THE PROBABILISTIC METHOD AND RANDOM GRAPHS

by

Brian Ketelboeter

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Mathematics
University of Manitoba
Winnipeg

UNIVERSITY OF MANITOBA

DEPARTMENT OF DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "*The probabilistic method and random graphs*" by *Brian Ketelboeter* in partial fulfillment of the requirements for the degree of *Master of Science*.

Dated: ____________________________

Supervisor: ____________________________

David Gunderson

Readers: ____________________________

Dr Ben Li

____

Dr M. Doob

UNIVERSITY OF MANITOBA

Date: *September 2012*

Author: *Brian Ketelboeter*

Title: *The probabilistic method and random graphs*

Department: *Department of Mathematics*

Degree: *M.Sc.*

Convocation: *October*

Year: *2012*

Signature of Author

Abstract

The probabilistic method in combinatorics is a nonconstructive tool popularized through the work of Paul Erdős. Many difficult problems can be solved through a relatively simple application of probability theory that can lead to solutions which are better than known constructive methods.

This thesis presents some of the basic tools used throughout the probabilistic method along with some of the applications of the probabilistic method throughout the fields of Ramsey theory, graph theory and other areas of combinatorial analysis.

Then the topic of random graphs is covered. The theory of random graphs was founded during the late fifties and early sixties to study questions involving the effect of probability distributions upon graphical properties. This thesis presents some of the basic results involving graph models and graph properties.

Contents

Chapter 1

Introduction

This chapter introduces the topics of the probabilistic method and random graphs, along with an outline of this thesis.

1.1 The probabilistic method

The probabilistic method is a powerful tool in discrete mathematics. The essential idea is as follows: Trying to prove a structure with a collection of properties exists, define an appropriate sample space, Ω, a probability measure $\mathbb{P}$, and show that the desired properties hold with positive probability. Here is an example of the probabilistic method at work.

For every positive integer k, the Ramsey number $R(k,k)$ is the smallest integer n such that every two-colouring of the edges of the complete graph, K_n, there is a complete subgraph, K_k, such that all of the edges are either red or all of the edges are blue. F. P. Ramsey [32] showed that for every positive integer, k, $R(k,k)$ exists.

Claim 1. If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ then $R(k,k) > n$.

A proof by Erdős [8] showed that if n satisfies the hypothesis, then with positive probability there exists a two-colouring of the edges of K_n, with no monochromatic K_k. One advantage of Erdős' proof is that the proof contains no constructions and few calculations; whereas a constructive proof of such a theorem would involve finding a colouring that fails to have a monochromatic K_k. For more on constructive methods in Ramsey theory, see Radziszowski [31]. See Chapter 4 for the proof of Claim 1 and other probabilistic results in Ramsey theory.

1.2 Random graphs

In their papers (see *e.g.* [14]) of 1959-60, the Hungarian mathematicians Paul Erdős and Alfred Rényi popularized the methods that underlie the foundations of the theory of random graphs. Their idea was to use probabilistic and statistical methods to study limiting behavior of graph theoretic properties.

One approach to better understand the notion of random graphs is to think of a random graph as a living organism that evolves with time. Let p be the probability that any two vertices in a graph are joined by an edge. Upon being born, $p = 0$, the graph is just a collection of isolated vertices with no other structure. As p starts to grow, the graph gains edges with the 'typical' graph having $\binom{n}{2}p$ edges. While $p < \frac{1}{\binom{n}{2}}$, the average graph is a forest of small trees. As p continues to grow, the trees get larger along with small and then larger cycles start to appear. Around $p = \frac{1}{n}$ the graph starts forming larger connected components; coalescing into yet larger components eventually forming a giant component made up of nearly all the vertices.

Part of what made these papers so fundamental to the development of random graphs was the ability to present many different and interesting questions to work on, some of which are studied in this thesis. The first topic covered is threshold functions. Let Q be a property of graphs. The function $r\colon \mathbb{Z}^+ \to [0,1]$ is called a threshold function if for large positive integers n, if $p < r(n)$, the probability of a graph having Q goes to zero, while for $r(n) < p$, the probability of a graph having Q goes to one.

As an example of a threshold function, let Q be the property 'G has an isolated vertex. For every n, the graph on n vertices with no edges, $\overline{K_n}$, has isolated vertices while the complete graph, K_n, has no isolated vertices and there are many graphs between these two extremes that both have and don't have Q. When $0 \leq p$ is smaller than $r(n)$, the probability of a graph having an edge is less than 1/2 so the graphs with fewer edges have higher probabilitygraphs with higher probability has few edges so is likely to have isolated vertices. As p gets closer to one, the typical graph has a larger number of edges and therefore is less likely to have isolated vertices.

As Erdős and Rényi [9] explain, while the study of random graphs is interesting in it's own right, the evolutionary behavior of graphs may be considered a simplified model of the evolution of more sophisticated structures consisting of vertices and connections such as railways or communication networks. In a recent paper by Newmann *et al.* [29] some of the more recent applications of the random graphs phenomena are explored. The writers explain that random graphs "have been employed extensively as models of real world networks of various types", with particular success in the field of epidemiology. The spread of a disease through a "community depends strongly on the pattern of contacts" between the individuals infected with the disease and those who are susceptible to it. A good model for this pattern is a network, "with individuals represented by vertices and contacts capable of transmitting the disease by edges". This 'class of epidemiological models are known as susceptible/infectious/recovered (or SIR) models' often use the so-called "fully mixed approximation", in which the assumption is that "contacts are random and uncorrelated, *i.e.*, that they form a random graph". Thus the concept of random graphs has moved on from being observations regarding properties of graphs to form a basis to the study of real world problems in such fields as sociology, biology and computer science.

1.3 Layout of Thesis

The second chapter introduces basic terminology used throughout this thesis along with mathematical necessities. The third chapter presents elements of probability theory along with several theorems that are the basis for the methods displayed in this thesis. Chapter four presents an outline of four different methods commonly deployed within combinatorial probabilistic arguments. The fifth chapter presents some examples of applications of probability to various topics in general combinatorics including Ramsey theory, combinatorial set theory and combinatorial geometry and graph theory.

Chapters six through eight present some results in the theory of random graphs. Chapter six introduces the probability models $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,q}$ of graphs, along with a 0-1 law, a combination of logic, graph theory and probability theory. Chapter seven presents the idea of threshold functions and applies them to some graphical properties along with a theorem for balanced graphs, allowing for the calculation of thresholds for a large class of properties. Chapter 8.1 examines what occurs in the neighborhood of threshold functions and shows that some graph properties converge in distribution to the Poisson distribution. Section 8.2 then introduces the theory of graphical evolution, which sumarises the ideas of random graphs into one theory (see *e.g.* [4]). Chapter nine presents the material to complete the computations in 8.1 and 8.2.

Chapter 2

Notation and mathematical tools

This chapter contains the definitions, symbols and mathematical tools necessary for the entire text. Throughout this text, the integers, rational and real numbers are denoted by $\mathbb{Z}, \mathbb{Q}$, and $\mathbb{R}$, respectively. The collection of positive integers $\{1, 2, 3, \ldots\}$ is denoted by $\mathbb{Z}^+$. For integers $a < b$, the notation $[a, b] = \{x \in \mathbb{Z} : a \leq x \leq b\}$ is used for an interval of integers. In particular, $[1, n] = \{1, 2, \ldots, n\}$, is often abbreviated by $[n]$. Given $n \in \mathbb{Z}^+$, the permutation group of n elements is denoted by S_n.

For the most part, functions in this script are real-valued, (often integer-valued), and have $\mathbb{Z}^+$ or $[n]$ as their domain. For example, $t(n)$ might be the number of graphs on n vertices that have no triangle. To compare two such functions and their asymptotic rates of growth, certain notation is helpful. For two functions f and g, write $f = \mathrm{o}(g)$ [read "f is little oh of g"] or $f(n) = \mathrm{o}(g(n))$ if and only if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. For example, $\ln x = \mathrm{o}(x^2)$, and $\frac{1}{n} = \mathrm{o}(1)$. Hence, the notation $f(n) = (1 + \mathrm{o}(1))g(n)$ means that f and g are approximately equal for large n, that is, $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$, in which case one often writes $f \sim g$.

When f is eventually bounded above by some fixed multiple of g, another notation, called the "big oh" notation is used: write $f(n) = \mathrm{O}(g(n))$ if there exists a positive constant $C \in \mathbb{R}$ and an $n_0 \in Z^+$ so that for all $n > n_0$, $f(n) \leq Cg(n)$. For example, $x^2 + 1 = \mathrm{O}(3x^2 + 14x)$ (*e.g.*, take $C = 1$, and $n_0 = 10$). Turning the big oh notation inside out, define $f = \Omega(g)$ if $g = \mathrm{O}(f)$. If both $f = \mathrm{O}(g)$ and $f = \Omega(g)$, write $f = \Theta(g)$; this essentially describes the situation where for some constants c and C, f and g satisfy $cg \leq f \leq Cg$. It is often convenient to abbreviate an expression like $\lim_{n\to\infty} f(n) = L$ by $f(n) \to L$. The notation $f(n) \to \infty$ is reserved to describe functions whose values get arbitrarily large as n gets large.

2.1 Set theory

In this section, the definitions and notation are given and is not intended to be a set theory primer; for further information on set theory, the text [22] is a comprehensive source. For any set Ω, the *power set* of Ω is the set of all subsets of X; throughout this text the notation $\mathcal{P}(\Omega)$ is used for the power set of Ω. For any set $A \in \mathcal{P}(\Omega)$, by A^c is meant the set of $x \in \Omega$ such that $x \notin A$ and if $A, B \in \mathcal{P}(\Omega)$ denote $A \setminus B = \{x \in A : x \notin B\}$. Thus $A^c = \Omega \setminus A$. For any set X, and any $k \in \mathbb{Z}^+$, $[X]^k = \{S \subseteq X : |S| = k\}$.

Given a set X and an $r \in \mathbb{Z}^+$, an *r-colouring* is a function $\chi : X \to [r]$; while for $i \in [r]$, the set $\chi^{-1}(i)$ is the i-th colour class. A set $Y \subset X$ is said to be monochromatic under a colouring χ iff there exists an i such that $Y \subseteq \chi^{-1}(i)$. If $\chi : X \to [2]$ is a two-colouring of X, it is common practice to replace the set $\{1, 2\}$ by the set $\{\text{red, blue}\}$; if $Y \subseteq X$, the statement "Y is monochromatic" is often replaced with the statement "Y is red" or "Y is blue".

Definition 2.1.1. *(Total Ordering)*
A set X is totally ordered by the relation $\leq$ iff $\leq$ is a reflexive, antisymmetric and transitive binary relation.

2.2 Graph theory

Many of the applications of the probabilistic method spotlighted here involve graph theory. In this section, the notation and basic graph theory used in this thesis is presented. For further information, see any basic text on graph theory (*e.g.* [39] or [7]).

A *graph* is an ordered pair $G = (V, E)$, where $V = V(G)$ is a non-empty set whose elements are called *vertices*, and $E = E(G) \subseteq [V]^2$ is a set of unordered pairs of vertices; elements of E are called *edges*. An edge $e = \{x, y\} \in E(G)$ is said to join x and y; also x and y are end points of e, x and y are incident with e, or x and y are adjacent. For each $n \in \mathbb{Z}^+$, if $|V| = n$ then the graph $K_n = (V, [V]^2)$ is called the *complete graph* on n vertices. A *directed graph* (or digraph) is a pair $D = (V, E)$ of vertices, V and a set E consisting of ordered pairs of vertices, each such pair in E is called a directed edge or arc.

If $G = (V, E)$ is a graph, a graph $H = (V', E')$ is called a *subgraph* of G if $V' \subseteq V$ and $E' \subseteq ([V']^2) \cap E$; if $E' = ([V']^2) \cap E$, H is called the subgraph of G induced by V'. If $S \subset V$, denote K_S as the complete subgraph on S. A set $S \subseteq V$ is called *clique* (or a *k-clique* if $|S| = k$) when the subgraph induced by S is K_S. A set $V' \subseteq V$ is *independent* if $[V']^2 \cap E = \emptyset$ and $\alpha(G)$ is the size of the largest independent subset of V.

A *walk* in a graph $G = (V, E)$ is an alternating sequence $v_1 e_1 v_2 e_2 \ldots e_{m-1} v_m$ of vertices and edges (not necessarily distinct) so that for each $i = 1, 2, \ldots m-1$, $e_i = \{v_i, v_{i+1}\}$; such a walk has length $m - 1$. A *closed walk* is a walk where $v_1 = v_m$. A *trail* is a walk with no edge repeated. A *path* is a trail with no vertex repeated. A *cycle* is a closed walk with no repeated vertices (except the first and last). A graph is *connected* if there is a path between every pair of vertices. A subset, S, of vertices is called a *clique* if the induced subgraph is complete.

A *tournament* is an ordered pair $\mathcal{T} = (V, D)$ where $V = V(\mathcal{T})$ is set whose elements are called vertices (or players) and $D = D(\mathcal{T}) \subset V \times V$ where for every $v_i \neq v_j \in V$, either $(v_i, v_j) \in D$ or $(v_j, v_i) \in D$ but not both. A graph $G = (V, E)$ is *bipartite* if there exists $U, W \subseteq V$, disjoint such that $U \cup W = V$ and for every edge e $= \{x, y\} \in E$ either $x \in U$ and $y \in W$ or vice versa. The *girth* of a graph, $G = (V, E)$, (denoted by $\text{girth}(G)$) is the length of the shortest cycle in G. For every graph G, the nonnegative integer $\chi(G)$, called the *chromatic number of* G, is the least integer such that there exists a $\chi(G)$-colouring of $E(G)$ such that no pair of adjacent vertices are monochromatic. The following well known fact is needed later.

Lemma 2.2.1. *For every graph G,*

$$\chi(G) \geq \frac{n}{\alpha(G)}. \tag{2.1}$$

Proof. If χ is any $\chi(G)$ colouring of V, the colour classes partition V into independent sets of size at most $\alpha(G)$. Thus

$$\chi(G)\alpha(G) \geq n.$$

□

2.3 Useful approximations

As Alon and Spencer explain in [1], part of the art of the probabilistic method is the deduction of bounds that may not be the best possible but allow for cleaner and simpler proofs. As is seen throughout this document, approximations are often necessary in the proofs.

2.3.1 Approximating the exponential

To approximate the exponential, recall

$$\forall x \in \mathbb{R} \quad e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{2.2}$$

is the Taylor series at x of the exponential function.

Lemma 2.3.1. *For all $x \in [0, \infty)$, $1 + x \leq e^x$.*

Proof. Truncate equation (2.2) after the second term. $\square$

Lemma 2.3.2. *For all $d \in (0, \infty)$, $\left(1 - \frac{1}{d+1}\right)^d \geq e^{-1}$.*

Proof. Assume $d > 0$ and put $x = 1/d$ in inequality (2.3.1) to get

$$\frac{d+1}{d} = 1 + \frac{1}{d} \leq e^{1/d}.$$

Thus,

$$\left(\frac{d+1}{d}\right)^d \leq e.$$

Hence

$$e^{-1} \leq \left(\frac{d}{d+1}\right)^d = \left(1 - \frac{1}{d+1}\right)^d. \qquad \square$$

2.3.2 Approximating the logarithm

Another function for which approximations are necessary is the natural log, denoted throughout this text as ln. Since

$$1 - y^{n+1} = (1 - y)(1 + y + \ldots + y^n)$$

or

$$\frac{1}{1-y} = \frac{1 + y + \ldots + y^n}{1 - y^{n+1}}. \tag{2.3}$$

When $|y| < 1$, then the limit as n goes to infinity in equation (2.3) exists so that if $|y| < 1$,

$$\begin{aligned} \frac{1}{1-y} &= \lim_{n \to \infty} \sum_{k=0}^{n} y^k \\ &= \sum_{k=0}^{\infty} y^k. \end{aligned}$$

Integrating both sides yields

$$\ln(1 - y) = -\sum_{k=1}^{\infty} \frac{y^k}{k} \tag{2.4}$$

or replacing y by $-y$,

$$\ln(1+y) = -\sum_{k=1}^{\infty} \frac{(-y)^k}{k}. \tag{2.5}$$

2.3.3 Binomial coefficients

For the positive integers, $k \leq n$, recall the following formulas:

$$n! = n(n-1)(n-2)\cdots 2 \cdot 1;$$

$$(n)_k = n(n-1)(n-2)\cdots(n-k-1);$$

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}.$$

The notation $(n)_k$ is often called a falling factorial. To conclude this section, here are a few more useful results whose proofs won't be provided here.

Lemma 2.3.3. *If $k = \mathrm{o}(n^{1/2})$, then*

$$\frac{(n)_k}{n^k} \sim \mathrm{e}^{-(\frac{k^2}{n} + \frac{k^3}{2n^2})}. \tag{2.6}$$

Lemma 2.3.4. (See for example [16])(Stirling's approximation of $n!$)

$$n! = (1 + \mathrm{o}(1))\sqrt{2n\pi}\left(\frac{n}{\mathrm{e}}\right)^n. \tag{2.7}$$

Theorem 2.3.5. (Binomial theorem) *Let $x, y \in \mathbb{R}$, then for every $n \in \mathbb{Z}^+$;*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{2.8}$$

Lemma 2.3.6. (See for example [7]) *For all $t \in \mathbb{Z}^+$, $\binom{n}{t} \leq \left(\frac{ne}{t}\right)^t$.*

Proof. (Of Lemma 2.3.6) Let $t \in \mathbb{Z}^+$ be given. Theorem 2.3.5 implies

$$\left(1 + \frac{t}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{t}{n}\right)^k \leq \mathrm{e}^t.$$

Therefore as $t > 0$,

$$\binom{n}{t} \left(\frac{t}{n}\right)^t \leq \mathrm{e}^t.$$

Hence

$$\binom{n}{t} \leq \left(\frac{n\mathrm{e}}{t}\right)^t.$$

□

Chapter 3

Elementary probability

This chapter covers the probability basics that are used throughout this thesis. As it is assumed that a reader knows nothing about probability theory, the definitions needed, along with properties, are given. This chapter concludes with a few probability inequalities that are necessary for this thesis.

3.1 Measure theory

Measure theory is the basis of modern probability theory, so this section provides the needed measure theoretic ideas.

Definition 3.1.1. *Given a set Ω, a collection $\Sigma \subseteq \mathcal{P}(\Omega)$ is called an algebra* iff*:*

i. $\Omega \in \Sigma$.

ii. *If $A \in \Sigma$ then $A^c \in \Sigma$.*

iii. *If $A, B \in \Sigma$ then $A \cup B \in \Sigma$.*

An algebra Σ is called a σ-algebra if Σ is closed under countable unions.

Definition 3.1.2. *If $\Sigma \subseteq \mathcal{P}(\Omega)$ is a σ-algebra, a function $\mu : \Sigma \to [0, \infty]$ is called a measure* iff*:*

i. $\mu(\emptyset) = 0$.

ii. *If* $A \subseteq B$*, then* $\mu(A) \leq \mu(B)$*. In this case,* μ *is said to be monotone.*

iii. *If* $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{P}(\Omega)$ *are pairwise disjoint, then* $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$*. In this case,* μ *is said to be countably additive.*

Lemma 3.1.3 (Properties of a measure)**.** *Let* Σ *be a* σ*-algebra. Suppose* $\mu : \Sigma \to [0, \infty]$ *is a measure and* $A, B \in \Sigma$*.*

1. If $\{A_i\}_{i=1}^{n} \subset \Sigma$ *are disjoint then* $\mu(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$*.*

2. If $B \subset A$ *and* $\mu(B) < \infty$ *then* $\mu(A \setminus B) = \mu(A) - \mu(B)$*.*

3. (a measure is countably subadditive) If $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ *(not necessarily disjoint) then*

$$\mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

4. If $A \in \Sigma$ *then* $\mu(A) + \mu(A^c) = \mu(\Omega)$*.*

Proof. 1. Assume $\{A_i\}_{i=1}^{n} \subseteq \Sigma$ are mutually disjoint and for all $i > n$, define $A_i = \emptyset$, thus $\{A_i\}_{i=1}^{\infty} \subseteq \Sigma$ are mutually disjoint. Since the union of a set with any number of empty sets doesn't change the set,

$$\cup_{i=1}^{n} A_i = \cup_{i=1}^{\infty} A_i.$$

As μ is a measure,

$$\begin{aligned}
\mu(\cup_{i=1}^{n} A_i) &= \mu(\cup_{i=1}^{\infty} A_i) \\
&= \sum_{i=1}^{\infty} \mu(A_1) && \text{(Definition 3.1.2 } (iii)\text{)} \\
&= \sum_{i=1}^{n} \mu(A_i) + \sum_{i=n+1}^{\infty} \mu(A_i) \\
&= \sum_{i=1}^{n} \mu(A_i) + 0 && \text{(Definition 3.1.2 } (i)\text{)} \\
&= \sum_{i=1}^{n} \mu(A_i)
\end{aligned}$$

as claimed.

2. If $B \subset A$ then $A = B \cup (A \setminus B)$ is a disjoint union. Property 1 implies

$$\begin{aligned} \mu(A) &= \mu(B \cup (A \setminus B) \\ &= \mu(B) + \mu(A \setminus B). \end{aligned}$$

Therefore

$$\mu(A \setminus B) = \mu(A) - \mu(B) \qquad\qquad (\mu(B) < \infty),$$

showing Lemma 3.1.3 (2).

3. Let $B_1 = A_1$ and for $1 < k$, $B_k = A_k \setminus (\cup_{i=1}^{k-1} A_i)$.

$$\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i.$$

Definition 3.1.2 (ii) implies:

$$\forall i \quad \mu(B_i) \leq \mu(A_i).$$

Thus

$$\begin{aligned} \mu(\bigcup_{i=1}^{\infty} A_i) &= \mu(\bigcup_{i=1}^{\infty} B_i) \\ &= \sum_{i=1}^{\infty} \mu(B_i) \qquad\qquad \text{(Definition 3.1.2 (iii))} \\ &\leq \sum_{i=1}^{\infty} \mu(A_i), \end{aligned}$$

as claimed.

4. As $\Omega = A \cup A^c$ is a disjoint union, property 1 implies:

$$\mu(\Omega) = \mu(A) + \mu(A^c),$$

as claimed.

$\square$

3.1.1 Probability theory

Definition 3.1.4. *A probability space is a triple* $(\Omega, \Sigma, \mathbb{P})$ *where* Ω *is a set,* Σ *is a* $\sigma-$*algebra of subsets of* Ω *and* $\mathbb{P} : \Sigma \to [0,1]$ *is a measure such that* $\mathbb{P}(\Omega) = 1$.

If Ω is a finite set, the probability space $(\Omega, \Sigma, \mathbb{P})$ is called a *finite probability space.*

Example 3.1.5. Given a two sided coin, with one side 'heads', (h), and one side 'tails', (t), the ordered triples of (h) and (t)

$$\Omega = \{(h,h,h),(h,h,t),(h,t,h),(t,h,h),(h,t,t),(t,h,t),(t,t,h),(t,t,t)\}$$

is a representation of the sample space of three flips of the coin.
Let $\Sigma = \mathcal{P}(\Omega)$ and define $\mathbb{P} : \Sigma \to [0,1]$ by $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$. The triple, $(\Omega, \Sigma, \mathbb{P})$, is a probability space.

Example 3.1.6. Let $\Omega = [5]$ and $\Sigma = \mathcal{P}(\Omega)$. Define $\mathbb{P} : \Sigma \to [0,1]$ by $\mathbb{P}(A) = \frac{|A|}{5}$. The triple $(\Omega, \Sigma, \mathbb{P})$ is a probability space.

Unless otherwise specified, the probability spaces used in this thesis are finite and $\Sigma = \mathcal{P}(\Omega)$. When $\Sigma = \mathcal{P}(\Omega)$ is assumed, the notation $(\Omega, \mathbb{P})$ is used to denote a probability space.

In order to better explain the probabilistic method, one technique is to use the terminology of statistics. To apply the probabilistic method, first define an experiment and sample space, Ω, which is the collection of all possible outcomes. Define a measure $\mathbb{P}$ on Ω representing the probability that $G \in \Omega$ is a result of the experiment. Next, define the σ-algebra, $\Sigma \subseteq \mathcal{P}(\Omega)$ of desired possible events that could result in the given experiment. The event $A \in \Sigma$ is said to have occurred if the resultant of the experiment is some $G \in A$.

The probability of the event A occurring as a result of the experiment is the probability of some outcome $G \in A$ being the resultant of the experiment; calculated by $\mathbb{P}(A) = \sum_{G \in A} \mathbb{P}(G)$. The event $A = \emptyset$ expresses the result the event that A is impossible and $A = \Omega$ is the event of any possible outcome occurring; explaining why in Definition 3.1.4, $\mathbb{P}[\emptyset] = 0$, as an impossible event can not occur; while $\mathbb{P}[\Omega] = 1$ as the experiment must yield some outcome.

Given any two events $A, B \subset \Omega$, some ways in which other events can be formed include $A \wedge B$, the event that both A and B occur; $A \vee B$, the event of outcomes either in A or B (or both). The event corresponding to all of the outcomes not in A is denoted as $\overline{A}$; $A \wedge \overline{B}$, the event of outcomes in A and not in B; while the event of all outcomes only in A or only in B and not both is denoted as $A \wedge \overline{B} \vee B \wedge \overline{A}$ or $A \triangle B$. In the case that $A \wedge B = \emptyset$, then A and B are called *mutually exclusive*.

Example 3.1.7. Using Example 3.1.5, let A, B and C be the events 'first flip is a head', 'second flip is a head' and 'third flip is a head' respectively. The event $A \vee B$ is 'either the first or second flip is a head' while $A \wedge B$ is the event 'both the first and second flip is a head' and $A \wedge B \wedge C$ is the event 'all three flips are heads'. Clearly, if D is the event 'first flip is a tail', A and D are mutually exclusive while B and D are not. Furthermore, observing that $A = \{(h,h,h),(h,h,t),(h,t,h),(h,t,t)\}$, $B = \{(h,h,h),(h,h,t),(t,h,h),(t,h,t)\}$, $A \vee B = \{(h,h,h),(h,h,t),(h,t,h),(h,t,t),(t,h,h),(t,h,h)\}$ and $A \wedge B = \{(h,h,h),(h,h,t)\}$ showing that

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{1}{2} = \mathbb{P}[B]$$

and

$$\mathbb{P}[A \vee B] = \frac{3}{4}$$

while

$$\mathbb{P}[A \wedge B] = \frac{1}{4}.$$

Lemma 3.1.8. *Given any probability space* $(\Omega, \Sigma, \mathbb{P})$

$$\forall A, B \in \Sigma, \quad \mathbb{P}[A \vee B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \wedge B].$$

Proof. As $A \vee B = (A \triangle B) \vee (A \wedge B)$ and $\mathbb{P}$ is a measure,

$$\begin{aligned}
\mathbb{P}[A \vee B] &= \mathbb{P}[(A \triangle B) \vee (A \wedge B)] \\
&= \mathbb{P}[A \triangle B] + \mathbb{P}[A \wedge B] && \text{(Definition 3.1.2(2))} \\
&= \mathbb{P}[(A \wedge \overline{B}) \vee (B \wedge \overline{A})] + \mathbb{P}[A \wedge B] && \text{(definition of } A \triangle B) \\
&= \mathbb{P}[A \wedge \overline{B}] + \mathbb{P}[B \wedge \overline{A}] + \mathbb{P}[A \wedge B] && \text{(Lemma 3.1.3).}
\end{aligned}$$

Since $A = (A \wedge \overline{B}) \vee (A \wedge B)$ are disjoint,

$$\begin{aligned}
\mathbb{P}[A] &= \mathbb{P}[A \wedge \overline{B}] + \mathbb{P}[A \wedge B] && \text{(Lemma 3.1.3)} \\
\mathbb{P}[A \wedge \overline{B}] &= \mathbb{P}[A] - \mathbb{P}[A \wedge B] && (\mathbb{P}[A \wedge \overline{B}] \leq 1).
\end{aligned}$$

Put both equations together yields

$$\begin{aligned}
\mathbb{P}[A \vee B] &= (\mathbb{P}[A] - \mathbb{P}[A \wedge B]) + (\mathbb{P}[B] - \mathbb{P}[B \wedge A]) + \mathbb{P}[A \wedge B] \\
&= \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \wedge B].
\end{aligned}$$ □

3.1.2 Conditional probability and independence of events.

Definition 3.1.9. *Given the probability space* $(\Omega, \mathbb{P})$ *and two events* $A, B \in \Omega$. *If* $\mathbb{P}[B] \neq 0$, *define*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \wedge B]}{\mathbb{P}[B]}, \tag{3.1}$$

called the conditional probability of A *given that* B *has occurred.*

Example 3.1.10. Continuing Example 3.1.6, define event $B = \{x \in \Omega : x \text{ is odd}\}$ and events $A_1 = \{y \in \Omega : y \text{ is divisible by } 3\}$ and $A_2 = \{z \in \Omega : z \text{ is divisible by } 2\}$ then

$$P[A_1|B] = 1/3 \quad \text{and} \quad P[A_2|B] = 0.$$

Definition 3.1.11. *Given a probability space* $(\Omega, \mathbb{P})$, *the events* A, B *are called independent events if*

$$\mathbb{P}(A \wedge B) = \mathbb{P}(A)\mathbb{P}(B).$$

For $1 < n \in \mathbb{Z}^+$, a collection of events $\{B_i\}_{i=1}^n \subset \Omega$ are mutually independent if for every $1 \leq k \leq n$ and every $S \in [n]^k$

$$\mathbb{P}[\bigwedge_{i \in S} B_i] = \prod_{i \in S} \mathbb{P}[B_i]$$

If events A, B are independent, equation (3.1) becomes $\mathbb{P}[A|B] = \mathbb{P}[A]$.

Example 3.1.12. (See [34]) For the collection of events $\{A_i\}_{i=1}^n$ to be independent, it is not enough that each pair be mutually independent. Consider the following game: an urn contains four balls numbered $1, 2, 3, 4$ respectively. One ball is pulled at uniformly and randomly. Each player picks two numbers between 1 and 4 and a player wins an award if one of their numbers are chosen. Three people Tom, Dick and Harry decide to play. Tom chooses 1 and 2, Dick chooses 1 and 3 while Harry chooses 1 and 4. Let T be the event Tom wins, D be the event Dick wins and H be the event Harry wins. Then

$$\mathbb{P}[T \wedge D] = \mathbb{P}[T \wedge H] = \mathbb{P}[D \wedge H] = 1/4$$

while

$$\mathbb{P}[T \wedge D \wedge H] = 1/4 \neq \mathbb{P}[T] \cdot \mathbb{P}[D] \cdot \mathbb{P}[H] = 1/8,$$

thus the above events are pairwise independent but not mutually independent.

The following is an example of three events, A, B_1 and B_2 such that A is independent of both B_1 and B_2 but not of $B_1 \wedge B_2$.

Example 3.1.13. (See [23, p.222])Flip a fair coin twice. Let B_1 be the event of heads on the first flip, B_2 be the event of heads on the second flip and A be the event where both flips come up the same then

$$\mathbb{P}[A|B_1] = \mathbb{P}[A|B_2] = 1/2,$$

while

$$\mathbb{P}[A|B_1 \wedge B_2] = 1.$$

3.2 Random variables, expectation and variance

3.2.1 Random variables

Definition 3.2.1. *A real valued random variable X on a probability space $(\Omega, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$. If X is random variable on the probability space $(\Omega, \mathbb{P})$ and $a \in \mathbb{R}$, the event $A = \{G : X(G) = a\}$ is denoted by $\{X = a\}$.*

In general, a random variable is defined as a measurable function. As this thesis covers finite probability spaces $(\Omega, \mathbb{P})$, every function on Ω is measurable and so the added definition of what it means for a function to be measurable won't be discussed here. For a general discussion of this and other probability theory topics, please see most books on real analysis or measure theory such as [35] or [21].

If X, Y are two random variables on $(\Omega, \mathbb{P})$, call X and Y *independent* if for every $a, b \in \mathbb{R}$ and $A = \{X = a\}$, $B = \{Y = b\}$ then A and B are independent.

Example 3.2.2. Let $(\Omega, \mathbb{P})$ be a probability space.

(i) For any event A, define

$$X_A(G) = \begin{cases} 0 & : \text{ if } G \notin A; \\ 1 & : \text{ if } G \in A; \end{cases}$$

called the indicator random variable of A.

(ii) If X has a finite range, $\{x_i\}_{i=1}^n$, and let $A_i = \{X(G) = x_i\}$ then $X = \sum_{i=1}^n x_i X_{A_i}$.

(iii) Let Ω be as in Example 3.1.5, let $X : \Omega \to \mathbb{R}$ be defined by $X(G)$ is the number of heads in G. If $G_1 = (h, h, t)$, $G_2 = (t, h, t)$, then $X(G_1) = 2$ and $X(G_2) = 1$.

