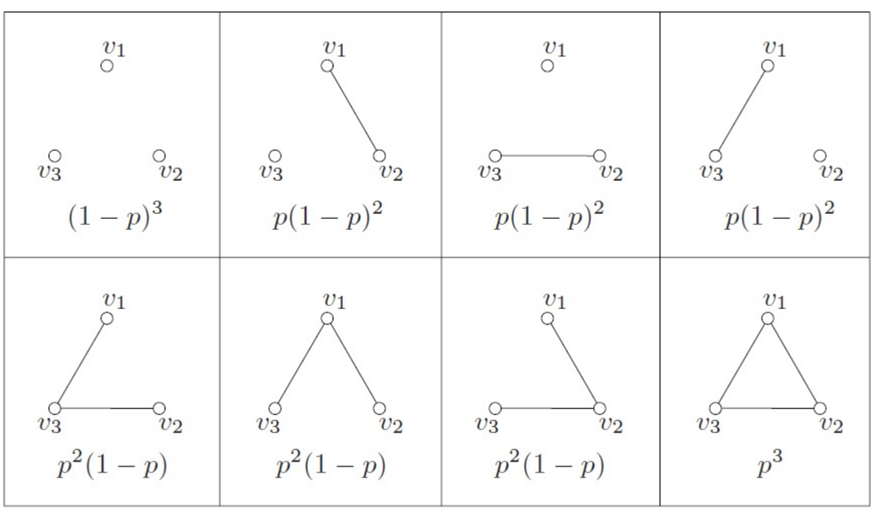

Figure 13.1. The probability space $\mathcal{G}_{3,p}$.

With X as the random variable denoting the number of components $X(G)$ of graph $G \in \mathcal{G}_{3,p}$ we have the expectation

$$E(X) = 3 \times (1-p)^3 + 2 \times 3p(1-p)^2 + 1 \times (3p^2(1-p) + p^3) = 2 - 2p - p^3.$$

Note. Let X and Y be random variables on probability space (Ω, P) and let $r, s \in \mathbb{R}$. Then $rX + sY$ is a random variable on (Ω, P) and

$$E(rX + sY) = \sum_{\omega \in \Omega} (rX(\omega) + sY(\omega))P(\omega)$$

$$= r \sum_{\omega \in \Omega} X(\omega)P(\omega) + s \sum_{\omega \in \Omega} Y(\omega)P(\omega) = rE(X) + sE(Y). \quad (13.4)$$

So the expectation is linear. Also, if X_A is an indicator random variable then

$$E(X_A) = \sum_{\omega \in \Omega} X_A(\omega)P(\omega) = \sum_{\omega \in \Omega, X_A(\omega)=1} P(\omega) = P(X_A = 1). \quad (13.5)$$

Note. Recall that for a drawing $\tilde{G}$ of a graph G in the plane, two edges of $\tilde{G}$ *cross* if they meet at a point other than a vertex of $\tilde{G}$. Each such point is called a *crossing* of the two edges. The *crossing number* of G, denoted $\mathrm{cr}(G)$, is the least number of crossings in a drawing of G in the plane (see Exercise 10.1.8).

Note 13.2.A. Let G be a simple planar graph (notice that random graphs are, by definition, simple). By Corollary 10.21, if $n \geq 3$ then $m \leq 3n - 6$. In particular, if $n \geq 3$ then $m \leq 3n$. In fact, if $n = 1$ then $m = 0$ and this inequality holds. If $n = 2$ then $m \in \{0, 1\}$ and the inequality holds. So for all simple planar graphs we have

$$m \leq 3n, \qquad (*)$$

In Exercise 10.3.1, the inequality $m \leq 3n - 6$ of Corollary 10.21 is to be used to show that for any simple graph G, $\mathrm{cr}(G) \geq m - 3n + 6$. We can similarly use $(*)$ to show that for any simple graph G, $\mathrm{cr}(G) \geq m - 3n$ (which we can also derive from Exercise 10.3.1). Bondy and Murty call the inequality $\mathrm{cr}(G) \geq m - 3n$ the "trivial lower bound" on the crossing number. We use this in the proof of the following stronger lower bound on the crossing number.

Lemma 13.1. THE CROSSING LEMMA.
Let G be a simple graph with $m \geq 4n$. Then $\mathrm{cr}(G) \geq \dfrac{1}{64}\dfrac{m^3}{n^2}$.

Note. The Crossing Lemma was given in 1982 by M. Ajtai, V. Chvátal, M. M. Newborn, and E. Szemerédi and independently in 1983 by F. T. Leighton. The proof given here is due to N. Alon; see N. Alon and J. H. Spencer's *The Probabilistic Method* 2nd Edition, Wiley-Interscience Series in Discrete Mathematics and Optimization (2000). Recently, Eyal Ackerman in "On Topological Graphs with at Most Four Crossings per Edge," *Computational Geometry*, **85**: 101574, 31 pages (2019) proved the related results:

- Let G be a simple graph with $m \geq 7.5n$. Then $\mathrm{cr}(G) \geq \dfrac{1}{33.75}\dfrac{m^3}{n^2}$.
- Let G be a simple graph with $m \geq 7n$. Then $\mathrm{cr}(G) \geq \dfrac{1}{29}\dfrac{m^3}{n^2}$.