3.2.2 Expectation of random variables

Definition 3.2.3. *Given a finite probability space* $(\Omega, \mathbb{P})$*, the real valued function defined on the collection of random variables (on* Ω*) by*

$$\mathbb{E}[X] = \sum_{G \in \Omega} X(G)\mathbb{P}(G)$$

is called the expectation value of X *or the first moment of* X*.*

Properties. If $(\Omega, \mathbb{P})$ is a finite probability space, $a, b \in \mathbb{R}$ and X and Y are random variables on Ω, then:

(i) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. This property is denoted as the *linearity of expectation.*

Proof.

$$\begin{aligned}
\mathbb{E}[aX + bY] &= \sum_{G \in \Omega} (aX + bY)(G)\mathbb{P}[G] \\
&= \sum_{G \in \Omega} (aX+bY)(G)\mathbb{P}(G) \\
&= \sum_{G \in \Omega} (aX(G)+bY(G))\mathbb{P}(G) \\
&= \sum_{G \in \Omega} aX(G)\mathbb{P}(G)+\sum_{G \in \Omega} bY(G)\mathbb{P}(G) \\
&=a\sum_{G \in \Omega} X(G)\mathbb{P}(G) + b\sum_{G \in \Omega} Y(G)\mathbb{P}(G) \\
&=a\mathbb{E}[X] + b\mathbb{E}[Y].
\end{aligned}$$

□

(ii) If $a \leq X \leq b$ then $a \leq \mathbb{E}[X] \leq b$.

Proof. If $a \leq X$ then for all $G \in \Omega$, $a \leq X(G)$. Therefore

$$\begin{aligned}
a = a\sum_{G \in \Omega} \mathbb{P}(G) & \qquad (\textstyle\sum_{G \in \Omega} \mathbb{P}(G) = 1) \\
\leq \sum_{G \in \Omega} X(G)\mathbb{P}(G) & \qquad (a \leq X(G))
\end{aligned}$$

$$= E[X].$$

The proof is similar for $X \leq b$. $\square$

(iii) $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.

Proof. Let $Y = X - E[X]$. Then

$$\begin{aligned} 0 &\leq E[Y^2] \\ &= \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \qquad \text{(linearity of expectation)} \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2. \end{aligned}$$ $\square$

(iv) If $\{x_i\}_{i=1}^n \subset \mathbb{R}$ is the range of X, then $\mathbb{E}[X] = \sum_{i=1}^n x_i \mathbb{P}(X = x_i)$.

Proof. Let $A \subset \Omega$ and X be the indicator random variable for A. Then

$$\begin{aligned} \mathbb{E}[X] &= \sum_{G \in \Omega} X(G)\mathbb{P}(G) \\ &= \sum_{G \in A} X(G)\mathbb{P}(G) + \sum_{G \in \bar{A}} X(G)\mathbb{P}(G) \\ &= \sum_{G \in A} 1 \times \mathbb{P}(G) = \sum_{G \in \Omega} 1 \times \mathbb{P}(X = 1). \end{aligned}$$

Assume $\{x_i\}_{i=1}^n$ is the range of X and $A_i = \{X = x_i\}$ then $X = \sum_{i=1}^n x_i X_{A_i}$. Linearity of expectation implies

$$\mathbb{E}[X] = \sum_{i=1}^n x_i \mathbb{E}(X_{A_i}) = \sum_{i=1}^n x_i \mathbb{P}(X_{A_i}).$$ $\square$

(v) If X and Y are independent random variables then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof. Let $X = \sum_{i=1}^n x_i X_{A_i}$ and $Y = \sum_{j=1}^m y_j X_{B_j}$. Then

$$XY = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} x_i y_j X_{A_i \wedge B_j}.$$

Thus

$$
\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[\sum_{\substack{1\le i\le n\\1\le j\le m}} x_i y_j X_{A_i\wedge B_j}] \\
&= \sum_{\substack{1\le i\le n\\1\le j\le m}} x_i y_j \mathbb{E}[X_{A_i\wedge B_j}] && \text{(linearity of expectation)} \\
&= \sum_{\substack{1\le i\le n\\1\le j\le m}} x_i y_j \mathbb{P}(A_i\wedge B_j) \\
&= \sum_{\substack{1\le i\le n\\1\le j\le m}} x_i y_j \mathbb{P}(A_i)\mathbb{P}(B_j) && \text{(independence of } A_i,\ B_j\text{)} \\
&= \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}
$$

□

Lemma 3.2.4. *Suppose X is a random variable on the finite probability space $(\Omega, \mathbb{P})$ and $\mathbb{E}[X] = a$. Then there is a $G_1, G_2 \in \Omega$ such that $X(G_1) \le a \le X(G_2)$.*

Proof. The proof is nearly trivial by contradiction. □

Corollary 3.2.5. *If $X \ge 0$ is an integer valued random variable and $\mathbb{E}[X] < 1$ then there is a $G \in \Omega$ such that $X(G) = 0$.*

3.2.3 Conditional expectation

Definition 3.2.6. *Given an event A in a finite probability space and a discrete random variable, X, define the conditional expectation of X conditioned A by*

$$\mathbb{E}[X|A] = \sum_{G\in\Omega} X(G)\mathbb{P}[G|A].$$

Lemma 3.2.7. *For any random variables X and Y,*

$$\mathbb{E}[Y] = \sum_{x} \mathbb{P}[X = x] \cdot \mathbb{E}[Y|X = x]. \tag{3.2}$$

Proof.

$$\sum_x \mathbb{P}[X = x]\mathbb{E}[Y|X = x] = \sum_x \mathbb{P}[X = x] \sum_{G\in\Omega} Y(G)\cdot\mathbb{P}[G|X = x]$$

$$
\begin{aligned}
&= \sum_{G\in\Omega} Y(G) \sum_{x} \mathbb{P}[X = x] \cdot \mathbb{P}[G|X = x] \\
&= \sum_{G\in\Omega} Y(G)\mathbb{P}(G) \\
&= \mathbb{E}[Y].
\end{aligned}
$$

□

3.2.4 Variance and covariance of random variables

Definition 3.2.8. *If X is a random variable on the probability space $(\Omega, \mathbb{P})$ the variance of X is*

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

Definition 3.2.9. *If X and Y are random variables on the probability space $(\Omega, \mathbb{P})$, define the covariance of X and Y by*

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Through multiplication and the properties of expectation, the following lemma is seen.

Lemma 3.2.10. *If X and Y are random variables on the probability space $(\Omega, \mathbb{P})$, then*

$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Lemma 3.2.11. *Suppose X and Y are random variables on the probability space $(\Omega, \mathbb{P})$ then*

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y].$$

Proof.

$$
\begin{aligned}
\mathrm{Var}[X + Y] &= \mathbb{E}[X + Y) - \mathbb{E}[X + Y]]^2 \\
&= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])
\end{aligned}
$$

□

While variance of a random variable is an important tool in the probabilistic method, the following lemmas reduce the need for covariance.

Lemma 3.2.12. *If X and Y are independent random variables on the probability space $(\Omega, \mathbb{P})$, then*

$$\mathrm{Cov}[X, Y] = 0.$$

Lemma 3.2.12 is a direct consequence of Property (v) of expection of random variables.

Corollary 3.2.13. *If X and Y are independent random variables on a probability space $(\Omega, \mathbb{P})$ then*

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Corollary 3.2.13 combines Lemmas 3.2.11 and 3.2.12 together.

Lemma 3.2.14. *Suppose A and B are two events of a probability space, $(\Omega, \mathbb{P})$*

$$\mathrm{Cov}[X_A, X_B] \leq \mathbb{P}[A \wedge B].$$

Proof. Observe

$$X_A X_B = X_{A \wedge B}$$

so that

$$\begin{aligned}\mathrm{Cov}[X_A, X_B] &= \mathbb{P}[A \wedge B] - \mathbb{E}[X_A]\mathbb{E}[X_B] \\ &\leq \mathbb{P}[A \wedge B].\end{aligned}$$

□

3.2.5 Distributions

The following are a few examples of distributions. The probability function $\mathbb{P}$ can be described by the *probability distribution function* $\mathbb{P}[X = x]$ or the *cumulative distribution function* $\mathbb{P}[X \leq k]$.

Example 3.2.15. (Uniform)

Let $\emptyset \neq \Omega$ be a finite set. For $G \in \Omega$, define $\mathbb{P}(G) = \frac{1}{|\Omega|}$. Then $(\Omega, \mathbb{P})$ is a probability space called the *uniform distribution.*

Claim 1. Let $X = \sum_{i=1}^{n} X_{A_i}$ be a random variable on the finite set Ω with the uniform distribution. Then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \frac{|A_i|}{|\Omega|}.$$

Proof. For all $A \subseteq \Omega$,

$$\mathbb{E}[X_A] = \mathbb{P}[A] = \frac{|A|}{|\Omega|}.$$

Linearity of expectation implies

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_{A_i}] = \sum_{i=1}^{n} \frac{|A_i|}{|\Omega|}.$$ □

Example 3.2.16. (Binomial)

Suppose that n independent trials are to be run, each with probability of 'success' to be p and 'failure' to be $1-p$. If X is the random variable counting the number of successes that occur in the n trials, then X is called a *binomial random variable* with parameters (n, p), given by

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Note that by the binomial theorem,

$$\sum_{k=0}^{n} \mathbb{P}[X = k] = 1.$$

Observe that in Example 3.2.2(iii), X is a binomial random variable with parameters $n = 3$ and $p = 1/2$.

Claim 1. Let X be a random variable on $(\Omega, \mathbb{P})$ with binomial distribution of parameters (n, p), then $\mathbb{E}[X] = np$.

Proof. From the definition of X and $\mathbb{E}[X]$,

$$\begin{aligned}
\mathbb{E}[X] =& \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
=& \sum_{i=1}^{n} \binom{n}{i-1} p^i (1-p)^{n-i} \\
=& np \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-1-(i-1)} \\
=& np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} && (k = i-1) \\
=& np(p + (1-p))^{n-1} && \text{(Theorem 2.3.5)} \\
=& np.
\end{aligned}$$

□

Claim 2. Let X be a random variable on $(\Omega, \mathbb{P})$ with binomial distribution of parameters (n, p), then $\mathrm{Var}[X] = np(1-p)$.

Proof. As

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{i=1}^{n} i^2 \binom{n}{i} p^i (1-p)^{n-i} \\
&= np \left(\sum_{i=1}^{n-1} i \binom{n-1}{j} p^{i-1} (1-p)^{(n-1)-(i-1)} \right) \\
&= np \left(\sum_{i=1}^{n-1} (i-1) + 1 \binom{n-1}{j} p^{i-1} (1-p)^{(n-1)-(i-1)} \right) \\
&= np \left(\sum_{i=1}^{n-1} i \binom{n-1}{i-1} p^{i-1} (1-p)^{(n-1)-(i-1)} \ldots \right. \\
&\qquad \left. \ldots + \sum_{i=1}^{n-1} i \binom{n-1}{p}^{i-1} (1-p)^{(n-1)-(i-1)} \right) \\
&= np \left(\sum_{i=1}^{n-1} (\binom{n-1}{i-1} p^{i-1} (1-p)^{(n-1)-(i-1)} + 1 \right) && \text{(Theorem 2.3.5)} \\
&= np((n-1)p + 1) && \text{(applying 1)} \\
&= (np)^2 + np(1-p).
\end{aligned}$$

Therefore

$$\mathbb{E}[X^2] - \mathbb{E}[X]^2 = (np)^2 + np(1-p) - (np)^2 = np(1-p).$$

□

Example 3.2.17. **(Multinomial)**

If a sequence of n independent trials are run where each trial has $K > 2$ possible outcomes, each with probabilities $p_1, p_2, \ldots, p_k$. Then the probability that the n-th trial has x_1 outcomes of the first kind, x_2 outcomes of the second kind,$\ldots, x_k$ outcomes of the k-th kind, is

$$\frac{n!}{x_1!x_2!\ldots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}.$$

Example 3.2.18. **(Geometric)**

Suppose a sequence of independent trials are run, each with the probability, p of 'success', and $1-p$ of 'failure'. Let X measure the number of trials needed until the first 'success'. Then

$$\begin{aligned}
\mathbb{P}[X=1] &= p \\
\mathbb{P}[X=2] &= (1-p)p \\
\mathbb{P}[X=3] &= (1-p)(1-p)p \\
\vdots \quad &\vdots \quad \vdots \\
\mathbb{P}[X=k] &= (1-p)^{k-1}p
\end{aligned}$$

since if the first 'success' is on the $k-$th trial, the experiment must fail for the $k-1$ prior trials.

Claim 1. Let X be a random variable on the $(\Omega, \mathbb{P})$ with a geometric distribution with parameter p. Then $\mathbb{E}[X] = 1/p$.

Proof. Let $u = 1-p$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= p\sum_{k=1}^{\infty} ku^{k-1} \\
&= p\sum_{k=1}^{\infty} \frac{\mathrm{d}}{\mathrm{du}}u^{k-1} && (\tfrac{\mathrm{d}}{\mathrm{du}}u^k = ku^{k-1}) \\
&= p\frac{\mathrm{d}}{\mathrm{du}}\sum_{k=1}^{\infty} u^{k-1}
\end{aligned}$$

$$
\begin{aligned}
&= p\frac{\mathrm{d}}{\mathrm{d}u}\frac{1}{1-u} \\
&= p\frac{1}{(1-u)^2} \\
&= \frac{1}{p}.
\end{aligned}
$$

□

Example 3.2.19. (Hypergeometric)

Let Ω be a set with $n = a+b$ objects, a of type A, b of type B. If $N \leq n$ objects are chosen without replacement from Ω, let X count the number of elements selected of type A. Then X is said to have the *hypergeometric distribution* if for each k satisfying $0 \leq k \leq a$ and $N - k \leq b$,

$$\mathbb{P}[X = k] = \frac{\binom{a}{k}\binom{b}{N-k}}{\binom{n}{N}}.$$

Example 3.2.20. (Poisson)

A random variable X, with $\mathbb{E}[X] = \mu$ has the *Poisson distribution* if $\mathbb{P}[X = k] = \mathrm{e}^{-\mu}\frac{\mu^k}{k!}$.

Example 3.2.21. (Normal) The *standard normal distribution* is given by the cumulative distributive function

$$\mathbb{P}[X \leq k] = \int_{-\infty}^{k} \frac{e^{-x^2/2}}{\sqrt{\pi}}\mathrm{d}x.$$

As an exercise in calculus, it can be shown that

$$\lim_{k\to\infty} \mathbb{P}[X \leq k] = 1.$$

3.3 Necessary inequalities

Part of the art of the probabilistic method is in the estimation of bounds. Often the bounds need not be great, just shown to exist or to be of a certain type. To this end, the following inequalities are often useful.

Definition 3.3.1. *An inner product on a real vector space V is a function*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

such that for all $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in V$ *and all* $k \in \mathbb{R}$ *satisfies the following properties:*

(1) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

(2) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

(3) $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$

(4) $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ *and* $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ iff $\mathbf{v} = 0$. *The value* $||\mathbf{v}||$ *is called the norm (or length) of* $\mathbf{v}$.

An inner product space is a the pair $(V, \langle \cdot, \cdot \rangle)$ *where* V *is a real vector space and* $\langle \cdot, \cdot \rangle$ *is an inner product on* V.

As an abuse of language, in what follows the inner product space $(V, \langle \cdot, \cdot \rangle)$ is referred to as an inner product space V, without reference to the inner product.

Example 3.3.2. Let $(\Omega, \mathbb{P})$ be a finite probability space. Let V be the vector space of all random variables on $(\Omega, \mathbb{P})$. Then

$$\langle X, Y \rangle = \mathbb{E}[XY]$$

is an inner product on V.

Property (1) of Definition 3.3.1 follows trivially from the definition of expectation. For property (2), let X, Y and Z be random variables on $(\Omega, \mathbb{P})$. Then

$$\begin{aligned}
\langle (X+Y), Z \rangle &= \mathbb{E}[(X+Y)Z] = \mathbb{E}[XZ + YZ] \\
&= \mathbb{E}[XZ] + \mathbb{E}[YZ] && \text{(linearity of expectation)} \\
&= \langle X, Z \rangle + \langle Y, Z \rangle
\end{aligned}$$

as needed. Similarly, property (3) also follows from linearity of expectation. For property (4),

$$0 = \mathbb{E}[X^2] = \sum_{G \in \Omega} X^2(G)\mathbb{P}[G].$$

Then for all $G \in \Omega$, $X(G) = 0$ which is exactly $v = 0$ in V.

Lemma 3.3.3 (Cauchy-Schwarz inequality)**.** *Let V be an inner product space then for all $\mathbf{u}, \mathbf{v} \in V$,*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| \cdot ||\mathbf{v}||.$$

The proof appears in such books as [2].

Corollary 3.3.4. *Suppose X and Y are random variables on a finite probability space $(\Omega, \mathbb{P})$ then*

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

Proof. Let V be as in Example 3.3.2. Lemma 3.3.3 implies

$$|\mathbb{E}[XY]| = |\langle X, Y \rangle| \leq ||X|| \cdot ||Y|| = \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}. \qquad \square$$

Definition 3.3.5. *A function $f : \mathbb{R} \to \mathbb{R}$ is convex* iff *for all $0 \leq \lambda < 1$ and for all $x, y \in \mathbb{R}$,*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{3.3}$$

The following shows that inequality (3.3) is true for certain linear combinations also.

Theorem 3.3.6 (Jensen's inequality)**.** *Let $\lambda_1, \lambda_2, \ldots \lambda_n \in (0,1)$ such that $\sum_{k=1}^{n} \lambda_k = 1$ and $x_1, x_2, \ldots x_n \in \mathbb{R}$. If $f : \mathbb{R} \to \mathbb{R}$ is convex then*

$$f(\lambda_1 x_1 + \lambda_2 x_2 \ldots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_n f(x_n).$$

Proof. The proof is by induction on n. For $m = 1$ the statement is trivial. For $m = 2$, the theorem is the definition of a function being convex.
Assume for $n \geq 2$, the statement of the theorem is true for $m = n$. Suppose $\lambda_1, \lambda_2, \ldots \lambda_{n+1} \in (0,1)$ satisfies $\sum_{k=1}^{n+1} \lambda_k = 1$ and $x_1, x_2 \ldots x_{n+1} \in \mathbb{R}$. Let

$$y = \left(\frac{\lambda_1 x_1}{\lambda_1 + \lambda_2} + \frac{\lambda_2 x_2}{\lambda_1 + \lambda_2} \right) \quad \text{and} \quad \hat{\lambda} = \lambda_1 + \lambda_2.$$

As

$$\sum_{k=1}^{n+1} \lambda_k x_k = \hat{\lambda} y + \sum_{k=3}^{n+1} \lambda_k x_k,$$

the induction hypothesis implies

$$\begin{aligned}
f(\sum_{k=1}^{n+1}\lambda_k x_k) &= f(\hat{\lambda}y + \sum_{k=3}^{n+1}\lambda_k x_k) \\
&\leq \hat{\lambda}f(y) + \sum_{k=3}^{n+1}\lambda_k f(x_k) \\
&\leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \sum_{k=3}^{n+1}\lambda_k f(x_k) \qquad \text{(convexity of } f\text{)}. \qquad \square
\end{aligned}$$

3.3.1 Markov and Chebychev inequalities

Theorem 3.3.7 (Markov's inequality)**.** *Let* $X \geq 0$ *be a random variable on the finite probability space* $(\Omega, \mathbb{P})$ *and* $0 < t \in \mathbb{R}$*. Then*

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}. \tag{3.4}$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{G\in\Omega} X(G)\mathbb{P}[G] \\
&\geq \sum_{\substack{G\in\Omega \\ X(G)\geq t}} X(G)\mathbb{P}[G] \\
&\geq \sum_{\substack{G\in\Omega \\ X(G)\geq t}} t\mathbb{P}[G] \\
&\geq t \sum_{\substack{G\in\Omega \\ X(G)\geq t}} \mathbb{P}[G] \\
&= t\mathbb{P}[X \geq t]. \qquad \square
\end{aligned}$$

Theorem 3.3.8 (Chebychev's inequality)**.** *Let* X *be a random variable on the finite probability space* $(\Omega, \mathbb{P})$ *and* $t > 0 \in \mathbb{R}$*. Then*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}. \tag{3.5}$$

Proof. Let $Y = (X - \mathbb{E}[X])^2$. Equation (3.4) implies

$$\mathbb{P}[Y \geq t^2] \leq \frac{\mathbb{E}[Y]}{t^2}.$$

Since $\mathbb{E}[Y] = \mathrm{Var}(X)$ the result follows by taking square roots of both sides.

$\square$

For the next inequality (from [1]), let $\{A_i\}_{i=1}^n$ be a finite collection of events in a probability space $(\Omega, \mathbb{P})$ with the respective indicator random variables $\{X_i\}$. If $X = \sum_{i=1}^n X_i$ then

$$\begin{aligned}\mathrm{Var}(X) =& \sum_{i=1}^n V[X_i] - \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)\\ =& \sum_{i=1}^n \mathbb{E}[X_i] - (\mathbb{E}[X_i])^2 - \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)\\ =& \mathbb{E}[X] - \sum_{i=1}^n (\mathbb{E}[X_i])^2 - \sum_{i \neq j} \mathrm{Cov}(X_i, X_j). \qquad (3.6)\end{aligned}$$

Define a relation '$\sim$' on $[n]$ by $i \sim j$ if A_i and A_j are not independent. Let

$$\Delta = \sum_{i \sim j} \mathbb{P}[A_i \wedge A_j].$$

Then

$$\mathrm{Var}[X] \leq E[X] + \Delta \qquad \text{(by Lemma 3.2.14).} \qquad (3.7)$$

Corollary 3.3.9. *If $X \geq 0$ is a finite random variable then*

$$\begin{aligned}\mathbb{P}[X = 0] \leq& \mathbb{P}[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]] \qquad (3.8)\\ \leq& \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{\mathbb{E}[X] + \Delta}{\mathbb{E}[X]^2}. \qquad (3.9)\end{aligned}$$

Proof. The first part of inequality (3.2.14) follows from noting that

$$|X - \mathbb{E}[X]| \geq \mathbb{E}[X]$$

is the same as

$$\{X - \mathbb{E}[X] \leq -\mathbb{E}[X]\} \vee \{X - \mathbb{E}[X] \geq \mathbb{E}[X]\}.$$

Thus

$$\mathbb{P}[X \leq 0] \leq \mathbb{P}[\{X - \mathbb{E}[X] \leq -\mathbb{E}[X]\} \wedge \{X - \mathbb{E}[X] \geq \mathbb{E}[X]\}]$$

$$
\begin{aligned}
&= \mathbb{P}[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]] \\
&\leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2}.
\end{aligned}
$$

While the second inequality of (3.2.14) follows from equation (3.6). □

For the definition of almost surely, please see Definition 6.1.6.

Corollary 3.3.10. *If* $\mathrm{Var}[X] = \mathrm{o}(\mathbb{E}[X]^2)$ *then almost surely* $X > 0$.

Corollary 3.3.11. *If* $\mathrm{Var}[X] = \mathrm{o}(\mathbb{E}[X]^2)$ *then almost surely,* $X \sim \mathbb{E}[X]$.

Proof. Theorem 3.3.8 implies for any $\epsilon > 0$,

$$
\mathbb{P}[|X - \mathbb{E}[X]| > \epsilon\mathbb{E}[X]] \leq \frac{\mathrm{Var}[X]}{\epsilon^2\mathbb{E}[X]^2}.
$$

The condition on $\mathrm{Var}[X]$ implies the corollary. □

The following inequality is similar to the Chebychev inequality, but gives a bound on one side of the probability. It can ber found in a number of places, including [17, p. 152].

Theorem 3.3.12 (Chebychev-Cantelli)**.** *Let* $t > 0$ *and* X *be a random variable on a probability space* $(\Omega, \mathbb{P})$ *with* $\mathrm{Var}[X] < \infty$. *Then*

$$
\mathbb{P}[X - \mathbb{E}[X] \geq t] \leq \frac{\mathrm{Var}[X]}{\mathrm{Var}[X] + t^2}.
$$

Proof. Without loss of generality, assume $\mathbb{E}[X] = 0$. For all $t > 0$,

$$
\begin{aligned}
0 = \mathbb{E}[X] &= \sum_{G \in \Omega} X(G)\mathbb{P}[G] \\
&= \sum_{G \in \Omega} (XI_{\{X \leq t\}} + XI_{\{X > t\}})(G)\mathbb{P}[G] \\
&= \sum_{G \in \Omega} (XI_{\{X \leq t\}}\mathbb{P}[G] + \sum_{G \in \Omega} XI_{\{X > t\}}\mathbb{P}[G] \\
&\geq \sum_{G \in \Omega} (XI_{\{X \leq t\}})(G)\mathbb{P}[G] + t\sum_{G \in \Omega} I_{\{X > t\}}\mathbb{P}[G] \\
&= \mathbb{E}[XI_{\{X \leq t\}}] + t\mathbb{P}[\{X > t\}]
\end{aligned}
$$

$$= \mathbb{E}[XI_{\{X\leq t\}}] + t(1 - \mathbb{P}[X \leq t].$$

Rearranging the terms gives

$$\begin{aligned}
\mathbb{E}[(t-X)I_{\{X\leq t\}}] &= \mathbb{E}[tI_{\{X\leq t\}}] - \mathbb{E}[XI_{\{X\leq t\}}] \\
&= t\mathbb{P}[X \leq t] - \mathbb{E}[XI_{\{X\leq t\}}] \\
&\geq t - \mathbb{E}[X] \\
&= t. && \text{(since } \mathbb{E}[X] = 0)
\end{aligned}$$

Thus

$$\begin{aligned}
t^2 &\leq \mathbb{E}[(t-X)I_{\{X\leq t\}}] \\
&\leq \mathbb{E}[(t-X)^2]\mathbb{E}[I^2_{\{X\leq t\}}] && \text{(Lemma 3.3.4)} \\
&= (t^2 - 2t\mathbb{E}[X] + \mathbb{E}[X^2])\mathbb{E}[I_{\{X\leq t\}}] \\
&= (t^2 - \mathbb{E}[X^2])\mathbb{E}[T_{\{X\geq t\}}] && (\mathbb{E}[X] = 0) \\
&= (t^2 + \mathrm{Var}[X])\mathbb{E}[T_{\{X\leq t\}}].
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{t^2}{(t^2 + \mathrm{Var}[X])} &\leq \mathbb{E}[T_{\{X\leq t\}}] \\
&= \mathbb{P}[X \leq t] = 1 - \mathbb{P}[X \geq t].
\end{aligned}$$

Rewriting:

$$\begin{aligned}
\mathbb{P}[X \geq t] &= \mathbb{P}[X - \mathbb{E}[X]] \\
&\leq \frac{\mathrm{Var}[X]}{t^2 + \mathrm{Var}[X]}.
\end{aligned}$$

□

3.3.2 Chernoff's inequalities

This section covers a pair of inequalities related to the paper [6] by Hermann Chernoff.

Lemma 3.3.13. *For $a > 0$,*

$$\left(\frac{\mathrm{e}^a + \mathrm{e}^{-a}}{2}\right) \leq \mathrm{e}^{a^2/2}. \tag{3.10}$$

Proof. Using the Taylor series (2.2) for e^a and e^{-a}, gives

$$\left(\frac{e^a + e^{-a}}{2}\right) = \sum_{n=0}^{\infty} \frac{a^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{a^{2n}}{2^n n!} = e^{a^2/2}. \qquad \square$$

Theorem 3.3.14 (Chernoff's inequality [6])**.** *Let $\{X_i\}$ be mutually independent indicator random variables on a finite probability space $(\Omega, \mathbb{P})$ such that $\mathbb{P}[\{X_i = 1\}] = \mathbb{P}[\{X_i = -1\}] = \frac{1}{2}$. If $S_n = \sum_{i=1}^{n} X_i$ then for any $\lambda > 0$,*

$$\mathbb{P}[\{S_n > \lambda\}] < e^{\frac{-\lambda^2}{2n}}. \tag{3.11}$$

Proof. Observe that $-n \leq S_n \leq n$. Therefore if $\lambda > n$, (3.11) is trivial. So assume $0 < \lambda < n$. Let $\alpha = \frac{\lambda}{n}$. If $1 \leq i \leq n$ then $\mathbb{E}[e^{\alpha X_i}] = \frac{e^{\alpha}+e^{-\alpha}}{2}$ and since for every $\{i,j\} \in [n]^2$, X_i and X_j are independent, e^{X_i} and e^{X_j} are also. Therefore

$$\begin{aligned}
\mathbb{P}[\{\alpha S_n > \alpha\lambda\}] &= \mathbb{P}[\{e^{\alpha S_n} > e^{\alpha\lambda}\}] \\
&= \mathbb{P}[\{\prod e^{\alpha X_i} > e^{\alpha\lambda}\}] \\
&= \prod \mathbb{P}(\{e^{\alpha X_i} > e^{\alpha\lambda}\}) \prod \mathbb{P}[\{e^{\alpha X_i} > e^{\alpha\lambda}\}] \\
&\leq \prod \frac{\mathbb{E}[e^{\alpha X_i}]}{e^{\alpha\lambda}} && \text{(inequality (3.4))} \\
&= \frac{\prod \frac{1}{2}(e^{\alpha} + e^{-\alpha})}{e^{n\alpha\lambda}} \\
&\leq \frac{\prod e^{\alpha^2/2}}{e^{n\alpha\lambda}} && \text{(Lemma 3.3.13)} \\
&= e^{n(\frac{\alpha^2}{2} - \alpha\lambda)} \\
&= e^{n(\frac{\lambda^2}{2n^2} - \frac{\lambda^2}{n})} \\
&= e^{\lambda^2(\frac{1}{2n} - 1)} \\
&\leq e^{-\lambda^2/2n}. \qquad \square
\end{aligned}$$

The next theorem is a second example of a Chernoff inequality.

Theorem 3.3.15. *Assume $X_1, X_2, \ldots X_n \in \{0,1\}$ are independent random variables on some finite probability space $(\Omega, \mathbb{P})$. Suppose for each $1 \leq i \leq n$,*

$$\mathbb{P}[X_i = 1] = p_i \in (0,1).$$

If $S_n = \sum_{i=1}^n X_i$ *and* $\mu = \mathbb{E}[S_n]$ *then for all* $\delta \in (0,1)$,

$$\mathbb{P}[S_n > (1+\delta)\mu] \leq \mathrm{e}^{-\frac{\delta^2\mu}{2}}.$$

Proof. Let $f(S_n) = \mathrm{e}^{S_n}$. Then for all $t \in \mathbb{R}$

$$\begin{aligned}
\mathbb{E}[\mathrm{e}^{tS_n}] &= \pi_{i=1}^n \mathbb{E}[\mathrm{e}^{tX_i}] && \text{(independence of } X_i\text{)}\\
&= \pi_{i=1}^n (p_i\mathrm{e}^t + (1-p_i)1)\\
&= \pi_{i=1}^n \big(1 + p_i(\mathrm{e}^t - 1)\big)\\
&\leq \mathrm{e}^{p_i(\mathrm{e}^t-1)}. && \text{(by (2.3.1))}
\end{aligned}$$

Thus for all $\delta > 0$,

$$\begin{aligned}
\mathbb{P}[S_n > (1+\delta)\mu] &= \mathbb{P}[tS_n > (1+\delta)t\mu]\\
&= \mathbb{P}[\mathrm{e}^{tS_n} > \mathrm{e}^{(1+\delta)t\mu}]\\
&\leq \frac{\mathbb{E}[\mathrm{e}^{tS_n}]}{\mathrm{e}^{(1+\delta)t\mu}} && \text{(by 3.4)}\\
&\leq \frac{\prod_{i=1}^n \mathrm{e}^{p_i(\mathrm{e}^t-1)}}{\mathrm{e}^{(1+\delta)t\mu}}\\
&= \frac{\mathrm{e}^{(\mathrm{e}^t-1)\sum_{i=1}^n p_i}}{\mathrm{e}^{(1+\delta)t\mu}}\\
&= \frac{\mathrm{e}^{(\mathrm{e}^t-1)\mu}}{\mathrm{e}^{(1+\delta)t\mu}}\\
&= \mathrm{e}^{((\mathrm{e}^t-1)-t(\delta+1))\mu}. && (3.12)
\end{aligned}$$

To minimize equation (3.12), set $t = \ln(\delta+1)$ so that

$$\begin{aligned}
\mathrm{e}^{\mu((\mathrm{e}^t-1)-t(\delta+1))} &= \mathrm{e}^{\mu((\delta+1)+1-1)-(\ln(\delta+1)(\delta+1))}\\
&= \mathrm{e}^{\mu(\delta-(\delta+1)(\ln(\delta+1)))}. && (3.13)
\end{aligned}$$

As $0 < \delta < 1$, equation(2.4) implies:

$$\begin{aligned}
(\delta+1)\sum_{1\geq 1}(-1)^{i+1}\frac{(\delta)^i}{i} &= \sum_{1\geq i}\frac{(-1)^{i+1}(\delta)^{i+1}}{i} + \sum_{1\geq i}(-1)^{i+1}\frac{(\delta)^i}{i}\\
&= \delta + \sum_{2\geq i}(-\delta)^i\left(\frac{1}{i-1} - \frac{1}{i}\right) && (3.14)
\end{aligned}$$

$$
\begin{aligned}
&= \delta + \sum_{2 \geq i} (-\delta)^i \left(\frac{1}{i(i-1)} \right) \\
&\leq \delta + \frac{\delta^2}{2} - \frac{\delta^3}{6} \\
&\leq \delta + \frac{\delta^2}{2}. \qquad (3.15)
\end{aligned}
$$

Equations (3.14) and (3.13) give

$$\mathbb{P}[S_n > (1+\delta)\mu] \leq \mathrm{e}^{[\mu(\delta - (\delta + \frac{\delta^2}{2})]} = \mathrm{e}^{-\mu\delta^2/2}. \qquad \square$$

The next example illustrates the use of Chernoff's inequality (3.3.14).