Note. L. A. Székely in "Crossing Numbers and Hard Erdös Problems in Discrete Geometry," *Combinatorics, Probability and Computing* **6**, 353–58 (1997) used The Crossing Lemma to easily derive a number of theorems in combinatorial geometry (the existing proofs at the time were complicated). We'll present two of these. Since the proofs are based on The Crossing Lemma, it can be argued that they are based ultimately on the expectation of random variables (though their statements, like The Crossing Lemma, make no explicit mention of random variables or expectations).

Note. To introduce some ideas from combinatorial geometry, consider a set of n points in the plane. For a given $k \in \mathbb{N}$, we could ask how many lines can pass through at least k points. For example, if n is a perfect square and the points are arranged in a square $\sqrt{n} \times \sqrt{n}$ grid, there are $2\sqrt{n} + 2$ lines which pass through $\sqrt{n}$ points (namely, $\sqrt{n}$ horizontal lines passing through each "row" of $\sqrt{n}$ points, $\sqrt{n}$ vertical lines passing through each "column" of $\sqrt{n}$ points, and 2 lines passing through the diagonals of the grid). The next result considers a bound on the number of lines which pass through more than k points.

Theorem 13.2. Let P be a set of n points in the plane, and let ℓ be the number of lines in the plane passing through at least $k+1$ of these points, where $1 \leq k \leq 2\sqrt{2n}$. Then $\ell < 32n^2/k^3$.

Note. The next result concerns the maximum number of pairs of points there can be, among a set of n points in the plane, whose distance is exactly one.

Theorem 13.3. Let P be a set of n points in the plane, and let k be the number of pairs of points of P at unit distance. Then $k < 5n^{4/3}$.

Note. We now consider a sequence of probability spaces (Ω_n, P_n) for $n \in \mathbb{N}$. In particular, we consider the probability spaces $\mathcal{G}_{n,p}$ where p is a function of n and $p(n) \to 0$ as $n \to \infty$ (because "it is with sparse graphs that we are mostly concerned"—see Bondy and Murty page 339).

Definition. Given a sequence (Ω_n, P_n) of probability spaces (where $n \in \mathbb{N}$), a property A is said to be satisfied *almost surely* if $P_n(A_n) \to 1$ as $n \to \infty$, where $A_n = A \cap \Omega_n$.

Definition. If $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ are two functions such that $g(n) > 0$ for n sufficiently large, we write:

$f \ll g$ if $f(n)/g(n) \to 0$ as $n \to \infty$,

$f \gg g$ if $f(n)/g(n) \to \infty$ as $n \to \infty$, and

$f \sim g$ if $f(n)/g(n) \to 1$ as $n \to \infty$.

Note. Markov's Inequality is seen in Mathematical Statistics 1 (MATH 4047/5047). See my online notes for Mathematical Statistics 1 on 1.10. Important Inequalities where a proof of Markov's Inequality is given for a continuous random variable (the case for a discrete random variable being left as an exercise).

Proposition 13.4. Markov's Inequality

Let X be a nonnegative finite random variable on probability space (Ω, P) and $t > 0$. Then $P(X \geq t) \leq \dfrac{E(X)}{t}$.

Note. Since we will usually consider random variables X_n in probability space (Ω_n, P_n) as a quantity, then the following version of Markov's Inequality is sufficient for our purposes.

Corollary 13.5. Let X_n be a nonnegative integer-valued random variable in a probability space (Ω_n, P_n) where $n \in \mathbb{N}$. If $E(X_n) \to 0$ as $n \to \infty$, then $P(X_n = 0) \to 1$ as $n \to \infty$.

Note 13.2.B. Let $G \in \mathcal{G}_{n,p}$ be a random graph and let X be the number of triangles of G. For $S \subseteq V$ with $|S| = 3$, let A_S be the event that the induced subgraph of G, $G[S]$, is a triangle and let X_S be the indicator random variable for A_S. Then $X = \displaystyle\sum_{S\subseteq V,|S|=3} X_S$. Also, $P(A_S) = p^3$ (the probability that the three edges of the triangle with vertex set S are in G). By linearity of expectation and equation (13.5),

$$E(X) = E\left(\sum_{S\subseteq V,|S|=3} X_S\right) = \sum_{S\subseteq V,|S|=3} E(X_S)$$

$$= \binom{n}{3} p^3 = \frac{n(n-1)(n-2)}{6} p^3 = \frac{n^2 - 3n + 2}{6} np^3 < (np)^3$$

since $f(n) = n^2 - (n^2 - 3n + 2)/6$ is strictly increasing (since $f'(n) = 2n - (2n-3)/6 = (n+3)/6 > 0$ for $n \geq 1$) and $f(1) = 1$, so that $n^2 - (n^2 - 3n + 2)/6 \geq 1 > 0$ for $n \geq 1$. So if $pn \to 0$ as $n \to \infty$ then G almost surely is triangle-free.

Note. Recall that a stable set (or an "independent set") in a graph G is a set of vertices no two of which are adjacent. The cardinality of a maximum stable set in G is the stability number of G, denoted $\alpha(G)$ (see Section 12.1. Stable Sets).

Theorem 13.6. A random graph in $\mathcal{G}_{n,p}$ almost surely has stability number at most $\lceil 2p^{-1} \log n \rceil$.

Note. Other applications of the probabilistic method are given in the exercises (especially applications to hypergraphs). In commemoration of Dr. Teresa Haynes, a frequent instructor of this class at ETSU between 2000 and 2018, notice that a "dominating set" in a graph is defined in Exercise 13.2.13 (Dr. Haynes' research specialty centers on domination theory; see 1.5. Directed Graphs for a reference to a couple of her books).

Revised: 3/6/2021