Example 3.3.16. Consider flipping a fair coin n times. Let $X_i = 1$ if and only if the i-th flip is a head and 0 otherwise. Then $\mathbb{E}[X_i] = 1/2$. Let $X = \sum_{i=1}^n X_i$ be the number of heads. Thus $\mu = \mathbb{E}[X] = n/2$. Inequality (3.3.15) implies for all $0 < \delta < 1$,

$$\mathbb{P}[X > (1+\delta)\mu] \leq \mathrm{e}^{-\delta^2\mu/2} = \mathrm{e}^{-\delta^2 n/4}.$$

The next example compares the bounds given by Chebychev's inequality (3.3.8) and Chernoff's inequality (3.3.14) and therefore gives an indication of when Chernoff's inequality is preferable to Chebychev.

Example 3.3.17. [38] Consider Example 3.3.16 and assume $\delta = 1/2$. To ensure that $\mathbb{P}[X > 3n/4] \leq 0.1$, suppose $\mathrm{e}^{-(1/2)^2 n/4} = \frac{1}{10}$. Taking natural logarithms yields: $\frac{-n}{64} = -\ln 10$ or $n = 64 \ln 10$ or $n \approxeq 148$.

Alternatively, as the X_i's are independent, inequality (3.3.8) and Lemma 3.2.13 imply

$$\mathrm{Var}[X] = \sum_{i=1}^n \mathrm{Var}[X_i] = n/4$$

thus

$$\mathbb{P}[|X - n/2| > 3n/4] \leq \frac{n/4}{3n/4} = 1/3$$

independent of n. Thus the Chernoff bounds allows for for an estimate on the size of n to ensure that the probability $X > (1+\delta)\mu$ is small, whereas Chebychev's inequality only gives as upper bound the probability, regardless of n.

Chapter 4

Essential methods

This section outlines a few of the methods commonly used in applications of the probabilistic method.

4.1 First moment method

Suppose $X_1, X_2, \ldots X_n$ are random variables, $c_1, c_2, \ldots c_n$ are real numbers and $X = c_1 X_1 + \ldots + c_n X_n$, then Lemma i implies

$$\mathbb{E}[X] = c_1 \mathbb{E}[X_1] + \ldots + c_n \mathbb{E}[X_n].$$

The first moment method uses random variables and expectation to show the existence or nonexistence of a certain entity $\mathfrak{F}$. Often, calculating expected values achieved through a decomposition into simple indicator random variables. The power of this principle lies in the fact that linearity of expectation has no restrictions regarding dependence or independence of the X_i's. The first example is viewed as one of the first examples of using probability theory to prove a purely combinatorial result.

Theorem 4.1.1. [37] *There is a tournament on n players with at least $n!2^{-(n-1)}$ Hamiltonian paths.*

Proof. Fix n and let Ω be the set of all tournaments on n players. If $T \in \Omega$, let $X(T)$ count the number of Hamiltonian paths in T.

For $\sigma \in S_n$, define the path $P_\sigma = \{v_{\sigma(1)}, v_{\sigma(2)}, v_{\sigma(3)}, \dots v_{\sigma(n)}\}$. These P_σ have the property that for every Hamiltonian path $P \subseteq T$, there is a unique $\sigma \in S_n$ such that $P = P_\sigma$. Say T has A_σ if $P_\sigma \subseteq T$ and let X_σ be the indicator random variable for A_σ.

Write A_σ as the events:

$$A_\sigma = \{(v_{\sigma_1}, v_{\sigma_2}) \in T\} \wedge \{(v_{\sigma_2}, v_{\sigma_3}) \in T\} \wedge \cdots \wedge \{(v_{\sigma(n-1)}, v_{\sigma(n)}) \in T\}.$$

The definition of Ω implies that each event occurs independently with probability $\frac{1}{2}$, hence

$$\begin{aligned}
\mathbb{E}[X_\sigma] &= \mathbb{P}[A_\sigma] \\
&= \mathbb{P}[\{(v_{\sigma_1}, v_{\sigma_2}) \in T\} \wedge \cdots \wedge \{(v_{\sigma(n-1)}, v_{\sigma(n)}) \in T\}] \\
&= \mathbb{P}[\{(v_{\sigma(1)}, v_{\sigma(2)}) \in T\}] \cdots \mathbb{P}[\{(v_{\sigma(n-1)}, v_{\sigma(n)}) \in T\}] \qquad \text{(events indep.)} \\
&= 2^{-(n-1)}.
\end{aligned}$$

As $X = \sum_{\sigma \in S_n} X_\sigma$ and $|S_n| = n!$, $\mathbb{E}[X] = \sum_{\sigma \in S_n} \mathbb{E}[X_\sigma] = n!2^{(1-n)}$. Lemma 3.2.4 implies there is a $T \in \Omega$ with at least $\mathbb{E}[X]$ Hamiltonian paths. $\square$

The next theorem is folklore.

Theorem 4.1.2 ([12]). *Every graph with m-edges has a (not necessarily induced) bipartite subgraph of at least $m/2$ edges.*

Proof. The proof is from [1]. Define the sample space

$$\Omega = \{(U, W) : \quad U \text{ and } W \text{ partition } V.\}$$

For every $(U, W) \in \Omega$, let

$$T_{(U,W)} = \{\mathrm{e} = \{x, y\} \in E : \{x \in U \wedge y \in W\} \vee \{x \in W \wedge y \in U\}\}.$$

For every $e \in E$, let A_e be the event '$e \in T_{(U,V)}$' with the respective indicator random variable X_e.

$$\mathbb{E}[X_e] = \mathbb{P}[A_e]$$

$$
\begin{aligned}
&= \mathbb{P}[\{(x \in U) \wedge (y \in W)\} \vee \{(x \in W) \wedge (y \in U)\}] \\
&= \mathbb{P}[(x \in U) \wedge (y \in W)] + \mathbb{P}[(x \in W) \wedge (y \in U)] \qquad \text{(disjoint events)} \\
&= \mathbb{P}[x \in U]\mathbb{P}[y \in W] + \mathbb{P}[x \in W]\mathbb{P}[y \in U] \qquad \text{(indep. events)} \\
&= (1/2)(1/2) + (1/2)(1/2) = 1/2.
\end{aligned}
$$

If $X[(U,V)] = |T_{(U,V)}| = \sum_{e \in E} X_e[(U,V)]$ so that

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{\mathrm{e} \in E} \mathbb{E}[X_{\mathrm{e}}] \\
&= m\mathbb{E}[X_{\mathrm{e}}] \qquad (\forall \mathrm{e}_1, \mathrm{e}_2 \in E \quad \mathbb{E}[X_{\mathrm{e}_1}] = \mathbb{E}[X_{\mathrm{e}_2}]) \\
&= m/2.
\end{aligned}
$$

Lemma 3.2.4 implies there is a pair $(U, W) \in \Omega$ with at least this many "crossing" edges. $\square$

4.2 Alterations and deletions

A second probabilistic tool useful in combinatorial research is often called alterations or the deletion method. As in the first moment method, first define a probability space and a "nice" random variable. Next calculate the expected value and choose an element with at most that much of the property. Then alter this element in such a way as to eliminate what is not desired.

Theorem 4.2.1. [9] *Let $k, \ell \in \mathbb{Z}^+$, $\ell \geq 3$. There exist a graph G such that* $\mathrm{girth}(G) > \ell$ *and* $\chi(G) \geq k$.

Proof. (This proof is based [1].) Assume n is an unspecified, large, nonnegative integer and let $0 < \theta < \dfrac{1}{\ell}$. Let $p = n^{\theta - 1}$. Assume $G = (V, E) \in \mathcal{G}_{n,p}$ (for the definition of $\mathcal{G}_{n,p}$), please see Chapter 6) and let $X(G)$ be the number of cycles of length at most ℓ.

Let $S = \{v_1, v_2 \ldots v_s\} \in [V]^s$, and A_S be the event "S induces a cycle". There are $(n)_s$ sequences of distinct $S \in [V]^s$, and each cycle can be identified by $2s$ of those sequences: there are two ways to choose direction and s ways to choose a first vertex. Furthermore, given any ordering of S, the probability of a graph having this cycle is p^s. Putting all of this together yields,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=3}^{l} \frac{(n)_i}{2i} p^i \\
&= \sum_{i=3}^{\ell} (n)_i \frac{p^i}{2i} \\
&= \sum_{i=3}^{\ell} (n)_i \frac{n^{(\theta-1)i}}{2i} \\
&\leq \sum_{i=3}^{\ell} n^i n^{(\theta-1)i} \\
&= \sum_{i=3}^{\ell} n^{\theta i} \leq l n^{\theta \ell} = \mathrm{o}(n) \qquad\qquad (\theta\ell < 1).
\end{aligned}$$

From (3.4),

$$\mathbb{P}[X > \frac{n}{2}] \leq \frac{\mathbb{E}[X]}{\frac{n}{2}} = \mathrm{o}(1), \tag{4.1}$$

which implies there is an n large enough so that $\mathbb{P}[X > \dfrac{n}{2}] < \frac{1}{2}$.

Next, observe Lemma 2.2.1 implies bounding $\alpha(G)$ from above bounds $\chi(G)$ from below. Since for any integer $x > 0$,

$$\mathbb{P}[\alpha(G) = x] \leq \binom{n}{x}(1-p)^{\binom{x}{2}} \leq n^x \mathrm{e}^{-px\frac{(x-1)}{2}} = (ne^{-p\frac{(x-1)}{2}})^x, \tag{4.2}$$

if $x = \lceil \frac{3}{p} \ln(n) \rceil$, (4.2) implies

$$\mathbb{P}[\alpha(G) = x] \leq n^{-\frac{x}{2}} = \mathrm{o}(1). \tag{4.3}$$

Furthermore,

$$\mathbb{P}[\alpha(G) \geq x] \leq \sum_{i=x}^{n} \mathbb{P}[\alpha(G) = i] = \mathrm{o}(1). \tag{4.4}$$

Note that the choice of x is made to simplify the subsequent calculations; the proof only requires x to be large enough so that $\mathbb{P}([\alpha(G) \geq x]) = \mathrm{o}(1)$, to guarantee the existence of an n so that $\mathbb{P}[\alpha(G) \geq x] < 1/2$.

Pick n large enough to make both (4.1) and (4.4) less than 1/2. Since

$$\mathbb{P}[\overline{\{X(G) \leq n/2\}} \vee \overline{\{\alpha(G) < x\}}] \leq \mathbb{P}[\overline{\{X(G) \leq n/2\}}] + \mathbb{P}[\overline{\{\alpha(G) < x\}}]$$

$$< 1/2 + 1/2 = 1$$

then

$$\begin{aligned}\mathbb{P}[\{X(G) \le n/2\} \wedge \{\alpha(G) \le x\}] &= \mathbb{P}[\{X(G) \le n/2\} \wedge \{\alpha(G) \le x\}] \\ &> 1 - 1 = 0\end{aligned}$$

ensuring there is a graph, $G \in \mathcal{G}_{n,p}$, with

$$X(G) \le \frac{n}{2} \text{ and } \alpha(G) < x.$$

From G, create a new graph G^* by deleting a vertex from each cycle of length at most ℓ. Thus, $\operatorname{girth}(G^*) > \ell$.

By the choice of G, the number of cycles of length ℓ or less is at most $n/2$ so that the deletion of vertices from G leaves

$$|V(G^*)| \ge n - \frac{n}{2} = \frac{n}{2}.$$

The construction of G^* ensures $\alpha(G^*) \le \alpha(G)$. Hence, by Lemma 2.2.1

$$\chi(G^*) \ge \frac{|V(G^*)|}{\alpha(G^*)} \ge \frac{n}{\frac{3}{p}\ln n} = \frac{\frac{n}{2}}{\frac{3}{n^{\theta-1}}\ln n} = \frac{n^\theta}{6\ln n}. \tag{4.5}$$

Equation (4.5) gives a second lower bound on n as choosing n large enough so that $\frac{n^\theta}{6\ln n} > k$ implies $\chi(G)$ is high enough showing that the created graph has the requisite properties. □

4.3 Second moment method

Assume X is a random variable with $\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. The second moment method is any proof that uses variance. The sections on random graphs (specifically Chapter 7 on threshold functions) illustrate the use of the second moment method. Therefore any further explanation is delayed until Chapter 7.

4.4 The Lovász local lemma

Let $A_1, A_2, \ldots A_k$ be events in a probability space. Often in combinatorial applications, it is necessary to show

$$\mathbb{P}[\wedge_{i=1}^{k} \overline{A_i}] > 0. \tag{4.6}$$

If it is possible to show $\mathbb{P}[\bigvee_{i=1}^{k} A_i] < 1$, then the definition of probability implies equation (4.6). To accomplish this, one method is to calculate $\sum_{i=1}^{k} \mathbb{P}[A_i]$ then using the relationship $\mathbb{P}[\bigvee_{i=1}^{k} A_i] \leq \sum_{i=1}^{k} \mathbb{P}[A_i]$ for the desired result. In the case that the events are mutually independent, this result is the best possible, while in the case of dependence between events, such a relationship may be useless.

This section introduces a result useful to guarantee equation (4.6) when there is dependence between events. The Lovàsz local lemma is best applied when there are few dependencies between events. While the Lovàsz local lemma is purely a probabilistic result, it was developed by Lovàsz and Erdős in 1975 to handle problems in combinatorics. First a definition used in the proof is given.

Definition 4.4.1. *Given a finite collection of events $\{A_i\}_{i=1}^{n}$ in any probability space, a dependency digraph $D = (V, E)$ is defined by*

$$V = [n] \quad E = \{(i, j) | A_i \textit{ is dependent on } A_j\}.$$

While there are several variants of the proof of the Lovász local lemma (*e.g.* [1], [4]), each is more or less the same. This version and its corollary are from [1].

Theorem 4.4.2. Lovász local lemma [13]
Given a finite collection of events $\{A_i\}_{i=1}^{n}$ in any probability space and the respective dependency digraph $D = (V, E)$, if there exists real numbers $0 < x_i < 1$ such that for every event

$$\mathbb{P}[A_i] \leq x_i \prod_{(i,j) \in E} (1 - x_j) \tag{4.7}$$

then

$$(a)\forall_{1\le i\le n}\mathbb{P}[A_i|\bigwedge_{(i,j)\in E}\overline{A_j}]\le x_i \quad (b)\mathbb{P}[\bigwedge_{i=1}^{n}\overline{A_i}]\ge\prod_{i=1}^{n}(1-x_i)$$

.

Proof. The proof is by induction on the size of $S \subsetneq [n]$. For all $0 \le k \le n-1$ let $P(k)$ be the proposition

$$\forall S\in[n]^k,\quad \forall i\in[n]\setminus S\quad \mathbb{P}[A_i|\bigwedge_{j\in S}\overline{A_j}]\le x_i. \tag{4.8}$$

For $k = 0$, $S = \emptyset$ hence (4.8) is for all $i \in [n]$ $\mathbb{P}[A_i] \le x_i$ which follows from the assumptions on A_i.

Let $0 < m < n$ be given and assume for every $0 \le k < m$ $P(k)$ holds.

Let $S \in [n]^m$ (without loss of generality, assume $S = [m]$) and $i \in [n] \setminus S$. Let $S_1 = \{j \in S|(i,j) \in E\}$ and $S_2 = S \setminus S_1$; then

$$\begin{aligned}
\mathbb{P}[A_i|\bigwedge_{j\in S}\overline{A_j}] &= \frac{\mathbb{P}[A_i\wedge(\wedge_{l\in S_1}\overline{A_l})\wedge\left(\bigwedge_{j\in S_2}\overline{A_j}\right)]}{\mathbb{P}[\left(\bigwedge_{l\in S_1}\overline{A_l}\right)\wedge\left(\bigwedge_{j\in S_2}\overline{A_j}\right)]}\\
&\le \frac{\mathbb{P}[A_i\wedge\left(\bigwedge_{l\in S_1}\overline{A_l}\right)\wedge\left(\bigwedge_{j\in S_2}\overline{A_j}\right)]}{\mathbb{P}[\overline{A_1}]\mathbb{P}[\overline{A_2}|\overline{A_1}]\cdots\mathbb{P}[\overline{A_m}|\bigwedge_{j\in S\setminus\{m\}}\overline{A_j}]}\\
&\le \frac{\mathbb{P}[A_i|\bigwedge_{j\in S_2}\overline{A_j}]}{\prod_{j\in S}(1-x_j)}\\
&\le \frac{\mathbb{P}[A_i]}{\prod_{j\in S_1}(1-x_j)}\le x_i,
\end{aligned}$$

which is (a). From which it follows:

$$\mathbb{P}[\overline{A_i}|\bigwedge_{(i,j)\in E}\overline{A_j}]\ge 1-x_i.$$

To complete the proof, observe that

$$\begin{aligned}
\mathbb{P}[\bigwedge_{i=1}^{n}\overline{A_i}] &= \mathbb{P}[\overline{A_1}]\mathbb{P}[\overline{A_2}|\overline{A_1}]\cdots\mathbb{P}[\overline{A_m}|\bigwedge_{i=1}^{n}\overline{A_i}]\\
&= (1-\mathbb{P}[A_1])(1-\mathbb{P}[A_2|A_1])\cdots(1-\mathbb{P}[A_m|\bigwedge_{i=1}^{n}\overline{A_i}])\\
&\ge \prod_{i=1}^{n}(1-x_i)
\end{aligned}$$

as needed. $\square$

A corollary of the local lemma is often useful in the case of some symmetry.

Corollary 4.4.3. *Given a finite collection of events $\{A_i\}_{i=1}^m$ that are mutually independent of all except at most d separate events, if there exists a $0 < p < 1$ such that for every $i \in [n]$, $\mathbb{P}[A_i] \leq p$ and $\mathrm{e}p(d+1) \leq 1$ then $\mathbb{P}[\bigwedge_{i=1}^{m} \overline{A_i}] > 0$.*

Proof. If $d = 1$, then the result follows from

$$\mathbb{P}[\wedge \overline{A_i}] \geq \prod (1-p)^n.$$

Otherwise, observe that if $x = 1/(d+1)$ then:

$$x(1-x)^d \geq x\mathrm{e}^{-1} \geq p$$

from the hypothesis. Theorem 4.4.2 gives the result. □

In general, the probabilistic method uses the symmetric version, Corollary 4.4.3, of the Lovǎ́sz local lemma rather than the generalized version, Theorem 4.4.2. The following examples represent but a few of the many applications of this lemma.

An *arithmetic progression of length k* (in symbols, AP_k) is a string of positive integers of the form $\{a, a+d, a+2d, \ldots a+(k-1)d\}$. The positive integer $N = W(k)$ is called the *k-van der Waerden's number* iff for every $n \geq N$ and every two-colouring, χ, of $[n]$ there is a monochromatic $AP_k \subset [n]$. Using Corollary 4.4.3 of the Lovasz local lemma, provides a method to get a lower bound.

Example 4.4.4. [23, p. 244] For every $k \geq 2$, if

$$\mathrm{e}2^{1-k}(n^2+1) \leq 1 \tag{4.9}$$

then $W(k) > n$.

Proof. Let $n = 2^{k-1}$ and χ be any random and uniform two-colouring of $[n]$ (*i.e.* for $x \in [n]$, the probability $\chi(x)$ is red=1/2=the probability $\chi(x)$ is blue).

Let $A \subset [n]$ be any AP_k, and let S be the event "A is monochromatic". Using $p = \mathbb{P}[S] = 2^{1-k}$, since S is dependent on at most $d = n^2$ events, Corollary 4.4.3, along with the conditions on n, give the result. $\square$

The next example demonstrates the use of the general Lovász local Lemma 4.4.2.

Example 4.4.5. [3] Let $\mathcal{F}$ be a family of sets, each of which has at most $k \geq 2$ points. Also suppose that for each point v,

$$\sum_{S \in \mathcal{F} \,:\, v \in S} \left(1 - \frac{1}{k}\right)^{-|S|} 2^{1-|S|} \leq \frac{1}{k}.$$

Then $\mathcal{F}$ is 2-colourable.

Proof. Let $\mathcal{F} = \{S_1, S_2, \ldots, S_m\}$ and colour the points red and blue independently and randomly with a probability of 1/2. For every $S_t \in \mathcal{F}$, let A_t be the event that S_t is monochromatic; thus

$$\mathbb{P}[A_t] = 1 - 2^{1-|S_t|}.$$

Events A_t and A_s are dependent if and only if $0 < |S_t \cap S_s|$. Define the relation '$\sim$' on $[m]$ by $s \sim t$ iff A_s are dependent on A_t.

To complete the proof, it must be shown that for every $t \in [m]$,

$$\mathbb{P}[A_t] \leq x_t \prod_{s \sim t} (1 - x_s).$$

To this end,

$$x_t \prod_{s \sim t} (1 - x_s) \geq x_t \prod_{v \in S_t} \left(\prod_{j : v \in S_j} 1 - x_j \right) \tag{4.10}$$

$$\geq x_t \prod_{v \in S_t} \left[1 - \sum_{j : v \in S_j} x_j \right] \tag{4.11}$$

$$\geq x_t \left(1 - \frac{1}{k}\right)^{|S_t|} \tag{4.12}$$

since the assumptions on the theorem state $\sum_{j:v\in S_j} x_j \leq \frac{1}{k}$. Thus

$$x_t \prod_{s\sim t}(1-x_s) \geq x_t \left(1-\frac{1}{k}\right)^{|S_t|} = 2^{1-|S_t|} = \mathbb{P}[A_t].$$

Lemma 4.4.2 implies that

$$\mathbb{P}[\overline{A_1} \wedge \overline{A_2} \wedge \ldots \wedge \overline{A_m}] > 0.$$

Thus there is a two-colouring such that no S_t is monochromatic. $\square$

Chapter 5

Combinatorial applications

This chapter demonstrates applications of the probabilistic method throughout combinatorics.

5.1 Ramsey theory

As Ramsey numbers have proven to be difficult to calculate using deterministic methods, the use of probability theory has given some of the best known results.

Definition 5.1.1. *Let $k, \ell \in \mathbb{Z}^+$. The Ramsey number $R(k, \ell)$ is the smallest positive integer n such that every red-blue colouring of the edges of K_n, there is either a red K_k or a blue K_ℓ.*

Ramsey (1929) showed that for every two positive integers, k, ℓ, $R(k, \ell)$ is finite. The numbers $R(k, k)$ are some times called the diagonal Ramsey numbers.

Theorem 5.1.2. (Erdős, 1947) *Let $n \geq k$ be positive integers. If*

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

then $R(k, k) > n$.

Proof. Let $G = (V, E)$ be a complete graph on n vertices. Let Ω be the collection of two-colourings of $E(G)$. For every $S \in [V]^k$, let A_S be the event that "the edges of K_S are monochromatic." Call $\chi \in \Omega$ a good colouring if no A_S occurs.

Let $\chi \in \Omega$ be such that for every $\{x, y\} \in E$,

$$\mathbb{P}[\{x, y\} \text{ is red}] = 1/2 = \mathbb{P}[\{x, y\} \text{ is blue}]$$

independently and randomly chosen. Let $S \in [n]^k$ be given. For all $\{x, y\} \in [S]^2$, let $B_{\{x,y\}}$ be the event '$\{x, y\}$ is blue' and $R_{\{x,y\}}$ be the event '$\{x, y\}$ is red'. As

$$A_S = (\bigwedge_{\{x,y\} \in [S]^2} B_{\{x,y\}}) \vee (\bigwedge_{\{x,y\} \in [S]^2} R_{\{x,y\}})$$

so that

$$\begin{aligned}
\mathbb{P}[A_S] &= \mathbb{P}[(\bigwedge_{\{x,y\} \in [S]^2} B_{\{x,y\}}) \vee (\bigwedge_{\{x,y\} \in [S]^2} R_{\{x,y\}})] \\
&= \mathbb{P}[\bigwedge_{\{x,y\} \in [S]^2} B_{\{x,y\}}] + \mathbb{P}[\bigwedge_{\{x,y\} \in [S]^2} R_{\{x,y\}}] \quad \text{(mutu. disj. events)} \\
&= \prod_{\{x,y\} \in [S]^2} \mathbb{P}[B_{\{x,y\}}] + \prod_{\{x,y\} \in [S]^2} \mathbb{P}[R_{\{x,y\}}] \quad \text{(ind. of events)} \\
&= \prod_{\{x,y\} \in [S]^2} 2^{-1} + \prod_{\{x,y\} \in [S]^2} 2^{-1} \\
&= 2\left(2^{-\binom{k}{2}}\right) \quad (|[S]^2| = \binom{k}{2}).
\end{aligned}$$

If every colouring of G is bad then $R(k, k) \leq n$, but,

$$\begin{aligned}
\mathbb{P}(\bigvee_{S \in [V]^k} A_S) &\leq \sum_{S \in [V]^k} \mathbb{P}(A_S) \\
&= 2^{1-\binom{k}{2}} \sum_{S \in [V]^k} 1 \\
&= 2^{1-\binom{k}{2}} \binom{n}{k} < 1.
\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{P}(\overline{\bigvee_{S \in [V]^k} A_S}) &= \mathbb{P}(\bigwedge_{S \in [V]^k} \overline{A_S}) \\ &= 1 - \mathbb{P}(\bigvee_{S \in [V]^k} A_S) > 0.\end{aligned}$$

Therefore there is a two-colouring of G for which no A_S occurs. □

The next proof is very similar but uses expectation and random variables.

Proof. (of Theorem 5.1.2) Let Ω be the sample space of all two-colourings of the edges of K_n such that for all $e \in E(K_n)$,

$$\mathbb{P}[e \text{ is red}] = \mathbb{P}[e \text{ is blue}] = \frac{1}{2}$$

determined independently. Let X be the nonnegative, integer valued random variable on Ω defined by $X(\chi)$ is the number of monochromatic K_k. For $S \in [V]^k$, let A_S be the event "K_S is a monochromatic" and X_S be the indicator random variable for A_S thus

$$\begin{aligned}\mathbb{E}[X_S] &= \sum_{G \in \Omega} X_S(G)\mathbb{P}(K_S \subset G \text{ is monochromatic}) \\ &= 2^{1-\binom{k}{2}}.\end{aligned}$$

As $X = \sum_{S \in [n]^k} X_S$,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{S \in [V]^k} \mathbb{E}[X_S] \\ &= \sum_{S \in [V]^k} 2^{1-\binom{k}{2}} \\ &= \binom{n}{k} 2^{1-\binom{k}{2}} < 1.\end{aligned}$$

Lemma 3.2.4 implies there exists $G \in \Omega$ with $X(G) = 0$ *i.e.* a two-colouring of K_n with no monochromatic K_k. □

Before moving on to the lower bound these proofs produce, here is an illustration of the counting argument which both proofs of Theorem 5.1.2 formalize. Observe that there for all $K \in [n]^k$, there are $2^{\binom{n}{2}+1-\binom{k}{2}}$ 2-colourings of G for which $[K]^2$ is monochromatic, while there are $2^{\binom{n}{2}}$ two-colourings of G. Therefore, the percentage of two-colourings that leave $[K]^2$ monochromatic is (at most) $2^{1-\binom{k}{2}}$. As $|[n]^k| = \binom{n}{k}$, if

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1,$$

there are two-colourings with no monochromatic $[K]^2$.

Both of these give the following bound on $R(k,k)$.

Theorem 5.1.3. *For every positive integer k,*

$$R(k,k) \geq (1+\mathrm{o}(1)) \frac{k 2^{k/2}}{\sqrt{2}}.$$

Proof. Starting with the relationship:

$$\binom{n}{k} 2^{1-\binom{k}{2}} < \frac{n^k}{k^k} 2^{-\binom{k}{2}+1} \qquad (k \text{ is fixed, } n \text{ varies}).$$

Assume $\frac{n^k}{k^k} 2^{-\binom{k}{2}+1} \leq 1$. Then

$$n^k \leq k^k 2^{\frac{k^2}{2}} 2^{-\frac{k}{2}} 2^{-1} \qquad \text{(simplifying)}.$$

So

$$\begin{aligned} n &\leq k 2^{\frac{k}{2}} 2^{-1/2} 2^{-1/k} && \text{(taking the } k\text{-th root)} \\ &= (1+\mathrm{o}(1)) \frac{k}{\sqrt{2}} 2^{\frac{k}{2}} && (2^{-1/k} = \mathrm{o}(1)). \end{aligned}$$

Thus

$$(1+\mathrm{o}(1)) \frac{k}{\sqrt{2}} 2^{\frac{k}{2}} \leq R(k,k).$$

□

Theorem 5.1.4. *If*

$$\mathrm{e}\left(\binom{k}{2}\binom{n}{k-2} + 1\right) \cdot 2^{1-\binom{k}{2}} < 1 \tag{5.1}$$

then $R(k,k) > n$.

Proof. Assume n satisfies equation (5.1). Let Ω be the collection of two-colourings of K_n as in the second proof of Theorem 5.1.2. Let $p = 2^{1-\binom{k}{2}}$ and $d = \binom{k}{2}\binom{n}{k-2}$. For any k-set $S \in [n]^k$, let A_S be the event that "S is a monochromatic k-clique".

As in Theorem 5.1.2,

$$\mathbb{P}[A_S] = p.$$

To bound the number of events A_S is dependent upon, observe that if $|S \cap T| \leq 1$, then the complete subgraphs generated by S and T share no edges, therefore A_S is independent of A_T, as the colour of one edge is independent of another. For any 2-set $\hat{S} \in [S]^2$ there are at most $\binom{n}{k-2}$ k-sets which contain $\hat{S}$. As there are $\binom{k}{2}$ such $\hat{S}$, there are at most d events dependent on S. While this might be an over count of the number of dependent events, the proof only requires a uniform bound on d. Therefore

$$\mathrm{e}(d+1)p = \mathrm{e}(\binom{k}{2}\binom{n}{k-2} + 1) \cdot 2^{1-\binom{k}{2}} < 1$$

hence Corollary 4.4.3 implies

$$\mathbb{P}[\bigwedge_{S \in [n]^k} \overline{A_S}] > 0$$

i.e. there is a two-colouring with no monochromatic k-clique. $\square$

While Theorem 5.1.4 presents a second method to prove a lower bound on $R(k,k)$, the reason this theorem is useful because Theorem 5.1.4 gives a better lower bound on $R(k,k)$.

Theorem 5.1.5. *For every positive integer, k,*

$$R(k,k) \geq (1 + \mathrm{o}(1))\sqrt{2}k2^{k/2},$$

where $\mathrm{o}(1) \to 0$ *as* $k \to \infty$.

Proof. As k is fixed and n can vary. The approximation

$$\left(\binom{k}{2}\binom{n}{k-2} + 1\right) \sim \left(\binom{k}{2}\binom{n}{k-2}\right)$$

$$\sim \left(\frac{k(k-1)}{2}\frac{n^{k-2}}{(k-2)^{k-2}}\right)$$

is acceptable. Therefore

$$\mathrm{e}\left(\frac{k(k-1)}{2}\frac{n^{k-2}}{(k-2)^{k-2}}\right)2^{1-(k^2-k)/2} < 1,$$

and so

$$\begin{aligned}
\left(\frac{k^2n^{k-2}}{(k-2)^{k-2}}\right) &\leq \frac{2^{(k^2-k)/2-1}}{\mathrm{e}} \\
&\leq 2^{(k^2-k-4)/2} && (2 < \mathrm{e} \text{ and simplifying}).
\end{aligned}$$

Thus

$$\begin{aligned}
n^{k-2} &\leq \frac{(k-2)^{k-2}2^{(k-2)(\frac{k+1}{2}-\frac{1}{k-2})}}{k^2} && (\text{divide } k^2-k-4 \text{ by } k-2) \\
n &\leq (k-2)2^{(k+1)/2}2^{-(k-2)^{-2}} && (\text{taking the } k-2 \text{ root}) \\
&\sim k2^{(k+1)/2}2^{-(k-2)^{-2}} && (k-2 \sim k) \\
&\sim k2^{(k+1)/2}(1+\mathrm{o}(1)) && (2^{-(k-2)^{-2}} = \mathrm{o}(1)) \\
&= (1+\mathrm{o}(1))\sqrt{2}k2^{k/2}.
\end{aligned}$$

Thus

$$(1+\mathrm{o}(1))\sqrt{2}k2^{k/2} \leq R(k,k). \qquad \square$$

Theorem 5.1.6. (See for example [1, pp. 25-26]) *Let $k \in \mathbb{Z}^+$. For all n,*

$$R(k,k) > n - \binom{n}{k}2^{1-\binom{k}{2}}.$$

Proof. For every $n \in \mathbb{Z}^+$, let Ω be the sample space of 2-colourings of $E(K_n)$ that uniformly and randomly colour every $e \in E(K_n)$ with $\mathbb{P}(e \text{ is red}) = \mathbb{P}(\{x,y\} \text{ is blue}) = 1/2$. Let $X : \Omega \to \mathbb{R}$ be the number of monochromatic K_k's. As in the second proof of Theorem 5.1.2,

$$\mathbb{E}(X) = \binom{n}{k}2^{1-\binom{k}{2}}.$$

Lemma 3.2.4 implies there exists an $\chi \in \Omega$ so that $X(\chi) \leq \mathbb{E}(X)$.

For each monochromatic (w.r.t. χ) K_k, remove one vertex. It is possible that the K_k's overlap, but this can only reduce the number of vertices removed, hence improving the result. If m is the remaining number of vertices, then

$$m \geq n - \binom{n}{k} 2^{1-\binom{k}{2}}.$$

The induced colouring on the K_m subgraph has no monochromatic K_k's (as removing one vertex from each monochromatic K_k's means a monochromatic subgraph can be at most a K_{k-1}) as needed. $\square$

Theorem 5.1.7. (See for example [1, p. 16]) *Let $k \in \mathbb{Z}^+$. For $n \in \mathbb{Z}^+$, there is a two-colouring of the edges of K_n with at most $\binom{n}{k}2^{1-\binom{k}{2}}$ monochromatic K_k's.*

Proof. Define the sample space Ω of two-colourings of $E(K_n)$. Choose $\chi \in \Omega$ such that the colour of every edge is determined independently with probability 1/2. The use of $p = 1/2$ is mainly for expediency, while also reflecting the fact that the desired two objects (either a red or a blue K_k) are of the same size, which under these conditions, have an equal probability of occurring.

For an accurate counting of the desired subgraphs, define the random variable X on Ω by

$$X(\chi) = \text{ the number of monochromatic } K_k's.$$

For all $S \in [n]^k$, define the event A_S to be "the edges of K_S are monochromatic" with the respective indicator random variable, X_S.

Claim 1. : $\mathbb{E}[X_S] = \mathbb{P}[A_S] = 2^{1-\binom{k}{2}}$.

The proof of the claim is as the first proof of Theorem 5.1.2.

Since $X = \sum_{S \in [n]^k} X_S$,

$$\mathbb{E}[X] = \mathbb{E}[\sum_{S \in [n]^k} X_S]$$

$$= \sum_{S \in [n]^k} \mathbb{E}[X_S] \qquad \text{(linearity of expectation)}$$
$$= \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Theorem 3.2.4 implies there is a two-colouring with at most $\binom{n}{k}2^{1-\binom{k}{2}}$ monochromatic K_k's, thus completing the proof of Theorem 5.1.7. □

The following example illustrates the same idea, but using a different probability.

Theorem 5.1.8. (See for example [1]) *Let $k \in \mathbb{Z}^+$ be given. For every $3 \leq n \in \mathbb{Z}^+$, there is a 3-colouring of $E(K_n)$ with at most $\binom{n}{3}3^{1-\binom{k}{3}}$ monochromatic K_k's.*

The only difference between this example and the last is that every edge can be coloured in three different ways. Since the desired events, monochromatic K_k's, are still of equal size, changing the sample space to be uniform three-colourings where the probability of any edge having any one colour is $\frac{1}{3}$. Everything else is as in Example 5.1.7.

For the diagonal Ramsey numbers, the desired objects are both copies of a monochromatic K_k, justifying the use of $p = 1/2$ in the above arguments. When $k \neq \ell$, one of the difficulties in using probability theory to get lower bounds on $R(k, \ell)$ is the necessity of using a general value of p.

Theorem 5.1.9. *Let k, ℓ be positive integers. If there exists a probability p such that*

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{\ell} (1-p)^{\binom{\ell}{2}} < 1 \tag{5.2}$$

then $n < R(k, \ell)$.

Proof. Assume p and n satisfy equation (5.2). Let Ω be the collection of two-colourings of K_n. For $K \in [n]^k$, let B_K be the event "all of the edges of K_K are blue" and for $L \in [n]^\ell$, let R_L be the event "all of the edges of K_L are red".

$$\mathbb{P}[B_K] = p^{\binom{k}{2}}$$

$$\mathbb{P}[R_L] = (1-p)^{\binom{\ell}{2}}.$$

There are $\binom{n}{k}$ such K and $\binom{n}{\ell}$ such L, therefore

$$\begin{aligned}
\mathbb{P}[A_\chi] &= \mathbb{P}[(\bigvee_{K\in[n]^k} B_K) \vee (\bigvee_{L\in[n]^\ell} R_L)] \\
&\leq \mathbb{P}[\bigvee_{K\in[n]^k} B_K] + \mathbb{P}[\bigvee_{L\in[n]^\ell} R_L] && \text{(Definition 3.1.2)} \\
&\leq \textstyle\sum_{K\in[n]^k} \mathbb{P}[B_K] + \sum_{L\in[n]^\ell} \mathbb{P}[R_L] \\
&= \binom{n}{k} p^{\binom{k}{2}} + \binom{n}{\ell}(1-p)^{\binom{\ell}{2}} < 1 && \text{(assumption on } n \text{ and } p\text{).}
\end{aligned}$$

Therefore there is a two-colouring $\chi \in \Omega$ that satisfies

$$\begin{aligned}
\overline{\left(\bigvee_{K\in[n]^k} B_K\right) \vee \left(\bigvee_{L\in[n]^\ell} R_L\right)} &= \left(\overline{\bigvee_{K\in[n]^k} B_K}\right) \wedge \left(\overline{\bigvee_{L\in[n]^\ell} R_L}\right) \\
&= \left(\bigwedge_{K\in[n]^k} \overline{B_K}\right) \wedge \left(\bigwedge_{L\in[n]^\ell} \overline{R_L}\right).
\end{aligned}$$

Hence $\chi(G)$ has no blue K_k and no red K_ℓ, which implies $R(k,\ell) > n$ as claimed. □

5.2 Extremal Set theory

In this section, results using the probabilistic method in set theory are shown.

Definition 5.2.1. *A collection $\mathfrak{F}$ of sets is called k-uniform* iff *for every $S \in \mathfrak{F}$, $|S| = k$.*

Definition 5.2.2. *For $k \in \mathbb{Z}^+$, let $B(k)$ be the minimum possible number of sets in a k-uniform family which is not 2-colourable.*

To put it another way, if n is a positive integer, $B(k) < n$ if there is a family $\mathfrak{F} \subseteq X$ of sets with $|\mathfrak{F}| = n$ where every $A \in \mathfrak{F}$ satisfies $|A| = k$ such that for every 2-colouring of X there is an element A of $\mathfrak{F}$ which is monochromatic.

Theorem 5.2.3. [10] *If k is sufficiently large, then*

$$B(k) \leq (1+\mathrm{o}(1))\frac{e\ln 2}{4}k^2 2^k.$$

Proof. Let X be an arbitrary set with $|X| = \lfloor k^2/2 \rfloor = n$ and $\Omega = [X]^k$ with the uniform probability measure. Let $\mathbf{A} \in \Omega$ be a random element (*i.e* $\forall A \in \Omega$, $\mathbb{P}[\mathbf{A} = A] = \frac{1}{\binom{n}{k}}$). Let

$$\mathfrak{F} = \{\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A_b}\}$$

be an independent and random subset of Ω.

Let χ be any two-colouring of X with a red elements and $n-a$ blue elements. For $\mathbf{A} \in \mathfrak{F}$, let $\mathbf{A}_\chi$ denote the event '$\mathbf{A}$ is monochromatic with respect to χ' and

$$B_\chi^F = \bigwedge_{\mathbf{A} \in \mathfrak{F}} \overline{\mathbf{A}_\chi}.$$

Thus

$$\begin{aligned}
\mathbb{P}[\mathbf{A}_\chi] &= \frac{\binom{a}{k} + \binom{n-a}{k}}{\binom{n}{k}} \\
&\geq 2\frac{\binom{\lfloor n/2 \rfloor}{k}}{\binom{n}{k}} && \text{(by Theorem 3.3.6)} \\
&\sim \mathrm{e}^{-1}2^{1-k} && \text{(by Theorem 2.3.4)} \\
&= p.
\end{aligned}$$

The independence of $\mathbf{A}_i \in \mathfrak{F}$ implies

$$\mathbb{P}[B_\chi^F] \leq (1-p)^b \leq e^{-pb}.$$

If $B^F = \bigvee_\chi B_\chi^F$ then

$$\begin{aligned}
\mathbb{P}[B^F] &= \mathbb{P}[\bigvee_\chi B_\chi^F] \\
&\leq \sum_\chi \mathbb{P}[B_\chi^F] \leq 2^n e^{-pb} && (2^n \text{ two-colourings of } X).
\end{aligned}$$

If $2^n e^{-pb} < 1$, there is a family $\mathfrak{D} = \{A_1, A_2 \dots A_b\} \subset \Omega$ witness to the event

$$\begin{aligned}
\overline{B^D} &= \overline{\bigvee_{\chi} B^D_{\chi}} = \bigwedge_{\chi} \overline{B^D_{\chi}} \\
&= \bigwedge_{\chi} \overline{\bigwedge_{A \in \mathfrak{D}} \overline{A_{\chi}}} \\
&= \bigwedge_{\chi} \bigvee_{A \in \mathfrak{D}} A_{\chi},
\end{aligned}$$

i.e., for every 2-colouring χ of X, there is a $A \in \mathfrak{D}$ such that A is monochromatic with respect to χ thus $B(k) < b$.

To complete the proof, it must be shown that for $b = (1 + \mathrm{o}(1))\frac{e \ln 2}{4} k^2 2^k$, $2^n e^{-pb} < 1$. Replace $\mathrm{o}(1)$ by $c_k > 0$, and observe

$$\begin{aligned}
-pb &= -e^{-1} 2^{1-k} \left(\frac{e \ln 2}{4} k^2 2^k + c_k \frac{e \ln 2}{4} k^2 2^k \right) \\
&= -\frac{k^2 \ln 2}{2} - c_k \frac{k^2 \ln 2}{2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
2^n e^{-pb} &= 2^n 2^{-k^2/2 - c_k \frac{k^2 \ln 2}{2}} \\
&= 2^{\lfloor \frac{k^2}{2} \rfloor - \frac{k^2}{2} - c_k \frac{k^2 \ln 2}{2}} && (n = \lfloor k^2/2 \rfloor) \\
&\leq 2^{-c_k \frac{k^2 \ln 2}{2}} < 1 && (c_k > 0)
\end{aligned}$$

as desired. □

Definition 5.2.4. *A family $\mathfrak{F}$ of subsets of X has property B if there exists a two-colouring of X such that no element $A \in \mathfrak{F}$ is monochromatic.*

The next example appears in [1, pp. 65-66].

Example 5.2.5. Let $H = (V, E)$ be a hypergraph in which every edge has at least k-elements and suppose every edge of H intersects at most d other edges. If

$$\mathrm{e}(d + 1)2^{1-k} < 1$$

then H has property B.

Proof. This proof uses Corollary 4.4.3 with the probability $p = 2^{1-k}$ being a bound on the probability that any one edge is monochromatic. Since if $f, g \in E$ and A_f is the event that f is monochromatic, then A_f is dependent on $A_g \iff f$ and g share vertices, the conditions given imply A_f is dependent on at most d events. Observing that

$$\mathbb{P}[\bigwedge_{f \in E} \overline{A_f}] > 0$$

implies there is a two-colouring with no monochromatic edge. □

Let $k \in \mathbb{Z}^+$. A family of sets, $\mathfrak{F}$, is called *k-uniform* iff for all $A \in \mathfrak{F}$ $|A| = k$.

Theorem 5.2.6. [13] *If every member of a k-uniform family intersects at most 2^{k-3} other members, then the family is 2-colourable.*

Proof. This is an application of the Lovász local lemma.

Assume $\mathfrak{F} = \{A_1, A_2 \ldots A_f\}$ is a k-uniform family of subsets of some set X in which every member $A \in \mathfrak{F}$ intersects at most 2^{k-3} other members of $\mathfrak{F}$.

Independently and uniformly colour X ($i.e$ for all $x \in X$; $\mathbb{P}[x \text{ is red}] = 1/2 = \mathbb{P}[x \text{ is blue}]$). For $A \in \mathfrak{F}$, let $\mathbf{A}$ be the event 'A is monochromatic'. As the colouring is independent,

$$\mathbb{P}[\mathbf{A}] = 2^{1-k} = p.$$

Using $d + 1 = 2^{k-3} + 1$, then

$$\begin{aligned} \mathrm{e}(d+1)p &= e(2^{k-3}+1)2^{1-k} \\ &= e(2^{-2} + 2^{1-k}) < 1 \end{aligned}$$

by applying Corollary 4.4.3,

$$\mathbb{P}[\overline{\mathbf{A_1}} \wedge \overline{\mathbf{A_2}} \wedge \cdots \wedge \overline{\mathbf{A_f}}] > 0 \tag{5.3}$$

such that there is two-colouring for which none of the $A_i \in \mathfrak{F}$ is monochromatic. Hence $\mathfrak{F}$ is 2-colourable as needed. □

Before proceeding to the next theorem, the following lemma is needed.

Lemma 5.2.7. *Let $n \in \mathbb{Z}^+$ and $p \in (0,1)$. Let $np \leq s$. Then*

(i) $\alpha = \dfrac{n-s}{s+1}\dfrac{p}{1-p} < 1$

(ii) $\displaystyle\sum_{k=s+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \leq \binom{n}{s} \frac{\alpha^s}{1-\alpha}.$

Proof. For part (i), if $np \leq s$, then $n - s \leq n(1-p)$ while $\dfrac{1}{s+1} < \dfrac{1}{s} < \dfrac{1}{np}$. Therefore

$$\alpha = \frac{n-s}{s+1}\frac{p}{1-p} < \frac{n(1-p)}{np}\frac{p}{1-p} = 1,$$

completing the proof of (i).

For part (ii), note that

$$\frac{\binom{n}{s+1}}{\binom{n}{s}} \frac{p^{s+1}(1-p)^{n-s-1}}{p^s(1-p)^{n-s}} = \alpha$$

so that for $s+1 \leq v \leq n$,

$$\frac{\binom{n}{v}}{\binom{n}{s}} \frac{p^v(1-p)^{n-v}}{p^s(1-p)^{n-s}} = \frac{\binom{n}{v}}{\binom{n}{v-1}} \frac{p^v(1-p)^{n-v}}{p^{v-1}(1-p)^{n-v-1}} \frac{\binom{n}{v-1}}{\binom{n}{v-2}} \frac{p^{v-1}(1-p)^{n-v-1}}{p^{v-2}(1-p)^{n-v-2}} \cdots$$

$$\cdots \frac{\binom{n}{s+1}}{\binom{n}{s}} \frac{p^{s-1}(1-p)^{n-s-1}}{p^s(1-p)^{n-s}} \leq \alpha^v.$$

Therefore

$$\binom{n}{v} \leq \binom{n}{s}\alpha^v;$$

hence

$$\begin{aligned}
\sum_{k=s+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} &\leq \binom{n}{s} \sum_{k=s+1}^{n} \alpha^{k-s} \\
&\leq \binom{n}{s} \alpha^s \sum_{k=1}^{\infty} \alpha^k \\
&= \binom{n}{s} \alpha^s \frac{1}{1-\alpha} \qquad\qquad (\alpha < 1).
\end{aligned}$$

□

If $\mathfrak{Y}$ is a k-uniform family of subsets of X, the next result in this section shows that under certain conditions, it is possible to colour X in $r = \lfloor \frac{k}{\ln k} \rfloor$ colours so that every set $A \in \mathfrak{F}$ has at most $t = \lceil 2e \ln k \rceil$ monochromatic elements.

Theorem 5.2.8. [18] *Let $k \in \mathbb{Z}^+$. Let $\mathfrak{F}$ be a k-uniform family of subsets of X and suppose that no point belongs to more than k elements of $\mathfrak{F}$. Then for k sufficiently large, there exists an $r = \lfloor \frac{k}{\ln k} \rfloor$ colouring of X such that no $A \in \mathfrak{F}$ has more than $t = \lceil 2e \ln k \rceil$ points of the same colour.*

Proof. (This proof is from [23, p. 242]) Colour the points of X with r colours such that each point is coloured randomly and independently with a probability of $1/r$. For $i \in [r]$ and $S \in \mathfrak{F}$, let $A(S, i)$ be the event 'S has more than t points coloured i'. Any two such events, $A(S, i)$, $A(S', i')$ are dependent if $S \cap S' \neq \emptyset$. Define a relation $\sim$ on $\mathfrak{F}$ by $S \sim S'$ iff $A(S, i)$ is dependent on $A(S', i')$. For any event $A(S, i)$, the first step is to bound the number d of dependent events. As S has k elements, each of which are in (at most) $k - 1$ other $S' \in \mathfrak{F}$ and there are r colours,

$$d \leq (1 + k(k-1))r \leq k^3.$$

To apply Corollary (4.4.3), the next step is to bound $\mathbb{P}[A(S, i)]$. For every $S \in \mathfrak{F}$ and every $i \in [r]$,

$$\begin{aligned}
\mathbb{P}[A(S, i)] &= \sum_{j=t+1}^{k} \binom{k}{j} \left(\frac{1}{r}\right)^j \left(1 - \frac{1}{r}\right)^{k-j} \\
&\leq \binom{k}{t} \left(\frac{1}{r}\right)^t && \text{(Lemma 5.2.7)} \\
&\leq \left(\frac{\mathrm{e}k}{t}\right)^t \left(\frac{1}{r}\right)^t && \text{(Lemma 2.3.6)} \\
&= \left(\frac{k\mathrm{e}}{tr}\right)^t \\
&< 2^{-t} < k^{-4} && \text{(assumption on } k\text{).}
\end{aligned}$$

Using $d = k^3$, $p = k^{-4}$, then

$$\mathrm{e}p(d+1) \sim \mathrm{e}k^{-1} < 1.$$

Hence, Corollary (4.4.3) implies

$$\mathbb{P}[\bigwedge_{\substack{S\in\mathfrak{F}\\ i\in[r]}} \overline{A(S,i)}] > 0.$$

Thus there is an r-colouring of X such that no element of $\mathfrak{F}$ has more than t elements of the same colour. $\square$

Definition 5.2.9. *A family $\mathfrak{Y}$ of sets is called an anti-chain if no set of $\mathfrak{Y}$ is contained in another.*

The following result was independently discovered by Lubell [26], Meshalkin [28] and Yamamoto [40]; consequently the result is often called "LYM"-inequality.

Theorem 5.2.10. (LYM-inequality) *Let $n \in \mathbb{Z}^+$. Let X be a set of n elements. Let $\mathfrak{F} \subseteq \mathcal{P}(X)$ be an anti-chain. Then*

$$\sum_{A\in\mathfrak{F}} \frac{1}{\binom{n}{|A|}} \leq 1.$$

Proof. (This proof is from [1, p. 197-198]) Uniformly choose $\sigma \in S_n$ (*i.e.* for all $\sigma \in S_n$, the probability of choosing σ is $\frac{1}{n!}$). Let $\mathcal{C}_\sigma = \{\{\sigma(j) : 1 \leq j \leq i\} \quad 0 \leq i \leq n\}$. To clarify, $\mathcal{C}_\sigma$ is the collection $\{\emptyset\}$ (when $i = 0$, the set is empty), $\{\sigma(1)\}, \{\sigma(1), \sigma(2)\} \ldots \{\sigma(1), \sigma(2), \ldots \sigma(n)\} = \{1, 2 \ldots n\}$ as σ is a permutation. This collection has one set of every cardinality between 0 and n.

For every $A \in \mathcal{P}([n])$, observe that the event '$A \in C_\sigma$' occurs when the set $\{\sigma(1), \sigma(2) \ldots \sigma(|A|)\} = A$, while the elements

$$\sigma(|A| + 1), \sigma(|A| + 2) \ldots \sigma(n) \in [n] \setminus A$$

for which there are $|A|!(n - |A|)!$. Therefore, the uniform choice of σ implies

$$\mathbb{P}[A \in \mathcal{C}_\sigma] = \frac{|A|!(n - |A|)!}{n!} = \frac{1}{\binom{n}{|A|}}. \tag{5.4}$$

For every $\mathbf{A} \subset \mathcal{P}([n])$, let

$$X(\mathbf{A}) = \begin{cases} |\mathbf{A} \cap \mathcal{C}_\sigma| & \text{if } \mathbf{A} \subseteq \mathfrak{F} \\ 0 & \text{otherwise.} \end{cases}$$

and for each $A \in \mathfrak{F}$, let X_A be the indicator random variable for the event "$A \in \mathcal{C}_\sigma$".

Since $\mathfrak{F}$ is an anti-chain and $\mathcal{C}_\sigma$ is a chain,

$$|\mathfrak{F} \cap \mathcal{C}_\sigma| \leq 1$$

so that

$$X = \sum_{A \in \mathfrak{F}} X_A \leq 1.$$

Hence

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{A \in \mathfrak{F}} \mathbb{E}[X_A] \\
&= \sum_{A \in \mathfrak{F}} \mathbb{P}_\sigma[A] \\
&= \sum_{A \in \mathfrak{F}} \frac{1}{\binom{n}{|A|}} \leq 1 \qquad\qquad (\text{as } \mathbb{E}[X] \leq 1.)
\end{aligned}$$

□

5.3 Probabilistic proofs in graph theory

To motivate the next example, here is a lemma. Recall that if $G = (V, E)$ is a graph, a set $V' \subseteq V$ is called independent iff $[V']^2 \cap E = \emptyset$ and $\alpha(G)$ is the cardinality of the largest independent subset of V.

Lemma 5.3.1. *Let G be a graph on n vertices and $\Delta(G) = d$. Then G contains an independent set of at least $n/(d+1)$ vertices.*

Proof. Partition V as follows: Choose a vertex v_1 with smallest degree and let V_1 be the set of v_1 along with v_1's neighbors.
Choose a vertex $v_2 \in V \setminus V_1$ of smallest degree and let V_2 be the set of v_2 along with v_2's neighbors in $V \setminus V_1$.
Continue in this manner until all of V is used up, with sets $V_1, V_2 \ldots V_k$ partitioning V and for $1 \leq i \leq k$,

$$|V_i| = |V \setminus \cup_{j \leq i-1} V_j| \geq n - i(d+1).$$

Let

$$U = \{v_1, v_2, \ldots, v_k.$$

As $n - k(d+1) \leq 0$, then $\frac{n}{d+1} \leq k$. From the construction of U, $U \subset V$ is an independent set of k elements. $\square$

The next two theorems guarantee the existence of an independent set of a certain size. Furthermore, the next theorem, appearing in [1, p. 70-71], guarantees the maintenance of structure.

Theorem 5.3.2. *Let $H = (V, E)$ be a graph with maximum degree d and let $V = V_1 \cup V_2 \cup \ldots \cup V_r$ be a partition of V. Suppose for all $i \in [r]$, $|V_i| \geq 2ed$. Then there is an independent set $W \subset V$ that contains a vertex from each set.*

Proof. Let $g = \lceil 2ed \rceil$. Without loss of generality, assume for all i, $|V_i| = g$ (otherwise choose $V_i^* \subseteq V_i$ with $|V_i^*| = g$. Examine the subgraph induced by $\{V_i^*\}_{i=1}^r$. If two vertices are independent in the induced subgraph then the two vertices are independent in the larger graph).

Form a set W by randomly and uniformly choosing a vertex from each V_i. Using Corollary 4.4.3 of the Lovász local lemma, it is shown below with positive probability, no edge has both of its endpoints in W.

Define the events

- for all $f \in E$, let A_f be the event "both end points of f are in W"
- if $x \neq y$ let $A_{x,y}$ be "$x \in V_i \wedge y \in V_j \quad \text{then} \quad i \neq j$"
- For all $z \in V$ let A_z be "$z \in W$".

By the definition of W, for all $f = \{x, y\} \in E$,

$$A_f = A_{x,y} \wedge A_x \wedge A_y.$$

By the uniformity of the choice of x and y,

$$\mathbb{P}[A_x] = \mathbb{P}[A_y] = \frac{1}{g}.$$

Thus

$$\mathbb{P}[A_f] \leq \mathbb{P}[A_x]\mathbb{P}[A_y] = \frac{1}{g^2}.$$

To estimate the number of events that A_f is dependent upon, if f has endpoints in V_i and V_j then A_f is independent of all events with endpoints not in $V_i \cup V_j$ so that A_f is dependent on at most $2gd - 1$ events. By the choice of g,

$$\frac{\mathrm{e}2gd}{g^2} = \frac{\mathrm{e}((2gd-1)+1)}{g^2} = \frac{2\mathrm{e}d}{g} < 1.$$

Therefore Corollary 4.4.3 implies with positive probability, none of the A_f occur. Thus with positive probability there is a collection W for which no A_f occur. Hence there is an independent set with a vertex in each set as needed. □

Theorem 5.3.3. [36] *Let $n \in \mathbb{Z}^+$. Suppose $G = (V, E)$ is a graph on n vertices. If, for some $1 \leq k \leq n-1$, $\frac{nk}{2} \leq |E(G|$ then $\alpha(G) \geq \frac{n}{2k}$.*

Proof. Let $p = \frac{1}{k}$. Let $S \subset V$ be chosen such that $\mathbb{P}[v \in S] = p$.
For $e \in E$, let A_e be the event "both endpoints of e are in S" and let Y_e be the indicator random variable for A_e. Let $Y = \sum_{e \in E} Y_e$. As $Y_e = 1$ if and only if both end points are in S which occurs with probability p^2, since $|E(G)| = \frac{nk}{2}$, linearity of expectation implies

$$\mathbb{E}[Y] = \frac{nk}{2}p^2.$$

Let $X = X(S) = |S|$ such that $\mathbb{E}[X] = np$. Define $Z = Z(S) = X(S) - Y(S)$. If $0 < Z(S) = \ell$ then S has ℓ more vertices containing both end points in S than edges. This implies that S has at least ℓ independent vertices (as they share no edges).
By linearity of expectation,

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[X] - \mathbb{E}[Y] \\ &= np - \frac{nk}{2}p^2\end{aligned}$$

$$= \frac{n}{k} - \frac{nk}{2}\frac{1}{k^2}$$
$$= \frac{n}{2k}.$$

Thus there is a $S \subseteq [n]$ with $Z(S) \geq \frac{n}{2k}$ which implies S has an independent set of size

$$\frac{n}{2k} \leq \alpha(G). \qquad \square$$

Theorem 5.3.4. *Let $G = (V, E)$ be a graph on $0 < n$ vertices. For $v \in V$, let $\deg(v) = d_v$, then*

$$\alpha(G) \geq \sum_{v \in V} \frac{1}{1 + d_v}.$$

Proof. Let Ω be the collection of total orders of V. Let

$$I_< = \{v \in V : \{v, w\} \in E \to v < w\}$$

so that $x \in I_<$ iff for all $w \in N(x)$ $x < w$.

Claim 1. For all $v \in V$,

$$\mathbb{P}[v \in I_<] = \frac{1}{d_v + 1}.$$

Proof (of claim): If b is the number of total orderings of V and a is the number of total orderings of V where v is the least element in $\{v\} \cup N(v)$. The observation that if $<_1$ is a total ordering of V with $v <_1 x$ then there is a second total ordering $<_2$ with $x <_2 v$, thus for every $w \in N(v)$, the number of orderings where w is the least element of $\{v\} \cup N(v)$ is also a. Hence

$$\sum_{x \in \{v\} \bigcup N(x)} a = (d_v + 1)a = b,$$

which implies

$$a = \frac{b}{d_v + 1}.$$

Together this yields

$$\mathbb{P}[v \in I] = \mathbb{P}[v \text{ is the least element of } \{v\} \cup N(v)]$$

$$= \frac{a}{b} = \frac{1}{d_v + 1},$$

thus concluding the proof of Claim 1.

Returning to the proof of Theorem 5.3.4, let X be a random variable defined by $X(<) = |I_<|$. Let X_v be the indicator random variable for the event '$v \in I_<$'. Hence

$$X = \sum_{v \in V} X_v$$

and therefore

$$\mathbb{E}[X] = \sum_{v \in V} \mathbb{E}[X_v] = \sum_{v \in V} \frac{1}{d_v + 1}.$$

Corollary 3.2.5 implies there is a total ordering $<$ so that:

$$|I_<| = X(I_<) \geq \mathbb{E}[X] = \sum_{v \in V} \frac{1}{d_v + 1}.$$

To complete the proof, it is shown that for any total ordering $<$, $I_<$ is an independent set. Assume $x, y \in I_<$, $x \neq y$ and $y \in N(x)$. As $<$ is a total ordering of V, $x < y$ or $y < x$. Suppose $y < x$ then x is not the least element in $\{x\} \cup N(x)$ which implies $x \notin I_<$, a contradiction, and if $x < y$, then y is not the least element in $N(y) \bigcup y$, a contradiction. Therefore if $x, y \in I_<$ then x and y are independent which implies $\alpha(G) \geq |I_<|$. $\square$

Denote the complete m-partite graph on n vertices partitioned as evenly as possible by $T^m(n)$, called the Turán graph. From this definition of $T^m(n)$, $\alpha(T^m(n)) = m$.

Theorem 5.3.5. *Let $m \leq n \in \mathbb{Z}^+$. Suppose q and r satisfy $n = mq + r$ and $0 \leq r < m$. Let $e = r\binom{q+1}{2} + (m - r)\binom{q}{2}$. Let $H = (W, F)$ be a graph on n vertices and e edges. Then $\alpha(H) \geq m$ and with equality* iff *$H \simeq T^m(n)$.*

Proof. Let $T^m(n) = (V, E)$ be the Turán graph on n vertices.

Claim 1. Assume m is as in Theorem 5.3.5. Then $\sum_{v \in V} \frac{1}{d_v + 1} = m$.

Proof (of Claim 1): Let $V_1, V_2, \ldots, V_m$ be the m classes that V is split into. From the construction of $T^m(n)$, the subgraph induced from each component is complete so that each connected component is a complete subgraph. Thus for all $v \in V_i$, $\deg(v) = |V_i| - 1$. Therefore

$$\sum_{v \in V_i} \frac{1}{\deg(v)+1} = \frac{|V_i|}{\deg(v)+1} = \frac{|V_i|}{|V_i|} = 1.$$

Hence, summing over the m components proves Claim 1.

For the next claim, let $H = (W, F)$ be any graph on n vertices and define

$$D(H) = \sum_{v \in W} \frac{1}{\deg(v)+1}.$$

Claim 2. $D(H)$ is minimized when every vertex has the same degree.

Proof (of Claim 2): Let

$$\mathbf{v} = \left(\frac{1}{\sqrt{(d_{v_1}+1)}}, \frac{1}{\sqrt{(d_{v_2}+1)}}, \ldots, \frac{1}{\sqrt{(d_{v_n}+1)}} \right)$$

and

$$\mathbf{u} = \left(\sqrt{(d_{v_1}+1)}, \sqrt{(d_{v_2}+1)}, \ldots, \sqrt{(d_{v_n}+1)} \right).$$

Lemma (3.3.3) implies

$$\begin{aligned}
n^2 &= (\langle \mathbf{u}, \mathbf{v} \rangle)^2 \\
&\leq \left(\sum_{j=1}^{n} \frac{1}{d_{v_1}+1} + \ldots \frac{1}{d_{v_n}+1} \right) \\
&((d_{v_1}+1) + \ldots (d_{v_n}+1)) \\
&= D(H)(2\mathbb{E}[H] + n) \qquad\qquad (\text{as } \textstyle\sum_{j=1}^{n} d_{v_j} = 2\mathbb{E}[H].)
\end{aligned}$$

Lemma (3.3.3) is minimized when there is a nonzero $t \in \mathbb{R}$ such that $\mathbf{v} = t\mathbf{u}$ or for all $i \in [n]$ $\quad \frac{1}{\sqrt{(d_{v_i}+1)}} = t\sqrt{(d_{v_i}+1)}$. Hence $1/t = d_{v_i} + 1$; thus all of the vertices have the same degree. Completing the proof of Claim 2.

The next part of the proof uses the same idea as Lemma 1. Recall that for any $H = (W, F)$ with $|W| = n$ and $|F| = e$,

$$\alpha(H) \geq \sum_{v \in W} \frac{1}{d_v + 1} \geq m.$$

If '$<$' is any total ordering of W, let $I_<$ and $X(<) = |I_<|$ be as in the previous proof.

Claim 3. If $\alpha(H) = \mathbb{E}[X]$ then X is constant.

Proof (of Claim 3): Assume $\alpha(H) = \mathbb{E}[X]$ and X is not constant. Lemma 1 implies for any total ordering $<$ of W, $\alpha(H) \geq X(I_<)$. If X were not constant, there would be a total ordering $<_1$ of W such that

$$\alpha(H) \geq X(I_{<_1}) > \mathbb{E}[X]$$

which is a contradiction. Thus Claim 3 is shown.

Claim 4. If $\alpha(H) = m$, then H is the union of cliques.

Proof (of Claim 4): Assume otherwise; there is vertex $v \in W$ such that

$$\{x, y\}, \{x, z\} \in F \quad \text{but} \quad \{y, z\} \notin F.$$

Let $<_1$ be an ordering of W that begins with x, y, z and $<_2$ that begins y, z, x and otherwise is the same as $<_1$ then $I_{<_1} = I_{<_2} \cup \{x\}$ (since x is the least element in $<_1$, x is the least element in its neighborhood, whereas in $<_2$, x is greater than two of its neighbors and nothing else is changed) thus X is not constant and $\alpha(H) \neq \mathbb{E}[X]$ which is a contradiction to the assumption. Thus completing the proof of Claim 4.

Returning to the proof of Theorem 5.3.5, assume $H = (W, F)$ is a graph with $|W| = n$ and $|F| = e$ and $T^m(n) = (V, E)$ is the Turán graph. If $H \simeq T^m(n)$, then there is an graph isomorphism $\beta : H \to T^m(n)$ such that for $x, y \in W$, $\{x, y\} \in F$ iff $\{\beta(x), \beta(y)\} \in E$. Hence any pair $x, y \in W$ are independent iff $\beta(x), \beta(y)$ are independent in $T^m(n)$. Therefore $V_1, V_2, \ldots, V_m$ partition V into independent sets iff $\beta^{-1}(V_1), \beta^{-1}(V_2), \ldots, \beta^{-1}(V_m)$ partition W into independent sets. Hence $\alpha(H) = m$.

Suppose $\alpha(H) = m$. Claim 4 implies the H is the union of cliques. Partition W into $W_1, W_2, \ldots, W_k$ so that for each i, W_i is a maximal clique.

Claim 5. If $W_1, W_2, \ldots, W_k$ are a partition of W into maximal cliques, then $k = m$.

As H is the union of cliques and $W_1, W_2, \ldots, W_k$ are maximal cliques, each of the W_i's are independent sets. Therefore, $k \leq m$.

Suppose $k < m$. Partition W into independent sets $U_1, U_2, \ldots, U_m$. As $k < m$ and both collections partition W, there is a $1 \leq i \leq k$ and $1 \leq s < t \leq m$ such that $W_i \cap U_s \neq \emptyset$ and $W_i \cap U_t \neq \emptyset$. This is a contradiction as $W_1, W_2, \ldots, W_k$ are maximal cliques while $U_1, U_2, \ldots, U_m$ are independent sets. Therefore $m \leq k$ and Claim 5 is shown.

As $\alpha(H) = \sum_{v \in W} \frac{1}{1+d_v} = m$ is minimized, Claim 2 implies every vertex has as close to the same degree as possible. As this is also true in $T^m(n)$, there is an isomorphism between H and $T^m(n)$. □

The last result in this section involves extremal graph theory. Suppose H is some graph on k vertices and $n \geq k$. The number $ex(n, H)$ is the largest number of edges such that for any graph on n vertices with $ex(n, H) + 1$ edges has a copy of H as a subgraph.

Theorem 5.3.6. *There is a constant c such that for n sufficiently large,*

$$ex(n, K_{2,2}) > cn^{\frac{4}{3}}.$$

Proof. Assume n is large and $p \in (0, 1)$. For $G \in \mathcal{G}_{n,p}$, let $X(G)$ count the number of (not necessarily induced) copies of $K_{2,2}$ in G and let $Y(G)$ be the number of edges in G. Then

$$\mathbb{E}[X] = 3\binom{n}{4}p^4 \sim \frac{n^4p^4}{8}$$

and

$$\mathbb{E}[Y] = \binom{n}{2}p \sim \frac{n^2p}{2}.$$

Let $Z = Y - X$. Then Z is the random variable in $\mathcal{G}_{n,p}$ measuring the difference between the number of edges in a graph and the number of $K_{2,2}$'s. Linearity of expectation implies

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[Y] - \mathbb{E}[X] \\ &\sim \frac{n^2p}{2} - \frac{n^4p^4}{8} \\ &= \frac{n^2p}{2}\left(1 - \frac{n^2p^3}{4}\right).\end{aligned}$$

When $p(n) = n^{-2/3}$, then

$$\begin{aligned}\mathbb{E}[Z] &= \frac{n^{4/3}}{2}\left(1 - \frac{1}{4}\right) \\ &= \frac{3n^{4/3}}{8}.\end{aligned}$$

By Lemma 3.2.4, fix a graph G such that $Z(G) \geq \mathbb{E}[Z]$. Removing one edge from every $K_{2,2}$ forms a new graph G' with no $K_{2,2}$'s and $Y(G) - X(G) \geq \frac{3n^{4/3}}{8}$ edges as needed. □

The actual bounds for $ex(n, K_{2,2}) = cn^{3/2}$ (see [11] for the upper bound and [33] for the lower bound). Thus Theorem 5.3.6 is an example where the probabilistic bound is not the best possible.

5.4 Other examples

For $\mathbf{v} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, let $||\mathbf{v}|| = \sqrt{x_1^2 + x_2^2 \cdots + x_n^2}$ be the Euclidean norm on $\mathbb{R}^n$. The next theorem is a famous result in combinatorial geometry.

Theorem 5.4.1. [25] *Let $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n} \in \mathbb{R}^n$, where for each i, $||\mathbf{v_i}|| = 1$. There exists $\epsilon_1, \epsilon_2, \ldots, \epsilon_n \in \{-1, 1\}$ such that*

$$||\epsilon_1\mathbf{v_1} + \epsilon_2\mathbf{v_2} + \ldots + \epsilon_n\mathbf{v_n}|| \leq \sqrt{n}$$

, and there exists $\alpha_1, \alpha_2, \ldots, \alpha_n \in \{-1, 1\}$ such that

$$||\alpha_1\mathbf{v_1} + \alpha_2\mathbf{v_2} + \cdots + \alpha_n\mathbf{v_n}|| \geq \sqrt{n}.$$

Proof. Select $\beta_1, \beta_2 \ldots \beta_n$ independently and uniformly from $\{-1, +1\}$. Let

$$X(\beta_1, \ldots \beta_n) = ||\beta_1 \mathbf{v_1} + \ldots + \beta_n \mathbf{v_n}||^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j \mathbf{v_i} \cdot \mathbf{v_j}.$$

From the definition,

$$\begin{aligned} E[\beta_i] =& -1\mathbb{P}[\beta_i = -1] + \mathbb{P}[\beta_i = 1] \\ =& (-1)\frac{1}{2} + \frac{1}{2} && \text{(uniformity of the choice of } \beta_i) \\ =& 0. \end{aligned}$$

Furthermore, as the β_i are chosen independently,

$$\mathbb{E}[\beta_i \beta_j] = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases}$$

Hence

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}[\epsilon_i \epsilon_j] \mathbf{v_i} \cdot \mathbf{v_j} \\ &= \sum_{i=1}^{n} \mathbf{v_i} \cdot \mathbf{v_i} = n. \end{aligned}$$

Thus there exists an $\epsilon_1, \epsilon_2 \ldots \epsilon_n$ and an $\alpha_1, \alpha_2, \ldots \alpha_n$ such that

$$\begin{aligned} X(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) &= ||\epsilon_1 \mathbf{v_1} + \ldots + \epsilon_n \mathbf{v_n}||^2 \\ &\leq \mathbb{E}[X] = n. \end{aligned}$$

While

$$\begin{aligned} X(\alpha_1, \alpha_2, \ldots, \alpha_n) &= ||\alpha_1 \mathbf{v_1} + \ldots + \alpha_n \mathbf{v_n}||^2 \\ &\geq \mathbb{E}[X] = n. \end{aligned}$$

Taking square roots yields the result. □

Definition 5.4.2. *A set $x_1, x_2, \ldots x_k$ of positive integers is said to have distinct sums* iff *for all $S \subseteq [k]$,*

$$\sum_{i \in S} x_i$$

are distinct.

The proof of the next theorem uses the second moment method.

Theorem 5.4.3. ([1, pp. 52-53]) *Let $f(n)$ denote the maximal k for which there exists a set*

$$\{x_1, x_2, \ldots, x_k\} \subset [n]$$

with distinct sums. Then

$$f(n) \leq log_2(n) + \frac{1}{2}\log_2 \log_2 n + \mathrm{O}(1).$$

Proof. Fix $\{x_1, x_2, \ldots, x_k\} \in [n]^k$ with distinct sums. Let $\epsilon_1, \epsilon_2, \ldots, \epsilon_k \in \{0, 1\}$ be chosen independently so that for all $j \in [k]$,

$$\mathbb{P}[\epsilon_j = 0] = \mathbb{P}[\epsilon_j = 1] = \frac{1}{2}$$

and define

$$X = \epsilon_1 x_1 + \epsilon_2 x_2 + \ldots + \epsilon_k x_k \qquad (X \text{ is a random sum}).$$

As the ϵ's are chosen independently,

$$\mathbb{E}[X] = \frac{x_1 + x_2 + \ldots + x_k}{2}$$

To calculate the variance,

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{i=i}^{k} \mathrm{Var}[\epsilon_i x_i] && (\epsilon_i\text{'s are independent}),\\
&= \sum_{i=1}^{k} \mathbb{E}[(\epsilon_i x_i)^2] - \mathbb{E}[\epsilon_i x_i]^2\\
&= \sum_{i=1}^{k} \mathbb{E}[(\epsilon_i)^2] x_i - \frac{1}{4} x_i^2.
\end{aligned}$$

Since every $x_i \in [n]$,

$$\mathrm{Var}[X] = \frac{\sum_{i=1}^{k} x_i^2}{4} \leq \frac{kn^2}{4}.$$

Chebychev's inequality implies for every $\lambda > 1$,

$$\mathbb{P}\left[|X - \mu| \geq \lambda \frac{n\sqrt{k}}{2}\right] \leq \frac{\mathrm{Var}[X]}{(\lambda \frac{n\sqrt{k}}{2})^2}$$

$$\leq \frac{n^2k/4}{\lambda^2 n^2 k/4} = \lambda^{-2}$$

or

$$\mathbb{P}[|X-\mu| < \lambda\frac{n\sqrt{k}}{2}] \geq 1-\lambda^{-2}. \tag{5.5}$$

Since the x_i's were chosen to have distinct sums, to any $t \in \mathbb{R}$,

$$\mathbb{P}[X=t] = 0 \quad \text{or} \quad 2^{-k}$$

(as there is one distinct event corresponding to every t, which occurs with probability 2^{-k} or 0 depending on whether or not X can equal t). Thus,

$$\mathbb{P}[|X-\mu| < \lambda\frac{n\sqrt{k}}{2}] \leq 2^{-k}\left(\lambda\frac{n\sqrt{k}}{2}+1\right) \tag{5.6}$$

since $|X-\mu| < \lambda\frac{n\sqrt{k}}{2}$ is the event $-\lambda\frac{n\sqrt{k}}{2} < X-\mu < \lambda\frac{n\sqrt{k}}{2}$ which can occur $\lambda n\sqrt{k}+1$ ways each with at most a probability of 2^{-k}.

Putting together inequalities (5.5) and (5.6) yield

$$1-\lambda^{-2} \leq \mathbb{P}[|X-\mu| \leq \lambda\frac{n\sqrt{k}}{2}] \leq 2^{-k}(\lambda n\sqrt{k}+1) \tag{5.7}$$

or

$$1-\lambda^{-2} \leq 2^{-k}(\lambda n\sqrt{k}+1),$$

from which it follows that

$$\frac{2^k(1-\lambda^{-2})-1}{\lambda\sqrt{k}} \leq n.$$

Take $\lambda = \sqrt{3}$ yields:

$$\frac{2^k\frac{2}{3}-1}{\sqrt{3k}} < n.$$

Let $k_0 = \log_2(n) + \frac{1}{2}\log_2\log_2(n) + c$. Then

$$2^{k_0} = 2^c n\sqrt{\log_2(n)}.$$

While

$$\begin{aligned}\sqrt{3k_0} &= \sqrt{3(\log_2(n) + \frac{1}{2}\log_2\log_2(n) + c)} \\ &= \sqrt{3\log_2(n) + \frac{3}{2}\log_2\log_2(n) + 3c}.\end{aligned}$$

Therefore

$$\begin{aligned}\frac{2^{k_0+1} - 3}{2\sqrt{3k_0}} &= \frac{2^{c+1} n\sqrt{\log_2(n)} - 3}{3\sqrt{3\log_2(n) + \frac{3}{2}\log_2\log_2(n) + 3c}} \\ &= \frac{n}{3}\sqrt{\frac{2^{2(c+1)}\log_2(n)}{3\log_2(n) + \frac{3}{2}\log_2\log_2(n) + 3c}} - \frac{3}{\sqrt{3\log_2(n) + \frac{3}{2}\log_2\log_2(n) + 3c}}.\end{aligned}$$

Picking a proper value for $0 < c$ gives the result. $\square$

Chapter 6

Probability models for graphs

In this chapter, two probability models for graphs are introduced along with some properties of each. At the end of the chapter, a result combining graph theory, logic and probability is shown providing the basis for the chapter on threshold functions.

Definition 6.0.4. *Let $n \in \mathbb{Z}^+$. The set Ω is the collection of all graphs on an n-set, V.*

6.1 Model A

Definition 6.1.1 (Model A)**.** *For $0 < p < 1$, and $n \in \mathbb{Z}^+$, let $\mathcal{G}_{n,p}$ denote the model (called Model A) of graphs on n vertices in which the edges are chosen independently, each with probability p.*

To clarify, $\mathcal{G}_{n,p} = (\Omega, \mathbb{P})$ where for every edge, e, and each graph, $G \in \Omega$,

$$\mathbb{P}[e \in E(G)] = p,$$

and each edge is independently chosen. The next lemma shows this is a probability space.

Lemma 6.1.2. *For every $n \in \mathbb{Z}^+$, $0 < p < 1$, $\mathcal{G}_{n,p} = (\Omega, \mathbb{P})$ is a probability* *space such that for all $G_0 \in \Omega$, if $|E(G_0)| = m$ then*

$$\mathbb{P}[G_0] = p^m(1-p)^{\binom{n}{2}-m}.$$

Proof. It must be shown that $\mathbb{P}$ is a probability measure on $\mathcal{G}_{n,p}$.

(i) Let $G_0 \in \Omega$ have m edges. For every $e \in [V]^2$, let A_e be the event "$e \in E(G)$" so that

$$\begin{aligned}
\mathbb{P}[\{G_0\}] &= \mathbb{P}[G = G_0] \\
&= \mathbb{P}[(\bigwedge_{e \in E(G)} A_e) \wedge (\bigwedge_{e \notin E(G)} \overline{A_e})] \\
&= p^m(1-p)^{\binom{n}{2}-m} && \text{(indep. of } A_e\text{)}
\end{aligned}$$

as claimed.

(ii) Since for every $0 \leq e \leq \binom{n}{2}$, there are $\binom{\binom{n}{2}}{e}$ graphs with $|E(G)| = e$, Theorem 2.3.5 implies

$$\begin{aligned}
\mathbb{P}[\Omega] &= \sum_{G \in \Omega} \mathbb{P}[G] \\
&= \sum_{m=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{m} p^m q^{\binom{n}{2}-m} = 1.
\end{aligned}$$

(iii) If $A \subset B \subset \Omega$, then $0 \leq \mathbb{P}[A] \leq \mathbb{P}[B] \leq 1$.
Notice $B = A \wedge (B \setminus A)$ therefore

$$\begin{aligned}
\mathbb{P}[B] &= \sum_{G \in B} \mathbb{P}[G] \\
&= \sum_{G \in A} \mathbb{P}[G] + \sum_{G \in B \setminus A} \mathbb{P}[G] && (A \text{ and } B \setminus A \text{ are disjoint}) \\
&\geq \sum_{G \in A} \mathbb{P}[G] \\
&= \mathbb{P}[A]
\end{aligned}$$

as needed.

(iv) $\mathbb{P}[\emptyset] = 1 - \mathbb{P}[\Omega] = 1 - 1 = 0.$

Thus $\mathbb{P}$ is a probability measure. $\square$

Example 6.1.3. Expected degree of a vertex

For $G \in \mathcal{G}_{n,p}$, let $X(G) = \deg_G(v_1)$, and for $2 \le i \le n$, let $X_i(G)$ be the indicator random variable for the event '$\{v_1, v_i\} \in E(G)$' so that $\mathbb{E}[X_i] = p$ and $X = \sum_{i=2}^{n} X_i$. Hence linearity of expectation implies

$$\mathbb{E}[X] = (n-1)p.$$

Example 6.1.4. Expected number of edges

Let $V = \{v_1, v_2, \ldots, v_n\}$. Enumerate the pairs $\{v_i, v_j\} \in [V]^2$ in any manner. For $G \in \mathcal{G}_{n,p}$, let $X(G) = |E(G)|$ and for $1 \le i \le \binom{n}{2}$, let X_i be the indicator random variable of the event '$i \in E(G)$'. From the definition of $\mathcal{G}_{n,p}$, $\mathbb{E}[X_i] = p$ and

$$X = \sum_{i=1}^{\binom{n}{2}} X_i,$$

so by linearity of expectation,

$$\mathbb{E}[X] = \binom{n}{2} p.$$

These two examples illustrate properties of Model A: the larger p is, the larger the expected number of edges of graph is (see Theorem 6.3.5). Recall that a vertex v is isolated iff there is no edge incident to v.

Example 6.1.5. Expected number of isolated vertices

Let X be a random variable on $\mathcal{G}_{n,p}$ counting the number of isolated vertices in a graph; for $v \in V$, let A_v be the event 'v is isolated' and let X_v be the respective indicator random variable. Then

$$\mathbb{E}[X_v] = (1-p)^{n-1}.$$

As $X = \sum_{v \in V} X_v$,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{v \in V} \mathbb{E}[X_v] \\ &= n(1-p)^{n-1} && \text{(linearity of expectation).} \end{aligned}$$

Some terminology is needed for the next definition. For every $n \in \mathbb{Z}^+$, let Ω_n be a model of random graphs of order n so that $\Omega_n = G_{n,p(n)}$ for some $p(n) \in [0,1]$ or $\Omega_n = G_{n,M(n)}$ for some $0 \leq M(n) \leq \binom{n}{2}$.

Definition 6.1.6 (Almost surely)**.** *Let $\{\Omega_n\}_{n=1}^{\infty}$ be a sequence of models of random graphs. A property of graphs Q is said to hold almost surely* iff $\lim_{n\to\infty} \mathbb{P}[Q] = 1$.

As an example of the calculations in later sections, the next example considers the property of connectedness in graphs. Recall that a graph is connected iff between any two vertices there is a path.

Theorem 6.1.7. (See [30]) *Let $0 < p < 1$ be given. In $\mathcal{G}_{n,p}$ almost surely every graph is connected.*

Proof. Assume n and p are given. For $G \in \mathcal{G}_{n,p}$, let

$$X(G) = \begin{cases} 1 & G \text{ is disconnected} \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathbb{E}[X]$ is the probability that a graph is disconnected in $\mathcal{G}_{n,p}$.

Let $\mathcal{P} = \{U, W\}$ be a nontrivial (*i.e.* $U \neq \emptyset \neq W$) partition of V. Say $\mathcal{P}$ is a *good* partition if there are no edges between vertices in U and vertices in W. Thus

$$\mathbb{P}[\mathcal{P} \text{ is good}] = (1-p)^{|U||W|}.$$

For every $k \in [n]$, there are $\binom{n}{k}$ partitions with $|U| = k$ and $|W| = n-k$, thus

$$\mathbb{E}[X] \leq \sum_{k=1}^{n-1} \binom{n}{k} (1-p)^{k(n-k)},$$

the inequality being necessary as a disconnected graph may be partitioned in multiple ways. Inequality (3.4) implies

$$\mathbb{P}[X \geq 1] \leq \mathbb{E}[X]$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{n-1} \binom{n}{k} (1-p)^{k(n-k)} \\
&\leq \sum_{k=1}^{n-1} \binom{n}{k} \mathrm{e}^{k(-p(n-k))} \\
&\leq \sum_{k=1}^{n-1} n^k \mathrm{e}^{k(-p(n-k)} \\
&= \sum_{k=1}^{n-1} (\mathrm{e}^p n \mathrm{e}^{-pn})^k \\
&\sim \sum_{k=1}^{n-1} (n\mathrm{e}^{-pn})^k \\
&\leq (n-1)(n\mathrm{e}^{-pn}) = \mathrm{o}(1) \quad (p \text{ fixed and } n \text{ large} \Rightarrow n^2\mathrm{e}^{-pn} < 1).
\end{aligned}
$$

$\square$

Theorem 6.1.8. (See [30]) *Let $p \in (0,1)$ be given. Define the random variable κ on $\mathcal{G}_{n,p}$ by $\kappa(G)$ is the size of the largest clique. Then*

$$\lim_{n\to\infty} \mathbb{P}\left[\kappa(G) > \frac{2\ln n}{-\ln p}\right] = 0.$$

Proof. For $n \in \mathbb{Z}+$, let $r = \left\lceil \dfrac{(-2\ln n - \ln p)}{\ln p} \right\rceil$. For every $R \in [n]^r$, let A_R be the event 'the subgraph induced by R is a clique' and κ_R be the respective indicator random variable. Let $\kappa_r = \sum_{R\in[n]^r} X_R$ be the random variable on $\mathcal{G}_{n,p}$ counting the number of r-cliques. Note that if $\kappa_r(G) \geq 1$, then $\kappa(G) \geq r$. As the subgraph generated by R has $\binom{r}{2}$ edges,

$$\mathbb{E}[\kappa_R] = p^{\binom{r}{2}}.$$

Thus

$$
\begin{aligned}
\mathbb{E}[\kappa_r] &= \sum_{R\in[n]^r} \mathbb{E}[\kappa_R] && \text{(linearity of expectation)} \\
&= \binom{n}{r} p^{\binom{r}{2}} \\
&\sim (ne)^r p^{\binom{r}{2}} && \text{(Lemma 2.3.4)}
\end{aligned}
$$

$$= (nep^{\binom{r-1}{2}})^r.$$

Since $p \in (0,1)$ is fixed, equation (3.3.8) implies

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}[\kappa(G) > r] &\leq \lim_{n\to\infty} \mathbb{P}[\kappa_r(G) \geq 1] \\
&\leq \lim_{n\to\infty} \mathbb{E}[\kappa_r] \\
&\leq \lim_{n\to\infty} \left[nep^{\binom{r-1}{2}}\right]^r = 0.
\end{aligned}$$

□

6.2 Model B

Definition 6.2.1. *(Model B)*

For $n \in \mathbb{Z}^+$, and $0 \leq q \leq \binom{n}{2}$, let $\Omega_q = \{G \in \Omega : |E(G)| = q\}$. Define

$$\forall G_o \in \Omega_q \qquad \mathbb{P}[G = G_0] = \binom{\binom{n}{2}}{q}^{-1}$$

and

$$\mathcal{G}_{n,q} = (\Omega_q, \mathbb{P}).$$

Lemma 6.2.2. *$\mathcal{G}_{n,q}$ is a uniform probability space.*

Proof. (i) $\mathbb{P}[\Omega_q] = 1$. There are $\binom{\binom{n}{2}}{q}$ graphs in Ω_q, thus

$$\begin{aligned}
\mathbb{P}[\Omega_q] &= \sum_{G\in\Omega_q} \mathbb{P}[G] \\
&= \frac{1}{\binom{\binom{n}{2}}{q}} \sum_{G\in\Omega_q} 1 = 1.
\end{aligned}$$

(ii) If $A \subseteq B \subseteq \Omega_q$ then

$$\begin{aligned}
\mathbb{P}[B] &= \sum_{G\in B} \mathbb{P}[G] \\
&= \sum_{G\in A} \mathbb{P}[G] + \sum_{G\in B\setminus A} \mathbb{P}[G] \\
&\geq \sum_{G\in A} \mathbb{P}[G] = \mathbb{P}[A]
\end{aligned}$$

as needed.

(iii) $\mathbb{P}[\emptyset] = 1 - \mathbb{P}[\Omega_q] = 1 - 1 = 0$.

Thus $\mathcal{G}_{n,q}$ is a probability space. $\square$

Example 6.2.3. [30, p. 9-10]

In $G_{4,2}$, the probability of a graph being connected is zero as no graphs with four vertices and two edges are connected. In $G_{4,3}$, the probability of a graph being connected is

$$\begin{aligned} 1 - \frac{\binom{4}{3}}{\binom{\binom{4}{2}}{3}} &= 1 - \frac{4}{\binom{6}{3}} \\ &= 1 - 4/20 = 4/5. \end{aligned}$$

Example 6.2.4. For $G \in \mathcal{G}_{n,q}$ with $V(G) = \{v_1, v_2, \ldots, v_n\}$, let $X(G) = \deg(v_1)$. Then $\mathbb{E}[X] = 2q/n$.

Proof. Let $2 \leq i \leq n$. Let X_i be the indicator random variable for the event "$\{v_1, v_i\} \in E(G)$". As there are $\binom{n}{2}$ edges to choose from, picking one (say $\{v_1, v_i\}$) leaves the remaining $\binom{n}{2} - 1$ edges from which $q - 1$ are chosen. Therefore for all $2 \leq i \leq n$,

$$\mathbb{E}[X_i] = \binom{\binom{n}{2} - 1}{q - 1} \binom{\binom{n}{2}}{q}^{-1}.$$

Linearity of expectation gives

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=2}^{n} \mathbb{E}[X_i] = (n-1)\mathbb{E}[X_2] \\ &= (n-1) \binom{\binom{n}{2} - 1}{q - 1} \binom{\binom{n}{2}}{q}^{-1} \\ &= (n-1) \left(\frac{(\binom{n}{2} - 1)!}{(q-1)!(\binom{n}{2} - q)!} \right) \left(\frac{q!(\binom{n}{2} - q)!}{\binom{n}{2}!} \right) \\ &= \frac{(n-1)q}{n(n-1)/2} = \frac{2q}{n} \end{aligned}$$

$\square$

6.3 Comparative results in $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,q}$

Throughout the rest of these notes, it will occasionally be necessary to compare two measure spaces G_{n,p_1} and G_{n,p_2} (or G_{n,q_1} and G_{n,q_2}) with respect to some property Q. In these cases, the notation $\mathbb{P}_{p_1}(Q)$ and $\mathbb{P}_{p_2}(Q)$ (or $\mathbb{P}_{q_1}(Q)$ and $\mathbb{P}_{q_2}(Q)$ respectively) to refer to the measure of Q in G_{n,p_1} versus the measure of Q in G_{n,p_2} (or G_{n,q_1} versus G_{n,q_2} respectively). Theorem 6.3.5 is an example of comparing a property in two measure spaces.

Definition 6.3.1. *A property Q of graphs is called monotone increasing if whenever G satisfies Q and G is a subgraph of H then H satisfies Q.*

Example 6.3.2. Let Q be the property "G contains a triangle". Then Q is monotone increasing since if a subgraph contains a triangle, so does the larger graph.

Example 6.3.3. The property that G contains a cycle is monotone increasing. If G is a subgraph of H containing a cycle then so does H.

Example 6.3.4. Before the next theorem, here is a algorithm to produce a graph in $\mathcal{G}_{n,q}$. Enumerate all of the possible $\binom{n}{2}$ edges for a graph in $\mathcal{G}_{n,q}$, then form a graph H by choosing q edges randomly and uniformly (*e.g.*, via a random number generator where the probability of every number is equal). The first part of the proof of the next theorem uses this idea.

Theorem 6.3.5. *Let $n \in \mathbb{Z}^+$ be given. Suppose Q is a monotone increasing property. If $0 \leq q_1 < q_2 \leq \binom{n}{2}$ and $0 \leq p_1 < p_2 \leq 1$ then*

$$\mathbb{P}_{q_1}(Q) \leq \mathbb{P}_{q_2}(Q)$$

and

$$\mathbb{P}_{p_1}(Q) \leq \mathbb{P}_{p_2}(Q).$$

Proof. As in Example 6.3.4, form a graph $H \in \mathcal{G}_{n,q_1}$ and extend H to a graph $G \in \mathcal{G}_{n,q_2}$ in the same manner by choosing the remaining $q_2 - q_1$ edges in the same manner. If H has Q then G has Q. As every $G \in \mathcal{G}_{n,q_2}$ can be generated in such fashion,

$$\mathbb{P}_{q_1}[Q] \leq \mathbb{P}_{q_2}[Q].$$

Let $p = \frac{p_2 - p_1}{1-p_1}$. Choose any $G_1 \in \mathcal{G}_{n,p_1}$ and randomly independently choose a second graph $G_2 \in \mathcal{G}_{n,p}$. Form a new graph $G \in \Omega$ by $E(G) = E(G_1) \cup E(G_2)$. The next step is to show that $G \in G_{n,p_2}$. As $\mathbb{P}[e \in E(G_1)] = p_1$ and $\mathbb{P}[e \in E(G_2)] = p$ and the edges are chosen independently,

$$\begin{aligned}\mathbb{P}[\{e \in E(G_1\} \wedge \{\mathrm{e} \in E(G_2)\}] &= \mathbb{P}[\{e \in E(G_1\}]\mathbb{P}[\{e \in E(G_2)\}] \\ &= p_1 p\end{aligned}$$

so that

$$\begin{aligned}\mathbb{P}[e \in E(G_1) \cup E(G_2)] &= \mathbb{P}[\{e \in E(G_1)\} \vee \{e \in E(G_2)\}] \\ &= \mathbb{P}[\{e \in E(G_1)\}] + \mathbb{P}[\{e \in E(G_2\}] \\ &- \mathbb{P}[\{e \in E(G_1)\} \wedge \{e \in E(G_2)\}] \\ &= p_1 + p - p_1 p = p_2\end{aligned}$$

implying $G \in G_{n,p_2}$. As Q is monotone increasing, if G_1 has Q so does G. Therefore

$$\mathbb{P}_{p_1}(Q) \leq \mathbb{P}_{p_2}(Q). \qquad \square$$

Example 6.3.2 and Theorem 6.3.5 give the following.

Corollary 6.3.6. *Let Q be the property "G has a triangle" and q_1, q_2, p_1, p_2 be as in Theorem 6.3.5. Then*

$$\mathbb{P}_{q_1}(Q) \leq \mathbb{P}_{q_2}(Q)$$

and

$$\mathbb{P}_{p_1}(Q) \leq \mathbb{P}_{p_2}(Q).$$

6.4 Zero-One laws in graph theory

This section discusses some of the results regarding the first order theory of graphs. This theory of logic consists of variables $\{x, y, z, u, v, \ldots\}$ representing vertices of a graph, the operations of equality and adjacency ($x = y, x \sim y$), the usual Boolean connectives ($\wedge, \vee, \ldots$), along with universal and existential quantification ($\forall, \exists$).

Example 6.4.1. Some examples of first order statements in theory of graphs. In the first order theory of graphs, the statement:

$$\exists x \exists y \exists x [x \sim y \wedge y \sim z \sim x]$$

is the property "there exists vertices x, y, z such that $\{x, y\}, \{y, z\}, \{z, x\} \in E(G)$"; *i.e.* G contains the triangle formed by the vertices x, y, z.

In the first order theory of graphs, the statement:

$$\forall x \exists y [x \sim y]$$

is the property "for all vertices x, there exists a vertex y such that $\{x, y\} \in E(G)$; *i.e.* G contains no isolated vertices.

A theory of logic is called *consistent* if the theory has a model. A theory is *complete* if, for every formula in its signature, either that formula or its negation is a logical consequence of the axioms of the theory. In what follows, the next definition is neccessary.

Definition 6.4.2. *A finite set Γ of sentences truth-functionally entails a sentence P (denoted Gamma $\vDash P$) if and only if there is no truth-value assignment on which every member of Γ is true and P is false.*

Definition 6.4.3. *For an integer $k > 0$, and a graph G with $n > 2k$ vertices, let P_k be the proposition "for every disjoint $W_1, W_2 \subset [n]$ with $|W_1|, |W_2| \leq k$, there is a $z \in [n] \setminus (W_1 \cup W_2)$ adjacent to every $v \in W_1$ and not adjacent to any $v \in W_2$".*

The following two lemmas can be found in *e.g.* [4].

Lemma 6.4.4. *Choose $k \in \mathbb{N}$. Enumerate the prime numbers in $\mathbb{Z}^+$ as $p_1, p_2, p_3 \ldots$ in increasing order. Define*

$$E = \{\{i, j\} : (i < j) \wedge (p_i \text{ divides } j)\}.$$

Then $G = (\mathbb{N}, E)$ has P_k.

Proof. Let $W_1, W_2 \subset \mathbb{N}$ be disjoint with $|W_i| \leq k$ for $i = 1, 2$. Let

$$x = \prod_{i \in W_1} p_i.$$

From the definition of G, for every $i \in W_1$, $\{i, x\} \in E(G)$ while for every $j \in W_2$, $\{j, x\} \notin E(G)$, as W_1 and W_2 are disjoint and for $j \in W_2$, p_j does not divide x. As W_1, $W_2 \subset \mathbb{N}$ with $|W_i| \leq k$ arbitrary implies G has P_k. $\square$

Lemma 6.4.5. *Let $k \in \mathbb{N}$ be fixed. Then*

$$\lim_{n \to \infty} \mathbb{P}[G_{n,p} \text{ satisfies } P_k] = 1.$$

Proof. Observe that for $k > 1$, P_k implies P_{k-1}; therefore it is enough to prove the statement is true for k-sets and the general statement follows.

Let k and $n \geq 2k$ be given. For every disjoint $W_1, W_2 \subset [n]^k$ and $z \in [n] \setminus (W_1 \cup W_2)$, let $R_k(W_1, W_2; z)$ be the proposition that z is adjacent to every $v \in W_1$ and not adjacent to any $v \in W_2$ and $S_k(W_1, W_2)$ be the proposition that there exists a vertex z adjacent to every $v \in W_1$ and not adjacent to any $v \in W_2$. Therefore

$$S_k(W_1, W_2) = \bigvee_{z \in [n] \setminus (W_1 \cup W_2)} R_k(W_1, W_2; z).$$

Hence

$$P_k = \bigwedge_{\substack{W_1, W_2 \in [n]^k \\ W_1 \cap W_2 = \emptyset}} S_k(W_1, W_2).$$

For any $n > 2k$,

$$\mathbb{P}[G \in G_{n,p} \text{ has } \overline{R_k(W_1, W_2; z)}] \leq 1 - p^k (1-p)^k.$$

Therefore

$$\begin{aligned}\mathbb{P}[G \in G_{n,p} \text{ has } \overline{S_k(W_1, W_2)}] &= (1-(p(1-p))^k)^{n-2k} \\ &\leq \mathrm{e}^{-(p(1-p))^k(n-2k)}.\end{aligned}$$

Which implies

$$\begin{aligned}\mathbb{P}[G \in G_{n,p} \text{ has } \overline{P_k}] &\leq \binom{n}{2k}\mathrm{e}^{-(p(1-p))^k(n-2k)} \\ &\leq n^{2k}\mathrm{e}^{-(n-2k)(p(1-p))^k}.\end{aligned}$$

Let $c_0 = \mathrm{e}^{2k(p(1-p))^k}$ and $c_1 = (p(1-p))^k$. Then

$$\mathbb{P}[G \in G_{n,p} \text{ has } \overline{P_k}] \leq c_0 n^{2k}\mathrm{e}^{-c_1 n} \qquad (k \text{ is a constant}).$$

Therefore

$$\mathbb{P}[G \in G_{n,p} \text{ has } \overline{P_k}] = \mathrm{o}(1).$$

Hence

$$\lim_{n\to\infty} \mathbb{P}[G \in G_{n,p} \text{ has } P_k] = 1 - \lim_{n\to\infty} \mathbb{P}[G \in G_{n,p} \text{ has } \overline{P_k}] = 1. \qquad \square$$

The following theorem is a slight generalization of Theorem 6.4.5 as k remains fixed while p is allowed to vary with n.

Theorem 6.4.6. (See [4]) *Let $k \in \mathbb{Z}^+$ be given. Assume that for every $\epsilon > 0$,*

$$\lim_{n\to\infty} n^\epsilon p = \lim_{n\to\infty} n^\epsilon(1-p) = \infty.$$

Then $G \in \mathcal{G}_{n,p}$ almost surely has P_k.

Proof. The proof is similar to the proof of Lemma 6.4.5 combined with the observation that as k is fixed, $n - 2k \sim n$. The theorem assumptions imply

$$(n-2k)(p(1-p))^k \sim (n^{1/2k}p)^k(n^{1/2k}(1-p))^k \to \infty.$$

Therefore,

$$\mathbb{P}[G \in G_{n,p} \text{ has } \overline{P_k}] \leq n^{2k}\mathrm{e}^{-(n^{1/2k}p)^k(n^{1/2k}(1-p))^k)} = \mathrm{o}(1)$$

as needed. $\square$

The following theorem can be found in many places (*e.g.* [4] and [1]). The proof given here follows [36]).

Theorem 6.4.7. *For every first order statement, A, of graph theory*

$$\lim_{n\to\infty} \mathbb{P}[G_{n,p} \text{ satisfies } A] = 0 \text{ or } 1.$$

Proof. Let $G = (\mathbb{Z}^+, E(G))$ and $G^* = (\mathbb{Z}^+, E(G^*))$ be two graphs. Suppose for all $k \geq 1$, G and G^* satisfy P_k.

Claim 1. There is an isomorphism from G to G^*.

Let $f : G \to G^*$ be defined as follows: let $f(1) = 1^*$. Assume that f is defined on a finite set $V \subset V(G)$ and choose a vertex $v \in V(G) \setminus V$. Define $f(v) \in V(G^*) \setminus f(V)$ such that for $u \in V$, uv is an edge in G if and only if $f(u)f(v)$ is an edge in G^*. Since for every $k \in \mathbb{Z}^+$, both G and G^* are k-satisfied, f is an isomorphism, ending the proof of Claim 1.

Claim 2. Any logical theory T that contains $\{P_k\}_{k=1}^{\infty}$ is consistent.

Lemma 6.4.4 implies G is a model of T which shows that T is consistent.

Claim 3. Any logical theory T that contains $\{P_k\}_{k=1}^{\infty}$ is complete.

Suppose T is a logical theory containing $\{P_k\}_{k=1}^{\infty}$. If T were not complete, then there exists a theorem, B, which is undecidable in T. Let $T_0 = T \cup B$ and $T_1 = T \cup \overline{B}$. As T is consistent, T_0 and T_1 are consistent. Gödel's completeness theorem states that every consistent theory has a finite or countable model, thus there are countable models, G_1 and G_2 for T_1 and T_2 respectively. From Claim (1), G_1 and G_2 are isomorphic; thus both agree on B, which is a contradiction of the assumptions on T_1 and T_2 as B must be true in both models or false in both models, since G_1 and G_2 are isomorphic. Hence Claim (3) is proven.

To complete the proof of Theorem 6.4.7, let $\{P_k\} \subset T$ be a theory of graphs. Given a first order statement A, if $T \models A$, as proofs are finite, there are a finite $\{P_{k_i}\}_{i=1}^m$ such that $\{P_{k_i}\} \models A$ or $\{P_{k_i}\} \models \overline{A}$. Thus

$$\lim_{n\to\infty} \mathbb{P}[\overline{A}] \leq \lim_{n\to\infty} \sum_{i=1}^{m} \mathbb{P}[\overline{P_{k_i}}] = 0,$$

which implies

$$\lim_{n\to\infty} \mathbb{P}[A] = 1.$$

If $\overline{A}$ is a true statement in a theory of graphs, then apply the above proof to show

$$\lim_{n\to\infty} \mathbb{P}[\overline{A}] = 1$$

and the relationship

$$\lim_{n\to\infty} \mathbb{P}[A] = 1 - \lim_{n\to\infty} \mathbb{P}[\overline{A}].$$

$\square$

Chapter 7

Threshold Functions

In this chapter, instead of looking at one space such as $\mathcal{G}_{n,p}$ a sequence of spaces $\{\mathcal{G}_{n,p}\}$ or $\{\mathcal{G}_{n,q}\}$ where $p = p(n)$ and $q = q(n)$ are allowed to vary with n. The "random variable" examined here is actually a sequence of random variables, defined on the separate spaces $\mathcal{G}_{n,p(n)}$ or $\mathcal{G}_{n,q(n)}$. The goal is to examine the limiting behaviour of certain properties as the values of p are varied. Note that it is standard to abuse the notation of random variables and refer to X as a random variable on $\mathcal{G}_{n,p}$ where the sequence $p = p(n)$ is defined instead of examining the sequence of random variables $X_n : \mathcal{G}_{n,p} \to \mathbb{Z}^+$.

7.1 Thresholds in $\mathcal{G}_{n,p}$

Definition 7.1.1 (Threshold function)**.** *Given a graph theoretic property A, a function $r : \mathbb{Z}^+ \to [0, 1]$ is called a threshold function in $\mathcal{G}_{n,p}$ for A if either*

- *when* $0 < p(n) = \mathrm{o}(r(n))$ *then* $\lim_{n\to\infty} \mathbb{P}[G_{n,p(n)} \models A] = 0$*,*
- *when* $r(n) = \mathrm{o}(p(n)) < 1$*, then* $\lim_{n\to\infty} \mathbb{P}[G_{n,p(n)} \models A] = 1$*,*

or

- *when* $0 \leq p(n) = \mathrm{o}(r(n))$ *then* $\lim_{n\to\infty} \mathbb{P}[G_{n,p(n)} \models A] = 1$
- *when* $r(n) = \mathrm{o}(p(n)$ *then* $\lim_{n\to\infty} \mathbb{P}[G_{n,p(n)} \models A] = 0$*.*

Theorem 7.1.2. *Let A_3 be the property "G contains a copy of K_3". Then $r(n) = n^{-1}$ is the threshold function for A_3.*

Proof. Let $n \in \mathbb{Z}^+$ and $p \in (0,1)$ be given. For every $A = \{a_1, a_2, a_3\} \in [n]^3$ and every $G \in \mathcal{G}_{n,p}$, let $X_A(G)$ be the indicator random variable for the event "the subgraph induced by A is a copy of K_3". As the subgraph induced by A is a copy of K_3 requires the independent events $(\{a_1, a_2\} \in E(G)) \wedge (\{a_2, a_3\} \in E(G)) \wedge (\{a_1, a_3\} \in E(G))$,

$$\begin{aligned} \mathbb{E}[X_A] &= \mathbb{P}[X_A = 1] \\ &= \mathbb{P}[(\{a_1, a_2\} \in E(G)) \wedge (\{a_2, a_3\} \in E(G)) \wedge (\{a_1, a_3\} \in E(G))] = p^3. \end{aligned}$$

Define $X = X_n = \sum_{A \in [n]^3} X_A$ count the the number of K_3's that occur in $G \in \mathcal{G}_{n,p}$. As expectation is linear,

$$\mathbb{E}[X] = \binom{n}{3} p^3.$$

Let $0 \leq p(n) = \mathrm{o}(n^{-1})$. Then

$$\mathbb{E}[X] \leq n^3 p^3$$

therefore

$$\lim_{n \to \infty} \mathbb{E}[X] = 0.$$

Now assume $n^{-1} = \mathrm{o}(p(n))$. Then $\lim_{n \to \infty} \mathbb{E}[X] = \infty$ but this doesn't necessarily mean much about the behaviour of almost all graphs. To determine what (if anything) can be derived from these calculations, the second moment method is used.

Ennumerate $[n]^3 = \{S_1, S_2, \dots S_{\binom{n}{3}}\}$ and let A_k be the event "the subgraph of G induced by S_k is a K_3" and let X_k be the respective indicator random variable. Then $X = \displaystyle\sum_{k=1}^{\binom{n}{3}} X_k$. Calculating

$$\begin{aligned} \mathrm{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\{i,j\}\in[n]\times[n]} \mathbb{E}[X_iX_j] - \mathbb{E}[X]^2 \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{j=2}^{n}\sum_{i<j} \mathbb{E}[X_iX_j] - \mathbb{E}[X]^2 \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i] + 2\sum_{j=2}^{n}\sum_{i<j} \mathbb{E}[X_iX_j] - \mathbb{E}[X]^2.
\end{aligned}
$$

To get an estimate on the terms The next step in the second moment method is to get an estimate on $\sum_{j=2}^{n}\sum_{i<j} \mathbb{E}[X_iX_j]$, To this end, choose $j \in [n]$ to be given. If, for $i < j$, the events A_i, A_j are independent, then

$$
\mathbb{E}[X_iX_j] = \mathbb{E}[X_i]\mathbb{E}[X_j].
$$

To figure out when this is true, note that if $i < j$ and $|S_i \cap S_j| < 2$ then A_i occurring has no affect on whether or not A_j occurs, as they share no edges, therefore A_i and A_j are independent. If $i < j$ and $|S_i \cap S_j| = 2$ a triangle induced by S_i would have an edge in the subgraph induced by S_j thus increasing the probability of a triangle in S_j, thus A_i and A_j are dependent.

Note that $X_iX_j = X_{A_i \wedge A_j}$ so that

$$
\begin{aligned}
\mathbb{E}[X_iX_j] &= \mathbb{E}[X_{A_i\wedge A_j}] \\
&= \mathbb{P}[A_i \wedge A_j] \\
&= \mathbb{P}[A_j]\mathbb{P}[A_i|A_j] \\
&= p^3p^2 = p^5
\end{aligned}
$$

the second to last of these equalities follows from noting that if A_j occurs (*i.e.* the subgraph of G induced by S_j is a triangle) and $|S_i \cap S_j| = 2$ then A_i occurs if and only if the other two edges of the triangle induced by S_i occur, which happens with probability p^2. Note that there are

$$
\binom{3}{2}(n-3) = 3(n-3)
$$

events dependent on A_j. Define the relation '$\preccurlyeq$' on $[n]$ by $i \preccurlyeq j$ if and only if and A_i is dependent on A_j. Putting this all together yields

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}[X_iX_j] &= \mathbb{E}[X_j] + \sum_{i\not\preccurlyeq j}\mathbb{E}[X_i]\mathbb{E}[X_j] + \sum_{\substack{i\preccurlyeq j\\ i\neq j}}\mathbb{E}[X_iX_j]\\ &= p^3 + \left(\binom{n}{3} - 3n\right)p^6 + 3np^5.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{j=1}^{n}\sum_{i=1}^{n}\mathbb{E}[X_iX_j]\\ &= \sum_{j=1}^{n}\mathbb{E}[X_j] + \sum_{i\not\preccurlyeq j}\mathbb{E}[X_i]\mathbb{E}[X_j] + \sum_{i\preccurlyeq j}\mathbb{E}[X_iX_j]\\ &= \binom{n}{3}p^3 + \binom{n}{2}\left(\binom{n}{3} - 3n\right)p^6 + n(3np^5).\end{aligned}$$

Thus

$$\begin{aligned}\mathrm{Var}[X] &= \binom{n}{3}p^3 + \binom{n}{2}\\ &\qquad\left(\binom{n}{3} - 3n\right)p^6 + n(3np^5) - (\binom{n}{3}p^3)^2\\ &= \mathrm{O}(n^3p^3 + n^5p^6 + n^2p^5 - n^6p^6).\end{aligned}$$

Theorem 3.3.8 gives

$$\begin{aligned}\mathbb{P}[|X - \mathbb{E}[X]| > \mathbb{E}[X]] &\leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2}\\ &= \mathrm{O}(\frac{n^3p^3 + n^5p^6 + n^2p^5 - n^6p^6}{n^6p^6})\end{aligned}$$

From the choice of p, $\lim_{n\to\infty}\mathbb{P}[|X - \mathbb{E}[X]| > \mathbb{E}[X]] = 0$. Thus Corollary 3.3.11 implies almost surely $X \sim \mathbb{E}[X]$. □

This section covers techniques used to compute thresholds in $\mathcal{G}_{n,p}$; while Section 7.2 shows some threshold examples in $\mathcal{G}_{n,q}$, using similar properties to give a point of reference to the similarities and differences in the two models. Chapter 8 presents examples of more in-depth analysis. The next theorem is a general version of Theorem 7.1.2.

Theorem 7.1.3. *Let $3 \leq k \in \mathbb{Z}^+$ and A_k be the property "$V(G)$ has a k-clique". Let $r(n) = n^{-2/(k-1)}$. Then $r(n)$ is a threshold function for A_k.*

Proof. The first step is to show that if $p(n) = \mathrm{o}(r(n))$ then

$$\lim_{n\to\infty} \mathbb{P}[G_{n,p(n)} \models A_k] = 0.$$

Let $G \in G_{n,p(n)}$ and $X(G)$ be the number of copies of K_k in G. Thus

$$\mathbb{E}[X] = \binom{n}{k} p^{\binom{k}{2}} < n^k p^k = (np)^k \qquad (0 < p < 1 \text{ thus } p^{\binom{k}{2}} < p^k). \tag{7.1}$$

From the condition on $p(n)$ and Lemma (3.4),

$$\mathbb{P}[X > 0] \leq \mathbb{E}[X] \to 0 \qquad\qquad (\text{as } n \to \infty).$$

Next, assume $r(n) = \mathrm{o}(p(n))$ so that equation (7.1) implies

$$\lim_{n\to\infty} \mathbb{E}[X] = \infty.$$

To show that $X \neq 0$ almost surely, it is suffices to show that $\mathrm{Var}[X] = o[\mathbb{E}[X]^2$. For every $S \in [V]^k$, let A_S be the event "the subgraph induced by S is a k-clique" and X_S be the respective indicator random variable for A_S. As in the case of $k = 3$, two events A_S and A_R are dependent if and only if $2 \leq |S \cap R|$. Assume $2 \leq t = |S \cap R|$. If A_S occurs, then $\binom{k}{2} - \binom{t}{2}$ other edges are necessary for A_T to occur. Therefore

$$\begin{aligned}
\mathbb{P}[A_R \wedge A_S] &= \mathbb{P}[A_S]\mathbb{P}[A_R|A_S] \\
&= p^{\binom{k}{2}} p^{\binom{k}{2}-\binom{t}{2}} \\
&= p^{2\binom{k}{2}-\binom{t}{2}}
\end{aligned}$$

As A_S is dependent on at most $\binom{n-k}{k-t}$ events,

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X] + \sum_{K\in[n]^k} \sum_{t=2}^{k-1} \binom{n-k}{k-t} p^{2\binom{k}{2}-\binom{t}{2}} - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X] + \binom{n}{k} \sum_{t=2}^{k-1} \binom{n-k}{k-t} p^{2\binom{k}{2}-\binom{t}{2}} - \mathbb{E}[X]^2.
\end{aligned}$$

As k is fixed and n gets arbitrarily large,

$$\sim n^k p^{k^2} + n^k \sum_{t=2}^{k-1} (n-k)^{(k-t)} p^{(k^2 - t^2/2)} - n^{2k} p^{k^2}. \tag{7.2}$$

Therefore, as $n^{-2/(k-1)} = \mathrm{o}(p(n))$, $\mathrm{Var}[X] = o[E[X]^2)$ as needed. Lemma 3.3.8 implies that

$$\lim_{n\to\infty} \mathbb{P}[|X - \mathbb{E}[X]| > \mathbb{E}[X]] = 0.$$

Therefore Corollary 3.3.11 implies $X \sim \mathbb{E}[X]$ almost surely as needed. Thus $r(n) = n^{-2/(k-1)}$ is the threshold function for A_k. □

On the somewhat opposite end of the spectrum, recall that an isolated vertex has degree zero.

Theorem 7.1.4. *Consider the property A_0 of "containing isolated vertices". In $\mathcal{G}_{n,p}$, the function*

$$r(n) = n^{-\frac{1}{2}}$$

is a threshold for A_0.

Proof. Let $G = (V, E) \in \mathcal{G}_{n,p}$ and $v \in V$. Let A_v be the property "v is isolated" and let X_v be the indicator random variable for A_v. Then

$$\mathbb{E}[X_v] = (1-p)^{n-1}$$

as property A_v is equivalent to the property "for all $x \in V(G) \setminus v \quad \{v, x\} \notin E(G)$". If $X(G)$ counts the number of isolated vertices in G, then $X = \sum_{v\in V} X_v$ thus

$$\mathbb{E}[X] = \sum_{v\in V} \mathbb{E}[X_v] = n(1-p)^{n-1}.$$

Write

$$(1-p)^n = e^{\ln(1-p)^n} = e^{n\ln(1-p)}. \tag{7.3}$$

Equation (2.4) implies

$$\ln(1-p) = -p - \frac{p^2}{2} - \frac{p^3}{3} - \dots \tag{7.4}$$

therefore

$$(1-p)^n = e^{n\left(-p-\frac{p^2}{2}-\dots\right)} = e^{-np-\frac{np^2}{2}-\dots)} = e^{-np}\mathrm{O}(1). \tag{7.5}$$

Provided $p = p(n) = \mathrm{o}(r(n))$,

$$\mathbb{E}[X_v] \sim ne^{-pn} \to 0.$$

To apply the second moment method, assume $r(n) < p(n) \leq 1$. If $v, w \in V$, $v \neq w$ with events A_v and A_w and the indicator random variables X_v, X_w as above, then

$$\begin{aligned}\mathbb{E}[X_vX_w] &= \mathbb{P}[A_v \wedge A_w] \\ &= \mathbb{P}[A_v]\mathbb{P}[A_w|A_v] \\ &= (1-p)^{n-1}(1-p)^{n-2} = (1-p)^{2n-3}.\end{aligned}$$

For n large,

$$\mathbb{E}[X_vX_w] \sim (1-p)^{2n}.$$

Thus equations (7.3), (7.4) and (7.5) with $2n$ replacing n imply

$$\mathbb{E}[X_vX_w] = e^{-2np-\frac{2np^2}{2}-\dots} = e^{-2np}\mathrm{O}(1).$$

As every pair of events A_v, A_w are dependent,

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \sum_{v\in V}\mathbb{E}[X_v^2] + \sum_{\{v,w\}\in[V]^2}\mathbb{E}[X_vX_w] - \mathbb{E}[X]^2 \\ &= \sum_{v\in V}\mathbb{E}[X_v] + \sum_{\{v,w\}\in[V]^2}\mathbb{E}[X_vX_w] - \mathbb{E}[X]^2 \\ &\sim \mathbb{E}[X] + \binom{n}{2}e^{-2pn} - \mathbb{E}[X]^2.\end{aligned}$$

Then

$$\frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \leq \frac{1}{\mathbb{E}[X]} + \frac{\binom{n}{2}e^{-2pn}}{\mathbb{E}[X]^2} - 1$$

$$
\begin{aligned}
&= \frac{1}{\mathbb{E}[X]} + \frac{\binom{n}{2} e^{-2pn}}{n^2 e^{-2pn}} - 1 && (\mathbb{E}[X]^2 \sim n^2 e^{-2pn}) \\
&\sim \frac{1}{\mathbb{E}[X]} && \left(\frac{\binom{n}{2} e^{-2pn}}{n^2 e^{-2pn}} - 1 \to 0 \right).
\end{aligned}
$$

Inequality (3.3.8) implies

$$
\begin{aligned}
\mathbb{P}[X = 0] &\leq \mathbb{P}[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]] \\
&\leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \\
&\sim \frac{1}{\mathbb{E}[X]} \to 0 && (r(n) = \mathrm{o}(p(n)),\ \lim_{n\to\infty} \mathbb{E}[X] = \infty)\,.
\end{aligned}
$$

Hence for $r(n) = \mathrm{o}(p(n))$, Corollary 3.3.11 implies almost surely $X \sim \mathbb{E}[X]$. Thus $r(n)$ is the threshold function. $\square$

A natural question to ask next is about the appearance of cycles. Note that a graph is called *acyclic* if it contains no cycles. An acyclic graph with at least one tree is called a forest.

Theorem 7.1.5. *The threshold function for the property of 'G being a forest' is $r(n) = 1/n$*

Proof. Assume n, p and k are given. Let $G \in \mathcal{G}_{n,p}$. Let A_k be the event "G has a subgraph that is a copy of C_k" and $X_k(G)$ be the number of subgraphs of G that are copies of C_k's. As in Theorem 4.2.1,

$$
\mathbb{E}[X_k] = \binom{n}{k} \frac{p^k}{2k}.
$$

If $X(G)$ is the number of cycles in G of any length, then

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{k=3}^{n} \binom{n}{k} \frac{p^k}{2k} \\
&\leq \sum_{k=3}^{n} n^k p^k \\
&= \sum_{k=3}^{n} (np)^k.
\end{aligned}
$$

If $p(n) = \mathrm{o}(n^{-1})$, then for all $\epsilon > 0$,

$$\begin{aligned}\mathbb{E}[X] &< \sum_{k=3}^{n} \epsilon^k \\ &< \frac{\epsilon^3}{1-\epsilon} \to 0.\end{aligned}$$

If $n^{-1} = \mathrm{o}(p(n))$, Theorem 7.1.2 implies $X_3 \sim \mathbb{E}[X_3]$, hence $X \to \infty$. Thus almost surely, $G \in \mathcal{G}_{n,p(n)}$ is not a forest. $\square$

Definition 7.1.6. *Let H be a graph with* e *edges and* v *vertices. Define* $\rho(H) = \mathrm{e}/v$ *to be the density of* H*. The graph* H *is called balanced* iff *for every subgraph* H' *of* H $\rho(H') \leq \rho(H)$*. The graph* H *is called strictly balanced* iff *for every proper subgraph* H'*,* $\rho(H') < \rho(H)$*.*

Complete graphs and cycles are examples of balanced graphs.

Theorem 7.1.7. *Let H be a balanced graph with v vertices and* e *edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then* $p = n^{-v/\mathrm{e}}$ *is the threshold function for A.*

Proof. Assume n, p and k are given. Let H be a balanced graph on v vertices and e edges. For every $S \in [n]^v$, let A_S be the event that the subgraph induced by S (in G) contains H.

While the previous examples allowed for exact calculations of the probability of an event, the proof of this theorem requires properly behaved bounds on $\mathbb{P}[A_S]$.

For a lower bound, note that if the subgraph induced by S to contain a copy of H, S must contain at least e, thus $p^{\mathrm{e}} \leq \mathbb{P}[A_S]$. A good upper bound is (in general) difficult to calculate, due to the large number of overlapping potential copies of H, but noting that any particular placement of H (in the subgraph induced by S) has a probability of p^{e} of occurring and there are at most $v!$ possible placements of H then

$$\mathbb{P}[A_S] \leq v! p^{\mathrm{e}}. \tag{7.6}$$

The next step is to approximate the expected number of H's in any $G \in \mathcal{G}_{n,p}$. Let X_S be the indicator random variable for A_S and $X = \sum_{i=1}^{\binom{n}{v}} X_S$ be the number of copies of H. Then

$$\begin{aligned}\binom{n}{v} p^{\mathrm{e}} &\leq \mathbb{E}[X] \\ &\leq \binom{n}{v} v! p^{\mathrm{e}} \\ &= (n)_v p^{\mathrm{e}} < n^v p^{\mathrm{e}} = (n^{\frac{v}{\mathrm{e}}} p)^{\mathrm{e}}.\end{aligned}$$

As $p(n) = \mathrm{o}(n^{-v/\mathrm{e}})$,

$$\mathbb{P}[X > 0] \leq \mathbb{E}[X] \to 0.$$

To calculate $\mathrm{Var}(X)$, first choose $S \in [n]^v$. Define a relation '$\preccurlyeq$' on $[n]^v \setminus S$ by $T \preccurlyeq S$ iff A_S is dependent on A_T. To determine how many T satisfy $T \preccurlyeq S$, examine the subgraph induced by $S \cap T$, if there is an edge of any copy of H then $T \preccurlyeq S$.

To estimate $\mathbb{P}[A_T | A_S]$, note that if $S \neq T \in [n]^v$ with $2 \leq |S \cap T| = i \leq v-1$, then any induced subgraph $\hat{H} \neq \emptyset$ of H contained in the subgraph of G induced by $T \setminus S \cap T$ satisfies

$$\rho(\hat{H}) = \frac{|E(\hat{H})|}{v - |S \cap T|} \leq \frac{\mathrm{e}}{v} = \rho(H)$$

since H is a balanced graph. Therefore

$$|E(\hat{H})| \leq \frac{\mathrm{e}(v - |S \cap T|)}{v}.$$

Thus, as in equation (7.6),

$$\begin{aligned}p^{\mathrm{e}} p^{\frac{\mathrm{e}(v-i)}{v}} &\leq \mathbb{P}[A_T \wedge A_S] &\quad (7.7)\\ &= \mathbb{P}[A_S]\mathbb{P}[A_T | A_S] \\ &\leq p^{\mathrm{e}} (v-i)! p^{\frac{\mathrm{e}(v-i)}{v}}. &\quad (7.8)\end{aligned}$$

As equation (7.7) is true for every T with $|T \cap S| = i$ and every $2 \leq i \leq v-1$, then

$$\sum_{\{T | T \preccurlyeq S\}} \mathbb{P}[A_T \wedge A_S] = \sum_{\{T | T \preccurlyeq S\}} \mathbb{P}[A_S]\mathbb{P}[A_T | A_S]$$

$$\leq \sum_{i=2}^{v-1} \binom{n-v}{v-i} p^{\mathrm{e}}(v-i)! p^{\frac{\mathrm{e}(v-i)}{v}}$$
$$\leq \sum_{i=2}^{v-1} (n-v)^{v-i} p^{\mathrm{e}+\frac{\mathrm{e}(v-i)}{v}}$$
$$\sim \sum_{i=2}^{v-1} (n-v)^{v-i} p^{2\mathrm{e}-i\mathrm{e}/v}.$$

Therefore $\mathbb{E}[X]^2 \sim n^{2v} p^{2\mathrm{e}}$ and

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X] + \sum_{S\in[n]^v} \sum_{\{T|T\preccurlyeq S\}} \mathbb{E}[X_T X_S] - \mathbb{E}[X]^2 \\
&\leq E[X] + \binom{n}{v} \sum_{i=2}^{v-1} (n-v)^{v-i} p^{2\mathrm{e}-i\mathrm{e}/v} - \mathbb{E}[X]^2 \\
&\sim \mathbb{E}[X] + n^v \sum_{i=2}^{v-1} (n-v)^{v-i} p^{2\mathrm{e}-i\mathrm{e}/v} - \mathbb{E}[X]^2 \\
&\leq \mathbb{E}[X] + \sum_{i=2}^{v-1} n^{2v-i} p^{2\mathrm{e}-i\mathrm{e}/v} - \mathbb{E}[X]^2.
\end{aligned}$$

Let $\hat{V}(X) = \sum_{i=2}^{v-1} n^{2v-i} p^{2\mathrm{e}-i\mathrm{e}/v}$. For every $2 \leq i \leq v-1$,

$$\begin{aligned}
\frac{n^{2v-i} p^{2\mathrm{e}-i\mathrm{e}/v}}{\mathbb{E}[X]^2} &= \frac{n^{2v-i} p^{2\mathrm{e}-i\mathrm{e}/v}}{n^{2v} p^{2\mathrm{e}}} \\
&= n^{-i} p^{-i\mathrm{e}/v} \\
&< n^{-i} n^{(v/\mathrm{e})(i\mathrm{e}/v)} \to 0 \qquad \text{(as } n \to \infty).
\end{aligned}$$

Thus $\mathrm{Var}[X] = \mathrm{o}(\mathbb{E}[X]^2)$ as necessary so Corollary 3.3.11 gives the result. □

7.2 Examples of threshold functions in $\mathcal{G}_{n,q}$

In this section, threshold functions in $\mathcal{G}_{n,q}$ are considered.

Definition 7.2.1. *Let Q be a property of graphs and let $\mathcal{A} \subset \mathcal{G}_{n,q}$ be the set of graphs with Q. Then $t = t(n)$ is called a threshold function for property Q if either* *(i) while* $0 \le q(n) = o(t(n))$ *then* $\mathbb{P}[\mathcal{G}_{n,q} \models \mathcal{A}] \to 0$,

(ii) while $t(n) = o(q(n))$ *then* $\mathbb{P}[\mathcal{G}_{n,q} \models \mathcal{A}] \to 1$

or if

(i)' while $0 \le q(n) = o(t(n))$ *then* $\mathbb{P}[\mathcal{G}_{n,q} \models \mathcal{A}] \to 1$,

(ii)' while $0 \le t(n) = o(q(n))$ *then* $\mathbb{P}[\mathcal{G}_{n,q} \models \mathcal{A}] \to 0$.

Theorem 7.2.2. *Threshold of 3-cliques In* $\mathcal{G}_{n,q}$, $t(n) = n$ *is the threshold function for the property of "having a subgraph isomorphic to* K_3*".*

Proof. Let n be given. Let X count the number of subgraphs isomorphic to K_3.

Assume $0 \le q \le \binom{n}{2}$ is arbitrary for the moment. Consider any $S \in [n]^3$. Let A_S be the event "S is a clique". To calculate the probability of A_S, observe that graph $G \in \mathcal{G}_{n,q}$ for which A_S occurs, then the remaining $q-3$ edges are chosen from the remaining $\binom{n}{2} - 3$ edges. Therefore

$$\mathbb{P}[A_S] = \frac{\binom{\binom{n}{2}-3}{q-3}}{\binom{\binom{n}{2}}{q}}. \tag{7.9}$$

If X_S is the indicator random variable for A_S, then $X = \sum_{S \in [n]^3} X_S$ so that

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{S \in [n]^3} \mathbb{E}[X_S] \\
&= \binom{n}{3} \frac{\binom{\binom{n}{2}-3}{q-3}}{\binom{\binom{n}{2}}{q}} \\
&= \binom{n}{3} \frac{\binom{q}{3}}{\binom{\binom{n}{2}}{3}} \\
&\le n^3 \frac{\binom{q}{3}}{(\binom{n}{2})^3} \\
&\le n^3 \frac{\binom{q}{3}}{(n^2/2)^3}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{8\binom{q}{3}}{n^3} \\
&\leq \frac{8q^3}{6n^3}.
\end{aligned}
$$

Inequality (3.4) implies

$$\mathbb{P}[X > 0] \leq \mathbb{E}[X] \to 0 \qquad\qquad (\text{if } q = \mathrm{o}(n)).$$

Assume $n = \mathrm{o}(q)$. Then

$$
\begin{aligned}
\mathbb{E}[X] &= \binom{n}{3}\frac{\binom{q}{3}}{\binom{\binom{n}{2}}{q}} \\
&\geq \frac{\binom{n}{3}q^3}{\binom{\binom{n}{2}}{3}} \geq n^2 \to \infty.
\end{aligned}
$$

To use the second moment method, note that if $S, T \in [n]^3$, the events, A_S, A_T are dependent if and only $|S \cap T| = 2$. As there are $\binom{3}{2} = 3$ ways to choose an $S' \in [S]^2$ and $n - 3$ ways to choose the third element of

$$T = S' \cup \{v \in V \setminus S\},$$

there are $3(n-3)$ such $T \in [n]^3 \setminus S$. In the notation of equation (3.6),

$$\Delta = \binom{n}{3} 3(n-3)\mathbb{P}[A_S \wedge A_T] \leq n^4.$$

The conditions on q imply $\Delta = \mathrm{o}(\mathbb{E}[X]^2)$, thus Corollary 3.3.11 imply almost surely $X \sim \mathbb{E}[X]$. Thus

$$t(n) = n$$

is the desired threshold for 3 cliques in $\mathcal{G}_{n,q}$. $\square$

The next theorem is the $\mathcal{G}_{n,q}$ version of Theorem 7.1.7. See the paper [14] for the proof.

Theorem 7.2.3. *Assume H is a balanced connected graph on k vertices and $k - 1 \leq l \leq \binom{k}{2}$ edges. Let $\mathcal{A}_{k,l}$ be the property that that the random graph $\mathcal{G}_{n,q}$ has a subgraph isomorphic to H. The threshold function for $\mathcal{A}_{k,l}$ is $t(n) = n^{2-\frac{k}{l}}$.*

The next theorem represents a sharper notion of a threshold function.

Theorem 7.2.4. *Let $n \in \mathbb{Z}^+$. Let $0 < c \in \mathbb{R}$ and $q = \frac{c}{2} n \ln n$.*

(i) If $c > 1$ then almost all graphs in $\mathcal{G}_{n,q}$ have isolated vertices.

(ii) If $c < 1$ then almost all graphs in $\mathcal{G}_{n,q}$ have no isolated vertices.

Proof. Since $q = \mathrm{o}(\binom{n}{2}^{1/2})$ Stirling's approximation (2.3.4) implies

$$\binom{\binom{n}{2}}{q} = (1 + \mathrm{o}(1))(2\pi q)^{-1/2} \left(\frac{\mathrm{e}\binom{n}{2}}{q} \right)^q. \tag{7.10}$$

For $G \in \mathcal{G}_{n,q}$, let $X(G)$ be the random variable defined to be the number of isolated vertices in G and for $v \in V$, let $X_v(G)$ be the indicator random variable for the event "v is isolated in G".

For a graph $G \in \mathcal{G}_{n,q}$, a vertex $v \in V(G)$ is isolated implies that the q edges are distributed amongst the $n-1$ remaining vertices. Therefore in $\mathcal{G}_{n,q}$,

$$\begin{aligned} \mathbb{E}[X_v] &= \mathbb{P}[v \text{ is isolated}] \\ &= \frac{\binom{\binom{n-1}{2}}{q}}{\binom{\binom{n}{2}}{q}} \end{aligned} \tag{7.11}$$

Since for any two vertices, v and $\hat{v}$, $\mathbb{E}[X_v] = \mathbb{E}[X_{\hat{v}}]$ and

$$X = \sum_{v \in V} X_v$$

implies

$$\mathbb{E}[X] = n \frac{\binom{\binom{n-1}{2}}{q}}{\binom{\binom{n}{2}}{q}}. \tag{7.12}$$

Formula (7.10) for $\binom{\binom{n-1}{2}}{q}$ gives

$$\begin{aligned} \binom{\binom{n-1}{2}}{q} &\sim (2\pi q)^{-1/2} \left[\frac{\binom{n-1}{2}\mathrm{e}}{q} \right]^q \\ &\sim (2\pi q)^{-1/2} \mathrm{e}^q \left[\frac{n}{c \ln n} \right]^q \left[1 - \frac{1}{n} \right]^{nc \ln n} \qquad (n - 1 = n(1 - 1/n)). \end{aligned}$$

Approximate

$$\left(1-\frac{1}{n}\right)^{n(c\ln n)} \sim \mathrm{e}^{-c\ln n} = n^{-c}.$$

Let $C_n = (2\pi q)^{-1/2}\left[\frac{n}{c\ln n}\right]^q$; then

$$\binom{\binom{n-1}{2}}{q} \sim C_n n^{-c}\mathrm{e}^q. \tag{7.13}$$

Similar calculations on $\binom{\binom{n}{2}}{q}$ imply

$$\begin{aligned}\binom{\binom{n}{2}}{q} &\sim (2\pi q)^{-1/2}\left(\frac{n^2}{2q}\mathrm{e}\right)^q \\ &= C_n e^q. \end{aligned} \tag{7.14}$$

Combining formulas (7.13) and (7.14) gives

$$\mathbb{E}[X] \sim n^{1-c}.$$

Therefore, for $c > 1$, $\mathbb{E}[X] \to 0$. By Markov's inequality,

$$\mathbb{P}[X \geq 1] \leq \mathbb{E}[X]$$

such that almost surely, $X = 0$.

For $0 < c < 1$,

$$\mathbb{E}[X] \to \infty.$$

To show $\lim_{n\to\infty}\mathbb{P}[X = 0] = 0$, the second moment method is needed. The first step is to estimate $\mathrm{Var}(X)$. For every $i \in [n]$, let A_i be the event "i is isolated". Then for $i \neq j \in [n]$,

$$\begin{aligned}\mathbb{P}[A_i \wedge A_j] &= \mathbb{P}[A_i]\mathbb{P}[A_j|A_i] \\ &= \frac{\binom{n-1}{q}}{\binom{n}{q}}\frac{\binom{n-2}{q}}{\binom{n}{q}}.\end{aligned}$$

Using the notation of (3.6),

$$\Delta = \binom{n}{2}\frac{\binom{\binom{n-1}{2}}{q}}{\binom{\binom{n}{2}}{q}}\frac{\binom{\binom{n-2}{2}}{q}}{\binom{\binom{n}{2}}{q}}$$

$$
\begin{aligned}
&\sim \binom{n}{2} \frac{\binom{(n-1)^2/2}{q}}{\binom{n^2/2}{q}} \frac{\binom{(n-2)^2/2}{q}}{\binom{n^2/2}{q}} \\
&\sim \binom{n}{2} \frac{\left(\frac{\mathrm{e}(n-1)^2}{2q}\right)^q}{\left(\frac{\mathrm{e}n^2}{2q}\right)^q} \frac{\left(\frac{\mathrm{e}(n-2)^2}{2q}\right)^q}{\left(\frac{\mathrm{e}n^2}{2q}\right)^q} && \text{(Lemma 2.3.4)} \\
&\sim \binom{n}{2} \left(\frac{(n-1)(n-2)}{n^2}\right)^{2q} = \mathrm{o}((\mathbb{E}[X])^2).
\end{aligned}
$$

Inequality 3.4 implies

$$
\begin{aligned}
\mathbb{P}[X = 0] &\leq \mathbb{P}[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]] \\
&\leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} \to 0.
\end{aligned}
$$

Thus, almost surely $X = \mathbb{E}[X]$. □

7.3 Unbalanced graphs

Theorems 7.1.7 and 7.2.3 showed how to calculate threshold functions for the property of 'containing a subgraph H' in the case that H is balanced, but what can be said if H is not balanced? Alon and Spencer [1] provide an answer in this case. Here an outline of the solution is given.

In calculating threshold functions it is necessary to examine each subgraph separately. This is because when considering the second moment, any possible intersection of proper subgraphs can occur. In the previous section, when considering Theorem 7.2.3, it was necessary to calculate for every $S_1, S_2 \in [n]^k$,

$$\mathbb{P}[A_{S_1} \wedge A_{S_2}]$$

where each of the A_{S_i} are as in Theorem 7.2.3.

In the case that H is unbalanced, the same ideas are applied except that it is necessary to find the subgraph H' of highest density. The idea is captured in the following theorem.

Theorem 7.3.1. *In the notation of Theorem 7.1.7, if H is not balanced then $p = n^{-v/e}$ is not the threshold function.*

Proof. Let H_1 be a subgraph of H with v_1 vertices and e_1 edges and $e_1/v_1 > e/v$. Let α satisfy $v_1/e_1 < \alpha < v/e$ and $p = n^{-\alpha}$. If $X = X(G)$ is the random on $\mathcal{G}_{n,p}$ counting the number of subgraphs of G isomorphic to H_1,

$$\mathbb{E}[X] = \binom{n}{v} p^{e_1} \sim n^{v_1}(n^{-\alpha})^{e_1} = \mathrm{o}(1).$$

Inequality (3.3.8) implies almost surely G has no copy of H_1, hence no copy of H. $\square$

Thus graphs that are unbalanced have proper subgraphs with higher density so that a threshold function of higher value works.

Chapter 8

Random Graphs

8.1 Beyond thresholds: property distributions

This section goes beyond the notion of threshold functions and discusses a similar question:
"Given conditions on $p(n)$ (respectively, $q(n)$) and a random variable X on $\mathcal{G}_{n,p}$ (respectively, $\mathcal{G}_{n,q}$) , what can be said about the limit properties of the distribution function $\mathbb{P}[X = j]$."
Presented here are examples for $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,q}$.

8.1.1 Examples in $\mathcal{G}_{n,p}$

Recall Theorem 7.1.4 showed $r(n) = n^{-\frac{1}{2}}$ is a threshold function in $\mathcal{G}_{n,p}$ of the property 'G has an isolated vertex', it was shown that for $p(n) < r(n) = n^{-1/2}$, almost surely $\mathcal{G}_{n,p}$ has isolated vertices while for $r(n) < p(n) = p$, almost surely $\mathcal{G}_{n,p}$ has no isolated vertices. In this subsection, assume that for some constant $c > 0$,

$$p(n) = c\frac{\ln n}{n}$$

where $0 < c$. Since for large n, $0 < \ln n < n^{\frac{1}{2}}$, the proof of Theorem 7.1.4 shows that

$$\mathbb{E}[X] \sim ne^{-pn}$$

$$= ne^{-c\frac{\ln n}{n}n}$$
$$= n^{1-c}.$$

Thus if $1 < c$, $\mathbb{E}[X] \to 0$ such that almost surely $X = 0$ and if $c < 1$ $\mathbb{E}[X] \to \infty$. At $c = 1$ something special happens. Is there a way to say more? This section shows a way to extend this analysis.

Theorem 8.1.1. [4] *Let $p(n) = \frac{\ln n}{n} + \frac{1}{n}$ and $X = X(G)$ be the number of isolated vertices in G. Then for $k = 0, 1, 2, \ldots,$*

$$\mathbb{P}[X = k] \to \frac{e^{-\lambda}\lambda^k}{k!}$$

where $\lambda = e^{-1}$.

This theorem states that X converges in distribution to the Poisson distribution with mean $\lambda = e^{-1}$.

Proof. The first step is to show $\lim_{n\to\infty} S_1 = \lim_{n\to\infty} \mathbb{E}[X] = \lambda$. From the proof of Theorem 7.1.4,

$$\begin{aligned}\mathbb{E}[X] &= n(1-p)^{n-1} \\ &\sim ne^{-p(n-1)} \\ &\sim ne^{-pn} \\ &= ne^{-n(\frac{\ln n}{n}+\frac{1}{n})} \\ &= ne^{-\ln n - 1} \\ &= e^{-1} = \lambda.\end{aligned}$$

The next step in the proof of Theorem 8.1.1 is to use the Theorem 9.3.1.

Claim 1. For all $2 \le r \in \mathbb{Z}^+$,

$$\lim_{n\to\infty} S_r = \frac{\lambda^r}{r!}$$

Proof (of Claim 1). Let $2 \leq r \leq n \in \mathbb{Z}^+$ be given and $V = \{v_1, v_2 \dots v_n\}$. For $i \in [n]$, let A_i be the event 'v_i is isolated' and X_i be the indicator random variable for A_i. Assume $1 \leq l_1 < l_2 < \dots l_r \leq n$ then

$$\begin{aligned}
\mathbb{E}[X_{l_1}X_{l_2}] &= \mathbb{P}[A_{l_1} \wedge A_{l_2}] \\
&= \mathbb{P}[A_{l_1}]\mathbb{P}[A_{l_2}|A_{l_1}] \\
&= (1-p)^{n-1}(1-p)^{n-2} \\
&= (1-p)^{2(n-2)+1} \\
&= (1-p)^{2(n-2)+\binom{2}{2}} \\
\mathbb{E}[X_{l_1}X_{l_2}X_{l_3}] &= \mathbb{P}[A_1]\mathbb{P}[A_2|A_1]\mathbb{P}[A_3|A_2 \wedge A_1] \\
&= (1-p)^{2(n-2)+1}(1-p)^{n-3} \\
&= (1-p)^{3(n-3)+3} \\
&= (1-p)^{3(n-3)+\binom{3}{2}}.
\end{aligned}$$

Thus the general formula is

$$\mathbb{E}[X_{l_1}X_{l_2}\cdots X_{l_r}] = (1-p)^{r(n-r)+\binom{r}{2}}.$$

Thus

$$S_r = \sum_{1} \mathbb{E}[X_{l_1}X_{l_2}\cdots X_{l_r}] = \binom{n}{r}(1-p)^{r(n-r)+\binom{r}{2}}. \tag{8.1}$$

To check the asymptotics of equation (8.1), observe

$$\begin{aligned}
S_r &= \binom{n}{r}(1-p)^{rn}(1-p)^{\binom{r}{2}-r^2} \\
&= \binom{n}{r}(1-p)^{rn}(1+o(1)) \qquad\qquad \text{(as } p(n) \to 0 \text{ and } r \text{ is fixed)} \\
&\sim \binom{n}{r}e^{-prn} \\
&= \binom{n}{r}e^{-rn(\frac{\ln n}{n}+\frac{1}{n})} \\
&= \binom{n}{r}e^{(-r\ln n - r)} \\
&\sim \frac{n^r}{r!}\frac{e^{-r}}{n^r}
\end{aligned}$$

$$= \frac{(e^{-1})^r}{r!}$$
$$= \frac{\lambda^r}{r!}.$$

Thus Claim 1 is proven. Theorem 9.3.1 implies Theorem 8.1.1. $\square$

The next theorem requires a lemma.

Lemma 8.1.2. *Let $H = (V, E)$ be a unlabeled graph on k vertices and automorphism group $aut(H)$ and $a = |aut(H)|$. If $l(H)$ is the number of ways to label H then*

$$l(H) = \frac{k!}{a}.$$

Proof. Label the vertices of H is all $k!$ ways. Any two labels are the same if they differ by an automorphism of H. Thus the $k!$ possible labels are partitioned into sets of size a, thus $l(H) = \frac{k!}{a}$. $\square$

Theorem 8.1.3. *Suppose H is a fixed strictly balanced graph with k vertices and $l \geq 2$ edges, and its automorphism group has a elements. Let c be a positive constant and set $p = cn^{-k/l}$. For $G \in \mathcal{G}_{n,p}$ denote by $X = X(G)$ the number of H-subgraphs of G then for $r = 0, 1, 2 \ldots$ and $\lambda = \frac{c^l}{a}$,*

$$\lim_{n\to\infty} \mathbb{P}[X = r] = \frac{e^{-\lambda}\lambda^r}{r!}. \tag{8.2}$$

Proof. Assume $V = \{v_1, v_2, \ldots v_n\}$ and H is a strictly balanced graph on k vertices and l edges. Let $\mathbf{H}^n$ be the collection of subgraphs generated by V that are isomorphic to H. For every $J \in \mathbf{H}^n$, let A_J be the event 'J is a subgraph' and X_J be the indicator random variable for A_J so that $\mathbb{E}[X_J] = p^l$.

To calculate $|\mathbf{H}_N|$, if $K \in [V]^k$, Lemma 8.1.2 implies there are $\frac{k!}{a}$ subgraphs on K that are isomorphic to H. As there are $\binom{n}{k}$ k sets, $|\mathbf{H}^N| = \binom{n}{k}\frac{k!}{a}$.

Since $X = \sum_{J\in\mathbf{H}^N} X_J$ then linearity of expectation implies

$$\mathbb{E}[X] = \sum_{J\in\mathbf{H}^n} \mathbb{E}[X_J]$$
$$= \binom{n}{k}\frac{k!}{a}p^l$$

$$\sim \frac{n^k}{k!}\frac{k!}{a}p^l$$
$$= \frac{n^k p^l}{a}.$$

Using $p^l = (cn^{-k/l})^l = c^l n^{-k}$ implies

$$\mathbb{E}[X] \sim \frac{c^l}{a} = \lambda. \tag{8.3}$$

Theorem 9.3.1 implies to complete the proof, it is enough to show for every $r \in \mathbb{Z}^+$,

$$\lim_{n\to\infty} \mathbb{E}_r[X] = \lambda^r.$$

In order to simplify the calculations of $\mathbb{E}_r[X]$, define the random variable $Y = Y(G)$ counting the number of isolated $J \in \mathbf{H}^n$ that are subgraphs of G. First show

$$\lim_{n\to\infty} \mathbb{E}_r[Y] = \lambda^r,$$

then approximate $\mathbb{E}_r[X]$ and prove

$$\mathbb{E}_r[X] - \mathbb{E}_r[Y] = \mathrm{o}(1).$$

Claim 1. Let $r \in Z^+$ be given. Then

$$\mathbb{E}_r[Y] = \binom{n}{k}\binom{n-k}{k}\binom{n-2k}{k}\cdots\binom{n-(r-1)k}{k}\left(\frac{k!}{a}\right)^r p^{kr}. \tag{8.4}$$

Proof (of Claim 1) To understand equation (8.4), choose a k-set K from $[V]$. As the desired graphs are to be disjoint, choose a k set from $[V] \setminus K$. Continue until r disjoint k sets are chosen. In each case, the probability there is an induced subgraph isomorphic to H is $\frac{k!p^l}{a}$. Thus Claim 1 is proven.

The next step is to calculate the asymptotics of equation (8.4). Observe that

$$\binom{n}{k}\binom{n-k}{k}\cdots\binom{n-(r-1)k}{k}\left(\frac{k!}{a}\right)^r p^{kr} \sim \left(\frac{n^k}{k!}\right)^r\left(\frac{k!}{a}\right)^r (p^l)^r$$
$$= \frac{n^{kr}}{(k!)^r}\frac{(k!)^r}{a^r}p^{lr}$$
$$= \left(\frac{n^k p^l}{a}\right)^r$$

$$= \left(\frac{n^k c (n^{-k/l})^l}{a} \right)^r$$
$$= \left(\frac{c}{a} \right)^r = \lambda^r.$$

Thus

$$\lim_{n \to \infty} \mathbb{E}_r[Y] = \lambda^r. \tag{8.5}$$

Ending the proof of Claim 1 The approximation of $\mathbb{E}_r[X]$, requires that H is strictly balanced. To this end, suppose A and B are graphs and B is isomorphic to H with exactly $1 \le t \le k-1$ vertices not in A.

Claim 2. $|E(A \bigcup B)| > |E(A)| + \frac{tl}{k}$.

Proof (of Claim 2): Let J be the subgraph of B induced by the $k-t$ vertices of B contained in A. As B is strictly balanced,

$$\frac{|E(J)|}{|V(J)|} = \frac{|E(J)|}{k-t} < \frac{l}{k}$$

or

$$|E(J)| < (k-t)\frac{l}{k}.$$

Thus if $e(B \setminus J)$ is the number of edges of B not in J,

$$e(B \setminus J) > l - \left(l - \frac{tl}{k} \right) = \frac{tl}{k}.$$

Hence

$$|E(A \cup B) > |E(A)| + \frac{tl}{k}.$$

Thus Claim 2is proven.

Now suppose $\{H_1, H_2, \ldots H_r\} \in [\mathbf{H}^n]^r$ such that

$$|V(H_1 \cup H_2 \cup \cdots \cup H_r)| = s.$$

Let $t_1 = k$ and for $2 \le j \le r$, let t_j be the number of vertices in H_j that are not in $H_1 \cup H_2 \cup \cdots H_{j-1}$ then

$$\sum_{j=2}^{r} t_j = s - k. \tag{8.6}$$

From Claim 2,

$$\begin{aligned}|E(H_1\bigcup H_2\bigcup\cdots\bigcup H_r)| &> |E(H_1\bigcup H_2\bigcup\cdots\bigcup H_{r-1})| + \frac{t_r l}{k}\\ &> |E(H_1\bigcup H_2\bigcup\cdots\bigcup H_{r-2})| + \frac{t_{r-1} l}{k} + \frac{t_r l}{k}.\end{aligned}$$

Continuing inductively, ...

$$\begin{aligned}&> |E(H_1)| + \frac{t_2 l}{k} + \frac{t_3 l}{k} + \ldots + \frac{t_r l}{k}\\ &= l + \frac{(s-k)l}{k}\\ &= \frac{sl}{k}. && \text{(by (8.6))}\end{aligned}$$

This implies for $\delta > 0$,

$$|E(H_1 \cup H_2 \cup \cdots \cup H_r)| = \frac{sk}{l} + \delta$$

where the 'δ' was added to make the strict inequality an equality.

Finally,

$$\begin{aligned}\mathbb{E}_r[X] - \mathbb{E}_r[Y] &< \sum_{s=k+1}^{rk-1} \binom{n}{s} \left(\binom{s}{k}\frac{k!}{a}\right)^r p^{\frac{sl}{k}+\delta}\\ &\sim \sum_{s=k+1}^{rk-1} \frac{n^s}{s!}\left(\frac{s^k}{k!}\frac{k!}{a}\right)^r p^{\frac{sl}{k}+\delta}\\ &= \sum_{s=k+1}^{rk-1} \frac{n^s}{s!}\left(\frac{s^k}{a}\right)^r p^{\frac{sl}{k}+\delta}\\ &= \sum_{s=k+1}^{rk-1} \frac{s^{kr}}{s!}\left(np^{\frac{l}{k}}\right)^s p^{\delta}\\ &= \sum_{s=k=1}^{rk-1} \frac{s^{kr}}{s!} c^2 (cn^{-k/l})^{\delta} && \text{(substituting } p = cn^{-k/l}\text{)}\\ &= \sum_{s=k+1}^{rk-1} \mathrm{O}(1) n^{-\frac{k\delta}{l}} = \mathrm{o}(1) && \text{(as } n \text{ is the only variable)}\end{aligned}$$

as claimed. Thus

$$\mathbb{E}_r[X] \simeq \mathbb{E}_r[Y] = \lambda^r.$$

Hence X has asymptotically Poisson distribution. $\square$

8.1.2 $\mathcal{G}_{n,q}$: Trees

Theorem 8.1.4. [14] *Suppose $N(n)$ is a sequence of positive integers such that for every n, $N(n) \le \binom{n}{2}$ and*

$$\lim_{n\to\infty} \frac{N(n)}{n^{\frac{k-2}{k-1}}} = \sigma > 0.$$

Let τ_k be the random variable on $G_{n,N(n)}$ counting the number of isolated trees of order k. Then for $j \in \mathbb{Z}^+ \cup \{0\}$,

$$\lim_{n\to\infty} \mathbb{P}[\tau_k = j] = \frac{\lambda^j e^{-\lambda}}{j!} \tag{8.7}$$

where

$$\lambda = \frac{(2\sigma)^{k-1} k^{k-2}}{k!}.$$

Proof. Let $N(n) = N$ and T_n^k be the set of k trees on n vertices. For every $T \in T_n^k$, let A_T be the event "T is an isolated tree in $G \in G_{n,N}$" and let X_T be the indicator random variable for A_T.

$$\begin{aligned}
\mathbb{E}[X_T] &= \frac{\binom{\binom{n-k}{2}}{N-k+1}}{\binom{\binom{n}{2}}{N}} \\
&\sim \frac{\binom{(n-k)^2/2}{(N-k+1)}}{\binom{n^2/2}{N}} \\
&= \frac{\left[(n-k)^2/2\right]^{N-k+1}}{\left[n^2/2\right]^N} \frac{N!}{(N-k+1)!} \\
&\sim \frac{\left[(n-k)^2/2\right]^{N}}{\left[n^2/2\right]^N} \frac{N^{k-1}}{\left[(n-k)^2/2\right]^{k-1}} \\
&\sim \left[1 - \frac{k}{n}\right]^{2N} \left[\frac{2N}{(n-k)^2}\right]^{k-1} \\
&\sim \left[1 - \frac{k}{n}\right]^{2N} \left[\frac{2N}{n^2}\right]^{k-1} \\
&\sim e^{-\frac{2kN}{n}} \left[\frac{2N}{n^2}\right]^{k-1}.
\end{aligned}$$

Let $2 \le r \in \mathbb{Z}^+$. The next step in the proof of Theorem 8.1.4 is to calculate the r-th factorial moment.

Claim 1. Let $\{T_1, T_2, \ldots T_r\} \in [T_n^k]^r$ with the respective indicator random variables $X_{T_1}, X_{T_2} \ldots$ then

$$\mathbb{E}[X_{T_1}, X_{T_2}, \ldots, X_{T_r}] = \mathbb{P}[A_{T_1} \wedge A_{T_2} \wedge \ldots \wedge A_{T_r}]$$
$$= \begin{cases} 0 & \text{if any pair of the } T_i\text{'s are not disjoint} \\ \frac{\binom{\binom{n-rk}{2}}{N-r(k-1)}}{\binom{\binom{n}{2}}{N}} & \text{otherwise.} \end{cases}$$

Proof (of Claim 1) Let $T_1, T_2 \ldots T_r \in T_n^k$ be distinct elements and A_{T_i} be the corresponding event "T_i is an isolated tree in $G \in G_{n,N}$". From the definition, if for $i \neq j$, $T_i \cap T_j \neq \emptyset$ then A_{T_i} is mutually disjoint from A_j, thus if any pair of $T_1, T_2, \ldots T_r$ are not disjoint,

$$\mathbb{E}[X_{T_1}, X_{T_2}, \ldots X_{T_r}] = \mathbb{P}[A_{T_1} \wedge A_{T_2} \wedge \ldots \wedge A_{T_r}] = 0. \tag{8.8}$$

Now assume that $T_1, T_2 \ldots T_r$ are pairwise disjoint. Then

$$\mathbb{E}[X_{T_1}, X_{T_2}, \ldots, X_{T_r}] = \mathbb{P}[A_{T_1} \wedge A_{T_2} \wedge \ldots \wedge A_{T_r}] \tag{8.9}$$
$$= \frac{\binom{\binom{n-rk}{2}}{N-r(k-1)}}{\binom{\binom{n}{2}}{N}}. \tag{8.10}$$

Combining equations (8.8) and (8.9) gives proves Claim 1

The next step in the proof of Theorem 8.1.4 is to prove the following claim.

Claim 2. Let $n \in \mathbb{Z}^+$ be given. Then

$$\sum \mathbb{E}[X_{T_1}, X_{T_2} \ldots X_{T_r}] \sim \binom{\binom{n}{k}k^{k-2}}{r} \frac{\left[\binom{\binom{n-rk}{2}}{N-r(k-1)}\right]}{\left[\binom{\binom{n}{2}}{N}\right]} \tag{8.11}$$

where the sum is over $[T_n^k]^r$.

Proof (of Claim 2) Notice that

$$|T_n^k| = \binom{n}{k} k^{k-2}.$$

To see this, choose a $K \in [n]^k$. As the vertices are labeled, Cayley's formula implies there are k^{k-2} distinct trees. As there are $\binom{n}{k}$ such K sets, there are

$\binom{n}{k}k^{k-2}$ trees. Equation (8.11) follows from Lemma 1, under the assumption that as k and r remain fixed, when n is large, the number of pairwise non-disjoint r-sets of k-trees is small in comparison to the total number of trees, thus Claim 2.

The next step in the proof of Theorem 8.1.4 is to examine the asymptotic behaviour of equation (8.11). As

$$
\begin{aligned}
\binom{\binom{n}{k}k^{k-2}}{r}\frac{\left[\binom{\binom{n-rk}{2}}{N-r(k-1)}\right]}{\left[\binom{\binom{n}{2}}{N}\right]} &\sim \binom{\frac{n^k k^{k-2}}{k!}}{r}\frac{\binom{(n-rk)^2/2}{N-r(k-1)}}{\binom{n^2/2}{N}} \\
&\sim \frac{\left[\frac{n^k k^{k-2}}{k!}\right]^r}{r!}\left[\frac{((n-rk)^2/2)^{N-r(k-1)}}{(N-r(k-1))!}\right]\left[\frac{(n^2/2)^N}{N!}\right] \\
&\sim \left[\frac{n^{rk}k^{r(k-2)}}{(k!)^r r!}\right]\frac{[(n-rk)^2/2]^{N-r(k-1)}}{(n^2/2)^N}\frac{N!}{(N-r(k-1))!} \\
&\sim \left[\frac{n^{rk}k^{r(k-2)}}{(k!)^r r!}\right]\frac{(n-rk)^{2(N-r(k-1))}}{n^{2N}}\frac{2^N}{2^{N-r(k-1)}(N)_{r(k-1)}} \\
&\sim \left[\frac{n^{rk}k^{r(k-2)}}{(k!)^r r!}\right]\left[\frac{n^{rk}k^{r(k-2)}}{(k!)^r r!}\right] \\
&\quad \left(1-\frac{rk}{n}\right)^N (n-rk)^{-2r(k-1)}\,2^{r(k-1)}(N)_{r(k-1)} \\
&\sim \left[\frac{n^{rk}k^{r(k-2)}}{(k!)^r r!}\right]\left(1-\frac{rk}{n}\right)^{2N}\frac{2^{r(k-1)}(N)_{r(k-1)}}{(n-rk)^{2r(k-1)}} \\
&\sim \left[\frac{n^{rk}k^{r(k-2)}}{(k!)^r r!}\right]\left(1-\frac{rk}{n}\right)^{2N}\frac{2^{r(k-1)}N^{r(k-1)}}{(n-rk)^{2r(k-1)}} \\
&= \left[\frac{k^{r(k-2)}}{(k!)^r r!}\right]\left(1-\frac{rk}{n}\right)^{2N} n^{rk}\left(\frac{(2N)}{(n-rk)^2}\right)^{r(k-1)} \\
&\sim \left[\frac{k^{r(k-2)}}{(k!)^r r!}\right] e^{\frac{-2Nrk}{n}} n^{rk}\left(\frac{2N}{(n-rk)^2}\right)^{r(k-1)} \\
&\sim \left[\frac{k^{r(k-2)}}{(k!)^r r!}\right] e^{\frac{-2Nrk}{n}} n^{rk}\left(\frac{2N}{n^2}\right)^{r(k-1)} \\
&\sim \frac{1}{r!}\left(\frac{k^{k-2}}{k!}\right)^r e^{\frac{-2Nrk}{n}}\left(\frac{2N^{r(k-1)}}{n^{r(k-1)-rk}}\right) \\
&= \frac{1}{r!}\left(\frac{k^{k-2}}{k!}\right)^r e^{\frac{-2Nrk}{n}}\left(\frac{2N^{r(k-1)}}{n^{r(k-2)}}\right) \\
&= \frac{1}{r!}\left(\frac{k^{k-2}}{k!}\right)^r e^{\frac{-2Nrk}{n}}\left(\frac{2N}{n^{\frac{r(k-2)}{r(k-1)}}}\right)^{r(k-1)}.
\end{aligned}
$$

Thus

$$\sum_{[T_n^k]^r} \mathbb{E}[X_{T_1}, X_{T_2}, \ldots, X_{T_r}] \sim \frac{1}{r!} \left(\frac{k^{k-2}}{k!} \right)^r e^{\frac{-2Nrk}{n}} \left(\frac{2N}{n^{\frac{k-2}{k-1}}} \right)^{r(k-1)}. \tag{8.12}$$

As

$$\lim_{n \to \infty} \frac{N(n)}{n^{\frac{k-2}{k-1}}} = \sigma > 0,$$

equation (8.12) implies

$$\begin{aligned} \lim_{n\to\infty} \sum_{[T_n^k]^r} \mathbb{E}[X_{T_1}, X_{T_2}, \ldots, X_{T_r}] &= \lim_{n\to\infty} \frac{1}{r!} \left(\frac{k^{k-2}}{k!} \right)^r e^{\frac{-2Nrk}{n}} \left(\frac{2N}{n^{\frac{k-2}{k-1}}} \right)^{r(k-1)} \\ &= \frac{1}{r!} \left(\frac{k^{k-2}}{k!} \right)^r (2\sigma)^{r(k-1)}. \end{aligned}$$

Using

$$\lambda = \frac{(2\sigma)^{k-1} k^{k-2}}{k!}$$

gives

$$\lim_{n\to\infty} \sum_{[T_n^k]^r} \mathbb{E}[X_{T_1}, X_{T_2}, \ldots, X_{T_r}] = \frac{1}{r!} \left(\frac{k^{k-2}(2\sigma^{k-1})}{k!} \right)^r = \frac{\lambda^r}{r!}. \tag{8.13}$$

Thus Theorem 9.3.1 implies Theorem 8.1.4. □

8.2 Graphical evolution

Up until now, the threshold functions that have been in calculated in this work have involved what might be called (rather vaguely) *local properties* versus what could be known as *global properties.* While a clear definition of these terms is unknown to the author, notice that if Q is the graph property of having a 3-clique, knowing that whether a graph has such a property says little about the overall structure of the graph, while knowing if a graph is connected, planar or if G is a forest gives information about the global structure of the graph.

To begin, here is an alternative viewpoint on random graphs. Let n be some large, yet unspecified number. Label the edges $1, 2, \ldots, \binom{n}{2}$. Starting with 1, in some manner (for instance a random binary number generator) choose an edge with probability p. This process takes $\binom{n}{2}$ steps and the resultant is a random graph. It is easy to see that this process is one way to view a random graph. For other methods, please see for example [30].

With this process in mind, it should be clear that the smaller p is, the fewer the expected number of edges would be. For very small p, there would most likely be many isolated vertices and just a smattering of edges here and there. As p is slowly increased, edges connect the isolated vertices into isolated trees. Cycles then start to appear. First 3,4 then k-cycles would start appearing; the higher the probability an edge being in the graph the more likely there is cycles. At some point, the graph would most likely be connected.

All of this was clear to Erdős and Rényi, (founders of the theory of random graphs), their papers in the late fifties and early sixties laid the ground work for most of these ideas. The size of the connected components of almost all graphs changes near

$$c > 0 \quad p = \frac{2c}{n}$$

which is where the phenomenon of the so called "double jump" occurs. As seen in Theorem 7.1.5, $p(n) = \frac{1}{n}$ is the threshold for acyclic graphs, and since this change is fundamental to the structure of graphs, it is an important area to research more closely. In this section, some of the results in this direction are shown.

In $\mathcal{G}_{n,p}$, let X be the number of cycles of any size. In an Theorem 7.1.5, it was shown that

$$\mathbb{E}[X] = \sum_{k=3}^{n} \binom{n}{k} \frac{p^k}{2k}. \tag{8.14}$$

If $X(G) = 0$ then G is acyclic hence G is a forest, while the larger $X(G)$ is, the less trees G has. Continuing, if

$$\lim_{n\to\infty} \mathbb{E}[X] = 0$$

Markov's inequality(3.4) implies $X = 0$ almost surely, and almost surely, $G \in \mathcal{G}_{n,p}$ is a forest.

8.2.1 Evolution of tree components

Theorem 8.2.1. [4] *For $k \geq 2$ and $G \in \mathcal{G}_{n,p}$, let $X_k = X_k(G)$ be the number of components in G that are trees of order k.*

(i) If $p(n) = \mathrm{o}(n^{-k/k-1})$, then almost surely $X_k = 0$.

(ii) If for some constant $c > 0$ and $p \sim cn^{-k/k-1}$ then X_k converges in distribution to the Poisson distribution with $\lambda = c^{k-1}k^{k-2}/k!$.

(iii) If $pn^{k/k-1} \to \infty$ and $pkn - \ln n - (k-1)\ln\ln n \to -\infty$ then for every $L \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}[X_k \geq L] = 1.$$

(iv) If $pn^{k/k-1} \to \infty$ and for some $x \in \mathbb{R}$, $pkn - \ln n - (k-1)\ln\ln n \to x$ then X_k converges in distribution to the Poisson distribution with $\lambda = \mathrm{e}^{-x}/(kk!)$.

(v) If $pn^{k/k-1} \to \infty$ and $pkn - \ln n - (k-1)\ln\ln n \to \infty$ then almost surely $X_k = 0$.

Proof. Let T_n^k be the collection of labeled trees of order k on n labeled vertices. Then $|T_n^k| = \binom{n}{k}k^{k-2}$.

(i) Assume $p(n) = \mathrm{o}(n^{-k/k-1})$. As trees of order k are balanced graphs with $v = k$ and $e = k-1$, it was shown in Theorem 7.1.7, $p(n) = n^{-k/k-1}$ is the threshold function for graphs having an isolated component in T_n^k. Therefore almost surely $X_k = 0$.

(ii) Assume that for some constant $c > 0$, $p(n) \sim cn^{-k/k-1}$.

First, observe that

$$\begin{aligned}
\mathbb{E}[X_k] &= \binom{n}{k} k^{k-2}p^{k-1}(1-p)^{\binom{k}{2}-(k-1)+k(n-k)} \\
&\sim \frac{k^{k-2}}{k!}n^k(cn^{-k/k-1})^{k-1}(1-p)^{kn}(1-p)^{\binom{k}{2}-(k-1)+k^2} \\
&\sim \frac{c^{k-1}k^{k-2}}{k!}\mathrm{e}^{-pkn} && (k \text{ fixed})
\end{aligned}$$

$$
\begin{aligned}
&\sim \frac{c^{k-1}k^{k-2}}{k!} && (pn \to 0)\\
&= \lambda.
\end{aligned}
$$

Let $1 < r \in \mathbb{Z}$ be given. Then

$$
\begin{aligned}
\mathbb{E}_r[X] &= \binom{n}{k}\binom{n-k}{k}\binom{n-2k}{k}\cdot\binom{n-(r-1)k}{k}\\
&\quad\cdot (k^{k-2})^r p^{r(k-1)}(1-p)^{\binom{rk}{2}-r(k-1)+rk(n-rk)}\\
&\sim \left(\frac{k^{k-2}}{k!}\right)^r n^{rk}(cn^{-k/k-1})^{r(k-1)}(1-p)^{nrk} && (r, k \text{ are fixed})\\
&\sim \left(\frac{c^{k-1}k^{k-2}}{k!}\right)^r \mathrm{e}^{-pnrk} && (1-p\sim e^{-p})\\
&= \lambda^r \mathrm{e}^{-crkn^{-k/k-1}n} && (p\sim cn^{-k/k-1})\\
&= \lambda^r \mathrm{e}^{-(crkn^{(-1/k-1)})}\\
&\sim \lambda^r
\end{aligned}
$$

as needed.

(iii) Assume $pn^{k/k-1} \to \infty$ and $pkn - \ln n - (k-1)\ln\ln n \to -\infty$. The second moment method is used to show for all $0 < L \in \mathbb{Z}$, $\mathbb{P}[X_k < L] \to 0$. As in (ii),

$$
\begin{aligned}
\mathbb{E}[X_k] &= \binom{n}{k} k^{k-2}p^{k-1}(1-p)^{k(n-k)+\binom{k}{2}-(k-1)}\\
&\sim n^k p^{k-1}(1-p)^{kn}\\
&\sim n^k p^{k-1}\mathrm{e}^{-kpn}.
\end{aligned}
$$

The idea for the proof that $\mathbb{E}[X_k] \to \infty$ comes from [4, p. 75]. Let $0 < x \in \mathbb{R}$ be arbitrary. Assume $p(n) = xn^{-k/k-1}$. Therefore

$$
\begin{aligned}
\mathbb{E}[X_k] &\sim n^k(xn^{-k/k-1})^{k-1}e^{-kxn^{-k/k-1}n}\\
&= x^{k-1}e^{-kxn^{-1/k-1}}
\end{aligned}
$$

As k and x are fixed,

$$
\mathbb{E}[X_k] \to x^{k-1}.
$$

Since x is arbitrary, is $pn^{k/k-1} \to \infty$ and $pkn - \ln n - (k-1) \ln \ln n \to -\infty$ then

$$\mathbb{E}[X_k] \to \infty.$$

To calculate $\mathbb{E}[X^2]$, let $\alpha \in T_n^k$. Let A_α be the event "α is an isolated component of $G \in \mathcal{G}_{n,p}$" and X_α be the respective indicator random variable. If $\{\alpha, \beta\} \in [T_n^k]^2$, then $X_\alpha X_\beta = 1$ iff α and β have no vertices in common.

As

$$\begin{aligned}
\mathbb{E}[X_\alpha X_\beta] &= \mathbb{P}[A_\alpha \wedge A_\beta] \\
&= \mathbb{P}[A_\beta | A_\alpha]\mathbb{P}[A_\alpha] \\
&= p^{k-1}(1-p)^{k(n-k)+\binom{k}{2}-(k-1)} p^{k-1}(1-p)^{k(n-2k)+\binom{k}{2}-(k-1)} \\
&= p^{2k-2}(1-p)^{2kn-3k^2+2\binom{k}{2}-2(k-1)}.
\end{aligned}$$

For every $\alpha \in T_n^k$, there are $\binom{n-k}{k}$ events independent of A_α, therefore

$$\mathbb{E}[X_k^2] = \binom{n}{k}\binom{n-k}{k} p^{2k-2}(1-p)^{2kn-3k^2+2\binom{k}{2}-2(k-1)}.$$

Hence $\mathbb{E}[X_k^2] = \mathrm{o}(\mathbb{E}[X]^2)$ hence Theorem 3.3.8 implies almost surely $X_k \sim \mathbb{E}[X_k]$, as needed.

(iv) Suppose $pn^{k/k-1} \to \infty$ and let $x \in \mathbb{R}$ such that $pkn - \ln n - (k-1) \ln \ln n \to x$. Then

$$\begin{aligned}
\mathbb{E}[X_k] &= \binom{n}{k} k^{k-2} p^{k-1}(1-p)^{\binom{k}{2}-(k-1)+k(n-k)} \\
&\sim \frac{k^{k-2}}{k!} n(np)^{k-1}(1-p)^{kn} \\
&\sim \frac{k^{k-2}}{k!} n(np)^{k-1} e^{-pkn}.
\end{aligned}$$

Let $\{c_n\} \subset \mathbb{R}$ such that $pkn - \ln n - (k-1)\ln\ln n + c_n = x$ or $c_n - x - \ln n - (k-1)\ln\ln n = -pkn$ so that

$$e^{-pkn} = e^{c_n - x - \ln n - (k-1)\ln\ln n}$$

$$= \frac{e^{c_n - x}}{n(\ln n)^{k-1}}.$$

Therefore

$$\begin{aligned}\mathbb{E}[X_k] &\sim \frac{k^{k-2}}{k!} n(np)^{k-1} \frac{e^{c_n - x}}{n(\ln n)^{k-1}} \\ &= \frac{k^{k-2} e^{c_n - x}}{k!} \left(\frac{np}{\ln n}\right)^{k-1}.\end{aligned}$$

As

$$\begin{aligned}\frac{np}{\ln n} &= \frac{x + \ln n + (k-1)\ln\ln n}{\ln n} \\ &= 1 + \mathrm{o}(1).\end{aligned}$$

Hence $\mathbb{E}[X_k] \sim \dfrac{k^{k-2} e^{-x}}{k!} = \lambda$.

Let $1 < r \in \mathbb{Z}$ be given. As in step (ii),

$$\begin{aligned}\mathbb{E}_r[X] &= \binom{n}{k}\binom{n-k}{k}\binom{n-2k}{k}\cdots\binom{n-(r-1)k}{k}\cdot \\ &\quad \cdot (k^{k-2})^r p^{r(k-1)}(1-p)^{\binom{rk}{2} - r(k-1) + rk(n-rk)} \\ &\sim \left(\frac{k^{k-2}}{k!}\right)^r n^{rk} p^{r(k-1)} (1-p)^{nrk} && (r, k \text{ are fixed}) \\ &\sim \left(\frac{k^{k-2}}{k!}\right)^r n^r (np)^{r(k-1)} (e^{-pnk})^r \\ &= \left(\frac{k^{k-2}}{k!}\right)^r n^r (np)^{r(k-1)} \left(\frac{e^{c_n - x}}{n \ln n^{k-1}}\right)^r \\ &\sim \lambda^r \left(\frac{np}{\ln n}\right)^{r(k-1)} \\ &\sim \lambda^r\end{aligned}$$

as needed.

(v) Assume $pn^{k/k-1} \to \infty$ and $pkn - \ln n - (k-1)\ln\ln n \to \infty$. As in step (ii),

$$\begin{aligned}\mathbb{E}[X_k] &= \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{\binom{k}{2} - (k-1) + k(n-k)} \\ &\sim n^k p^{k-1} e^{-kpn} \\ &\le n^k e^{-kpn}\end{aligned}$$

$$= e^{k \ln n - pkn} \qquad\qquad (n^k = e^{k \ln n}).$$

Therefore

$$\mathbb{E}[X_k] \le e^{k \ln n - knp} \to 0 \qquad (\text{as } pkn - \ln n - (k-1)\ln\ln n \to \infty).$$

Inequality (3.4) implies

$$\mathbb{P}[X_k > 0] \le \mathbb{E}[X_k] \to 0$$

as claimed.

□

Theorem 8.2.1 shows that for every $2 < k \in \mathbb{Z}$ and for $0 < p < n^{-k/k-1}$, the number of components of a graph $G \in \mathcal{G}_{n,p}$, that are k-trees is zero, while as p grows according to Theorem 8.2.1, the nature of the components that are k-trees changes until p is large so that almost all $G \in \mathcal{G}_{n,p}$ have no components that are k-trees. The next section illustrates more of the concept of graphical evolution.

8.2.2 Evolution of connected components

This section continues the synopsis of the evolution of random graphs; in particular the evolution of connected components. This will only be a summary, without any proofs. For more information, please see either Palmer [30] or Bollobás [4].

Section 8.2.1 shows the evolution of the number of components that are trees. A natural next question is how many vertices are in trees. Let $0 < c \in \mathbb{R}$ and $x(c) = \displaystyle\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(2ce^{-2c})^k$. Observe that $x(c)$ converges while $2ce^{-2c} \le e^{-1}$ and that $\displaystyle\lim_{c\to\infty} \frac{x(c)}{2c} = 0$.

Theorem 8.2.2. *Let X be the random variable counting the number of vertices in a random graph, $G \in Gnp$, that belong to components that are trees and*

$0 \leq c \in \mathbb{R}$. *Let* $p = \dfrac{2c}{n}$, *then*

$$\frac{E[X]}{n} \to \begin{cases} 1 & \text{for } c \leq 1/2 \\ \frac{x(c)}{2c} & \text{for } c > 1/2. \end{cases}$$

Which has the following corollary.

Corollary 8.2.3. *Let* $p = \frac{2c}{n}$ *and* $c < 1/2$, *almost all graphs have no components with more than one cycle.*

Thus almost all graphs are trees with one cycle.

Let Y be the random variable counting the size of the largest component of $G \in \mathcal{G}_{n,p}$. Notice that if for some $G \in \mathcal{G}_{n,p}$, $X(G) \sim n$, then G would be one large component with a smattering of small components. If for some $0 < \epsilon < 1$, $X(G) < (1-\epsilon)n$, then G would have two disconnected components of large size.

Theorem 8.2.4 ([9])**.** *Suppose* $1/2 < c$ *and* $p = \frac{2c}{n}$. *Let* $G(c) = 1 - \dfrac{x(c)}{2c}$. *then for all* $\epsilon > 0$,

$$\mathbb{P}\left[\left\{\frac{Y}{n} - G(c)\right\} < \epsilon\right] \to 1.$$

Note that as c gets large, $G(c) \to 1$. Thus, almost surely the largest component has at least $nG(c)$ vertices.

Chapter 9

Calculations for convergence in distribution

Definition 9.0.5 (Convergence in distribution). *Let $X, X_1, X_2, \ldots$ be nonnegative, integer valued random variables, if for every $0 \leq k \in \mathbb{Z}$,*

$$\lim_{n\to\infty} \mathbb{P}[X_n = k] = \mathbb{P}[X = k]$$

then $\{X_n\}$ converges in distribution to X. In notation, $X_n \xrightarrow{d} X$.

This chapter covers the material necessary in the calculations regarding the convergence of distribution.

9.1 Events

This section follows the presentation of Bollobás [4, pp. 15-25]. Here is an example that shows the idea for some of the results in this section. The statement and proof come from Jukna [23], but both are standard in most text books on elementary combinatorics. Let X be a finite set, and $A_1, A_2, \ldots, A_n \subset X$. For every $I \subset [n]$, define $A_I = \bigvee_{i\in I} A_i$, with the convention $A_\emptyset = X$.

Theorem 9.1.1 (Inclusion/exclusion principle)**.** *Let X be a finite set and $A_1, A_2, \dots A_n \subset X$. The number of elements of X not in any A_i is*

$$\sum_{I \subset [n]} (-1)^{|I|} |A_I|.$$

Definition 9.1.2. *Let the probability space $(\Omega, \mathbb{P})$ be given. Let $J \subset [n]$, a Boolean polynomial g_J on Ω is of the form*

$$g_J(A_1, A_2, \dots A_n) = \left(\bigwedge_{i \in J} A_i \right) \wedge \left(\bigwedge_{i \notin J} \overline{A_i} \right).$$

Theorem 9.1.3. *Let $f_1, f_2, \dots, f_k$ be Boolean polynomials in n variables $A_1, A_2, \dots A_n$ and let $b_1, b_2 \dots b_k$ be real constants. Suppose*

$$\sum_{i=1}^{k} b_i \mathbb{P}[\{f_i(B_1, B_2 \dots B_n\}] \geq 0 \tag{9.1}$$

whenever $B_1, B_2 \dots B_n$ are events in a probability space $(\Omega, \mathbb{P})$ such that

$$\forall 1 \leq i \leq n; \quad \mathbb{P}[B_i] = 0 \text{ or } 1.$$

Equation (9.1) holds for every choice of events $C_1, C_2 \dots C_n$ in $(\Omega, \mathbb{P})$.

Proof. From Definition 9.1.2, a Boolean polynomial can be written and a product of subsets of $[n]$. Therefore, if $g_J(A_1, A_2, \dots A_n) = \left(\bigwedge_{i \in J} A_i\right) \wedge \left(\bigwedge_{i \notin J} \overline{A_i}\right)$, define the constant

$$c_J = \begin{cases} 0 & g_J \text{ is not in inequality } (9.1) \\ b_i & \text{if } f_i = g_J. \end{cases}$$

Equation (9.1) can be rewritten as

$$\sum_{J \subset [n]} c_J \mathbb{P}[\{g_J(B_1, \dots B_n\}] \geq 0. \tag{9.2}$$

Given a set $J_0 \subseteq [n]$ choose events $B_1, B_2 \dots B_n$ such that

$$\mathbb{P}[B_i] = \begin{cases} 0 & \text{if } i \notin J_0 \\ 1 & \text{otherwise.} \end{cases}$$

For this choice of $\{B_1, B_2 \dots B_n\}$,

$$\mathbb{P}[g_J(B_1, B_2, \dots B_n)] = \begin{cases} 0 & \text{for all } J \neq J_0 \subseteq [n] \\ 1 & J = J_0. \end{cases}$$

Thus

$$c_{J_0} = \sum_{J \subset [n]} c_J \mathbb{P}[\{g_J(B_1, \dots B_n\}] \geq 0. \tag{9.3}$$

Hence for any collection of events $C_1, C_2 \dots C_n$ equation (9.2) holds. □

Theorem 9.1.3 has the following corollary.

Corollary 9.1.4. *If for every sequence of events $B_1, B_2 \dots B_n$ with $\mathbb{P}[B_i] = 0$ or 1,*

$$\sum_{i=1}^{k} b_i \mathbb{P}[\{f_i(B_1, B_2 \dots B_n\}] = 0,$$

then for any choice of events $C_1, C_2, \dots C_n$,

$$\sum_{i=1}^{k} b_i \mathbb{P}[\{f_i(C_1, C_2 \dots C_n\}] = 0.$$

For the rest of this section, let $A_1, A_2, \dots A_n$ be arbitrary events in a probability space $(\Omega, \mathbb{P})$. For $J \subseteq [n]$, define and $A_\emptyset = \Omega$ and $A_J = \bigwedge_{j \in J} A_j$. Next define $p_J = \mathbb{P}[A_J]$ and for $r = 0, 1, 2 \dots, n.$ let $S_r = \sum_{J \in [n]^r} p_J$.
The following corollary is similar Theorem 9.1.1 and allows for a calculation of the probabilities of events using combinations.

Corollary 9.1.5. *If $A_1, A_2, \dots A_n$ is a sequence of events in a probability space $(\Omega, \mathbb{P})$ then*

$$\mathbb{P}[\bigvee_{j=1}^{n} A_j] = \sum_{r=1}^{n} (-1)^{r+1} S_r.$$

Theorem 9.1.6. *Let $A_1, A_2, \dots, A_n$ be events in a probability space $(\Omega, \mathbb{P})$ and let $S_0, S_1, \dots S_n$ be as above and denote by p_k the probability that exactly k of these events occur. Then*

$$p_k = \sum_{r=k}^{n} (-1)^{r+k} \binom{r}{k} S_r. \tag{9.4}$$

Proof. As a result of Theorem 9.1.3, it is enough to show equation (9.4) holds for events $\{A_i\}_{i=1}^n$ such that for some $1 \leq l \leq n$,

$$\mathbb{P}[A_1] = \mathbb{P}[A_2] = \ldots = \mathbb{P}[A_l] = 1$$

and

$$\mathbb{P}[A_{l+1}] = \mathbb{P}[A_{l+2}] = \ldots = \mathbb{P}[A_n] = 0.$$

If B and C are events with $\mathbb{P}[B] = \mathbb{P}[C] = 1$ then

$$\mathbb{P}[\overline{B}] = \mathbb{P}[\overline{C}] = 1 - \mathbb{P}[B] = 0.$$

As $B \wedge \overline{C} \subseteq \overline{C}$ then

$$0 \leq \mathbb{P}[B \wedge \overline{C}] \leq \mathbb{P}[\overline{C}] = 0$$

hence

$$\mathbb{P}[B \wedge \overline{C}] = 0.$$

Since $B = (B \wedge \overline{C}) \vee (B \wedge C)$ is a disjoint union, then

$$\begin{aligned} 1 &= \mathbb{P}[B] \\ &= \mathbb{P}[B \wedge \overline{C} \vee B \wedge C] \\ &= \mathbb{P}[B \wedge \overline{C}] + \mathbb{P}[B \wedge C] \\ &= 0 + \mathbb{P}[B \wedge C] = \mathbb{P}[B \wedge C]. \end{aligned}$$

Thus for $1 \leq k \leq n$,

$$\mathbb{P}[\bigwedge_{j=1}^{k} A_j] = \begin{cases} 0 & \text{if for one of the } A_j\ \mathbb{P}[A_j] = 0 \\ 1 & \text{otherwise.} \end{cases}$$

All together this implies

$$S_r = \binom{l}{r}. \tag{9.5}$$

As a short explanation of why equation (9.5) holds,

$$S_r = \sum_{J \in [n]^r} p_J$$

and $p_J = 1$ if and only if for all $j \in J \quad \mathbb{P}[A_j] = 1$ which happens for all $J \in [l]^r$ by the assumptions on the A_j's. Using the standard counting argument then (as in the inclusion/exclusion principle) shows the result. $\square$

9.2 Random variables

9.2.1 Falling factorial

Let X be a random variable on some finite probability space, $(\Omega, \mathbb{P})$.
Denote

$$(X)_0 = X^0 = 1. \tag{9.6}$$

To clarify, as X is a random variable, for every $G \in \Omega, X(G)$ is a real number. Thus this formula says that for every $G \in \Omega$, $(X)_0(G) = 1$. For every positive integer r,

$$(X)_r = X(X-1)(X-2)\cdots(X-r+1). \tag{9.7}$$

Example 9.2.1. Let X be a random variable

$$\begin{aligned}
&(X)_0 = 1 \\
&(X)_1 = X \\
&(X)_2 = X(X-1) = X^2 - X \\
&(X)_3 = X(X-1)(X-2) = X^3 - 3X^2 + 2X \\
&(X)_r = (X)_{r-1}(X-r+1) = X(X)_{r-1} - (r-1)(X)_{r-1}.
\end{aligned}$$

These equations are the random variable version of the 'falling factorial notation' of real numbers (*i.e.*, $(x)_r = x(x-1)\cdots(x-r+1)$.)

If j is a positive integer, the notation X^j is the random variable $X^j(G) = (X(G))^j$, then by multiplying equation (9.7) shows that for all $0 \leq r \in \mathbb{Z}^+$,

$$(X)_r = \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} X^j. \tag{9.8}$$

Suppose X is a random variable that is expressed as the sum of random variables,

$$X = \sum_{j=1}^{n} X_j.$$

Then

$$\begin{aligned}
(X)_2 &= X(X-1) = X^2 - X \\
&= \sum_{(i,j)\in[n]\times[n]} X_i X_j - \sum_{i\in[n]} X_i \\
&= 2! \sum_{1\le i<j\le n} X_i X_j && (\text{as } X_i^2 = X_i). \\
(X)_3 &= (X)_2(X-2) \\
&= X\left(\sum_{1\le i<j\le n} X_i X_j\right) - 2\sum_{1\le i<j\le n} X_i X_j \\
&= 3! \sum_{1\le i<j<k\le n} X_i X_j X_k && (\text{as } X_i^2 = X_i).
\end{aligned}$$

The general formula is

$$(X)_r = r! \sum_{1\le l_1<l_2<\cdots<l_r\le n} X_{l_1} X_{l_2} \cdots X_{l_r}.$$

If, for every j, X_j is an indicator random variable for the event A_j, then

$$X_{l_1} X_{l_2} \cdots X_{l_r} = X_{A_{l_1}\wedge A_{l_2}\wedge\cdots\wedge A_{l_r}}.$$

Thus

$$\mathbb{E}[X_{l_1} X_{l_2} \cdots X_{l_r}] = \mathbb{P}[A_{l_1} \wedge A_{l_2} \wedge \cdots \wedge A_{l_r}].$$

Hence

$$\begin{aligned}
\mathbb{E}[(X)_r] &= r! \sum_{1\le l_1<l_2<\cdots<l_r\le n} \mathbb{E}[X_{l_1} X_{l_2} \cdots X_{l_r}] \\
&= r! \sum_{1} \mathbb{P}[A_{l_1} \wedge A_{l_2} \wedge \cdots \wedge A_{l_r}].
\end{aligned}$$

9.2.2 rth Factorial moments

For a positive integer r, define the r-factorial moment by

$$\mathbb{E}_r[X] = \mathbb{E}[(X)_r]. \tag{9.9}$$

Observe that if X counts the number of objects of a certain class, then $\mathbb{E}_r[X]$ is the expected number of *ordered r-tuples* of that class. From equation (9.8),

$$\mathbb{E}_r[X] = \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} \mathbb{E}[X^r], \tag{9.10}$$

as expectation is linear.

Corollary 9.2.2. *Let X be a random variable which takes values in $\{0, 1, 2, \ldots, n\}$ and let $\mathbb{E}_r[X]$ be the r-th factorial moment of X. Then*

$$\mathbb{P}[X = k] = \frac{\sum_{r=k}^{n} (-1)^{k+r} \frac{\mathbb{E}_r[X]}{(r-k)!}}{k!} \tag{9.11}$$

Proof. Let $k \in \{0, 1, 2, \ldots, n\}$ be given. For $i \in \{0, 1, 2, \ldots, n\}$, let $A_i = \{X \geq i\}$ and $B_i = \{X = i\}$.

Observe that B_k is the event that exactly k of the A_i occur. Using the notation of Theorem 9.1.6, $S_k = \dfrac{\mathbb{E}_k[X]}{k!}$ and

$$\begin{aligned}
p_k &= \mathbb{P}[B_k] \\
&= \sum_{r=k}^{n} (-1)^{r+k} \binom{r}{k} S_r && \text{(by (9.4))} \\
&= \sum_{r=k}^{n} (-1)^{r+k} \binom{r}{k} \frac{\mathbb{E}_r[X]}{r!} \\
&= \frac{\sum_{r=k}^{n} (-1)^{r+k} \frac{\mathbb{E}_r[X]}{(r-k)!}}{k!} && \text{(simplifying).}
\end{aligned}$$

□

9.2.3 Binomial moments

Definition 9.2.3 (Binomial moments)**.** *For nonnegative integers, r, the binomial moments S_r of X are*

$$S_r = \begin{cases} 1 & \textit{if } r = 0 \\ \sum \mathbb{E}[X_{l_1} X_{l_2} \cdots X_{l_r}] & \textit{if } r > 0 \end{cases} \tag{9.12}$$

where the sum in equation (9.12) *is over all* $1 \leq l_1 < l_2 < \cdots < l_r \leq n$.

Thus

$$S_r = \frac{\mathbb{E}_r[X]}{r!}. \tag{9.13}$$

The following theorems are from [30] and will not be proved here.

Theorem 9.2.4. *Suppose that for all* $i \in [m]$, X_i *is an indicator random variable and* $X = X_1 + X_2 + \ldots + X_m$. *Then*

$$\mathbb{P}[X = r] = \sum_{k=0}^{n-r} (-1)^{k-r} \binom{k+r}{k} S_{k+r}. \tag{9.14}$$

Corollary 9.2.5. *Suppose* X_i *and* X *are as in Theorem 9.2.4. Then*

$$\mathbb{P}[X = 0] = \sum_{k=0}^{n} (-1)^k S_k. \tag{9.15}$$

The following inequalities are also useful.

Theorem 9.2.6. *(Bonferroni inequalities)*

For all integers $0 < k$ *and* $0 < m$,

$$\mathbb{P}[X = k] \leq \sum_{j=0}^{2m} (-1)^j \binom{k+j}{j} S_{k+j}. \tag{9.16}$$

$$\mathbb{P}[X = k] \geq \sum_{j=0}^{2m-1} (-1)^j \binom{k+j}{j} S_{k+j}. \tag{9.17}$$

The proofs follow the same idea as Theorem 9.1.1.

9.3 Limit theorem

The theorem and proof are from [4].

Theorem 9.3.1. *Let* $X = \sum_{j=1}^{m} X_j$ *and let* S_k *be as in* (9.13). *Suppose there exists* $\mu \in (0, \infty)$ *such that for every* $k = 1, 2, \ldots,$

$$\lim_{n \to \infty} S_k = \mu^k. \tag{9.18}$$

Then

$$\lim_{n \to \infty} \mathbb{P}[X = k] = \frac{\mu^k e^{-\mu}}{k!}. \tag{9.19}$$

Proof. For every $m = 1, 2\ldots$, inequality (9.16) implies

$$\begin{aligned}
\mathbb{P}[X = k] &\leq \sum_{j=0}^{2m} (-1)^j \binom{k+j}{j} S_{k+j} \\
&\leq \sum_{j=0}^{2m} (-1)^j \binom{k+j}{j} \frac{\mu^{k+j}}{(k+j)!} \\
&= \frac{\mu^k}{k!} \sum_{j=0}^{2m} \frac{(-\mu)^j}{j!} \\
&\leq \frac{\mu^k}{k!} e^{-\mu}.
\end{aligned}$$

Similarly, inequality (9.17) implies

$$\begin{aligned}
\sum_{j=0}^{2m-1} (-1)^j \binom{k+j}{j} \frac{\mu^{k+j}}{(k+j)!} &= \frac{\mu^k}{k!} \sum_{j=0}^{2m-1} \frac{(-\mu)^j}{j!} \\
&\leq \mathbb{P}[X = k].
\end{aligned}$$

Putting these inequalities together yield for every $k = 1, 2\ldots$,

$$\lim_{n\to\infty} \mathbb{P}[X = k] = \frac{\mu^k e^{-\mu}}{k!}$$

as claimed. $\square$

Bibliography

[1] N. Alon and J. H. Spencer, *The probabilistic method*, John Wiley & Sons, Inc., New York, NY, 2000.

[2] H. Anton, *Elementary Linear Algebra*, John Wiley & Sons, Inc., Mississauga, Ontario, 2008.

[3] J. Beck, A remark concerning arithmetic progressions, *J. Combin. Theory Ser. A* **29** (1980), 376–379.

[4] B. Bollobás, *Random graphs*, Cambridge University Press, Cambridge, U.K., 2006.

[5] S. N. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, *Comm. Soc. Math. Kharkov* **13** (1912), 1–2.

[6] H. Chernov, A measure of the asymptotic efficiency tests of a hypothesis based on the sum of observations, *Ann. Inst. Statist. Math.* **23** (1952), 493–508.

[7] R. Diestel, *Graph theory*, Springer-Verlag, Berlin, Germany, 2006.

[8] P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.(N.S)* **53** (1947), 292–294.

[9] P. Erdős, Graph theory and probability, *Canad. J. Math.* **11** (1959), 34–38.

[10] P. Erdős, On a combinatorial problem II, *Acta. Math. Acad. Sci. Hungar.* **15** (1964), 445-447.

[11] P. Erdős, On extremal problems of graphs and generalized graphs, Israel J. Math. **2** (1964), 183–190. (Also reprinted in: P. Erdős, *The art of counting; selected writings*, J. Spencer ed., MIT Press, Cambridge, Massachusetts, 1973.)

[12] P. Erdős and D. J. Kleitman, On coloring graphs to maximize the proportion of multicolored k-edges, *J. Combinatorial Theory* **5** (1968), 164–169.

[13] P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, *Infinite and finite sets* (A. Hajnal *et al.*, editors), North Holland, Amsterdam, (1975), 609–628.

[14] P. Erdős and A. Rényi, On the evolution of random graphs, *Magyar. Tud. Akad. Mat. Kutato Int. Kőzl*, **5**, 17–61.

[15] R. Fagin, Probabilities on Finite Models, *J. Symbolic Logic*, **41** No. 1, (1976), 55–58.

[16] W. Feller, *An introduction to probability theory and its applications*, Vol. 1, 2^{nd}. Ed., John Wiley and Sons, Inc., New York, N.Y., 1966.

[17] W. Feller, *An introduction to probability theory and its applications*, Vol. 2, 2^{nd}. Ed., John Wiley and Sons, Inc., New York, N.Y., 1966.

[18] Z. Füredi and J. Kahn, On the dimension of ordered sets of bounded degree, *Order* **3** (1986), 15–20.

[19] A. Gut, *An intermediate course in probability*, Springer-Verlag, New York, NY, 1995.

[20] D. Gunderson, The probabilistic method for Math 8210, Unpublished notes, December, 2006.

[21] P. Halmos, *Measure theory*, D. van Nostrand Company, Inc., Princeton, NJ, 1950.

[22] T. Jech, *Set theory: the third millennium edition, revised and expanded*, Springer-Verlag, Berlin, Germany, 3rd rev edn (corrected 4th printing edn), 2006.

[23] S. Jukna, *Extremal combinatorics*, Springer-Verlag, Berlin, Germany, 2001.

[24] M. G. Kendall and B. Babbington-Smith, On the method of paired comparisons, *Biometrika* **33** (1940), 239–251.

[25] J.E. Littlewood, A.C. Offord, On the number of real roots of a random algebraic equation (III), *Rec. Math. (Mat. Sbornik) N.S*, **12 (54) (3)** (1943), 277–286.

[26] D. Lubell, A short proof of Sperner's lemma, *Journal of Combinatorial Theory 1*, **2** (1966), 299.

[27] J. Malitz, *Introduction to mathematical logic*, Springer-Verlag, New York, N.Y., 1975.

[28] L. D. Meshalkin, A generalization of Sperner's theorem on the number of subsets of a finite set, (in Russian), *Teor. of Veroj. Primen.* **8** (1963), 219–220. English translation in: *Theory Probab. Appl.*, **8** (1964), 204–205.

[29] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distribution and their applications, *Phys. Rev. E* **64** (2001).

[30] E. M. Palmer. *Graphical evolution*, John Wiley & Sons, Inc., New York, NY, 1985.

[31] S. P. Radziszowski, and X. Xu and Z. Shao, More Constructive Lower Bounds on Classical Ramsey Numbers, *SIAM J. Discrete Math.* **25** (2011), 394–400.

[32] F. P. Ramsey, On a problem of formal logic, *Proc. Lond. Math. Soc.*, 2, **30** (1929), 264–286.

[33] I. Reiman, *Acta Math. Acad. Sci. Hung.* **9** (1959), 269–279.

[34] S. Ross, *A first course in probability*, Prentice Hall, Upper Saddle River, NJ, 1998.

[35] H. L. Royden and P. M. Fitzpatrick, *Real analysis*, Prentice Hall, Boston MA, 2010.

[36] J. Spencer, *Ten Lectures on the probabilistic method*, SIAM, Philadelphia, PA, 1987.

[37] T. Szele, Kombinatorikai vizsgálatok az irányitott teljes gráffal kapcsolatban, *Mat. Fiz. Lapok* **50** (1943), 223–256. For an English translation see: T. Szele, *Publ. Mat.* **13** (1966), 145–168.

[38] K. Tangwongsan, "Probability bounds", Lecture notes: 15-492 Parallel algorithms, Sept. 27, 2007, Carnegie Mellon University.

[39] R. J. Wilson, and J. J. Watkins, *Graphs an introductory approach*, John Wiley & Sons, Inc., Hoboken, NJ, 1990.

[40] K. Yamamoto, Logarithmic order of maximum satisfiability, *J. Math Soc. of Japan* **6** (1954), 343–353.